The Contribution of miRNAs to Airway Inflammation

Lauren Alina Headley-Morris

National Heart and Lung Institute
Faculty of Medicine
Imperial College London

A Thesis Submitted for the Degree of Doctor of Philosophy (PhD)

Declaration of Originality

The clinical aspects of this thesis, including recruitment and sampling of patients, were conducted in collaboration with past and present researchers and clinicians within the Genomic Medicine Group based at the National Heart and Lung Institute, Imperial College London. All experimental design, laboratory work, and data analysis presented herein are my own, except where explicitly attributed to others through appropriate citations.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

This thesis explores the role of microRNAs (miRNAs) in asthma and eosinophilic lung disease (ELD) through a series of next generation sequencing (NGS) studies. The research aimed to identify differentially expressed (DE) miRNAs and their potential biological implications in these respiratory conditions.

Initially, a nasosorption sampling method was optimized for miRNA sequencing from nasal mucosal samples, demonstrating its feasibility as a non-invasive approach for planned subsequent respiratory studies. In addition to optimization of the sampling, extraction and sequencing library generation an analytical pipeline for downstream processing of generated sequencing data was also developed. Due to the impacts of the COVID pandemic the focus of the thesis had to alter. Consequently, research became focused on the analysis of circulating miRNA expression in whole blood samples from severe asthmatics, non-severe asthmatics, and healthy controls using NGS. These investigations identified distinct miRNA signatures associated with asthma severity, including miR-1304-3p, miR-32-5p, and members of the let-7 family.

Gene ontology analyses of the DE-miRNAs implicated their involvement in processes such as apoptosis regulation, cellular metabolism, and enzyme binding, providing insights into potential underlying molecular mechanisms of disease. The research also explored miRNA expression in both whole blood and lung biopsy samples from individuals with ELD and healthy controls, revealing potential associations between miRNAs and clinical features, as seen with miR-202-5p and Eosinophil counts.

While sample size limitations and technical challenges were encountered, particularly in relation to the ELD study, the research highlighted the complexity of miRNA regulatory networks in respiratory diseases. The findings lay the groundwork for future research aimed at developing novel diagnostic tools and personalized treatment approaches for respiratory diseases.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Professor Miriam Moffatt, Professor Trevor Hansel, Professor Mike Lovett, and Dr. Paul Lavender, for their invaluable contributions to my understanding of research. I am particularly indebted to Professor Moffatt for her exceptional mentorship, diplomacy, and unwavering support, especially during the challenges of the COVID-19 pandemic and the arrival of my son. Her guidance has been instrumental in my personal and professional development.

I also extend my thanks to the team at Genomic Medicine. Kenny Wong's meticulous organizational skills ensured the smooth running of the lab, while Leah Cuthbertson provided crucial experimental advice and assistance with R. I appreciate Verdiana Martello's expertise in sample processing and Anca Nastase for generously sharing her impeccable lab techniques. Thanks also to Claire McBrian for generously providing the samples used in a significant part of this thesis, which were collected during their research on Asthma and Eosinophilic lung disease. I am grateful for the opportunity to learn and apply advanced techniques, such as NGS, and to contribute to original research in this field.

This research was generously funded by Asthma UK (now Asthma + Lung UK), for which I am very grateful.

On a personal note, I am deeply thankful to my husband, Carl, for his unwavering support, particularly during challenging times. His encouragement, grammar skills, and shared enthusiasm for my research, even the "intricacies of scientific language," have been a constant source of strength. I am incredibly proud of his accomplishments as a teacher and entrepreneur, all while providing unwavering support for my studies.

I am eternally grateful to my parents for their continuous encouragement of my academic pursuits and their invaluable help with childcare, which allowed me to focus on my research.

Finally, I am thankful to my son, Harrison, who brought joy and laughter into our lives during a demanding period. His unconditional love and boundless energy were a constant reminder of the beauty beyond the lab.

Table of Contents

1. Chapter 1: Introduction 20
1.1. Asthma 20
1.1.1. The Global Impact of Asthma 20
1.1.2. Clinical Features and Diagnosis 21
1.1.3. Clinical Associations 28
1.1.4. The Role of Immune Cells in Asthma 29
1.1.5. Pathophysiology and Treatment Advances 32
1.1.6. Asthma Management and Disease Severity 38
1.2. Eosinophilic Lung Diseases 40
1.2.1. Classification 41
1.2.2. Demographics, Clinical Features, and Management 43
1.3. miRNAs 48
1.3.1. An Overview of miRNAs 48
1.3.2. Methods of miRNA Analysis 54
1.4. Practical Considerations 68
1.5. Asthma and Eosinophilic Lung Diseases: Associations with miRNAs 71
1.5.1. miRNA Expression in Asthma 71
1.5.1.1. Asthma Population Heterogeneity and miRNA Expression 74
1.5.1.2. miRNA Changes and Regulation of Inflammatory Pathways 76
1.5.2. miRNA Expression in ELD 78
1.5.3. Impact of Corticosteroids on miRNA Processes in Inflammatory Lung Conditions with a Focus on Asthma and ELDs 78
1.6. Hypothesis and Aims 80
2. Chapter 2: Methods and Materials 83
2.1. Methods 83
2.1.1. Clinical Sample Collection 83
2.1.1.1. Nasosorption Sample Collection 83
2.1.1.2. Blood Sampling 85
2.1.2. RNA Extraction Methods 86
2.1.2.1. Zymo RNA Extraction 87
2.1.2.2. Monarch RNA Extraction 88
2.1.2.3. PAXgene RNA Extraction 89

2.1.3. Quantification Methods for Nucleic Acids....91
2.1.3.1. Nanodrop (ND-1000) Spectrophotometer....92
2.1.3.2. QubitTM 3 Fluorometer....93
2.1.3.3. Agilent 2100 Bioanalyzer....94
2.1.4. RNA Sequencing Library Preparation....101
2.1.4.1. Size Selection Methods....106
2.1.4.2. Sample Concentration and Salt Removal Methods....110
2.1.5. Sequencing Methodology....112
2.1.5.1. Data Processing Methods....116
2.1.5.2. Data Analysis Methods....132
2.1.5.3. miRNet mRNA Target Analysis....133
2.1.5.4. Literature Review....137
2.1.5.5. Reference miRNA Analysis....137
3. Chapter 3: Method Development for Nasosorption sampling for miRNA sequencing. 140
3.1. Introduction....140
3.2. Methods....142
3.2.1. Study Participants....142
3.2.2. Library Preparation and Sequencing....144
3.2.3. Data Analysis....145
3.2.4. Literature Review....145
3.3. Results....146
3.3.1. RNA Isolation....146
3.3.1.1. Processing Protocol 1....147
3.3.1.2. Processing Protocol 2....150
3.3.1.3. Processing Protocol 3....153
3.3.1.4. Processing Protocol 3a....156
3.3.2. Library Synthesis....160
3.3.3. Size Selection....161
3.3.3.1. Library Concentration for Sequencing....165
3.3.4. Sequencing Analysis....169
3.3.4.1. Top Expressing miRNAs and Literature Review....175
3.4. Discussion....180
4. Chapter 4: miRNA Expression Comparison Between Severe Asthmatic and Healthy Control Whole Blood Samples....186
4.1. Introduction....186
4.2. Materials and Methods....189
4.2.1. Study Participants: The Celtic Fire Study....189

4.2.2. Sample Processing....191
4.2.3. Sequencing and Analysis....194
4.3. Results....194
4.3.1. Subject Demographics and Clinical Correlations....194
4.3.2. cDNA Library Processing....199
4.3.3. Sequencing Results....202
4.3.4. miRNet mRNA Target Analysis....215
4.3.4.1. Comparison of Gene Target Identification among GO Analysis Categories for Control vs Severe Asthma Pilot....222
4.3.5. Literature Comparison....226
4.4. Discussion....229
5. Chapter 5: miRNA Expression Comparison Among Severe Asthmatics, Non-Severe Asthmatics and Healthy Control Whole Blood Samples....239
5.1. Introduction....239
5.2. Materials and Methods....240
5.2.1. Study Participants and Study Conduct....240
5.2.2. Overview of Sample Processing....241
5.2.3. Sequencing and Analysis....244
5.3. Results....244
5.3.1. Subject Demographics and Clinical Correlations....244
5.3.2. Sample Processing....249
5.3.3. Sequencing Results....252
5.3.3.1. Differential Expression Analysis....261
5.3.3.2. DE-miRNAs and Clinical Outcome Correlation....270
5.3.4. miRNet mRNA Target Analysis....275
5.3.4.1. Control vs Severe Target Analysis....276
5.3.4.2. Non-Severe vs Severe Target Analysis....290
5.3.4.3. Control vs Non-Severe Target Analysis....297
5.3.5. Literature Comparison and Comparison with Pilot Study....302
5.3.6. mRNA/miRNA Expression Validation....305
5.4. Discussion....309
6. Chapter 6: Circulating and Bronchial miRNA Expression in Eosinophilic Lung Disease....323
6.1. Introduction....323
6.2. Methods....325
6.2.1. Study Participants and Study Conduct....325
6.2.2. General Sample Processing....327
6.2.3. Sequencing and Analysis....332

6.3. Results....334
6.3.1. Patient Demographics and Clinical Correlations....334
6.3.2. Sample Processing....338
6.3.3. Sequencing Analysis....342
6.3.4. Differential Expression Analysis....356
6.3.5. miRNA Correlation between Healthy Controls and ELDs....363
6.3.6. Target Identification....367
6.3.6.1. Blood Gene Ontology Analysis....368
6.3.6.2. Biopsy Gene Ontology Analysis....375
6.3.7. Reference Gene Analysis....382
6.4. Discussion....384
7. Chapter 7: Discussion....391
7.1. Nasosorption Sampling for small RNA Sequencing Analysis....391
7.1.1. Relevance of the Findings of Nasosorption sampling for Sequencing Analysis....392
7.2. Asthma....394
7.2.1. Relevance of Findings of miRNA Expression in Asthma Compared to in Health...395
7.3. Eosinophilic Lung Disease....398
7.3.1. Relevance of Findings of miRNA Expression in ELD Compared to in Health..399
7.4. miRNA Expression between Tissues....400
7.5. miRNA Expression between Inflammatory Respiratory Conditions....401
7.6. The Relationship between miRNA Expression and other Clinical Variables....403
7.7. The Relationship between miRNA Expression and Gene Ontology Results....404
7.8. Strengths and Limitations of the Studies....407
7.9. Future Work....411
7.10. Summary....413
8. References....417
9. Appendices....465

Abbreviations:

Af: *Aspergillus fumigatus*

AGO: Argonaut protein/gene

AHR: Airway hyperresponsiveness

ATP: Adenosine triphosphate

BD: Bronchodilator

BMI: Body Mass Index

BP: Biological Process

bp: Base pair

CC: Cellular Component

CCL5: Chemokine (C-C motif) ligand 5

cDNA: Synthesised library from reverse transcribed RNA samples

CDS: (Protein) Coding sequences

CF: Celtic Fire

CF1: Celtic Fire Pilot

CF2: Celtic Fire follow up study

CLASH: Cross-linking ligation and sequencing of hybrids

CLIP: Cross-linking and immunoprecipitation

CS: Corticosteroid

CT: Computed tomography

CV: Coefficient of variance

DE: Differential expression

DE-miRNA: Differentially expressed miRNA

DGCR8/PASHA: DiGeorge Syndrome Critical Region 8

DNA: Deoxyribonucleic acid

EGPA: Eosinophilic granulomatosis with polyangiitis (Churg-Strauss Syndrome)

ELDs: Eosinophilic Lung Diseases

FC: Fold change

FeNO: Fractional Exhaled Nitric Oxide

FEV1: Forced Expiratory Volume in one second

FVC: Forced Vital Capacity

g: Gram

gDNA: Genomic DNA

GINA: Global Initiative for Asthma

GO: Gene Ontology

GUI: Graphical user interface

GWAS: Genome wide association study

HITS-CLIP: High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation

HS: High sensitivity

ICEP: Idiopathic chronic eosinophilic pneumonia

ICS: Inhaled corticosteroids

IFNγ: Interferon gamma

IgE: Immunoglobulin E

IL-1: Interleukin-1

IL-1β: Interleukin-1 beta

IL-4: Interleukin-4

IL-5: Interleukin-5

IL-9: Interleukin-9

IL-10: Interleukin-10

IL-13: Interleukin-13

IL-25: Interleukin-25

IL-33: Interleukin-33

ILC2: Innate lymphoid cell type 2

iNOS: Inducible nitric oxide synthase

KEGG: Kyoto Encyclopedia of Genes and Genomes

LABA: Long-acting beta-agonist

lncRNA: Long non-coding RNA

M: Million

mcg: Microgram

MF: Molecular Function

miRNA: Micro RNA

miRNet: Gene target analysis tool for conducting Gene Ontology with miRNA expression data.

miRISC/RISC: miRNA/RNA induced silencing complex

mL: Millilitre

MRE: miRNA recognition element

M value: Stability score

ng: Nanogram

NGS: Next generation sequencing

NICE: National Institute of Care and Excellence

nm: Nanometer

NO: Nitric oxide

NS: Nasosorption

nt: Nucleotide

OCS: Oral corticosteroids

PAR-CLIP photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation

pri-miRNAs: Hairpin containing primary miRNA transcripts

pre-miRNAs: Precursor micro RNAs

PMID: PubMed Identifier

TSLP: Thymic stromal lymphopoietin

TNFα: Tumor Necrosis Factor (alpha)

QC: Quality Control

QPCR: Quantitative Polymerase Chain Reaction

RIN: RNA integrity number

RT: Reverse transcription

RT-PCR: Reverse transcription-polymerase chain reaction

SABA: Short acting beta-agonist

SAM: Synthetic absorptive matrix

SD: Standard deviation

SIGN: Scottish Intercollegiate Guidelines Network

siRNAs: Silencing RNAs

TCC: Tag count comparison

T_H1: T-helper 1

T_H2: T-helper 2

TMM: Trimmed mean of M values

TRBP: TAR (Trans-Activation Response) RNA-Binding Protein

tRF: tRNA fragment

tRNA: Transfer RNA

XPO5: Exportin 5

ΔCT: Change in threshold cycle

µg: Microgram

µL: Microlitre

vs: Versus

°C: degrees Celsius

Figures:

Figure 1.1. Early and Late Phase Asthmatic Response……*32*

Figure 1.2. Overview of Healthy and Asthmatic Airway…….….39

Figure 1.3. CT and Histology Features of Chronic Eosinophilic Pneumonia.…….….42

Figure 1.4. Key miRNA Biogenesis Pathways.……...……*54*

Figure 2.1. Photograph of the Nasosorption Device. …….…….….86

Figure 2.2. Photograph of Agilent Bioanalyzer Small RNA Chip and High Sensitivity DNA Chip.…….……..…….*99*

Figure 2.3. Example High Sensitivity DNA Bioanalyzer Readout..……101

Figure 2.4. Example Small RNA Bioanalyzer Readout.…….……102

Figure 2.5. Diagrammatic View of the Illumina Library Synthesis Protocol…….…..…...107

Figure 2.6. Pippin Prep Device with Cassette Inserted and Being Prepared for Experiment.. 109

Figure 2.7. Diagram of a Single Pippin Prep Cassette Channel…….110

Figure 2.8. Illumina Sequencing Process……117

Figure 2.9. Schematic of Data Processing from Raw Sequencing Reads to Visualization of Final Differentially Expressed (DE) miRNAs……119

Figure 2.10. FASTQC: Quality Control Summary....……121

Figure 2.11. FASTQC: An Example of Per-base Sequence Quality with Text Overlay.. ..……123

Figure 2.12. FASTQC: An Example of Per-sequence Quality Scores..…….124

Figure 2.13. FASTQC: Distribution of Sequence Length Across All Sequences.…….125

Figure 2.14. FASTQC: Sequence Content Across All Bases. …….. 126

Figure 2.15. An Example of FASTQC GC Distribution Histogram..……127

Figure 2.16. An Example of FASTQC Sequence Duplication Histogram……128

Figure 2.17. An Example of FASTQC GC Distribution Histogram..……129

Figure 2.18. An Example of FASTQC Overrepresented Sequence Table……130

Figure 2.19. ExceRpt Pipeline.……..…….131

Figure 2.20. miRnet Selection and Input Screen……135

Figure 2.21. miRnet Network Analysis Filtering Screen. . ……136

Figure 2.22. miRnet Network Viewer Mode . ……138

Figure 3.1. Outline of Sample Processing Protocol 1…….……150

Figure 3.2. Total RNA Extractions Conducted from Nasosorption samples Using two Different Total RNA Extraction Kits……..…….. .152

Figure 3.3. Outline of Sample Processing Protocol 2. ……..…….. .154

Figure 3.4. Small RNA Bioanalyzer Trace of Extracted Total RNA Sample from Protocol 2…155

Figure 3.5. Outline of Sample Processing Protocol 3. ……..…….. .157

Figure 3.6. Small RNA Bioanalyzer Trace of Nasosorption sample.. ….…….. .158

Figure 3.7. Outline of Sample Processing Protocol 3a……..……. 159

Figure 3.8. Summary RNA Concentrations of Samples from Sampling Protocol 3a.. ……. 160

Figure 3.9. Schematic of Sampling Handling and Stability Adaptations……. 161

Figure 3.10. cDNA Library Isolation Optimisation……..……165

Figure 3.11. High Sensitivity DNA Bioanalyzer Trace of cDNA Library from Nasosorption Samples…….……..166

Figure 3.12. cDNA Library Optimisation Clean-up and Size Selection Yields.……. 169

Figure 3.13. cDNA library of Sample Submitted for Sequencing…… .……..……170

Figure 3.14. FASTQC Result for Per Base Sequence Quality.……..……172

Figure 3.15. Scaled Percentage Filtered Alignment of Raw Reads from Nasosorption Sample…….174

Figure 3.16. Scaled Read Alignment of Healthy NS Sample.…….176

Figure 3.17. Bar Chart of miRNAs with Total Read Counts Greater than 1,000 from Sequencing of a Nasosorption Sample.……..178

Figure 3.18. Venn Diagram of miRNAs Comparison……..181

Figure 4.1: Sample Processing Flowchart.…….194

Figure 4.2. Clinical Correlation Analysis of Key Clinical Variables……200

Figure 4.3. cDNA Library Concentration after Pippin Prep Size Selection.. ……202

Figure 4.4. cDNA Library Concentration after Concentration Clean-up..……203

Figure 4.5. Boxplot Summary of Raw Reads for Control and Severe Asthma (Case) Sample Sequencing ……205

Figure 4.6. CF1 Control vs Severe Raw Read Alignment.. ……209

Figure 4.7. miRNA Diversity between Concentrated and Non-concentrated Samples…….211

Fig 4.8. Pilot Study Sequencing Results miRNAs per Sample vs White Cell Count...…….212

Figure 4.9. Volcano Plot of DE Analysis Result of miRNA Reads Compared by Severity. …….214

Figure 4.10. Correlation Heatmap of DE miRNAs and Clinical Features of Interest……216

Figure 4.11. Partial Gene Ontology Classification (Biological Process) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls.…219

*Figure 4.12. Partial Gene Ontology Classification (Molecular Function) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls.*221

Figure 4.13. Partial Gene Ontology Classification (Cellular Components) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls. ……223

Figure 4.14. Venn Diagram of Target Genes Control vs Severe Asthma.……...225

Figure 4.15. Gene Targets and Source miRNAs from GO Analysis Faceted by Analysis Interaction. ……227

Figure 4.16. Venn Diagram of Comparison of DE miRNAs with Literature and Control vs Severe…. ……230

Figure 5.1. Modified Protocol after Pippin Prep Replacement.……243

Figure 5.2. Correlation Analysis of Clinical Outcomes of Patient Cohort.……249

Figure 5.3. Patient BMI Class vs Asthma Severity..……250

Figure 5.4. Bioanalyzer Trace Result and Corresponding Sample Trace from Pippin Prep Size Selection….…… 253

Figure 5.5. Summary Boxplot of Raw Reads input into Alignment for Each Patient Category.……..256

Figure 5.6. Raw Read Alignment...……..258

Figure 5.7. miRNA Total Reads by RNA Sequencing Run by Patient Severity. ……260

Figure 5.8. Aligned miRNA Diversity Grouped by Sequencing Run and Patient Severity……261

Figure 5.9. DE miRNAs Identified from Multiple Comparison Analysis Among Control, Non-Severe, and Severe Asthma.. ……..264

Figure 5.10. Volcano Plot of Control vs Severe Asthma DE-miRNA Analysis……264

Figure 5.11. Volcano Plot of Non-Severe vs Severe Asthma DE-miRNA Analysis……268

Figure 5.12. Volcano Plot of Control vs Non-Severe Asthma DE-miRNA Analysis……269

Figure 5.13. Venn Diagram of DE-miRNAs Identified in DE Comparisons Between Groups.. .270

Figure 5.14. Correlation Analysis of Clinical Variables vs DE miRNAs from Control vs Severe. . ……..……273

Figure 5.15. Correlation Analysis of Clinical Variables vs DE miRNAs from Non-Severe vs Severe..……275

Figure 5.16. Correlation Analysis of Clinical Variables vs DE miRNAs from Control vs Non-Severe. ……276

Figure 5.17. Source miRNAs with Shared Targets.……279

Figure 5.18. Gene Targets and Top 10 Biological Process Category from DE-miRNA Input from Control vs Severe Asthma for Genes with 5 or more Source miRNAs……280

Figure 5.19. Gene Ontology Classification (Biological Process) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls.……282

Figure 5.20. Gene Ontology Classification (Molecular Function) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls and the Significance of Each Category..……284

Figure 5.21. Gene Ontology Classification (Cellular Component) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls and the Significance of Each Category..……286

Figure 5.22. Venn Diagram of Target Genes Identified as Significant in the Top 10 Categories of each GO Analysis Category in Control vs Severe Asthma Study Blood Sample miRNA Expression....……287

Figure 5.23. Gene Targets and Source miRNAs from Control vs Severe GO Analysis Faceted by Analysis Interaction……290

Figure 5.24. Gene Ontology Classification Summary Features and Number of Target Genes for Differentially Expressed miRNAs between Controls and Non-severe Asthmatics……294

Figure 5.25. Venn Diagram of Target Genes Identified as Significant in the Top 5 Categories of each GO Analysis Category in Non-Severe vs Severe Asthma Study Blood Sample miRNA Expression……..……295

Figure 5.26. Gene Targets and Source miRNAs from Non-severe vs Severe GO Analysis Faceted by Analysis Interaction.. .……298

Figure 5.27. Gene Ontology Classification Summary Features and Number of Target Genes for Differentially Expressed miRNAs between Controls and Non-severe Asthmatics……300

Figure 5.28. Venn Diagram of Target Genes Identified by GO Analysis in Control vs Non-severe Asthma Study……...……301

Figure 5.29. Venn Diagram of DE miRNAs Identified between Asthmatics and Healthy Controls in a Published Study and this Experimental Set (Severe Asthma, Non-severe asthma, and Healthy Control).……..……306

Figure 5.30. Graph of Top Stability Scores from geNorm Analysis.. .……308

Figure 5.31. Comprehensive Gene Stability Analysis Result..……309

Figure 5.32. miRNA Expression Stability Correlation for the Identification of a Pair of Reference miRNAs for qPCR Validation. ……310

Figure 6.1. Defrosted 2 mL Cryostorage Tube Containing RNAlater and Lung Biopsy Sample (red) Obtained by Bronchoscopy……..……330

Figure 6.2. Clinical Correlation Table Showing Key Clinical Outcome Measures.……339

Figure 6.3. Blood and Biopsy Total RNA Concentrations from ELD Patients and Healthy Control Subjects.. ……341

Figure 6.4. Overview of Individual Subject cDNA Library Information……343

Figure 6.5. Boxplot of Raw Reads Sequencing Results from ELD vs Healthy Control Whole Blood and Biopsy Samples..……345

Figure 6.6. Individual Sample Total RNA and Final cDNA Library Concentrations...……... 347

Figure 6.7. Alignment of Sequence Data From Two Libraries to Establish Presence of Adapter……... 349

Figure 6.8. Total Aligned miRNA Counts: Comparison of ELD vs Healthy controls (BEAM) Compared to Severe Asthma vs Healthy Controls (CF1).…….. 352

Figure 6.9. Comparison of Read Count from BEAM and CF1..…….. 354

Figure 6.10. Scaled Alignment of Reads..……... 356

Figure 6.11. Volcano Plot of DE Analysis Results of Blood Sample miRNA Expression between Controls and ELDs..…… 360

Figure 6.12. Volcano Plot of DE Analysis Results of Biopsy Sample miRNA Expression between Controls and ELDs..…… 362

Figure 6.13. Top 10 miRNAs with Highest Fold Change (FC) between Biopsy and Blood Sample..……... 364

Figure 6.14. Correlation of Clinical Measures vs Highest Ranking miRNAs Expressed in Blood Samples of ELD vs Healthy Controls…….... 366

Figure 6.15. Correlation of Clinical Measures vs Highest Ranking miRNAs Expressed in Biopsy Samples of ELD vs Healthy Controls…….... 368

Figure 6.16. ELD vs Healthy Control Blood sample miRNAs Gene Ontology Classification Summary Results (Biological Process, Molecular Function and Cellular Components).……..372

Figure 6.17. Venn Diagram of Target Genes Identified as Significant in the Top 5 Categories of each GO Analysis Category for Blood Samples...……373

Figure 6.18. Gene Targets and Source miRNAs from GO analysis Faceted by Analysis Interaction.……..376

Figure 6.19. ELD vs Healthy Control Biopsy sample miRNAs Gene Ontology Classification Summary Results (Biological Process, Molecular Function and Cellular Components).……..379

Figure 6.20. Venn Diagram of Target Genes identified as Significant in the Top 10 Categories of each GO Analysis Category for Biopsy Analysis.……381

Figure 6.21. Gene Targets and Source miRNAs from GO Analysis Faceted by Analysis Interaction.……..383

Figure. 6.22. Reference Gene Analysis for ELD vs Healthy Controls Blood and Biopsy Samples..……..*385*

Tables:

Table 1.1. GINA Guidelines on Suggested ICS Dose for "Low", "Medium" and "High" ICS Dosage…….…...*41*

Table 1.2. Classification of ELD in Clinical Practice.…….……...44

Table 2.1 Index and Adapter Sequence Details……..103

Table 2.2. Genboree Workbench ExceRpt Filtering and Alignment Details..……132

Table 2.3: Key Resources for Data Processing and Analysis of miRNA Sequencing Data……133

Table 3.1. Procesing Protocol 3 Concentrations at 20 Nucleotides from Small RNA Bioanalyzer Analysis…… … … ……..…. 157

Table 3.2. miRNAs Identified in Literature Compared to miRNA Hits from the Nasosorption sample Sequencing…….….180

Table 4.1. Demographics for Case and Control Groups……..…198

Table 4.2. Alignment and Read Processing Filters in ExeRpt Package……..…207

Table 4.3. Target Genes from GO Analysis Top 10 Most Significant (P < 0.05) Categories Compared Against One Another... ……..….225

Table 4.4. DE-miRNAs Identified in Literature (and the Results of Present Study)……..….229

Table 5.1. Index Sequences used for Study 2……..….245

Table 5.2. Patient Demographics of Patients with Samples that Completed Processing and Sequencing……..……247

Table 5.3. Table of DE-miRNAs Found Among Analysis Pairs... ……271

Table 5.4. Target Genes from GO Analysis top 10 Most Significant (P < 0.05) Categories Compared Against One Another.……288

Table 5.5. Differences in Gene Targets for miR-32-5p Identified between Pilot Study and Broader Inclusion Study in Control vs Severe Cohort Biological Process Analysis...……291

Table 5.6. Target Genes from GO Analysis Top 10 Most Significant (P < 0.05) Category Comparison in Non-Severe vs Severe Asthma Cohort.…….. 296

Table 5.7. Target Genes from GO Analysis Top 10 Most Significant (P < 0.05) Category Comparison in Control vs Non-severe Asthma Cohort.……..303

Table 5.8. Table of DE miRNAs Identified in Literature (and the Results of Present Study). ..305

Table 6.1. Index Sequences used for Samples in BEAM. ……333

Table 6.2. Distribution of Patients Across the Two Groups and Key Factors Considered in the Study…….…….. 337

Table 6.3. Total Aligned Reads for Each Library by Sample Type and by Group.……351

Table 6.4. P Values of T Tests Comparing Sample Type (Blood or Biopsy) for the BEAM Study.…….……...357

Table 6.5. ELD Target Genes from GO Analysis Categories (Blood).…….374

1. Chapter 1: Introduction

1.1. Asthma

1.1.1. The Global Impact of Asthma

Asthma is a chronic inflammatory respiratory disease affecting over 300 million individuals globally (Global Strategy for Asthma Management and Prevention, 2021). While some cases of childhood asthma resolve as the individual becomes an adult, for many there is no cure. There are various triggers for the disease including cold weather, exercise, dust particulates, and allergies (Vernon *et al.*, 2012; Edwards *et al.*, 2017). Disease management strategies aim to improve symptoms and prevent progression of asthma severity, exacerbations, hospitalisations, and death. Clinical management involves the chronic use of anti-inflammatory medications and bronchodilators, as well as avoidance of triggers. There are also genetic markers that preclude a predisposition for the condition (Moffatt *et al.*, 2007; Willis-Owen, Cookson and Moffatt, 2018; Herrera-Luis, Martin-Almeida and Pino-Yanes, 2024). For example, the gene *ORMDL3* was identified in a genome wide association study (GWAS), and allelic variants shown to be determinants of childhood asthma (Moffatt *et al.*, 2007). Deaths from the disease are predominantly in the elderly. In developing countries, however, where access to diagnostic facilities and long-term medicines for the condition can be limited, there are a greater number of deaths and across all age groups (Mattiuzzi and Lippi, 2020).

Together the impact of direct costs, such as those of medical intervention and support of well-controlled and poorly-controlled asthma; indirect costs, such as productivity and capacity to contribute to society; as well as the impact to the individual's quality of life, are predicted to cost the US nearly $988 billion alone over the next 20 years (Yaghoubi *et al.*, 2019). This emphasises the growing need for better understanding of unresponsive asthma beyond improved disease control through adherence.

1.1.2. Clinical Features and Diagnosis

Asthma is characterised by airway hyperresponsiveness (AHR) and at least partially reversible airway obstruction, resulting in recurring episodes of wheezing, shortness of breath and cough (Wenzel, 2012). Historic associations with asthma relate sensitisation to what is often something capable of being inhaled from the air (pollen or dust) followed by allergic responses due to sensitization to the component being the trigger for asthma symptoms. Non-allergic (or non-atopic) asthma includes asthma symptom responses to environmental or physiological responses such as hormonal changes or exercise, or external triggers that are not through the standard allergic process, such as environmental pollutants, or viral infections (Klain *et al.*, 2022). Thus, the disease is heterogeneous and categorised into phenotypes based on the inflammatory profile, allergic or non-allergic, and age of onset (Wenzel, 2012).

There is increased recognition that there is both innate and adaptive roles in asthma pathophysiology (Leite-de-Moraes, Hammad and Dy, 2012). As will be discussed in Section 1.1.4., asthma has a history of association with a subset of CD4+ T cells (Type 2 helper T

cells giving rise to Type 2 [T_H2] inflammation pathways). Type 2 inflammation is well studied with understanding of both downstream factors (granulocyte-macrophage colony-stimulating factor [GM-CSF], IL-3, IL-4, IL-5, IL-9, and IL-13) and upstream 'alarmins' cytokines (thymic stromal lymphopoietin [TSLP], IL-25 and IL-33) (Fahy, 2015; Calvén, Ax and Rådinger, 2020). The T_H2-high endotype is linked to recruitment, retention, and survival of eosinophils and is considered as an important type of inflammation in a considerable population of asthmatics. Other mechanisms include Th1 inflammation importantly mediated by the production of interferon-gamma (IFNγ) and Th17 inflammation characterized by IL-17A, IL-17E, IL-17F, and IL-22 cytokine production, which leads to neutrophil activation via IL-8 (Oliphant, Barlow and McKenzie, 2011). Both T_H2 and Th17 pathways can stimulate airway inflammation, tissue fibrosis, and AHR. Thus, a T_H2 high and T_H2 low classification has been widely accepted as two broad umbrella categories below which further subcategories are stratified.

To date, supposed biomarkers of type 2-low or neutrophilic asthma have had limited success. More recent understanding of how symbiotic factors within the lung and body contribute to immune responses, such as the microbiome are giving some indication of the complex interplay of how the airways respond as part of the organism as a whole (Smulders *et al.*, 2024). Differential cell counts may be used to measure the relative numbers of eosinophils and neutrophils in the airway, although eosinophilic and neutrophilic asthma are not mutually exclusive subtypes. The nature of inflammation may change in the same patient over time (D'silva *et al.*, 2007). However, blood neutrophil parameters are often cited to be poor surrogates for the proportion of neutrophils in the sputum (Zhang *et al.*, 2014). As with type 2 asthma, having a better understanding of the

underlying processes of this phenotype and identification of biomarkers will help direct clinical treatment and aid in developing novel therapies (Guida *et al.*, 2022).

In the UK, recommendations on the diagnosis, and subsequent therapeutic interventions based on said diagnosis, are made by the National Institute of Care and Excellence (NICE), and in Scotland, the Scottish Intercollegiate Guidelines Network (SIGN). Clinicians in the UK also use recommendations from an independent body, the Global Initiative for Asthma (GINA). A scientific committee of clinicians working around the globe create guidelines on an annual basis to improve asthma management of patients using the most up-to-date evidence-based methodologies. More details are available at the committee's website: www.ginasthma.org.

Current guidelines recommend that the diagnosis of asthma is based on a history of symptoms, typically wheezing and dyspnea, and objective findings of variable expiratory airflow limitation that is at least partially reversible with bronchodilators (Global Strategy for Asthma Management and Prevention, 2021). NICE guidelines expressly say not to diagnose based on atopic history or symptoms alone (Guideline, 2017). The guidelines, however, also conclude that in some cases examination results may not always be able to identify asthmatics.

Airflow limitation is measured using forced expiratory volume in one second (FEV1), and the ratio of FEV1 and Forced vital capacity (FVC) as a percentage reduced. FEV1/FVC ratio is preferable to FEV1 alone as other conditions may present with a reduced FEV1. In addition poor spirometry technique may be a source of error resulting in potential misdiagnosis. A reduced ratio, however, of FEV1/FVC strongly indicates airflow limitation.

An FEV1/FVC ratio that would be indicative of asthma would be a decrease in FEV1 > 12% or > 200 mL from baseline. This then enables comparison of results from tests such as the positive bronchodilator (BD) reversibility test, the exercise challenge test, or methacholine testing to understand other inflammatory profiles (for example atopic or non-atopic status) to select appropriate treatments or to rule out alternative diagnosis. For example, COPD is often associated with an FEV1/FVC ratio of <70% of predicted (Aaron *et al.*, 2024).

Bronchoprovocation challenge using methacholine to assess the sensitivity of the airway is often used. Methacholine is a non-selective agonist of the muscarinic receptor, present in smooth muscle in the airways. The Methacholine challenge has long been associated with a dose-dependent sensitivity correlating with disease severity whereby asthma sufferers will respond to low concentrations of methacholine with a significant reduction in FEV1 (>20%) compared to healthy controls (Davis, Blais and Cockcroft, 2018).

Lung function testing requires expertise and experience to execute appropriately, as well as the facilities and specialised equipment to conduct the tests (Gershon *et al.*, 2010). GINA has suggested that there is evidence that a positive BD test is more likely if reliever treatments, short-acting beta-agonist (SABA) or long-acting beta-agonist (LABA) medications, are withheld for greater than 4 and 15 hours respectively. It has been identified that over half of patients suspected of having asthma are not tested, leaving much diagnosis and management to be based on self-report (Sokol *et al.*, 2015). Asymptomatic asthmatics during examination have been shown to underreport and under perceive their condition (Gershon *et al.*, 2010).

As discussed above, historically, asthma was associated with being an allergic respiratory condition. There are various non-allergic asthma phenotypes that are often indicated through a differential metabolic state (Peters, Dixon and Forno, 2018). Obesity increases the risk of adult-onset asthma and, while males tend to have a higher prevalence of asthma in childhood, females are increasingly diagnosed through adolescence and adulthood. Much of this has been linked with hormonal changes and, in the case of obesity, the metabolic-immune interaction (Sinyor and Perez, 2023). Allergy testing can be conducted either through skin-prick testing to determine atopic status, or by use of measurement of specific Immunoglobulin E (IgE). It was noted, by GINA, that a positive skin test or serum IgE measurement does not confirm that the allergen is causing symptoms and GINA suggests confirming with the patient's history.

Integration of transcriptomic data has enabled identification of omics-associated clusters of similar expression patterns to identify cohorts of patients that in turn have enabled researchers to understand more about the gene expression related to patterns in phenotype. For example, a recent study conducted using collated data, from multiple measures of biological activity in sputum in a cohort of mild and severe asthma and healthy volunteers (mRNA microarrays, proteomics, and bacterial [16S]), identified five stable clusters that had key phenotype or endotype traits (Kermani *et al.*, 2024). One endotype trait in one cluster was a cohort with the highest sputum neutrophilia, while phenotypically the cluster had a higher incidence of nasal polyps, allergic rhinitis, and eczema relative to another cluster. This has great potential in informing a more comprehensive understanding of the complex endotypes seen in asthma, which could lead to breakthroughs in target identification and treatments.

Regarding endotyping, one approved biomarker by NICE for the diagnosis of asthma is fractional exhaled nitric oxide (FeNO). Nitric oxide (NO) production is an indicator of inflammation produced in response to inflammatory cytokines (TNF-α, IL-1, and IFN-γ) or exogenous inflammatory mediators such as LPS which activate inducible Nitric Oxide synthase (iNOS) producing nanomolar concentrations of NO (Gemicioglu *et al.*, 2014). Airway-associated cells that are potential sources of NO include epithelial cells, macrophages, neutrophils, eosinophils, and mononuclear cells (Moncada, Palmer and Higgs, 1991; Guo *et al.*, 1995; MacPherson *et al.*, 2001; Peake *et al.*, 2003; Odgers and Gow, 2014). NICE guidelines suggest a FeNO result above 40 parts per billion (ppb) as a positive test for asthma in adults, and 35 ppb for childhood asthma (Guideline, 2017). In a review of 43 studies, with data collectively from a total of 13,747 patients, an FeNO 40 (parts per billion) ppb for adults (corticosteroid naive, non-smokers) had a sensitivity of 41% and a specificity of 93%, whilst the cut off for children was set to a range between 20-29 ppb providing a sensitivity of 61% and specificity of 89% (Wang *et al.*, 2018). The authors concluded that FeNO measurement had a modest accuracy but may perform better in diagnosing steroid-naive patients. This suggests room for improvement with regard to biomarkers for asthma.

Biologically relevant inflammatory mediators (e.g. FeNO), however, are not routinely measured for the diagnosis or for the on-going patient assessment in relation to disease management. Immune cell counts in sputum have provided some contribution towards understanding asthma endotypes, however, it is likely that multiple mechanisms underlie these phenotypes. For example, eosinophil counts vary widely and can be increased short term by exercise (Wang *et al.*, 2007). As such, eosinophil counts are specifically mentioned

to be excluded from diagnosis by NICE guidelines. This might indicate that the distinct molecular biomarkers at the root of the mechanism are yet to be discovered. Cytokine measurements have also been proposed, however, low stability of cytokines in normal laboratory conditions and the resulting risk of false negatives make widespread application challenging (Liu *et al.*, 2021).

Periostin is an IL-13 and IL-4-inducible matricellular protein that is produced by bronchial epithelial cells and can be detected in peripheral blood (Jia *et al.*, 2012). Periostin has been implicated in several pathogenic processes in asthma including airway remodelling, subepithelial fibrosis, eosinophil recruitment and regulation of mucus production from goblet cells (Takayama *et al.*, 2006). It has been suggested that serum levels of periostin could be a useful predictor of type 2 eosinophilic airway inflammation in patients with uncontrolled, moderate-to-severe asthma (Novosad *et al.*, 2020).

All of the suggested biomarkers above are predominantly for Type 2 asthma. Identification of biomarkers of type 2-low or neutrophilic asthma have had limited success. As with Type 2 asthma, having a better understanding of the underlying processes of the type 2-low phenotype and identification of biomarkers will help direct clinical treatment and aid in developing novel therapies (Guida *et al.*, 2022). Biomarkers that enables a low-to-zero misclassification rate from other respiratory conditions is still therefore needed for asthma in order to improve clinical outcomes and help diagnose and manage patients more effectively, than the current self-report and lung function, and improve on the variability of success based on FeNO measures.

1.1.3. Clinical Associations

Several epidemiological studies have reported associations between asthma and demographic factors such as gender, body mass index (BMI), and age (Sunadome *et al.*, 2020; Chowdhury *et al.*, 2021). Regarding gender, as commented earlier asthma prevalence is generally higher in males during childhood but this trend reverses after puberty. This disparity has been attributed to hormonal influences as well as differences in airway structure and inflammation (Chowdhury *et al.*, 2021).

Obesity, defined as a BMI greater than 30 kg/m2, is a known risk factor for developing asthma in adults and can worsen asthma control and response to treatment. Proposed mechanisms include mechanical effects of excess weight on lung function, systemic inflammation, and hormonal influences (Peters, Dixon and Forno, 2018).

Age is another important factor in asthma epidemiology. Asthma is common in childhood, with onset often occurring before the age of 5 years (Beasley, Semprini and Mitchell, 2015). Asthma can, however, develop at any age and the prevalence remains significant in adulthood. Older adults with asthma may face additional challenges such as comorbidities, age-related physiological changes, and potential adverse effects of medications (Wardzyńska *et al.*, 2021).

These associations highlight the complex interplay between demographic factors, hormonal influences, and environmental exposures in the development and manifestation of asthma. Understanding these relationships undoubtedly will help inform personalized management strategies and targeted interventions for asthma patients. It is important, when conducting

research studies in patient groups, to ensure collection of relevant data to allow consideration of these recognised factors.

1.1.4. The Role of Immune Cells in Asthma

The asthmatic response is often broken down into two phases: an early response, immediately generating airway obstruction as a result of a localised immune cell reaction; followed by a later phase resulting in recruited-immune response airway obstruction and inflammation (Picado, 1992). Our understanding has advanced significantly since the early theories and the following is a brief summary of the background immunological components that are relevant to the research conducted in this thesis.

The key immune cells involved in the asthma response are mast cells, dendritic cells, eosinophils, activating group 2 innate lymphoid (ILC2) cells, and T and B cells (MacPherson *et al.*, 2001; Barnes, 2018). While not classified as immune cells, airway epithelial cells are also active participants in the immune response (Picado, 1992).

Allergen exposure triggers the release of alarmins (IL-25, IL-33, TSLP) from epithelial cells, activating ILC2s and promoting T_H2 immune responses (Calvén, Ax and Rådinger, 2020). In sensitized individuals, allergen-specific IgE on mast cells leads to degranulation upon re-exposure. The early phase (0-2 hours post exposure) involves mast cell release of histamine, prostaglandins, and cytokines, causing bronchoconstriction and increased vascular permeability (leading to oedema)(Picado, 1992; Sinyor and Perez, 2023)(Figure 1.1. Left panel) .

The late phase (3-12 hours post-exposure) is characterized by eosinophil recruitment and activation, primarily driven by IL-5 (Barnes, 2018) (Figure 1.1. Right panel). Eosinophils release cytotoxic proteins and cytokines, contributing to tissue damage and chronic inflammation (Rothenberg and Hogan, 2006). Sustained production of T_H2 cytokines (IL-4, IL-5, IL-13) perpetuates inflammation, mucus production, and AHR (Saeedavi *et al.*, 2022). IL-4 and IL-13 also promote IgE production creating a self-sustaining cycle (Lambrecht, Hammad and Fahy, 2019).

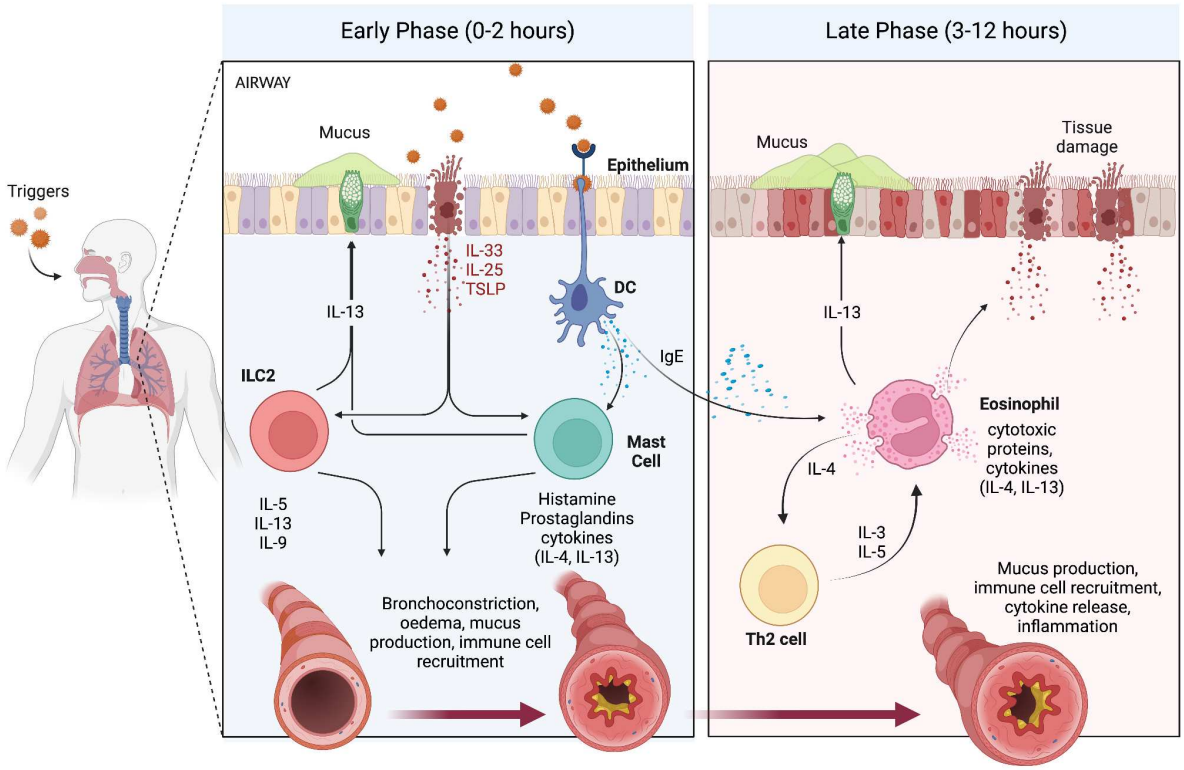


Figure 1.1. Early and Late Phase Asthmatic Response. *Left panel: The early phase (0-2 hours post exposure) - Allergen exposure triggers the release of alarmins (IL-25, IL-33, TSLP) from epithelial cells, activating ILC2s and promoting T_H2 immune responses. Allergen-specific Immunoglobulin E (IgE) on mast cells leads to degranulation upon re-exposure of trigger. Mast cell release of histamine, prostaglandins, and cytokines, causing bronchoconstriction and increased vascular permeability (leading to oedema). Right panel: The late phase (3-12 hours post-exposure) - IL-5 drives eosinophil recruitment and activation. Eosinophils release cytotoxic proteins and cytokines, contributing to tissue damage and chronic inflammation. Sustained production of T_H2 cytokines (IL-4, IL-5, IL-13) perpetuates inflammation, mucus production, and AHR. IL-4 and IL-13 also promote IgE production. Figure created with Biorender.com.*

The longstanding treatment framework for asthma has been to reduce airway constriction and obstruction by attenuating the hypersensitive inflammatory response through the use of glucocorticoids (corticosteroids [CS]) to prevent the early and late immune elements as well as a bronchodilator (BD) to reduce the smooth muscle contraction of the bronchi (Global Strategy for Asthma Management and Prevention, 2021). Research studies have shown that the corticosteroids ameliorate the late phase immune infiltration, including eosinophils, and lung inflammation (de Groot, Ten Brinke and Bel, 2015). Consequently, this has implicated eosinophils as a key contributor to the asthmatic inflammatory response. While the majority of patients do respond to inhaled CS and BD, 10% of patients are non-responsive to even systemic corticosteroid treatment (Bossley *et al.*, 2009). Furthermore, recent studies suggest these treatments may not fully protect against epithelial damage and airway remodelling caused by bronchoconstriction (Bagley *et al.*, 2019).

Eosinophils have been viewed as the main contributing immune cell in asthma for decades due to the elevated levels seen in the circulation and lung tissue of asthmatics (Picado, 1992). Eosinophils are segmented, or bi-lobed, granulocytic tissue resident immune cells generated from the bone marrow (Mahmoud, Austen and Simon, 1980). Although not exclusively resident within the lungs, their elevated presence within the lungs (from biopsy, bronchoalveolar lavage [BAL] or induced sputum) or in the circulation has been associated with asthma (Bousquet *et al.*, 1994). Murine and human eosinophils carry some similarities in distribution, homeostatic baseline locations, and increase in number in relation to certain disease states (Kips *et al.*, 2003). This has promoted the usage of murine models in research, although in more recent years there is growing a divide of belief of the models'

relevance to human asthma and airway disease, and the limitations of particular models (Holmes, Solari and Holgate, 2011).

Further evidence of the contribution of eosinophils, beyond their presence in the affected tissues, has been in relation to the T_H2 cytokine IL-5 that is important in mediating eosinophil expansion, priming, and recruitment, as well as prolonged tissue survival in response to allergic stimuli (Wynn, 2015). While eosinophils do not produce as much IL-5 as lymphocytes, IL-5 promotes the differentiation of terminally committed eosinophil precursors in the bone marrow. IL-5 works synergistically with IL-4 and IL-13, two other key asthma associated cytokines, as well as eotaxins that promote eosinophil-mediated inflammatory responses.

As noted earlier, eosinophils have been shown to be absent in some types of asthma, which has further added to phenotyping of the heterogeneous condition (Simpson *et al.*, 2005). As such, it could be useful to understand the underlying mechanisms or identify biomarkers to better define asthma in the presence or absence of eosinophilic airway inflammation and or elevated circulating eosinophils.

1.1.5. Pathophysiology and Treatment Advances

Asthma pathophysiology is complex due to the disease heterogeneity between individual sufferers and processes affected by repeated lung injury and insult as the disease progresses. As the core tenet of asthma is reversible airway constriction, the core physiological effects are focused around the tissues of the bronchial airways and interactions with immune response elements (detailed in Section 1.1.4.)

During acute exacerbations, bronchial smooth muscle constricts quickly, responding to airway allergens, sensitising agents, or environmental factors. Typically, this is associated with an IgE-dependent release of mediators from mast cells (histamine, tryptase, leukotrienes, and prostaglandins) (Figure 1.1. above) that can act directly on smooth muscle (Yssel *et al.*, 1998). Non IgE-dependent mechanisms that have been suggested to be related to the inflammatory state of the lung include some non-steroidal anti-inflammatory drugs e.g. Aspirin, environmental factors, such as exercise and stress, and cold air in contact with the airways (Beasley, Semprini and Mitchell, 2015).

Oedema through inflammation, mucus hypersecretion and the formation of mucus plugs contribute further to airway narrowing (Barnes, 2016). Inflammatory airway narrowing is the key target of preventative pharmacological intervention (Chua and Lai, 2007). Immune cell infiltration and innate cell activation (Section 1.1.1.4.) contribute to this process, although outcomes of selective targeting of associated immune cells through monoclonal antibodies of key inflammatory mediators such as IgE have not been effective in all cases in improving airway function, despite reducing overall numbers of hospitalisations (Humbert *et al.*, 2018).

Of the key cytokines and inflammatory molecules mentioned involved in asthma, several have been targets for pharmacological intervention since the advent of biologics. The nomenclature of which helpfully gives details to the type of antibody (a prefix chosen by the manufacturer e.g. Benra, the target type; for immune modulating = “li” or cytokine/cytokine receptor = “ki(n)”, the animal origin “zu” = humanized or “u” for human, and lastly details of the antibody class: “mab”= monoclonal antibody).

IgE being associated with T_H2 asthma through eosinophils and asthma inflammation has led to the development of the anti-IgE monoclonal antibody drug Omalizumab. Treatment of severe allergic asthma with Omalizumab was found to reduce exacerbations and reduce OCS usage (Busse *et al.*, 2001).

IL-5 is predominantly produced by T_H2 cells, ILC2 cells, mast cells, NK cells, NK T cells and eosinophils and mediates its effects mostly via controlling eosinophil development, maturation and activation in the bone marrow, as well as subsequent mobilization and survival (Pelaia *et al.*, 2019). IL-5 also modulates the development and function of mast cells and basophils. Three anti-IL-5 (or IL-5 receptor) biologics have been approved, (Rezlizumab [anti-IL-5R], Mepolizumab [anti-IL-5], and Benralizumab [anti-IL-5R]) all with trends of reducing exacerbation rates and making significant reductions of circulating eosinophils (Ghassemian *et al.*, 2021; Farne *et al.*, 2022).

Although the clinical therapies targeting IL-5 have had some success in attenuating asthma exacerbations, the main goal of targeting AHR, however, has not been met despite significant reductions in the presence of airway and circulating eosinophils (Mukherjee, Sehmi and Nair, 2014). By contrast a systematic review showed a decrease in exacerbations of approximately 50% (defined as systemic corticosteroid treatment for at least three days) in eosinophilic asthmatics with poorly controlled asthma on medium dose ICS treated with anti-IL5 therapies (Mepolizumab, Benralizuman, and Reslizumab) (Agache *et al.*, 2020). This study also identified that non-eosinophilic asthmatics on Benralizumab treatment also showed a significant reduction in exacerbation rates (something that was not seen in the single Reslizumab study). This might suggest that eosinophils are less

involved than previously believed, or that there are mechanisms or processes that solely reducing IL-5 or IL-5 receptor activation are not able to overcome, in particular, the sensitivity and responsiveness to the sensitised agent.

The anti-IgE biologic Omalizumab has been shown to attenuate airway remodelling. A cohort of asthma patients receiving Omalizumab over a 16-week period showed reduced airway wall thickening assessed via a validated computer tomography (CT) technique (Hoshino and Ohtawa, 2012). Of note, in the same study Omalizumab also demonstrated a similar reduction in airway and circulating eosinophils to that seen for the IL-5/IL-5R biologics. This may indicate that prevention of eosinophil infiltration of the lung tissue could avert increased extracellular matrix deposition the latter occurring in an attempt to repair damage from granulocyte release.

IL-4 binds to T helper cells, eosinophils, mast cells, B cells, bronchial epithelial cells, the endothelium, and airway smooth muscle cells (ASM cells) and promotes cellular inflammation in the lung. Dupilumab targets the receptor IL-4Ra and has been found to improve asthma control score (ACQ) and lung function (FEV1) in addition to reducing exacerbations in a persistent asthma cohort compared to placebo where LABA and ICS were discontinued for both groups (Wenzel *et al.*, 2013).

IL-13 has a multifunctional role that it plays in many aspects of asthma pathogenesis, including B-cell isotype switching, sub-epithelial fibrosis, recruitment and retention of eosinophils in inflamed airway tissue, and AHR (Nair and O'Byrne, 2019). Due to expression of the IL-13 receptors on ASM cells and epithelial cells, IL-13 enhances cholinergic-induced contractions of smooth muscle *in vitro* and promotes goblet cell

hyperplasia and mucus hypersecretion (Ladjemi *et al.*, 2018). One IL-13 targeting biologic, Lebrikizumab, was used (now discontinued by Elli Lilly) to treat moderate to severe uncontrolled asthma. In a large randomised, double-blind, placebo controlled Phase 3 trial while IL-13 was reduced there were no consistent reductions in exacerbations with high biomarkers (Hanania *et al.*, 2016). Another previously approved drug, Tralukinumab (another monoclonal antibody against IL13), was also discontinued by AstraZeneca after failing to meet endpoints in three separate trials (Keown, 2018).

Recent advances have led to the approval of an anti-TSLP biologic, Tezepelumab, in 2022 (Mullard, 2022). Findings since from the NAVIGATOR study suggest that treatment with Tezepelumab resulted in reduction of asthma exacerbations over a 1 year treatment period, although no statistical analyses were performed to compare effect sizes between the investigated subgroups (Carr *et al.*, 2024). A paper published the prior year considered the effect of Tezepelumab treatment on seasonal and perennial allergy and found reduction in exacerbations between 46% and 63% overall among the seasons (Pavord *et al.*, 2023). Both studies included placebo controls.

Put together there is a picture developing of steps being made towards addressing aspects of severe asthma pathology through targeting of these better studied inflammatory molecules, but little advancement in reducing AHR, despite in some cases the elimination/reduction of eosinophils or the inflammatory marker.

Repeated inflammatory insult and bronchoconstriction exacerbation result in: fibroblasts continually repairing resulting in airway basal membrane thickening, smooth muscle hyperplasia and hypertrophy, as well as airway nerve hyperalgesia and plasticity and these

all contribute to disease progression in asthma (Ward *et al.*, 2002; James *et al.*, 2012; Drake *et al.*, 2021).

Figure 1.2. highlights the key features of asthma pathophysiology discussed (in this section).

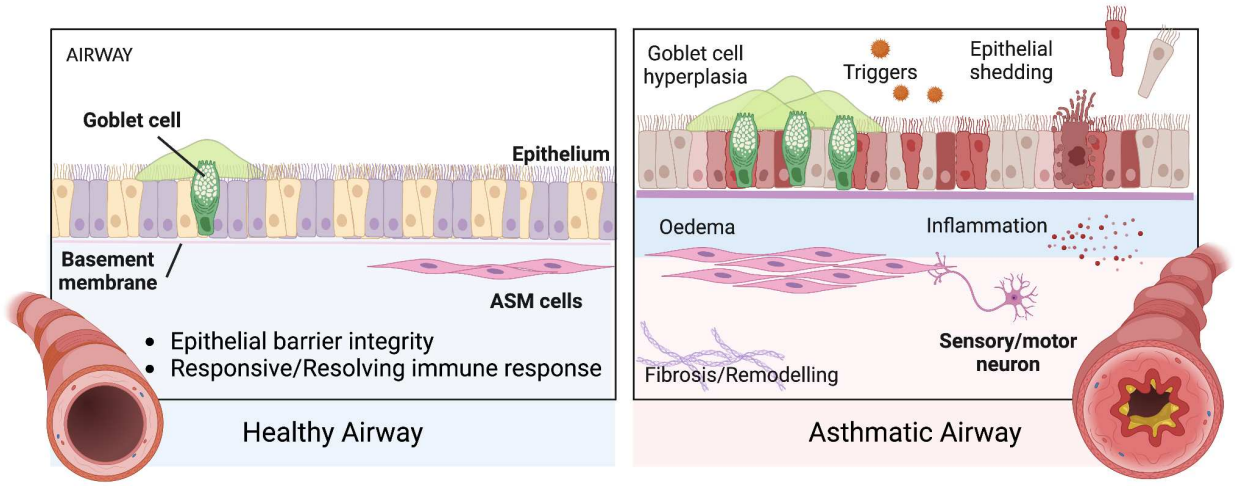


Figure 1.2. Overview of Healthy and Asthmatic Airway. *To the left is the healthy airway, with a thin basement membrane and cohesive epithelial barrier. An example cross section of a bronchiole is shown in the lower left of the pane. Immune challenge results in an adaptive response that resolves once the challenge is addressed. On the right panel an example of an asthmatic airway. Triggers, shown in orange (e.g. pollution, pollen, smoke) to which the body has become sensitised elicit an immune response. Epithelial barrier integrity is compromised, there is increased epithelial shedding and inflammation, the basement membrane is thickened. The airway smooth muscle hypertrophies and sensory and motor neurons become sensitised resulting in stronger bronchospasm as the condition develops. Immune cell activation promotes further inflammation, fibrosis and remodelling (including extra cellular matrix deposition), mucus production and oedema. Epithelial cell metaplasia into goblet cells occurs, and goblet cells undergo hyperplasia enabling increased quantities of mucus production. The lower right of the frame shows an example cross section of an asthmatic bronchiole. Created with Biorender.com.*

1.1.6. Asthma Management and Disease Severity

As discussed previously (Section 1.1.4) management of asthma is primarily through CS to suppress sensitised immune responses from occurring, as avoidance of triggers is not always possible, and the emphasis of use of what was considered the 'reliever' beta-agonist has now been reconsidered. Updated GINA guidelines (2023) have reinforced this. While maintaining the standard use of the brown (CS) inhaler, the blue (BD) short-acting beta2-agonist (SABA) reliever inhalers have been removed from standard initial treatment, in favour of a long-acting beta2-agonist (LABA) reliever for patients suffering with symptoms less than 4-5 days a week. There is now a more stepped approach where management of the condition dictates when to step up to the next level. There is still an arm of therapy that supports the use of a SABA where the patient is likely to be adherent to daily controller (ICS) therapy. This supports the treatment of mild and moderate asthma, whose symptoms are addressed by use of various combinations of inhalers.

Various inhaler delivery modes exist enabling patients to gain the most benefit from available device technologies for ICS asthma management strategies. In addition to the well-recognised aerosol inhaler, there are dry powder (DPI) varieties as well as those that deliver a metered dose in a pressurised inhaler (pMDI). Current guidelines on dosage rates of some examples of ICS formulations are given in Table 1.1. For brevity, only standard particle formulations have been included if available in both standard and extra fine particle formulations. Those with extra fine particles variations have an approximate 60% reduction in maximal comparable dose.

Table 1.1. GINA 2023 guidelines on suggested ICS dose for "low", "medium" and "high" ICS dosage.

	Total Daily ICS dose (mcg)		
Inhaled Corticosteroid	**Low**	**Mid**	**High**
Beclometasone dipropionate	200-500	>500-1000	>1000
Budesonide	200-400	>400-800	>800
Ciclesonide (extrafine particle)	80	>160-320	>320
Fluticasone furoate	100		200
Fluticasone propionate	100-200	>250-500	>500
Mometasone furoate	200-400		> 400

mcg = microgram

Treatment and management of patients with the severest form of asthma (Papiris *et al.*, 2002) is more challenging. Recent GINA guidelines (GINA 2023) have clarified and defined severe asthma from hard-to-treat asthma. Severe asthma is defined as being uncontrolled despite high dose ICS and LABA, or requires high dose ICS-LABA therapy to remain controlled. Hard-to-treat asthma by contrast is where potentially modifiable comorbidities or medication technique could be altered to achieve more control over asthma severity (an issue with adherence, or patient understanding of medication usage i.e. incorrect inhaler technique) (Darbà *et al.*, 2016; Roche *et al.*, 2022).

1.2. Eosinophilic Lung Diseases

Eosinophilic lung diseases (ELD) are a diffuse group of parenchymal lung diseases characterised by interstitial and alveolar infiltration of eosinophils. Although these conditions often present with significant impairment of lung function, a common feature of an ELD is that there is a resolution with systemic corticosteroid treatments without any sequelae in most cases (Cottin, 2016). Figure 1.3. highlights the consolidation within the lung as a result of eosinophilic infiltrates and inflammation in computed tomography (CT) and histology images.

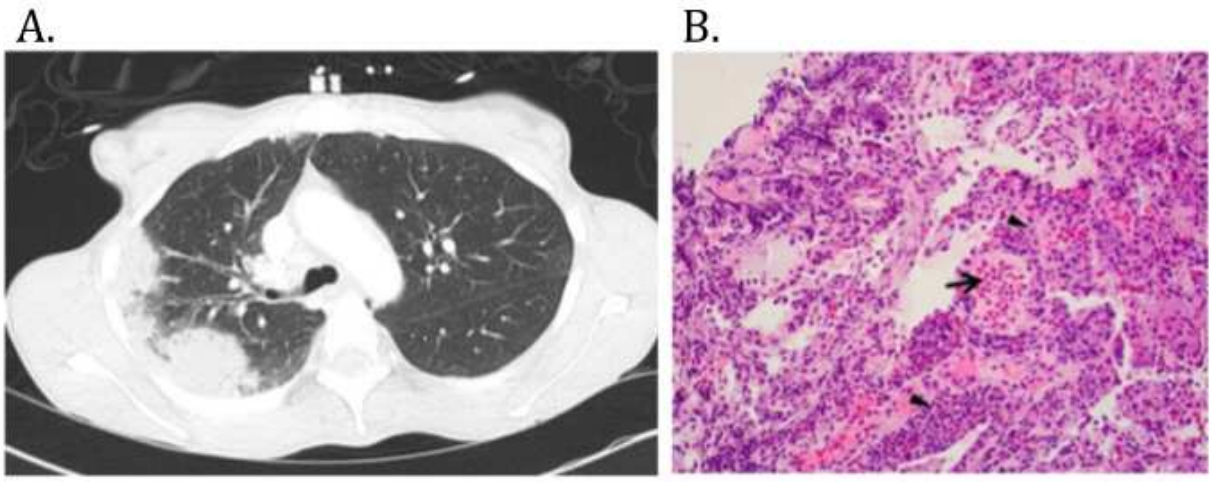


Figure 1.3. CT and Histology Features of Chronic Eosinophilic Pneumonia. *A. Axial CT showing left sided pleural consolidation. B. Haematoxylin and eosin stained lung biopsy sample showing Eosinophil infiltrates.* ***'Therapeutics and Clinical Risk Management 2019 15 397-403' Originally published by and used with permission from Dove Medical Press Ltd.'***

1.2.1. Classification

Typically, ELDs present as acute or chronic pneumonia or as a transient syndrome (Löffler syndrome - most commonly of parasitic origin), although not all ELDs exclusively affect the lung and in some cases, there may be no direct lung inflammation observed (Table 1.2.). The predominant causes are exposure to drugs, toxins, and fungal infection. In the cases of chronic ELD the cause is often idiopathic. When observed in conjunction with systemic inflammatory anomalies it often results in diagnosis of eosinophilic granulomatosis with polyangiitis (EGPA) or idiopathic hypereosinophilic syndromes.

Table 1.2. Classification of ELD in clinical practice.

Idiopathic ELD	Solitary pneumonia of unknown cause
	Idiopathic acute eosinophilic pneumonia
	Idiopathic chronic eosinophilic pneumonia (ICEP)*
Eosinophilic pneumonia in systemic syndromes	Eosinophilic granulomatosis with polyangiitis (EGPA)* (Churg-Strauss Syndrome)
	Idiopathic hypereosinophilic syndromes (lymphocytic or myeloproliferative variant)
ELD of known cause	Allergic bronchopulmonary aspergillosis (ABPA)* and related syndromes
	Eosinophilic pneumonia of parasitic origin
	Eosinophilic pneumonia of other infectious causes
	Drug induced eosinophilic pneumonias
Eosinophilic airway diseases	Eosinophilic asthma
	Hypereosinophilic asthma
	Idiopathic hypereosinophilic constrictive bronchiolitis

*Adapted from Cottin (2016) (Cottin, 2016). *Specific ELDs with features that overlap/have links with asthma.*

1.2.2. Demographics, Clinical Features, and Management

Some of the ELDs have features that overlap or have links with asthma and these will now be discussed further. Idiopathic Chronic eosinophilic pneumonia (ICEP) was first characterised by Carrington and colleagues as an onset of cough over several weeks, dyspnea, malaise, and weight loss with diffuse pulmonary infiltrates (Carrington *et al.*, 1969). The disease is rare, making up fewer than 3% of cases of interstitial lung diseases (disease that affect the interstitium – the tissue and space around the air sacs of the lungs, includes the pneumonias). It is more common in non-tropical areas where parasitic infection is low (Thomeer *et al.*, 2001). Up to 66% of patients have a prior history of asthma and 50% a history of atopy (from eczema, urticaria, to nasal polyposis) (Cottin, 2016). It is thought that fewer than 10% of those diagnosed are smokers, which contrasts with acute eosinophilic pneumonia where a smoking history is a more common occurrence (Marchand and Cordier, 2006).

ICEP has a female dominance of 2:1, with a mean age at time of diagnosis of 45 years (Marchand and Cordier, 2006; Cottin, 2016). Blood eosinophilia is often a diagnostic criterion, mean values of 5,000-6,000/mm3, whilst bronchoalveolar lavage reveals abnormally elevated eosinophils - between 12%-95%, with a mean of 58% (Crowe *et al.*, 2019). Total serum IgE is elevated in approximately 50% of cases (Marchand and Cordier, 2006). A hallmark feature of ICEP in X-ray imaging is peripheral opacities, most commonly bilaterally, representative of peripheral parenchymal infiltrates that are alveolar in nature

(Marchand and Cordier, 2006). ICEP has features that overlap with various other pulmonary conditions making absolute diagnostic criteria challenging. ICEP is typically diagnosed if there is a combination of >2 weeks of respiratory symptoms, BAL and/or blood eosinophilia, pulmonary infiltrates with a peripheral predominance, and exclusion of other known causes of eosinophilic lung diseases. The treatment goal is to induce remission and reduce the chance of relapse using oral corticosteroid (OCS) therapy (Cottin, 2016). Response to treatment occurs within days, followed by a tapered dose towards OCS discontinuation (Suzuki and Suda, 2018). As many as 50% of patients relapse, particularly during the tapering phase necessitating a further round/resumption of OCS therapy. Anti-IL-5 and anti-IL-5R monoclonal antibodies such as Mepolizumab, Reslizumab, Benralizumab and the anti-IgE monoclonal Omalizumab have been proposed as potential treatments for CEP due to their efficacy in reducing eosinophils (Garrett *et al.*, 2004; Kaya *et al.*, 2012; Mukherjee, Sehmi and Nair, 2014; Lim and Nair, 2015). Sarkis *et al*. reported the use of Mepolizumab followed by Reslizumab as a steroid sparing method to assist a patient achieve remission after several relapses on OCS alone (Sarkis *et al.*, 2020).

Considering ELDs of known cause there is again an overlap between ABPA and asthma. Allergic bronchopulmonary aspergillosis (ABPA) was first described by Hinson and colleagues in 1952 (Hinson, Moon and Plummer, 1952). Almost exclusively, ABPA affects subjects with a history of chronic bronchial disease (Geller *et al.*, 1999). Men and women are affected equally and it is usually seen in those in their 20s and 30s but has also been reported in children and infants (Shah, Bhagat and Panchal, 1993; Jat *et al.*, 2018; Bhankhur *et al.*, 2019). Occurrence is 1-2% in asthmatics and 7-10% in patients with cystic fibrosis, and there have been some reports of the disease in patients with chronic obstructive

pulmonary disease (Cottin, 2016). Patients experience chronic cough, dyspnea, brown or tan mucus plug expectorate, low fever, and chronic rhinitis. There are 5 classic stages of ABPA (acute, remission, recurrent exacerbations, corticosteroid dependent asthma, fibrotic end stage), however, despite recurrent exacerbations the 5 classic stages may not necessarily be followed. Typically, pulmonary infiltrates or peripheral blood eosinophilia may be present during the acute and exacerbation phases. Diagnostic criteria for ABPA were first proposed by Rosenberg and Patterson, and the criteria remain widely acknowledged today (Rosenberg *et al.*, 1977). That said, in 2002 an adaptation was proposed by Greenberger of minimally essential criteria, with a further amendment in 2013 (Greenberger, 2002, 2013). There is currently no conclusive agreement in the criteria required for diagnosis. Therefore the following are brief guidance on the type of diagnostic criteria included (Cottin, 2016). The four key criteria for diagnosis of ABPA (using the minimally essential criteria) are presence of asthma, immediate cutaneous reactivity to *Aspergillus fumigatus* (Af), total serum IgE > 1,000 ng / mL, and central bronchiectasis in the absence of distal bronchiectasis. Manifestation can present as mild asthma to extensive lung disease and respiratory failure. Unlike ICEP, ABPA left untreated often results in a chronic fibrotic lung disease reminiscent of post-tubercular fibrotic sequelae (Shah, 2007). Treatment goals for ABPA include management of asthma exacerbations and prevention of progression to bronchiectasis and fibrotic lung while minimising steroid effects. Corticosteroid treatment during exacerbations are the mainstay treatment, with periods of tapering and cessation. It is believed that treatment in this manner may prevent the disease from progressing to the fibrotic end-stage disease (Patterson *et al.*, 1987). Persistent airflow limitation may develop limiting the usage of inhaled corticosteroids. Monoclonal

antibody interventions such as anti-IgE (Omalizumab) may be of use in patients with difficult asthma and ABPA (Tillie-Leblond *et al.*, 2011; Sastre *et al.*, 2012).

Finally, considering eosinophilic pneumonia in systemic syndromes, Jacob Churg and Lotte Strauss first described eosinophilic granulomatosis with polyangiitis (EGPA) by (Churg and Strauss, 1951). It was therefore originally referred to (and is often still referred to in the published literature) as Churg-Strauss Syndrome. While today it is categorised within the umbrella of eosinophilic lung diseases, it affects multiple organs and tissues and, in some cases, may not cause airway inflammation at every stage. In the West there has been no identifiable increased prevalence in either gender, with a mean age at diagnosis of 45-50 years. In Japan, however, a slightly higher prevalence in women has been observed (Sada *et al.*, 2014). The presentation of EGPA is highly variable between individuals. It is divided into three possible stages which may or may not be presented in order or not all stages will be observed. The first stage is allergic or prodromal stage, which constitutes a period of allergy including asthma and/or allergic rhinitis. Asthma at this stage is difficult to manage often requiring systemic steroids and is present for approximately 9 years before other symptoms appear. For allergic rhinitis, rhinorrhea, nasal obstruction, nasal polyps (requiring surgical intervention), and sinusitis have been observed. The second stage is the eosinophilic stage, involving abnormally high eosinophil levels typically in the lungs or the digestive tract. Symptoms at this stage include weight loss, fever, abdominal pain, gastrointestinal bleeding, and asthma. Stage three can happen concurrently to stage two and is the hallmark of EGPA, the vasculitic stage. Stage three involves inflammation of the blood vessels affecting circulation to tissues and organs. Due to the involvement of the

circulatory system life threatening complications arising from impaired cardiac function may occur.

Putting together the rather complex differences in the underlying factors of this group of eosinophilic inflammatory respiratory conditions, the diagnostic methods involve a key focus on identification of elevated eosinophils particularly in the lung as well as ruling out other causes. The key diagnostic criteria for ELD are listed below:

1. Peripheral blood eosinophilia: Elevated levels of eosinophils in the blood, typically above 500-600 cells/μL or sometimes higher depending on the specific ELD subtype.

2. Pulmonary eosinophilia: Increased eosinophils in the lungs and airways, as detected by bronchoalveolar lavage (BAL) fluid analysis or lung biopsy. The percentage of eosinophils in BAL fluid is often greater than 25%.

3. Pulmonary infiltrates: Radiographic evidence of lung infiltrates or consolidations on chest X-ray or CT scans, which may be fleeting or migratory in nature.

4. Respiratory symptoms: Patients may present with cough, wheeze, shortness of breath, chest pain, or fatigue, depending on the specific ELD subtype and severity.

5. Elevated total serum IgE levels: Elevated IgE levels are commonly observed in ELD, although not universally present in all cases.

6. Exclusion of other causes: Ruling out other potential causes of eosinophilia and pulmonary infiltrates, such as infectious diseases, malignancies, or other pulmonary disorders.

Prognosis is dependent on a five-factor score risk stratification. These include, Reduced renal function (creatinine >1.58 mg/dl or 140 μmol/l), Proteinuria (>1 g/24h), Gastrointestinal haemorrhage, infarction, or pancreatitis, Involvement of the central nervous system and Cardiomyopathy. Patients with none of the risks are categorised as having mild disease and have an 11.5% 5-year mortality. Presence of one of the risks indicates severe disease and carries a 26% 5-year mortality rate. Those with two or more risks are classed as having very severe disease and carry a 45% 5-year mortality rate.

1.3. miRNAs

1.3.1. An Overview of miRNAs

Micro ribonucleic acids (miRNAs) are endogenous, short, noncoding, single stranded, RNAs of approximately 18-25 nucleotides in length (Booton and Lindsay, 2014). In 1993, miRNAs encoded by the Lin-14 gene of *Caenorhabditis elegans* (*C. elegans*) were the first miRNAs to be discovered. Initially they were believed to be molecules specific to lower organisms (Lee, Feinbaum and Ambros, 1993). Subsequently, seven years later, the miRNA Let-7 was discovered. Let-7 was found to be conserved among bilateral animals and as a consequence it was the first suggestion that such entities could have a regulatory capacity (Reinhart *et*

al., 2000). Later that same year Let-7 was identified as being conserved in humans (Pasquinelli *et al.*, 2000).

By the year 2002, there was further evidence of miRNAs being released from cells and of them being involved in human disease, as found in chronic myeloid leukemia (CLL) (Calin *et al.*, 2002). It is now acknowledged that miRNAs are evolutionarily conserved and regulate the expression of genes and gene networks (Alipoor *et al.*, 2016). miRNAs can induce the degradation of messenger RNA (mRNA) and/or repress the translation of mRNA (Gulyaeva and Kushlinskiy, 2016). Computational analysis suggests that 60% of all protein coding genes are targeted by miRNAs (Jansson and Lund, 2012). In addition, miRNAs have a role in regulating signalling pathways in multiple cell types and in the phenotype and development of immune cells, as well as the inflammatory response in multiple tissues (Hoefig and Heissmeyer, 2008).

miRNA biogenesis begins in the nucleus and ends with mature miRNA in the RNA-induced silencing complex (RISC) (Friedman, Balaga and Linial, 2013). As can be seen from Figure 1.4., RNA polymerase II/III transcribes miRNA genes into primary miRNAs (pri-miRNAs). Approximately 50% of known miRNAs are intragenic with relatively few exons of protein coding genes, while the remaining are intergenic and transcribed independently of the host gene and have independent promoters. Families of miRNAs are so called due to being transcribed as a single long cluster sharing a seed region. There are canonical and non-canonical synthesis processes resulting in the generation of miRNAs.

For the dominant canonical pathway, the microprocessor complex, comprising Drosha, a ribonuclease III enzyme, and DiGeorge Syndrome Critical Region 8 (DGCR8 or PASHA), processes pri-miRNAs into precursor miRNAs (pre-miRNAs). DGCR8 identifies N6-methyladenylated GGAC and other motifs within the pri-miRNA, Drosha then cleaves the pri-miRNA duplex at the base of the hairpin structure of pri-miRNA. This creates a 2 nucleotide (nt) 3′ overhang on the pre-miRNA. Exportin 5 (XPO5)/Ran GTP then transports pre-miRNAs to the cytoplasm where the RNase III endonuclease Dicer and TRBP (TAR [Trans-Activation Response] RNA-Binding Protein) cleave the terminal loop leaving the miRNA duplex.

The directionality of the miRNA strand gives the suffix of the mature miRNA form. With 5p being from the 5′ end of the pre-miRNA hairpin, and 3p being from the 3′ end. Both strands have the potential of being loaded onto the Argonaute (AGO) proteins (AGO 1-4 in humans) resulting in formation of a miRNA-induced silencing complex (miRISC or RISC). Loading is done in an ATP-dependent manner (Yoda *et al.*, 2010). As currently understood, loading of the 5p or 3p strand depends on various factors including cell type or cellular environment (Meijer, Smith and Bushell, 2014). The strand with a lower thermodynamic stability is often the leading guide strand and the unloaded (passenger) strand will be unwound. Passenger strands, with strong complementarity, are cleaved and degraded by cellular machinery whilst those with central mismatches will unwind and degrade passively (Ha and Kim, 2014).

To date, multiple non-canonical miRNA biogenesis pathways have been identified. These pathways use different combinations of the key proteins or from the same family of

proteins from the canonical pathway (DICER, Drosha, Exportin, and AGO). Two of the key non-canonical pathways are independent of key proteins: Drosha/DGCR8-independent and Dicer-independent (Figure 1.4.). The Drosha/DGCR8-independent pathway generates pre-miRNAs that resemble Dicer substrates - similar to mirtrons that are produced from mRNAs during splicing (Ruby, Jan and Bartel, 2007). Dicer-independent miRNAs are processed by Drosha from endogenous short hairpin RNA (shRNA) transcripts (Yang *et al.*, 2010). With this synthesis method AGO2 is required to complete maturation. The entire pre-miRNA is able to be loaded into AGO2 due to its short length, and the 3p strand is sliced by AGO2 (Cheloufi *et al.*, 2010).

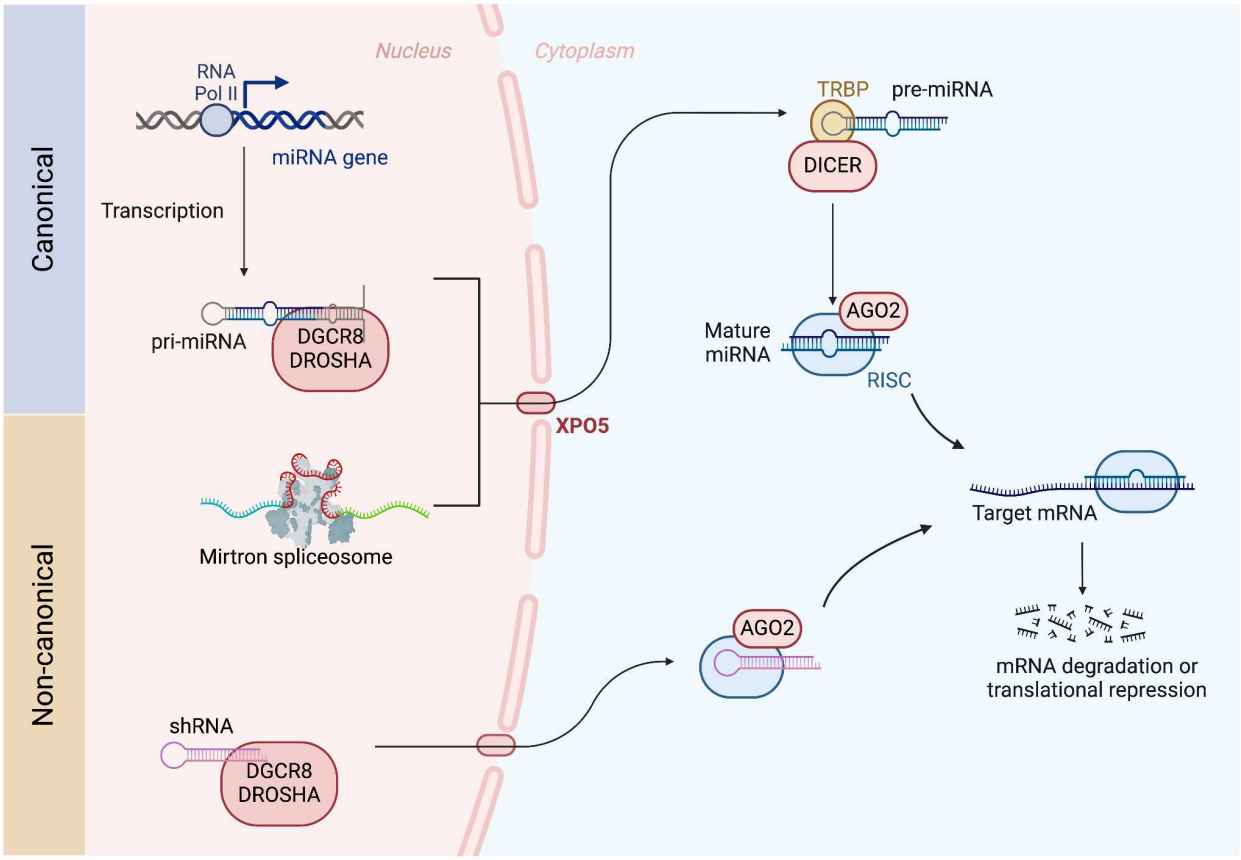


Figure 1.4. Key miRNA Biogenesis Pathways. *Canonical biosynthesis begins with RNA polymerase II/III transcribing miRNA genes into primary miRNAs (pri-miRNAs). Drosha and DiGeorge Syndrome Critical Region 8 (DGCR8 or PASHA) form the microprocessor complex that processes pri-miRNAs into precursor miRNAs (pre-miRNAs). Exportin 5 (XPO5)/Ran GTP then transports pre-miRNAs to the cytoplasm where the RNase III endonuclease Dicer and TRBP (TAR (Trans-Activation Response) RNA-Binding Protein) cleave the terminal loop leaving the miRNA duplex. Two canonical pathways are shown here. Firstly, the Drosha/DGCR8-independent pathway generates pre-miRNAs that resemble Dicer substrates - similar to mirtrons that are produced from mRNAs during splicing - and these then pass into the cytosol following the canonical pathway. Secondly, the Dicer-independent miRNAs are processed by Drosha from endogenous short hairpin RNA (shRNA) transcripts. The entire pre-miRNA can be loaded into AGO2 due to its short length, and the 3p strand is sliced by AGO2 to generate a final mature miRNA lead strand. Adapted from (O'Brien et al., 2018). Created with Biorender.com.*

The mature miRNA in RISC targets mRNAs through sequence complementarity, leading to mRNA silencing via degradation or translational repression. This necessitates complementarity of the sequences of the mRNA 3' untranslated region and the miRNA 5' region. Due to this complementarity the mature incorporated miRNA within the RISC is then able to bind and in so doing regulate expression of the gene(s) by translational repression or degradation of mRNA (Hoefig and Heissmeyer, 2008). One recent study, however, showed that protein coding sequences (CDS) containing miRNA recognition elements (MRE) may require minimal 5' base pairing and extensive 3' interaction with target miRNAs in some cases (K. Zhang *et al.*, 2018). This might indicate that there are various classes of miRNA recognising differing elements to consider. It is worth noting that each miRNA may target and regulate the expression of multiple genes, unlike silencing RNAs (siRNAs) which have a sole gene target (Friedman, Balaga and Linial, 2013).

Numerous papers over the last decade have been focused on the quantitative and qualitative assessment of miRNA expression with the identification of significant changes in expression profiles in various diseases (Gulyaeva and Kushlinskiy, 2016). The most notable changes have been observed in studies of cancer (Lu *et al.*, 2005; Raponi *et al.*, 2009; Moretti *et al.*, 2017). Such disease focused studies and their findings suggest that profiling miRNA expression may be a potential tool in both diagnosis and potential disease treatment. That said the exact mechanisms behind changes in miRNA expression, and the target effect of that change, are not yet fully understood.

1.3.2. Methods of miRNA Analysis

Early studies profiling miRNAs have been predominantly performed using Reverse transcription-polymerase chain reaction (RT-PCR) followed, more recently, by adoption of both microarray and bead-based hybridization (Luminex) platforms (Kim, 2015). RT-PCR or Quantitative-PCR (QPCR) are the gold standard for identification of miRNAs within samples as low concentrations can be identified from low input volumes. There are, however, limitations of this method and these will be further discussed later in this section.

Microarrays (for example those manufactured by Agilent) allow for high-throughput screening of miRNA expression but are limited by lower sensitivity and potential cross-hybridization (Fadiel and Naftolin, 2003). Microarrays work by having complimentary probes for tens or hundreds of specific miRNAs often organised for specific biological processes allowing the user to select a particular array e.g. inflammation, cancer (Prechl, Papp and Erdei, 2010). A significant limitation is the inability to identify novel miRNAs or miRNAs not already associated with a process with which the disease of interest may have an association. RT-PCR offers highly sensitive and specific quantification of miRNA levels. However, it has a much lower throughput (typically one test per miRNA) and can invariably be expensive hence the focus on a narrower range as opposed to a broad range of miRNAs. Bead-based hybridization, such as Luminex, combines high throughput with the sensitivity of RT-PCR but requires sophisticated equipment and complex data analysis and therefore has high associated costs (Lynch *et al.*, 2014). Each platform has distinct advantages and limitations influencing their application based on the research needs.

Newer methodologies such as Next Generation Sequencing (NGS), and Frac-seq, are being used to understand miRNA expression up to a single cell level (Martinez-Nunez *et al.*, 2018; Hücker *et al.*, 2021). NGS enables the identification of all miRNA transcripts (precursors) and mature miRNA sequences that are present within the sample tested and therefore is not biased to a particular group or set of miRNAs. Frac-seq, currently a somewhat less used technique, may provide useful information on regulatory aspects of miRNAs and in doing so enable a greater understanding of the secondary structures formed by small RNAs (Martinez-Nunez *et al.*, 2018).

There is an element of frequency bias that is being generated through miRNA research because of select miRNAs being identified as differentially expressed in QPCR. Historically, a clinical study would research a single, or small panel of miRNAs and publish the links associated (Razzak *et al.*, 2016). Subsequent research would invariably be focused on identifying the mechanisms and pathways that the particular miRNA acts upon to elicit the effect, often utilising murine models to dissect out its role (Kim *et al.*, 2017). This type of approach has worked well for gene (mRNA) analysis where large physiological effects could be generated through gene knock out, for example knock-out of the *ALPK3* gene was identified to cause cardiomyopathy in a murine model and was then linked with mutations in *ALPK3* gene mutations in humans with a cardiomyopathy phenotype (Brommage, Powell and Vogel, 2019). miRNAs are, however, responsible for fine-tuning gene expression and there is evidence that they often work in concert with other miRNAs (Friedman, Balaga and Linial, 2013). For example, deletion of miR-133a-1 or miR-133a-2 alone causes no observable effect, but deletion of both results in 50% mortality at the late embryonic stage of gestation from septal and cardiac defects (Liu *et al.*, 2008). That the effects of miRNAs are

also potentially more subtle, does not take away from the potential biological relevance that they may hold (Sevignani *et al.*, 2006).

Next generation (NG) sequencing methodologies enable miRNAs within a sample to be isolated and converted into cDNA and for all known as well as novel miRNA transcripts to be identified (Jansson and Lund, 2012). Sequencing of all expressed miRNAs within a sample could therefore provide an unbiased approach to understanding the differences between healthy controls and pathologies. As such, sequencing methodologies could be invaluable in the identification of miRNA panels for further mechanistic studies to feed back into clinical research in identifying potential diagnostic markers or therapeutic targets (Denayer, Stöhr and Van Roy, 2014).

There are some challenges and limitations, however, to consider. One of the current limitations is the small sample sizes being used to determine the selection of miRNAs that are then assessed within a larger validation cohort (Pandey *et al.*, 2018; Tian *et al.*, 2018). This is likely due to the cost of sequencing, however, with a larger cohort more sensitive fluctuations between groups may be identifiable (depending on phenotype definition), or panels of miRNAs may better represent patient populations with better precision. As yet, no paper in relation to miRNAs discussed in this Chapter or considered when reviewing the current literature was able to accurately differentiate between mild-moderate, or moderate-severe asthmatics, something that would be beneficial to clinicians in the form of a screening test (Maes *et al.*, 2016; Pandey *et al.*, 2018). Conversely, one group was able to identify a panel (not including miRNAs) that would enable patients with allergic rhinitis to be distinguished from asthmatics (Pandey *et al.*, 2018). An ideal test would not only be able

to differentiate asthma from other airway conditions, but also enable classification of asthma subtype regardless of inflammatory state at the time of testing.

Currently, one major limitation of traditional QPCR using selective primers is the lack of control miRNAs for normalization. It has been suggested that absolute quantitation using standard curves from equal input amounts of total RNA and the use of spike in exogenous miRNA should be conducted to achieve the most repeatable results. Examples in the literature of such spike-ins include the miRNAs of *C. elegans*: cel-mir-39 and cel-mir-54 (Li and Kowdley, 2012). Therefore, NGS methods demonstrate an additional level of quality of data output as controls are not essential (that said others have suggested inclusion of a spike-in may assist in normalisation of data), and the identification of novel miRNAs (with sufficient reads) is enabled as is an assessment of all present miRNAs in the sample.

Standard RNA sequencing (RNA-seq) workflows are primarily designed and optimized for the analysis of longer RNA molecules such as messenger RNAs (mRNAs), or for long non-coding RNAs (lncRNAs). mRNAs, being protein coding transcripts, vary in length between 2,500 and 5,000 base pairs and there are approximately 20-25,000 coding genes, (excluding splice variants). lncRNAs are non-coding similar to miRNAs, however, by comparison they are not as well conserved across species and are made of sequences greater than 200 nucleotides in length (with some up to 108 kB e.g. *Air*). Over 100,00 lncRNAs have been recorded (Uszczynska-Ratajczak *et al.*, 2018). The sequencing workflows may therefore not be particularly suited for the study of miRNAs, which are significantly shorter in length typically ranging from 18 to 25 nucleotides and only ~2,600 exist despite the occurrence of miRNA isomers. MiRNAs small size and limited number can

pose challenges (not only in the subsequent bioinformatics analysis) during various steps of the sequencing process including RNA extraction, library preparation, and library size selection (Buschmann *et al.*, 2016). As such sequencing pipelines must be put together based on the researcher's needs and with an understanding of the limitations of any particular model. For example, a multiple correction may overcorrect on a pool of only 2,600 possible miRNAs, where the model was originally designed for 25,000, thus generating false negative results.

Current methods of extraction have demonstrated low yields of miRNAs in plasma (Shirshova *et al.*, 2018). For miRNA research this has also been shown to be the case with one study demonstrating plasma ranges of 17-35ng/μL and serum ranges (from the same subject) between 32-108ng/μL (Wang *et al.*, 2012). It has been suggested that currently available methods of miRNA isolation may not provide sufficient yield for NGS (Qi *et al.*, 2011; Li and Kowdley, 2012). This has been shown to be untrue by Gautam *et al.* (2016) who compared multiple miRNA isolation methods from serum in a study that did result in successful NGS analysis. The Gautam study identified that although the yields of miRNA were often below the threshold for quantification, the samples were still successfully sequenced and mapped. The input of volumetric equivalents between samples may prevent true comparisons between subjects or diseases, but also may prevent too much normalization of what may already be dysregulated (Gautam *et al.*, 2016).

During library preparation, the protocols used for mRNA or total RNA sequencing may not efficiently capture and preserve the small miRNA molecules, leading to potential losses or biases in the resulting sequencing data. It is already understood that RNA integrity number

(RIN) values, used as a measure of RNA integrity, are not entirely relevant to miRNA. Therefore as a specific species of RNA a specific treatment of the data may be required (Ludwig *et al.*, 2017). Additionally the size selection step, which ideally should be employed so as to remove adapter dimers and fragments of incorrect size, can inadvertently remove a substantial portion of the miRNA population further exacerbating the potential for biases and reduced sensitivity (Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T Flåm, Leithaug, Dalland, Rayner, *et al.*, 2020).

As highlighted earlier there is then the issue of the bioinformatics tools and pipelines. For example, standard alignment algorithms may struggle to map accurately the short miRNA reads to the reference genome or transcriptome resulting in ambiguous or non-specific alignments (Zhao *et al.*, 2017). This in turn can impact downstream analyses such as differential expression analysis and target prediction.

To address these challenges, specialized miRNA sequencing workflows and bioinformatics pipelines need to be put together from available technologies and analytical programmes. These approaches often incorporate modifications to the library preparation protocols, optimized size selection strategies, and tailored alignment and analysis tools designed specifically for the unique properties of miRNA sequences.

One important aspect is the amplification step, where the adapter-ligated RNA molecules are reverse-transcribed and amplified via PCR. Standard amplification protocols may not be optimized for shorter miRNA sequences and therefore have the potential of introducing biases or inefficiencies (Li *et al.*, 2015). Adjustments to the PCR cycling conditions, primer design, or the use of miRNA-specific reverse transcription and amplification kits are a

critical consideration to ensure accurate and unbiased representation of miRNAs in the final sequencing library. To give specific details a standard small RNA cDNA library kit will have adapters and an index to attach to the miRNA (and any other RNA strands present) in the sample. These are crucial for the sequencing to occur. During the process of library synthesis adapters can join with themselves to create adapter dimers (Maqueda *et al.*, 2023). This is an artefact due to providing the reaction with a slight excess of reagents to ensure the entirety of the sample has adapters attached, and while it is something that can be minimised with optimisation it cannot be eliminated altogether. The length of the final cDNA strand of miRNA library will be only 18-25 bp longer than the unwanted adapter dimers.

Additionally, size selection strategies may need to be adapted to accommodate the small size range of miRNAs. Conventional size selection methods, such as gel or bead-based purification, may not provide sufficient resolution to accurately capture and separate the miRNA population from other small RNA species or adapter dimers (Keuls and Parchem, 2020). Alternative approaches, like the use of specialized size selection columns or electrophoretic techniques optimized for small RNA ranges, may therefore be beneficial (Munafo *et al.*, 2013).

Next generation sequencing of miRNAs is conducted on flow cell sequencing platforms that often take several tens of libraries alongside one another making the most cost-effective method of sequencing available (Campbell *et al.*, 2015). As such, unless filling a whole flow cell with miRNA library samples researchers must accept samples being combined with other users to keep costs down. Due to the short length of libraries relative to the length of

adapter dimers (a difference of approximately 20 base pairs) miRNA libraries are considered as high-risk for sequencing facilities and said facilities tend therefore to pool them with other "high-risk" samples (Novogene Ltd., Oxford. Personal Communication). One key risk of sequencing facilities implementing this strategy is that the yield of resulting data generated will be reduced or the data will be poor quality.

Other factors that categorise samples as high-risk at the pre-sequencing stage are unbalanced libraries where there are significant differences in input concentration. Some factors, such as high GC content or low diversity, cannot be identified as high-risk through standard sequencing facility quality control assessment (Tapestation or Agilent Bioanalyzer chips). miRNA libraries do also run the risk of having low diversity purely from the standpoint that the relative number of known miRNAs (~2,600) is significantly lower than the number of mRNAs (20,000-25,000 coding genes, excluding splice variants). High GC content and low diversity also impact the mechanism by which base pair calls are identified on the flow cell and this will be described in greater detail in Chapter 2.

Once sequencing data is obtained, target identification becomes a crucial step. However, the short length of miRNAs can lead to alignment to multiple genomic sequences, increasing the chance of false positive hits (Barbato et al., 2009). Further challenges arise during data analysis, particularly with potential contamination from exogenous RNA fragments, especially in samples like bronchial lavage or nasal airways. To address this, various methods are employed to identify and filter contaminant hits, such as mapping to vector databases (UniVec), and exogenous libraries available from Ensembl and NCBI.

Crucially, the accurate alignment of miRNA sequencing reads is itself a critical step in the analysis pipeline. One of the widely used aligners for small RNA sequencing data is STAR (Spliced Transcripts Alignment to a Reference) (Dobin *et al.*, 2013). While STAR is primarily designed for aligning longer RNA-seq reads, it can also be adapted for miRNA sequencing data by adjusting specific parameters.

Nevertheless, several challenges can arise when using STAR for miRNA alignment. First, the short sequence length of miRNAs can lead to an increased likelihood of multi-mapping or non-specific alignments. Additionally, the presence of sequence variations or post-transcriptional modifications can further complicate the alignment process. To address these issues, researchers may need to explore alternative alignment strategies or incorporate additional filtering steps. One approach that has incorporated the STAR aligner is the ExceRpt pipeline that was originally developed for the analysis of extracellular miRNAs. Its conservative approach to alignment and contaminant filtering makes it a valuable tool for miRNA sequencing studies, including those involving tissue or biofluid samples.

Differential expression analysis often requires normalisation as a key component before data is analysed. Appropriate normalization of miRNA sequencing data is crucial to account for technical biases and ensure accurate comparison of expression levels across samples (Tay *et al.*, 2017). The weighted Trimmed Mean of M-values (TMM) method is a widely used normalization approach that calculates scaling factors based on the assumption that most genes are not differentially expressed (Robinson and Oshlack, 2010). This assumption however may not hold true in certain cases, such as when a large proportion of genes are

differentially expressed or when there are significant compositional differences between samples. Quantile normalization, on the other hand, is a non-parametric method that forces the distribution of expression values to be the same across samples although it may introduce artifacts if the underlying distributions are significantly different (Suyundikov *et al.*, 2015). TMM is the standard normalisation strategy built into differential expression packages in R, the latter being commonly used due to its versatility.

That said, Garmire and Subramaniam identified that TMM performed poorly compared to quantile and Lowess normalisation on miRNA-seq data due to how highly conservative the processing was (Garmire and Subramaniam, 2012). As such, this may suggest the use of quantile or Lowess normalisation over the standard normalisation methodology included in the most used differential expression analysis workflows (TMM, log normalisation, or upper quartile scaling). Lowess and TMM were originally designed for the normalisation of microarray data. miRNA-seq has similarity to microarray in that the total potential genes available for alignment will be significantly smaller than that of RNAseq, whilst providing the advantage of providing an unbiased view of all present miRNAs within a sample. miRNA normalisation has, however, not typically been a consideration within the packages designed for RNA-seq read normalisation, whilst quantile and Lowess are not included as standard options within the workflow although can be conducted through separate packages during analysis.

For identifying differentially expressed (DE) miRNAs from sequencing data, three widely used statistical methods are DESeq2, BaySeq, and edgeR (Cordero *et al.*, 2012; Argyropoulos *et al.*, 2017; Lee *et al.*, 2023). DESeq2 employs negative binomial generalized

linear models with the Benjamini-Hochberg procedure to estimate significance of differential expression while accounting for biological and technical variability across samples. This approach assumes that most genes are not differentially expressed. BaySeq utilizes empirical Bayesian methods to compute likelihood distributions of expression values, enabling estimation of posterior probabilities of differential expression without making strong distributional assumptions. EdgeR also uses a Bayesian framework, as well as applying biological variability modelling using negative binomial distribution. All three methods have been successfully applied to miRNA-seq data analysis, albeit with some disagreements in the specific DE-miRNAs detected. The choice between these packages depends on the characteristics of the dataset, such as variability patterns between libraries, presence of outliers, and the assumptions that can be reasonably made. For example, EdgeR and BaySeq are well suited to handle comparing libraries with variance in sequencing depth due to accounting for technical and biological variability within the data, whilst DEseq is likely the most sensitive at handling overdispersion.

As highlighted earlier miRNAs are understood to exert their regulatory effects by binding to the 3' untranslated regions (3' UTRs) of target messenger RNAs (mRNAs), leading to translational repression or mRNA destabilization (Booton and Lindsay, 2014). Identifying the putative target genes of differentially expressed miRNAs is a crucial step in unravelling their functional roles and deciphering the underlying biological pathways they modulate. Several computational algorithms, such as TargetScan, miRanda, and DIANA-microT, have been developed to predict miRNA-target interactions based on sequence complementarity and conservation across species (Lin *et al.*, 2015; Ronchetti *et al.*, 2016; X. Ma *et al.*, 2023).

Algorithm choice aside, *in silico* predictions often result in large numbers of potential targets and further validation is necessary to ensure biological relevance. One approach to prioritize and interpret the predicted targets is through Gene Ontology (GO) enrichment analysis (Huntley *et al.*, 2016). The GO database provides a structured vocabulary for describing the functions of genes across three domains: Biological Processes, Molecular Functions, and Cellular Components. Molecular Function (MF) identifies the biochemical or molecular activity of the gene product; Biological Process (BP), corresponds to the wider biological area in which the gene product's MF acts; and Cellular Component (CC), which is the specific cellular location in which the gene product is active (Christophe Dessimoz, 2017).

The inclusion of specific records in the GO enrichment analysis can be tailored based on the study objectives and the level of confidence in the predicted targets (Pavlidis *et al.*, 2012). Researchers may choose to include only experimentally validated targets or apply stringent criteria based on the prediction scores or conservation levels (Peng *et al.*, 2017). Alternatively, a more inclusive approach, incorporating computationally or sequencing result predicted targets, can be employed to capture a broader landscape of potentially relevant pathways albeit with a higher risk of false positives. It has been suggested that filtering pathways by particular evidence codes may bias the analysis toward well-studied Biological Processes (Reimand *et al.*, 2019). This could be viewed as a strength when considering miRNAs as many are not well enough studied to have sufficient supporting evidence. Understanding how they might link to better known processes could be an effective method in developing a better understanding of the pathway with the highest confidence in the analysis.

Integrating miRNA target analysis with GO enrichment provides a comprehensive framework for elucidating the functional implications of dysregulated miRNA expression in various biological contexts, including respiratory diseases like asthma and the eosinophilic lung diseases. With some caveats, the GO domains can provide a comparison with expected pathways in the interest of identifying novel processes that might not have been considered. Combining information across domains (BP, MF, and CC) can help elucidate the interactions between different gene products (Christophe Dessimoz, 2017). For example, knowing that two proteins interact physically (Molecular Function) and are both localized to the cell membrane (Cellular Component) can suggest that they might be involved in the same Biological Process. The genes that were shared across the most domains are understood to be fundamental to the most processes within an organism, from cellular machinery to regulatory networks, thus identifying a potential hub of proteins from which future studies can develop from. In addition, looking for the most interactions between number of source miRNAs and gene targets and how that could contribute to an understanding of the endotype related to the phenotypes of asthma severities. GO analysis for miRNA analysis can currently be conducted on miRNet.ca. This platform enables miRNA-centric network visual analysis to be conducted. Query lists of miRNAs (e.g. DE-miRNAs) can be uploaded and filtered based on various requirements and visually inspected for connections as well as GO analysis. The results can then be downloaded and further analysed and graphs generated in R (Chang *et al.*, 2020).

For validation of miRNA expression QPCR is viewed as the gold standard and similar to when validating gene (mRNA) expression. QPCR requires inclusion of a reference with which to normalise other RNAs of interest. The core criteria when selecting said reference

is that it should have stable expression across all samples, and have a similar stability with target miRNAs, regardless of sample storage conditions (Kozera and Rapacz, 2013). There is an argument that they should be the same class of RNAs (Wong LLK, Lee K, Russell I, Chen C., 2010). Historically, QPCR enables analysis of relative gene (mRNA) expression through absolute quantification and relative quantification. Absolute quantification determines the copy number, often in relating the signal to a standard curve. Relative quantification relates the signal to a target transcript in the treatment group to that in the untreated or Control group, for example. One popular method enables analysis of relative changes in gene (mRNA) expression is using threshold cycle (CT), the ΔΔCt, which uses a reference gene to give relative quantities between a reference or control group and the test group of genes of interest (Livak and Schmittgen, 2001).

Multiple studies have demonstrated that the best choice of reference gene in the analysis of the relative miRNA expression level is either a miRNA or the geometric mean of several stably expressed miRNAs, previously other RNAs have been used for normalisation such as U6 or RNU6 (Vandesompele *et al.*, 2002). The most popular algorithms used for expression stability assessment are geNorm, NormFinder and Best Keeper (De Spiegelaere *et al.*, 2015; Anauate *et al.*, 2017; Xie, Wang and Zhang, 2023). While these were originally designed for use for gene stability analysis they have been applied to miRNA research, finding stably expressed miRNAs or other surrogate RNAs for the purpose of normalisation. One study, using Normfinder and GeNorm, identified that neither U6 and RNU6 were stable, while two miRNAs were identified as the most stable (Lamba *et al.*, 2014). This supports suggestions from Thermo Fisher, that RNA class may be important to reference selection (Wong LLK, Lee K, Russell I, Chen C., 2010).

It has been suggested that miRNA sequencing results are reliable enough that validation with QPCR may not be required where experiments conducted were of high quality and with sufficient replicates (Hughes, 2009). However, in the case of validation of identified differences between test groups, QPCR would be warranted (Coenye, 2021).

1.4. Practical Considerations

There are several challenges in allergy research including understanding the inciting events, the identification of the regulation of the allergy, defining the cellular interactions that are driving the response, and developing novel therapeutic strategies of treatment (Pua and Ansel, 2015). miRNA studies can be used to address fundamental questions around type 2 immunity, as there is an apparent robust effect on immune responses seen in other pathologies (Lindsay, 2008).

The study of miRNA expression in the respiratory tract has been challenging due to the difficulties in obtaining relevant samples. Traditional methods, such as bronchoscopy, are invasive and not clinically viable for ongoing assessment of the condition at a cellular level. Bronchoscopy, while providing direct access to the lower respiratory tract, involves inserting a bronchoscope through the nose or mouth into the lungs, which can cause discomfort, bleeding, and potential complications like infection or pneumothorax (Leiten *et al.*, 2016). This invasiveness makes it unsuitable for repeated sampling, especially in vulnerable populations such as paediatric or elderly patients, or those with severe respiratory conditions.

Additionally, bronchoscopy requires specialized equipment and trained personnel, making it logistically and financially burdensome. The need for hospital settings and potential sedation further complicates its use for longitudinal studies or large-scale screenings. The risk-to-benefit ratio is often unfavourable for patients with chronic conditions requiring regular monitoring.

Alternative methods, such as sputum induction, nasal swabs, exhaled breath condensate collection, and blood sampling (liquid biopsies), offer less invasive options but come with their own set of challenges. Sputum samples can be difficult to produce, especially in patients with dry coughs or poor mucus production, and may not always represent the lower respiratory tract accurately (Bakakos *et al.*, 2011). Exhaled breath condensate collection, while promising, often yields dilute samples with miRNA concentrations that are difficult to analyze reliably, although these technologies are advancing (Mitchell *et al.*, 2024). Blood sampling, or liquid biopsies, has emerged as a valuable tool for non-invasive monitoring of miRNA expression (Ramshankar and Krishnamurthy, 2013).

While it is less invasive than bronchoscopy, blood sampling still involves needle insertion, which can cause discomfort and anxiety in some patients. Additionally, circulating miRNA levels in the blood may not directly correlate with miRNA expression in specific tissues like the lungs, potentially complicating the interpretation of results (An *et al.*, 2020). Furthermore, variability in sample quality and quantity, along with potential contamination from the upper respiratory tract and oral cavity in the case of certain sample types, can complicate data interpretation (Li *et al.*, 2021).

One relatively new alternative to sampling of the airways is Nasosorption. Nasosorption is a novel, minimally invasive technique that allows for the sampling of the nasal mucosal lining fluid. This method involves the gentle absorption of the nasal fluid onto a synthetic absorptive matrix (SAM) which is then processed for downstream analysis.

While nasosorption has been previously used for the analysis of proteins and other biomolecules (Thwaites *et al.*, 2017, 2018), its application in miRNA profiling has not been extensively explored. In a study on Respiratory Syncytial Virus (RSV) in infants comparing nasosorption with nasopharyngeal aspirates (NPA), it was identified that nasosorption was well tolerated and was effective at detecting the stage of viral load between patients as well as other small molecule measures of inflammation. In contrast, NPA was not able to identify viral load between the patients tested (Thwaites *et al.*, 2017, 2018), 2018). Nasosorption was able to identify significant differences between RSV+ or RSV- subjects with regard to levels of the proteins interferon gamma (IFNγ), interleukin-1 beta (IL-1β), Chemokine (C-C motif) ligand 5 (CCL5), and interleukin-10 (IL-10), and measures were in agreement with those obtained from NPA. This suggests that nasosorption could be a sufficiently sensitive method for the sampling of miRNAs in the nose and potentially be a proximal surrogate for the airways.

Overall, the challenges of obtaining relevant respiratory samples for miRNA expression studies highlight the need for developing minimally invasive, reliable, and reproducible sampling methods. Innovations in this area could significantly advance our understanding of respiratory diseases and enhance the clinical applicability of miRNA-based diagnostics and therapeutics.

1.5. Asthma and Eosinophilic Lung Diseases: Associations with miRNAs

1.5.1. miRNA Expression in Asthma

Research into the involvement of miRNAs in asthma and links with other inflammatory and allergic respiratory conditions is in its infancy. In addition, the research conducted is predominantly via careful selection of specific miRNAs for analysis from previous experiments or literature searches and testing through QPCR and microarrays. miRNAs have been established as one of the post-transcriptional regulators that are important in the maintenance of immune homeostasis (O'Connell *et al.*, 2010).

In steroid insensitive patients lung function assessments have been shown to poorly correlate with airway inflammation (Tonga *et al.*, 2018). Subjective measures such as self-perception of a patient's asthma has been highlighted as a poor correlation with lung function parameters, and as such, though an important factor to consider for clinical work, it may not provide useful information towards an objective measure of disease (Steele *et al.*, 2012). Arguably, this may be because current objective measures do not marry up with perception but with more sensitive analysis it may not always be the case. The use of miRNA expression may be able to identify the differences between patients with steroid insensitivity and identify potential drivers of the resistance to treatment.

Evidence to support the role for miRNAs in allergic disease has been shown in experimental animal models of airway obstruction and allergic inflammation (Phipps *et al.*, 2009). A house dust mite (HDM) murine model was able to identify that Toll like receptor 4 /myeloid differentiation primary response 88 (TLR4/MyD88) signalling was associated with upregulation of miR-126 (Phipps *et al.*, 2009). The antagonism of miRNA function using antagomirs, directed towards miR-126, was then applied to the murine model and resulted in ablation of HDM-induced AHR, attenuation of immune cell recruitment to the lung (eosinophils and neutrophils), and a reduction in mucus hypersecretion (Mattes *et al.*, 2009). The antagonism of miR-126 also resulted in reduced production of T_H2 inflammatory cytokines IL-5 and IL-13 (Mattes *et al.*, 2009). Collinson *et al*. showed similar findings for miR-145 (Collison *et al.*, 2011) suggesting that inhibition of miRNAs could attenuate allergic responses. Within the same study the authors showed a comparable therapeutic effect of antagomirs compared to systemic glucocorticoid therapy. Both miR-126 and miR-145 target multiple mRNA targets which may mean that multiple pathways are affected to produce the anti-inflammatory effects observed (Fish *et al.*, 2008; Oglesby *et al.*, 2010; Zhu *et al.*, 2016). miRNAs are somewhat unique in their ability to target and regulate more than one gene, as shown by a study in gastric cancer where it was found that miRNAs have a multiple-to-multiple relationship with target genes (Hashimoto, Akiyama and Yuasa, 2013). Overexpression of two pairs of miRNA combinations (miR-224 in combination with miR-425, and miR-181 in combination with miR-340) downregulated their target genes *DPYSL2* and *KRAS*, and *KRAS* and *MECP2*, respectively (Hashimoto, Akiyama and Yuasa, 2013).

miR-126 has been identified as a regulator of vascular endothelial cell adhesion molecule 1 (VECAM1 – another protein involved in asthma pathophysiology), as decreasing levels of miR-126 resulted in increased VCAM1 expression on endothelial cells improving leukocyte adherence, and as such is associated with regulating leukocyte trafficking (Harris *et al.*, 2008). miRNAs may be directed towards the endothelium to reduce leukocyte trafficking or towards the airway lumen to instruct macrophages or localised epithelial cells in a particular inflammatory pathway. Circulating miRNAs and airway mucosal miRNA signatures may therefore differ significantly.

miR-21 has also been researched and shown to have association with asthma, particularly in experimental models it is one of the most upregulated miRNAs. IL-12 is known to drive a T_H1 response which may be beneficial in moving away from T_H2 inflammation. As the IL-12p25 3'UTR contains a highly conserved target sequence for miR-21 there has been some speculation that, in asthma, this may be key in the drive towards T_H2 polarisation by way of reducing the IL-12 T_H1 drive. Other studies support this T_H2 driving theory. In miR-21-deficient-CD4+ cells, although no change in IFNγ expression was observed attenuation in IL-4 resulted in a T_H2 response (Lu *et al.*, 2011). Upregulation of miR-21 has been identified in other T_H2 associated conditions including atopic dermatitis, eosinophilic esophagitis, and ulcerative colitis (Specjalski and Niedoszytko, 2020). Airway remodelling is one of the key changes of asthma pathology and miR-21 has been shown as a potential contributor to this process. In a murine model, blocking miR-21 expression resulted in attenuation of airway remodelling through the Transforming growth factor beta - Small Mothers Against Decapentaplegic 7 (TGFβ - SMAD7) pathway (Hur *et al.*, 2021). This suggests that overexpression of miR-21 in asthma could be a contributor or driver of

airway remodelling. This is in turn is supported by clinical findings where miR-21 (from serum) was identified as being more highly expressed in asthmatics compared to controls (ElKashef *et al.*, 2021).

Put together the current research suggests that miRNA dysregulation in asthma could be contributing to key features of asthma such as airway inflammation through promotion of immune cell recruitment to the airway, airway remodelling, mucus hypersecretion, airway hyperresponsiveness, and, in the case of severe asthma, steroid insensitivity.

1.5.1.1. Asthma Population Heterogeneity and miRNA Expression

As highlighted earlier (Section 1.1.5.), asthma is a very heterogeneous disease and this is an important consideration when conducting molecular studies including those of miRNA. miRNA studies where attempts to categorise patients into subpopulations of asthmatics have been made are the studies of Maes *et al*. (Maes *et al.*, 2016) and Panganiban *et al*. (Panganiban *et al.*, 2016). The Maes *et al.* study involved RT-PCR profiling of 755 miRNAs in sputum supernatants from healthy subjects (N = 10), mild asthmatics (N = 17) and severe asthmatics (N = 9). Only three miRNAs, miR-629-3p, miR-223-3p, and miR-142-3p, were found to be significantly differentially expressed between the healthy subjects and those with severe asthma. The three miRNAs identified were found to be negatively associated with lung function parameters (FEV1/FVC ratio and FEV). Using *in situ* hybridization the authors went on to show that miR-629-3p was specific to epithelial cells whilst miR-223-3p and miR-142-3p were localized to neutrophils. Interestingly, the authors found no miRNAs

associated with eosinophilia or T_H2 responses in the study cohort. Perhaps the issue currently is that relatively few miRNAs have sufficient published data to connect with airway processes. Mimetics are double stranded RNA molecules designed to mimic endogenous miRNAs, resulting in downregulation of target mRNA translation. Functionally, they replicate native miRNAs with guide and passenger strands allowing AGO loading and mRNA targeting (Wang, 2011). Although the mimetics were able to identify individual roles in regulating inflammatory processes, the miRNAs that were tested are unlikely to be in isolation in humans, and their promiscuity as molecules with their targets, might cause the bigger picture to be missed (Wang, 2011). It is plausible that groups of miRNAs act together to create effects. It has already been shown that miRNA mimics alter physiological miRNA effects at immune checkpoints by triggering alternate pathways in a non-specific manner (Prinz *et al.*, 2022).

The study of Panganiban *et al*., (2016) (Panganiban *et al.*, 2016), using an array approach, assessed the expression levels of 135 miRNAs in plasma from patients with allergic asthma and healthy individuals (N = 12) resulting in 30 miRNAs exhibiting differential expression being identified. The authors then validated the findings in a larger cohort (N = 79), this time to try to categorise between healthy individuals, asthmatics, and sufferers of airway rhinitis (AR) (Panganiban *et al.*, 2016). The results clustered into 5 distinct groups, Groups 1 and 3 showed miRNAs that were common to AR and asthma, which might be indicative of similarities in pathways of inflammation. Group 2 was specific to patients with asthma whilst Group 5 showed greater dysregulation in AR compared to asthmatics with Group 4 being specific to patients with AR. This was the first study to demonstrate that circulating

miRNAs could be used to identify asthma and differentiate between other airway diseases, which was done with an accuracy of 92.4% using machine learning algorithms.

The two studies contrasted in their findings, and the top 6 predictive miRNAs in the Panganiban *et al.* study showed only one match with the Maes *et al.* study - miR-223-3p which is exclusive to neutrophils (Maes *et al.*, 2016; Panganiban *et al.*, 2016). miR-145 identified in the Panganiban study (group panel exclusive to asthmatics) was previously reported in the Collinson *et al*. study to be associated with asthma as well as a more recent study in which Panganiban is also author (Collison *et al.*, 2011; Maes *et al.*, 2016; Panganiban *et al.*, 2016; Lambert *et al.*, 2018).

1.5.1.2. miRNA Changes and Regulation of Inflammatory Pathways

Further advances in understanding miRNA regulation of mRNAs has been through application of subcellular fractionation with high throughput RNA sequencing (Frac-seq). The technique has been used previously to identify links between alternative splicing and how translational regulation of mRNA isoforms is affected (Sterne-Weiler *et al.*, 2013). In a recent study involving a genome wide post-transcriptional study in bronchial epithelial samples from severe asthmatics (and healthy controls), 20 miRNAs were found to be differentially expressed between cytosolic and polyribosome associated fractions. miRNAs were cross referenced using TargetScan 7.0 (2018 release Section 1.3.2. above) for predicted miRNA targets in mammals. It was suggested by the authors that as there was an observed overlap of targets in each fraction that miRNAs in asthma may have more of an

impact on protein translation than on mRNA expression alone (Martinez-Nunez *et al.*, 2018).

Further assessment by Martinez-Nunez *et al*. identified that only 8 miRNAs controlled most of the dysregulation of cytosolic and polyribosome bound mRNA in severe asthmatics, and that 6 of those the authors validated in a larger patient cohort. The 6 miRNAs (miR-22-5p, miR-148a-3p, miR-342-3p, miR-495-3p, miR-543, and miR-197-3p) had the highest number of targets in both fractions, and two key pathways were identified: 'inositol compounds' and 'involved in the cell cycle'. The 6 miRNAs mentioned above were responsible for 89% of the regulation of polysome targets. This work demonstrates the value of integrating multiple omics datasets, and genome-wide dysfunction in the translation of mRNAs encoding structural and inflammatory factors in bronchial epithelial cells in severe asthmatics.

The findings of the study above are promising, although some researchers have raised concerns that less severe asthma may not be differentially identifiable due to small effect sizes. For example in the Martinez-Nunez *et al*. study, between the most dysregulated miRNAs there was only a very small difference: 0.0479 being the smallest effect size of expression of miR-22-5p between healthy controls and asthmatics, compared to 0.0038 for miR-543 under those conditions. This, therefore, may make differentiating patients difficult. With regard to the development of a picture of the relationships between miRNAs and the mRNAs that they regulate, this method of Frac-seq however appears robust.

1.5.2. miRNA Expression in ELD

Apart from the Cottin et al., paper discussed earlier (Section 1.2.) there have been few papers published regarding miRNA expression in ELD. To date, none have attempted an unbiased sequencing approach. Using other methodologies, one study considering the ELD variant ABPA identified differential expression of miR-132 in QPCR analysis of RNA isolated from whole blood had potential as a diagnostic marker of the condition between cases and controls (Esawy *et al.*, 2021).

1.5.3. Impact of Corticosteroids on miRNA Processes in Inflammatory Lung Conditions with a Focus on Asthma and ELDs

Corticosteroids are a key treatment method of managing inflammation and reducing the incidence of exacerbations in both asthma and eosinophilic lung disease (Chapter 1. Section 1.1.4.). Steroids are known to effectively switch off multiple inflammatory genes that have been activated during the inflammatory process. These include: cytokines, chemokines, adhesion molecules, inflammatory enzymes, receptors, and proteins (P. J. Barnes, 2006). Their efficacy in reducing inflammation is well-documented (Barnes and Pedersen, 1993; Zoorob and Cender, 1998; Peter J. Barnes, 2006). With an understanding that miRNAs act to regulate mRNA expression, an alteration of miRNA regulation through ongoing

pharmacological interventions for airway inflammation could then impact not only the expression of miRNAs but also the target mRNAs. Evidence of the alteration in miRNA expression has been found in a small number of studies (Solberg (Solberg *et al.*, 2012)(Li *et al.*, 2020), with only one having addressed the changes in ELDs specifically (Esawy *et al.*, 2021). One more broadly related study, in asthma, identified that steroid-naive subjects and steroid-using subjects both had abnormal bronchial epithelial miRNA expression patterns in microarray analysis (Ns of 217 and 200 respectively) compared to healthy controls (Solberg *et al.*, 2012). Treatment of steroid-naive patients with inhaled corticosteroids only changed significantly nine of the 217 differentially expressed miRNAs. This suggests corticosteroid treatment (in the short term at least) only modestly corrects abnormal miRNA expression.

A study by Li and co-workers identified that circulating miR-155-5p was associated with an increased response to dexamethasone treatment in a childhood asthma cohort (Li *et al.*, 2020). *In vitro* testing identified that miR-155-5p was involved in inhibiting the trans-repressive effects of corticosteroids on IL-1β. There may be other interactions occurring *in vivo* to explain this effect. With the little research available, it appears the interactions of corticosteroids and miRNA regulation of mRNAs in asthma are complex and the underlying processes regulating these factors are not well understood.

1.6. Hypothesis and Aims

In this thesis, three research studies will be described: the Nasosorption miRNA Study, the Celtic Fire miRNA Study and the Bronchoscopic evaluation of the Eosinophilic Airway Microbiome (BEAM) miRNA Study.

The Nasosorption miRNA Study was impacted by COVID and the focus of the thesis had to pivot. Samples and metadata were therefore sourced from two studies Celtic Fire and BEAM (that had been conducted by the Genomic Medicine Group, NHLI [McBrien, 2020]). This provided an opportunity to investigate circulating miRNA expression from a select set of asthmatic subjects and healthy controls, and biopsy and circulating miRNA expression from a group of subjects with ELDs together with a representative group of healthy control subjects.

Hypothesis for the Original Thesis Research Project- Nasosorption miRNA Study

The acute asthmatic response (that underpins asthmatic airway pathology) is influenced by the presence and interactions of epithelial cytokines (IL-33 and TSLP), miRNAs and the airway microbiota.

Aims

1. Develop reliable methods of respiratory miRNA quantification from Nasosorption samples and apply these methods on samples (both SAMs and oropharyngeal

[throat] swabs), obtained via subject recruitment, to define unbiased signatures of miRNA characteristic of the normal healthy and asthmatic airways.

2. Assess the inter-relationships between epithelial cytokines (IL-33 and TSLP), type II cytokines (including IL-5) (measured by MesoScale Discovery assays), the airway microbiome (determined by 16S QPCR and amplicon sequencing of DNA extracted from SAMs and oropharyngeal [throat]swabs) and miRNAs to the clinical features of asthma.

Hypothesis Post COVID: Celtic Fire and BEAM miRNA studies

The acute asthmatic response (that underpins asthmatic airway pathology) is influenced by the presence and interactions of miRNAs.

There will be a difference in miRNAs expressed in the lung compared to peripheral blood.

Aims

1. Conduct a pilot study to define (through implementation of next generation sequencing) unbiased signatures of circulating (blood) miRNA characteristic between Healthy Controls and asthmatic patients (Celtic Fire Study).

2. To identify differentially expressed (DE) miRNAs in whole-blood samples across a broader cohort of asthma severities: Healthy Controls, Non-Severe asthmatics, and Severe asthmatics (Celtic Fire Study).

3. To explore the potential target genes and associated biological pathways of the identified DE-miRNAs, providing insights into the molecular mechanisms underlying asthma pathogenesis and severity.

4. To compare and contrast the findings from the two studies and other relevant published literature, highlighting differences and similarities in expression targets or patterns.

5. To compare miRNA expression profiles between whole blood and lung biopsy samples from ELD patients and Healthy Controls (BEAM study) and how these compare with clinical parameters, such as eosinophil counts, IgE levels, and lung function measures including exploration of potential target genes and associated biological pathways.

6. To assess the feasibility of spike-in miRNA as a normalisation tool with low input samples (BEAM study).

2. Chapter 2: Methods and Materials

2.1. Methods

2.1.1. Clinical Sample Collection

2.1.1.1. Nasosorption Sample Collection

Nasosorption was conducted as previously described (Thwaites *et al.*, 2018). In brief, Nasosorption samples were collected using a specially designed sampling device (Figure 2.1). The device (Hunt Developments UK Ltd) consists of a flat quilted Synthetic Absorptive Matrix (SAM) which is attached to a thin filament of plastic with a manual applicator and screw fitting that fits into a sterile holder. Figure 2.1. below is a photograph of the Nasosorption device showing its features.

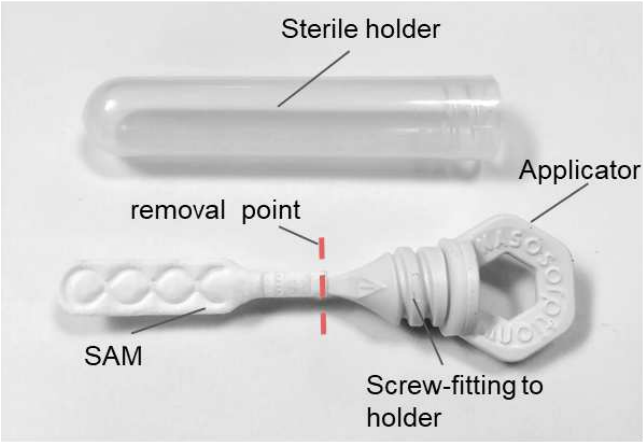


Figure 2.1. Photograph of the Nasosorption Device. *Applicator and sterile holder come in a single, sterile, sealed package. The applicator has a screw fitting allowing it to be unscrewed from the holder just prior to usage. The SAM is a synthetic absorptive matrix that is flat and wide (shaped like a cricket bat). Shown here is the flat side of the SAM with the matrix fixed to the applicator which extends from the screw fitting.*

The SAM was put into the nostril under direct vision and placed against the inferior turbinate. The patient then pressed their nostril closed with their index finger. The SAM was kept in place for 60 seconds exactly. The SAM was then gently removed and replaced back within its sterile holder with engagement of the screw fitting. The device was then placed on ice for a maximum of 30 minutes prior to the elution step.

Nasosorption samples were collected in duplicate from the left and the right nostril from each subject who had provided signed informed consent to participate in the study. Ethics details are provided in Chapter 3, Section 3.2.1. Within 30 minutes of collection, each SAM sample was removed from its holder and cut with sterile scissors at the removal point and placed into a Costar Spin-X centrifuge tube insert (without membrane) (Costar 9301, Corning Life Sciences Ltd) which was then placed into a microcentrifuge tube (Costar 3213, Corning Life Ltd). The same centrifuge (Labnet Prism R microcentrifuge) was used

throughout the isolation procedure unless otherwise stated. Next, a number of different RNA stabilization reagents were investigated (with a 200 µL volume added to the SAM in each case) including equal parts of RNAlater (Sigma-Aldrich, Darmstadt Germany) and ice cold phosphate buffered saline (PBS, Thermo Fisher, USA), RLT buffer (Qiagen, Hilden, Germany), RNAprotect (Qiagen), and RNA shield (Zymo Research, California, USA) (see Chapter 3 for full details of reagents tested). Irrespective of reagent added, post addition the sample was vortexed briefly before being centrifuged at 1,500 g for 20 minutes at 4℃ allowing a pellet to form. Post centrifugation and using a Gilson pipette to minimise disturbance of the pellet, the supernatant was transferred into a fresh sterile microcentrifuge tube and then centrifuged at 16,500 g for 20 minutes. The subsequent supernatant obtained was transferred into a fresh screw cap (with O-ring) 2mL tube (Thermo Scientific, Biobanking and Cell Culture Cryogenic Tubes) and frozen at -80℃ until miRNA extraction.

2.1.1.2. Blood Sampling

Blood samples were obtained as part of the Celtic Fire study and details of ethics and inclusion and exclusion criteria are detailed in Appendix: Figure 3. Blood samples for RNA extraction were collected by standard venous sampling into PAXgene® Blood RNA tubes (BD Biosciences, USA). Intracellular RNAs degrade within hours after blood collection but PAXgene tubes contain a proprietary reagent to prevent the degradation and stabilise RNA species as well as preventing, upon blood collection, gene induction. This was considered important for the sequencing results to best represent the *in vivo* expression profile.

Samples in PAXgene tubes were allowed to incubate at room temperature for 2 hours to achieve blood cell lysis (as per manufacturer's guidelines (Pre-Analytix, 2020). Samples were then stored at -80℃ until miRNA extraction.

2.1.2. RNA Extraction Methods

For the process of RNA extraction to include the very smallest RNA species, such as miRNAs, methods have been developed using proprietary combinations of salts that have ethanol and other reagents added in the laboratory. As with all RNA extraction protocols, RNases were minimised through laboratory hygiene protocols the latter also ensuring contamination sources of RNA and DNA, that might impact outcomes of downstream experiments, were also minimised. Sources of RNases are present in dust, skin, and moulds (Nilsen, 2013).

RNases pose a threat to the integrity of the RNA within the sample. RNases are often difficult to eliminate from surfaces as some can withstand autoclaving and are resistant to metal chelating agents. Prevention therefore is crucial. While there are often DNase steps added to protocols to degrade DNA, when performing RNA extraction it is recommended to ensure all surfaces and equipment to be used have been treated with RNAzap followed by a 70% ethanol wash to reduce the risk of RNase contamination of surfaces. A dedicated set of Gilson pipettes were used exclusively for RNA isolation work and were also cleaned (RNAzap and 70% ethanol wash). As well as these preventative measures, the use of sterile pipette tips with filters, and sterile RNase free (nuclease free) water were used (purchased from Sigma-Aldrich). Sterile eppendorfs and other single use nuclease free tubes (collection tubes, silica filter tubes) when not in use were kept in sealed packaging to reduce the risk of

environmental contaminants. Any plastic consumables used and not provided within kits, were also DNase and RNase free and stored under the same conditions.

2.1.2.1. Zymo RNA Extraction

The Zymo Total RNA (including miRNA) extraction kit was used (Cat#R1054). For liquid samples stored in RNAlater (Nasosorption samples), an equal volume of nuclease-free water was added to the sample. This was then followed by the addition of four volumes of Zymo RNA Lysis Buffer, with thorough mixing by flicking the tube, followed by the addition of one volume of 100% ethanol (Sigma-Aldrich, Darmstadt Germany) with mixing by pipetting (Gilson pipette and sterile pipette tips with filters). The sample was then transferred to a Zymo spin column in a collection tube and centrifuged at 20,000 g for 1 minute. The flow-through was discarded as the RNAs had been captured through binding to the silica filter of the column.

Next, with the RNAs bound to the silica filter, treatment with a DNase inactivator was performed by addition of a mixture of 5 µL DNase + 75 µL DNA digestion buffer mix as per the kit protocol. The RNA sample was then incubated at room temperature for 15 minutes.

Next the RNA sample was washed to remove proteins and other impurities by the addition of 400 µL RNA prep buffer to the column before centrifugation with the flow through discarded. This was then repeated twice more, first with 700 µL and then 400 µL of RNA prep buffer. The column was then centrifuged on full speed for one minute with the lid of the outer Eppendorf (that the column sits in) closed and another with the lid open to completely dry the column of any remaining ethanol (the latter being a component of the

RNA prep buffer). RNase free water was heated to 50 ℃ and 30-50 µL added to the column, the latter being placed into a fresh Eppendorf tube for final RNA collection. The column was incubated at room temperature for 1 minute before being centrifuged for 1 minute at full speed 20,000 g.

2.1.2.2. Monarch RNA Extraction

The Monarch Total RNA (including miRNA) extraction kit was used (Cat# T2010S). For liquid samples (Nasosorption) stored in RNAlater, an equal volume of nuclease-free water was added to the sample. This was followed by the addition of four volumes of RNA Lysis Buffer which was then mixed thoroughly by flicking the tube.

Next up to 800 µL of the sample to be extracted was transferred to a genomic DNA (gDNA) Removal column equipped with a collection tube. The collection tubes were labelled since the gDNA removal columns were destined for disposal after the centrifugation step. The sample was centrifuged for a 30 second spin at 16,000g to eliminate the majority of the gDNA. The flow through was retained and the gDNA column discarded.

An equal volume of ethanol (100 %) was added to the flow-through and mixed by pipetting. This addition of ethanol is crucial as it creates favourable conditions for RNA to bind to the RNA Purification column. Next, the mixture was carefully transferred to an RNA Purification column fitted into a collection tube. A 30 second centrifugation at 16,000g was carried out, followed by the disposal of the flow-through as the RNA was now bound to the matrix within the column.

A wash step was carried out to remove salts and contaminants prior to DNase treatment. In brief, 500 µL of RNA Wash Buffer was added and a 30 second centrifugation at 16,000g performed after which the flow-through was discarded. Then a mixture of 5 µL DNase I and 75 µL DNase I Reaction Buffer was prepared (per sample) in an RNase-free microfuge tube and pipetted directly onto the top of the matrix. The mixture was then incubated for 15 minutes at room temperature before the addition of 500 µL RNA Priming Buffer followed by a 30 second centrifugation after which the flow-through was discarded.

A wash with 500 µL RNA Wash Buffer, followed by a 30 second centrifugation at 16,000g and the disposal of the flow-through was then performed followed by a second wash with 500 µL RNA Wash Buffer and a 2 minute centrifugation at 16,000g. Following this the column was transferred to a RNase-free microfuge tube, ensuring that the tip of the column did not come into contact with the flow-through during the transfer.

Finally, 30 µL of Nuclease-free Water was directly added to the centre of the column matrix. A 30 second centrifugation at 16,000g was performed and the eluted RNA sample was transferred into a labelled cryotube and stored at -80℃ prior to quantification and downstream experiments.

2.1.2.3. PAXgene RNA Extraction

For the PAXgene blood samples, extraction was conducted using the PAXgene blood RNA extraction kit (QIAGEN) following the PAXgene blood miRNA extraction protocol.

In brief, the PAXgene tube was allowed to thaw at room temperature before centrifugation for 10 minutes at 3,270g using a swing out rotor (Allegra X-12 R, Beckman Coulter, USA) to

pellet the lysed sample. The supernatant was carefully removed by pipetting and a wash performed by resuspending the pellet in 4 mL of DNase/RNase free water after which the sample was centrifuged under the previous settings with a fresh Hemogard® lid. The supernatant was removed once more by pipetting after which 350 µL of loading buffer (BM1) was added and the pellet resuspended and transferred to a fresh 1.5 mL centrifuge tube (Thermo Scientific, Biobanking and Cell Culture Cryogenic Tubes). Protein lysis was then conducted by adding 300 µL Buffer BM2 and 40 µL Proteinase K to the sample followed by incubation in a shaking incubator (Eppendorf Thermomixer C) at 55℃ at 700 rpm for 10 minutes. The whole sample was then transferred into a shredder column placed in a 2 mL processing tube (both supplied with the kit) and centrifuged at 20,000 g for 3 minutes (Labnet Prism R microcentrifuge). The supernatant was carefully transferred by pipetting to a fresh 1.5 mL centrifuge tube, leaving the lysed red pellet undisturbed.

Next 700 µL of 100% isopropanol was added to the sample and the sample vortexed to mix. A 700 µL aliquot of the vortexed sample was then transferred into a PAXgene spin column in a processing tube. Unless otherwise stated all following centrifugations were performed using the Labnet Prism R microcentrifuge with centrifugation at 20,000 g for 1 minute. The spin column was centrifuged, before the remainder of the sample was loaded onto the column and centrifugation performed again. Next a wash step was performed by addition of 350 µL Buffer BM3 to the column which was then centrifuged after which the flow through was discarded. A DNase 1 mix was made up (70 µL Buffer RDD and 10 µL DNase stock per sample) and pipetted directly onto the spin column filter followed by incubation at room temperature for 15 minutes. Next 350 µL Buffer BM3 was added to the spin column which

was then centrifuged and flow through discarded. Five hundred µL Buffer BM4 was added to the column followed by centrifugation.

Next, and to prepare the column for sample elution, the spin column was placed into a fresh processing tube and centrifuged at 20,000 g with the Eppendorf lid open for 2 minutes to dry off all ethanol present (a component of Buffer BM4). The spin column was then placed into a fresh DNA free 1.5 mL centrifuge tube and 40 µL elution buffer (BM5) added directly onto the spin filter. The columns were incubated at room temperature for 1 minute before centrifugation after which the 1.5 mL centrifuge tube containing the eluted sample was sealed and incubated at 65℃ for 5 minutes on a heat block (Eppendorf Thermomixer C) before immediate placement on ice. The eluted sample was transferred into a labelled cryotube for storage at -80℃ prior to downstream experiments.

2.1.3. Quantification Methods for Nucleic Acids

Sample quantification forms an element of a quality control step. It allows one to confirm that the desired fragment size of nucleic acid species has been acquired and that a sufficient yield and/or concentration of said sample has been obtained. Various methodologies of sample quantification were applied within this thesis depending on the experiment being conducted and secondly cost effectiveness. Details of all methods used are provided below. In addition, within each results Chapter the method implemented is stated with reference back to the relevant section below.

The three methods used were the Nanodrop (ND-1000) spectrophotometer, the Qubit TM3 fluorometer and the Agilent 2100 Bioanalyzer. Each method varies in its level of precision,

sensitivity, and range of possible volume of input sample, with all three having a minimum input of 1 µL. For example, the Qubit has a high level of sensitivity and allows the input of a higher volume of sample to be tested (up to 10 µL) should the concentration at 1 µL be too low to be measured. The Nanodrop is less accurate (particularly if a sample is of low concentration) but only requires an input of 1 µL. The Qubit and the Nanodrop are only able to provide information on nucleic acid concentration of the tested solution. The Bioanalyzer enables assessment and concentration determination of different sizes of fragments.

2.1.3.1. Nanodrop (ND-1000) Spectrophotometer

The Nanodrop (ND-1000) spectrophotometer (ThermoFisher Scientific™, Massachusetts, USA) is an instrument that can be used for both DNA as well as RNA quantification. The instrument was first calibrated to zero absorbance by pipetting onto the measuring pedestal 1 µL of the blank solution (nuclease free water) after which the solution was removed from the pedestal with a single use lint free wipe. Next 1 µL of sample was pipetted onto the pedestal, the sample absorbance was then measured at 260 nm and the concentration of RNA determined using the Beer-Lambert law (within the software associated with the Spectrophotometer) which relates the absorbance of light through a sample to its concentration (Hansen and Desjardins, 2004).

After determining the RNA concentration, the quality of the RNA sample was assessed by measuring the absorbance at 260 nm and 280 nm. The ratio of absorbance at 260 nm to that at 280 nm indicates the purity of the RNA sample. A ratio of ~2.0 is generally

considered to be indicative of pure RNA, while a lower ratio may suggest the presence of contaminants such as proteins or organic compounds.

2.1.3.2. QubitTM 3 Fluorometer

The QubitTM 3 fluorometer (Thermo Fisher, USA) was also used for RNA quantification. First the Qubit RNA HS Assay reagents (Invitrogen, Thermo Fisher, USA) were allowed to equilibrate to room temperature (reagents are stored when not in use at 4°C). The so called 'working solution' was prepared by diluting the RNA HS reagent with the Qubit RNA HS buffer according to the kit instructions. This RNA HS Assay kit has a detection range of 0.2 to 200 ng/μL and requires two standards at these respective concentrations.

One hundred and 90 μL of working solution was added to 10 μL of RNA standard in a Qubit assay tube. For sample quantification, between 1 and 10 μL of sample was made up to 200 μL with working solution. The tube was then vortexed for several seconds and incubated at room temperature for 2 minutes.

Next RNA quantification mode was selected on the fluorometer and, after the standards or sample name was entered, the respective tube was inserted into the fluorometer and the 'measure' button pressed to obtain the RNA concentration result. The device records a result for each tube inserted and normalises the result based on the standards. Samples below or above the standards receive a too high or too low warning and if this occurs subsequent sample test tubes may be created using a further dilution or higher concentration of sample to identify a final concentration.

2.1.3.3. Agilent 2100 Bioanalyzer

The Agilent 2100 Bioanalyzer (Agilent, Santa Clara, USA) is a precision instrument that allows for the rapid and accurate analysis of nucleic acids and proteins. The instrument works by utilizing microfluidics technology to load and analyse samples. Samples are first loaded into a microfluidic chip. There are different chips depending on what is to be measured (DNA, RNA, protein) and size range (e.g. DNA fragment size) being assessed. The chip is then inserted into the Bioanalyzer which applies an electric field to the sample as pins are inserted into the wells of the chip. This causes the molecules to move through the microchannels and as they do so they encounter a series of sensors that detect their size and concentration. The data is then collected and analysed by the associated specialized software generating a highly accurate report of each sample's quality and integrity. This method allows for the precise quantification and characterization of DNA, RNA, and proteins, making it an essential tool for a wide range of biological applications. The Bioanalyzer can identify RNA as little as 5 ng/ µL in samples.

First, the Bioanalyzer pins are cleaned (using a dedicated cleaning chip) with RNAzap and then washed with water to remove any traces of RNAzap that might interfere with RNA sample quantification. To do this RNAzap was added to each well of the Bioanalyzer cleaning chip which was then loaded into the bioanalyzer, and the lid closed allowing the pins to be submerged into the RNAzap. The cleaning chip was then incubated within the Bioanalyzer for 2 minutes. The cleaning chip was removed from the Bioanalyzer, emptied by tapping the chip upside down onto a clean heavyweight tissue. Next, RNase-free water was added to each well of the cleaning chip which was then incubated within the

Bioanalyzer for a further 2 minutes in the same manner. The cleaning chip was then removed, emptied, and stored for future use.

For the experiments in this thesis, Small RNA chips and High Sensitivity DNA chips were used (Figure 2.2.). All reagents, for running 10 chips, are included in each kit. The protocol for running the Bioanalyzer for RNA versus DNA is slightly different. The following instructions are for running the Small RNA chips. Any differences in protocol for the High Sensitivity DNA chips are highlighted below. Included figures show differences in equipment as well as the expected trace. In the methods section of each Chapter the relevant kits used are indicated.

Whilst cleaning the pins of the Bioanalyzer, the Small RNA chip Bioanalyzer kit was taken out of –20°C storage and allowed to thaw at room temperature for 30 minutes. The ladder (provided in the kit) was denatured (essential to enable accurate size determination of RNA of samples) by heating for 2 minutes at 70°C using a thermocycler (DNA Engine PTC-200 [MJ Research, USA]) and then immediately placed on ice to ensure it remained denatured.

Next the dye concentrate (blue) was vortexed for 10 seconds and spun down for 1 minute at 10,000g (Labnet Prism R microcentrifuge). Next, 2 µL of the dye was pipetted into a 0.5 mL RNase-free tube, and 40 µL of filtered gel was added to the tube and mixed by pipetting. The tube was then centrifuged at 13,000 g for 10 minutes at room temperature (Labnet Prism R microcentrifuge). The priming station was set up ensuring that it was at the correct lever position 3 in the case of Small RNA chip (for HS DNA lever position 1) and that the plunger was at 1 mL. A timer was set for 1 minute. The run chip was then placed into the priming station and 9 µL of the prepared gel dye mix was added to well **G** (Figure 2.2.). The

priming station was then closed, and the plunger pressed to click and held for 1 minute before releasing. After allowing 5 seconds, the plunger was drawn back, and the station was then opened. The wells of the run-chip need to be filled to prevent bubbles forming between reagents/samples and potentially blocking the micro-channels linking the well to the gel dye. As such, pipetting needs to be done carefully with the tip of the pipette being placed directly in the bottom of the well edge.

Next, 9 μL of the gel dye was added to the two wells marked G and any remaining dye mix was discarded. RNA conditioning solution 9 μL was then added to the well labelled "CS", and then 5 μL small RNA marker was added to all sample wells (11 in total) as well as the ladder well. The marker contains a small quantity of the shortest and longest fragment size and is used by the Bioanalyzer software for normalising the sample to the ladder. Finally, 1 μL of the ladder was added into the ladder well and then into the dedicated sample wells 1 μL of sample was added. The chip was vortexed (on the specialized provided chip vortexer) for 1 minute at 2,400 rpm prior to being loaded and run on the Bioanalyzer. At times, small bubbles appear in sample or gel wells. These can be popped with single use 28G needles.



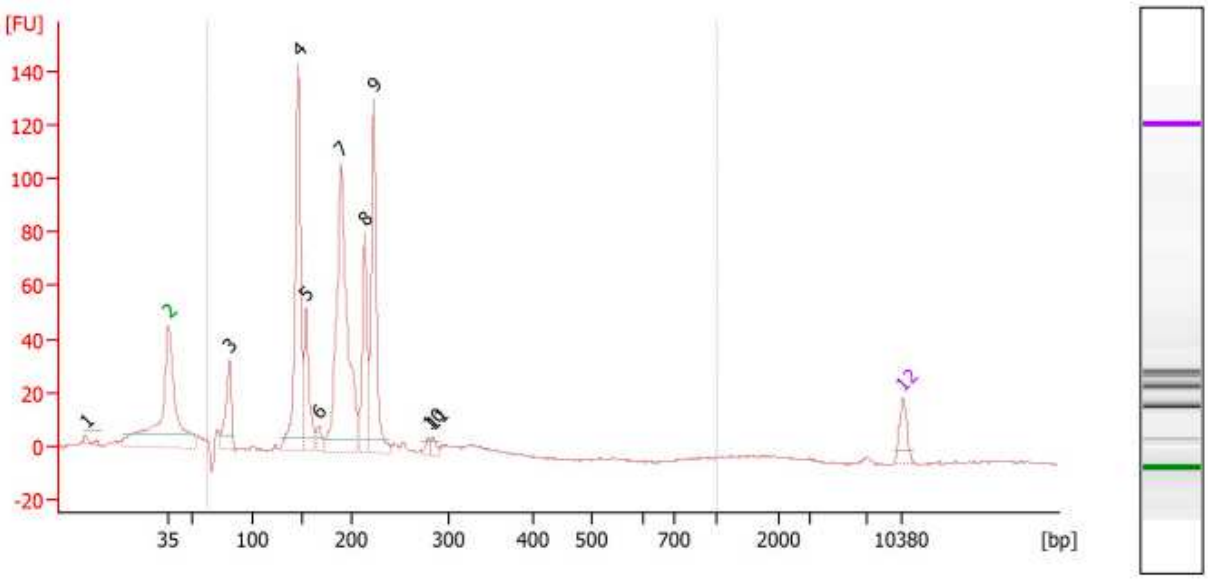
Figure 2.2. Photograph of Agilent Bioanalyzer Small RNA Chip and High Sensitivity DNA Chip. *Small RNA Bioanalyzer chip contains numbered wells for sample and marker addition, G stands for gel (this is the electrophoresis gel), and conditioning solution goes in the CS well. The HS DNA chip contains 4 G (gel) wells, and a ladder well, in addition to the 11 sample wells, the black circled* ***G*** *well is loaded directly by the syringe and plunger.*

During the Agilent Bioanalyzer run, to ensure it was performing smoothly, results from the ladder were monitored on screen. Once the run was complete the data was saved.

Unlike Small RNA chips, High Sensitivity (HS) DNA chips do not require a conditioning solution and the well instead is a gel 'G' well to which 9 µL of the gel dye is added. The ladder for DNA experiments does not require denaturing (as the sample to be quantitated is also double stranded). Figure 2.3. shows the resulting image file for a sample assessed by a High Sensitivity DNA chip with Figure 2.4. showing the readout for the Small RNA chip.

From Figures 2.3. and 2.4. below it can be seen that the traces differ depending on the chip used. The High Sensitivity DNA chip trace has two peaks of similar shape and size at the lower marker. As per the manufacturer's recommendations, at times it may be necessary to manually select the location of the lower marker in each lane as concentration of the

sample may be either very high or very low or fragments may be of similar size and skew the field. On occasion, the visual range markers (that select a desired range of interest within the measured field tested) also require adjustment. While the Small RNA chip shows a single large peak indicating the marker (for computational orientation) and a gradual increase in concentration of RNA of larger nucleotide lengths (Figure 2.4.).



Overall Results for sample 1 : **2.34**

Number of peaks found:	10	Corr. Area 1:	765.9
Noise:	0.3		

Peak table for sample 1 : **2.34**

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	7	0.00	0.0	
2	35	125.00	5,411.3	Lower Marker
3	81	402.25	7,484.4	
4	147	1,467.06	15,163.4	
5	155	468.79	4,593.9	
6	167	87.28	792.2	
7	189	1,732.39	13,872.1	
8	213	577.02	4,109.6	
9	222	922.80	6,285.6	
10	278	33.17	180.6	
11	284	43.77	233.8	
12	10,380	75.00	10.9	Upper Marker

Figure 2.3. Example High Sensitivity DNA Bioanalyzer Readout. *X axis identifies size of fragments, y axis identifies Functional Units (quantity of fragments of a specific size). Trace showing an example of a miRNA converted into cDNA with a small RNA library synthesis kit. The vertical bars are adjustable within the programme to select a range for the peak table (shown below the trace) to show. The bar to the right of the trace shows a gel image showing the distribution of DNA fragments across the measured range. Summarized sample results are detailed below the main trace. The green and purple arrows in the summary and numbers on the trace note the marker (a spike in that is added to each sample to enable the machine to automatically align the sample to the trace measurements).*

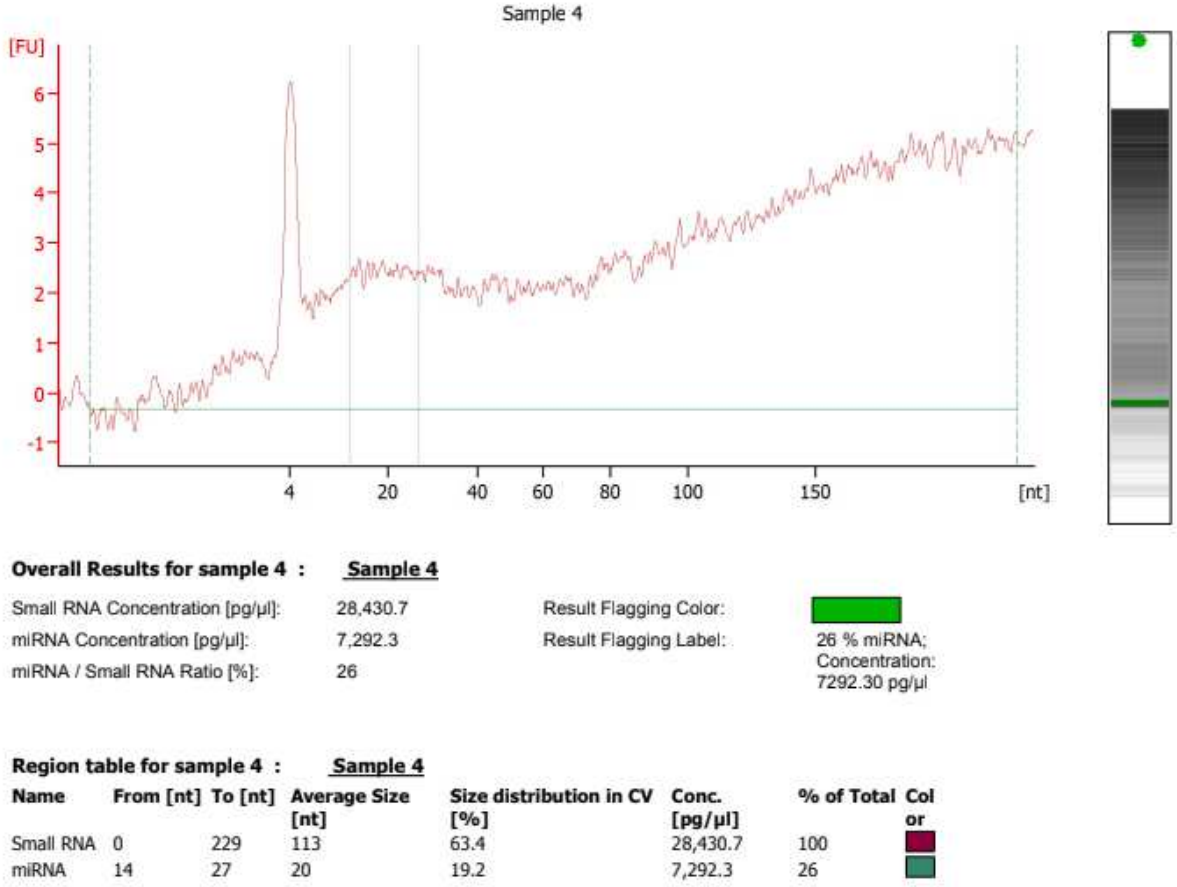


Figure 2.4. Example Small RNA Bioanalyzer readout. *X axis identifies size of fragments, y axis identifies Functional Units (quantity of fragments of a specific size). The vertical bars around the 20 nucleotide (nt) range indicate the range that small RNA (miRNA) will sit. The peak observed at 4 nt is the marker that is added to all sample wells and ladder well and enables the machine to align the sample to the scale. Summarized sample results are detailed below the main trace. The bar to the right of the trace is a gel image showing the distribution of RNA fragments across the measured range.*

2.1.4. RNA Sequencing Library Preparation

Small RNA libraries for each extracted RNA sample (that passed quality control) were prepared using the NEBnext Multiplex Small RNA library Prep Set for Illumina following the manufacturer's protocol. For all instances where a thermocycler is mentioned, a DNA Engine PTC-200 thermal cycler (MJ Research, USA) was used.

Table 2.1. gives details of some of the sequences (examples of adapters and primers) added to the RNA sample strands before reverse transcription into cDNA, as well as an example index sequence (in blue, index 1). Other index information is described in each Chapter as to the sequence associated with a particular sample.

Table 2.1. Sequencing Index and adapter sequence details *(Table taken from NEB Ltd., permission granted for replication).*

Product	**Index Primer**
NEBNext 3´ SR Adaptor for Illumina	5´-rAppAGATCGGAAGAGCACACGTCT-NH2 -3´
NEBNext SR RT Primer for Illumina	5´-AGACGTGTGCTCTTCCGATCT-3´
NEBNext 5´ SR Adaptor for Illumina 5´-	5´- rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrAr CrGrArUrC-3´
NEBNext Index 1 Primer for Illumina	5´-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTG GAGTTCAGACGTGTGCTCTTCCGATC*T-3´
NEBNext SR Primer for Illumina	5´- AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTT CTACAGTCCG*A-3´

The library protocol involves several steps (Figure 2.5.) starting with the ligation of the 3' SR adapter to the RNA. The 3' SR adapter for Illumina was first diluted 50:50 in nuclease-free water and then, in a sterile nuclease-free PCR tube, 1 µL of the diluted adapter was combined with 1-6 µL of input RNA sample (equating to 100 ng total RNA) with the addition of nuclease-free water to a total volume of 7 µL. Following mixing by pipetting, the tube was incubated in a preheated thermocycler for 2 minutes at 70°C and then transferred onto ice. Next 10 µL of 3' Ligation Reaction buffer (2X) and 3 µL of 3' Ligation enzyme mix were added to the tube, mixed thoroughly by pipetting, and then incubated for 1 hour at 25°C in a preset thermocycler. Next the reverse transcription (SR RT) primer was diluted 50:50 in nuclease-free water and 1 µL of said dilution plus 4.5 µL nuclease-free water added to the ligation mixture. The sample was run in a thermocycler (using the heated lid set to >85°C) for 5 minutes at 75°C followed by 15 minutes at 37°C, 15 minutes at 25°C, and then held at 4°C.

The 5' SR adapter was ligated by resuspending the 5' SR adapter in 120 µL of nuclease-free water and aliquoting 1 µL per sample + 10% wastage into separate, nuclease-free (200 µL capacity) PCR tubes. For total RNA inputs closer to 100 ng, the 5' SR adapter was diluted 50:50 in nuclease-free water. The adapter was incubated in the thermal cycler for 2 minutes at 70°C and then immediately placed on ice. Within 5 minutes, 1 µL denatured 5' SR adapter, 1 µL 5' Ligation Reaction buffer 10X, and 2.5 µL 5' Ligation enzyme mix were added to the ligation mixture. The mixture was mixed thoroughly and incubated for 1 hour at 25°C in a thermal cycler.

The 5' SR adapter was ligated by resuspending the 5' SR adapter in 120 µL of nuclease-free water and aliquoting 1 µL per sample + 10% excess into separate, nuclease-free (200 µL capacity) PCR tubes. For total RNA inputs closer to 100 ng, the 5' SR adapter was diluted 50:50 in nuclease-free water. The adapter was incubated in the thermal cycler for 2 minutes at 70°C and then immediately placed on ice. Within 5 minutes, 1 µL denatured 5' SR adapter, 1 µL 5' Ligation Reaction buffer 10X, and 2.5 µL 5' Ligation enzyme mix were added to the ligation mixture. The mixture was mixed thoroughly and incubated for 1 hour at 25°C in a thermal cycler.

The 5' SR adapter was ligated by resuspending the 5' SR adapter in 120 µL of nuclease-free water and aliquoting 1 µL per sample + 10% excess into separate, nuclease-free (200 µL capacity) PCR tubes. For total RNA inputs closer to 100 ng, the 5' SR adapter was diluted 50:50 in nuclease-free water. The adapter was incubated in the thermal cycler for 2 minutes at 70°C and then immediately placed on ice. Within 5 minutes, 1 µL denatured 5' SR adapter, 1 µL 5' Ligation Reaction buffer 10X, and 2.5 µL 5' Ligation enzyme mix were added to the ligation mixture. The mixture was mixed thoroughly and incubated for 1 hour at 25°C in a thermal cycler.

The next step was reverse transcription where 30 µL of the adapter ligated RNA was taken and, in a sterile nuclease-free tube, 8 µL First Strand Synthesis Reaction buffer, 1 µL red Murine RNase inhibitor, and 1 µL red Protoscript II RTase were added with mixing. The mixture was incubated for 60 minutes at 50°C in a thermal cycler.

After 60 minutes the RT reaction was curtailed through heat-inactivation by incubation at 70ºC for 15 minutes.

After completion of the reverse transcription, PCR amplification was performed by addition of the following components to the RT mix: Longamp Taq 2X Mastermix (50 µL), SR Primer for Illumina (2.5 µL), Index (x) primer (2.5 µL), and nuclease-free water (5 µL). The total volume was 100 µL and was mixed thoroughly by pipetting. PCR cycling conditions were: initial denaturation at 94°C for 30 seconds, then denaturation at 94°C for 15 seconds, annealing at 62°C for 30 seconds, extension at 70°C for 15 seconds for 12-15 cycles, followed by a final extension at 70°C for 5 minutes, and hold at 4°C until the next step.

It should be noted that individual index addition to each cDNA sample is important as it allows pooling of libraries generated from different RNA samples and post sequencing the reads associated with each RNA sample to be identified via the indexes.

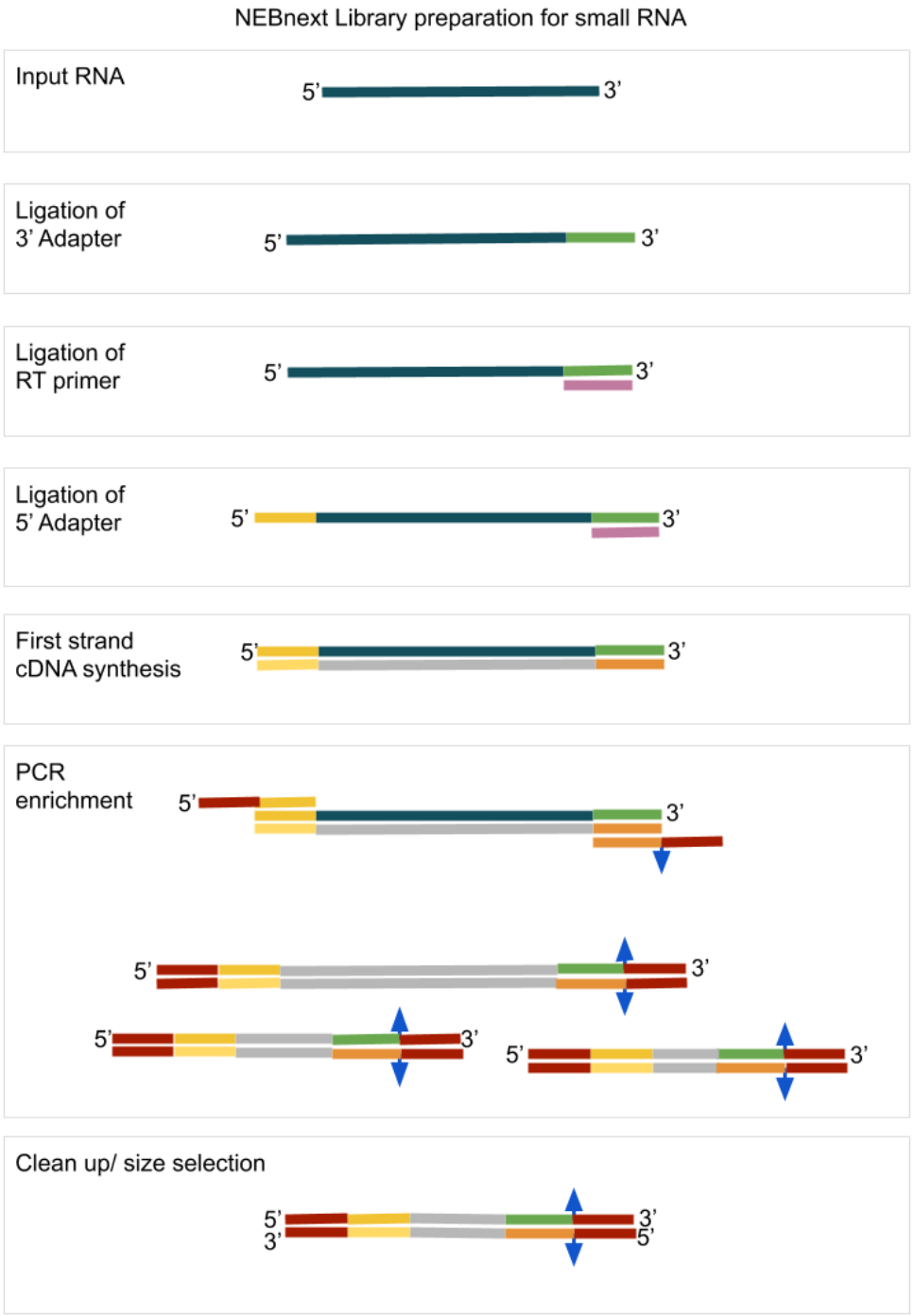


Figure 2.5. Diagrammatic View of the Illumina library synthesis protocol*. The lines represent strands of RNA within the sample and the coloured lines represent adapters and primers that are attached. The blue tick arrows represent the index sequence (6 codons) that enables sample differentiation. The strands become double after reverse transcription, and strands of undesired lengths are removed through size selection to keep only those that are of the correct length as miRNA. Diagram created in Biorender.*

Libraries were then quantified either using the Agilent 2100 Bioanalyzer, the Qubit 2 Fluorometer, or the Nanodrop (ND-1000) Spectrophotometer (see Section 2.1.3. above) before being stored at –20°C until size selection (see Section 2.3.5. below) and pooling prior to being sent off for next generation sequencing.

2.1.4.1. Size Selection Methods

Although kit choice influences the size of RNA species converted into a cDNA sequencing library, as total RNA (that contains all RNA species) is the input it is important to perform size selection on the generated library prior to sequencing. In this thesis focus was on miRNA species, hence use of the NEBNext Multiplex Small RNA Library Prep set for Illumina (see Section 2.1.4.), and the size target of the miRNA cDNA library (including adapters and index sequences) was therefore approximately 150 bp. It is important to note that the adapter sequences alone are approximately 120 bp.

As adapter sequences can result in oversaturation during the sequencing run and impact the number and quality of sequences obtained, it is essential that any excess unused adapters are removed through the size selection step. The relative difference in size between adapter sequences and the desired miRNA cDNA library is, however, very narrow. A very precise method of size selection is therefore required. Even applying such a method, as the size of a clean size-selected library is still close to an adapter sequence it is considered by sequencing facilities to be a high-risk sample. Consequently, facilities will only pool said samples (if they do not fill a full lane on the sequencer) with other customer samples deemed as high-risk. Size selection of such libraries with a Pippin Prep or a Pippin

Prep HT (both manufactured by Sage Science, Beverly, USA) is therefore desirable as the size selection criteria can be made to specific bp sizes.

Library size selection was therefore conducted on a Pippin Prep (Figure 2.6.) using the proprietary 3% Agarose dye free cassettes and associated reagents. Figure 2.6. shows the Pippin Prep with cassette inserted but lid open. When the setup is complete the lid is closed resulting in insertion of conductive electrodes into the appropriate wells enabling electrophoresis to occur.



Figure 2.6 Pippin Prep Device with Cassette Inserted and Being Prepared for Experiment. *As the electrophoresis buffer marked level to the top of the wells of the cassette. The final steps of the preparation are conducted while the cassette is already placed in the machine. The lid (dark blue, in the open position) then slides over the cassette inserting conductive electrodes into the appropriate wells to provide a current for the electrophoresis to occur.* (Image credit Sage Science).

The Pippin Prep is essentially a miniature electrophoresis gel with two positive electrode channels leading off from a main channel. As DNA is negatively charged (as are RNA and cDNA), when a current is passed through the Pippin Prep the DNA is drawn to the positive electrode. The DNA fragments meet resistance based on their size and so the smaller pieces

move more quickly through the gel towards the electrode. This enables the sizes of fragments to be segregated. For the specialised size selection to occur within the device, an optical laser identifies the size of passing fragments. Once the laser identifies the desired start length, the elution channel electrode is activated. This continues to draw the fragments through the channel into an elution well, until once again the optical laser identifies the desired finish length (as input by the user) and activates the main run channel of the gel. This effectively filters off the DNA of the desired length. Figure 2.7. below shows a representation of the features of a single channel of a Pippin Prep Cassette and how it works.

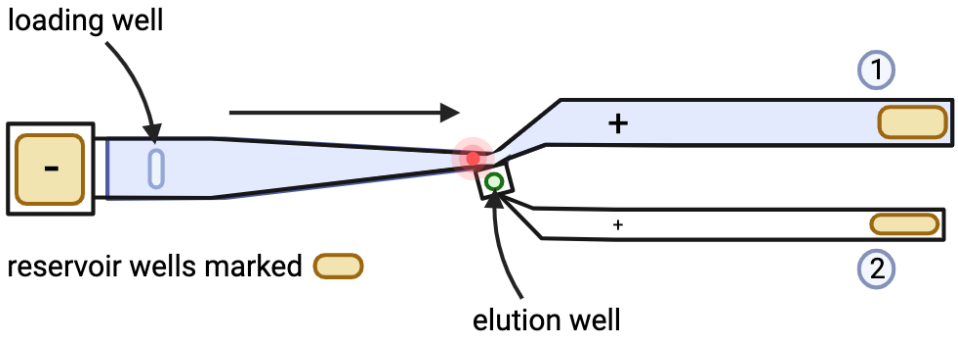


Figure 2.7. Diagram of a Single Pippin Prep Cassette Channel. *Five channels are present on each cassette and a single one is shown in this figure The sample and buffer are loaded into the loading well. A current is passed through the device drawing the negatively charged nucleic acid (DNA or RNA or cDNA) through the gel (blue) towards channel 1. When the optical detection (red) identifies the desired size of fragment, the electrode in lane 2 activates drawing the fragments into the elution well. When the upper desired range is met, as read by the optical detection electrode, lane 1 reactivates continuing drawing the nucleic acid fragments into the collection well for unwanted fragments. Reservoirs are marked. Diagram created in Biorender.*

To conduct the size selection, the Pippin Prep cassettes and reagents were allowed to come to room temperature before use. The Pippin Prep was calibrated by turning on the machine and placing the optical magnet side down to cover the lights. On the computer [calibrate] was selected with a target phMA of 0.8 and [calibrate ok] was clicked before selecting [Exit]. The calibration fixture was then removed.

A 3% agarose dye free cassette was opened and tapped to remove any bubbles in the column. The level of buffer was checked to ensure it was even in all wells, and the cassette then placed into the Pippin Prep with the elution wells on the right.

Forty µL of electrophoresis buffer was added to each well of the cassette ensuring all reservoir wells were filled to the top level, and the elution well was secured and sealed with adhesive strips. The prior steps occurred whilst the cassette was within the machine (with the lid of the Pippin Prep slid in the open position). The continuity test was performed by selecting [continuity test] and then [Test] on the computer interface connected to the Pippin. If perchance the test failed, the lid of the Pippin Prep was opened and either further elution buffer added or completely replaced in the elution well. The test was repeated until all wells passed.

The sample was prepared by taking 30 µL of the prepared library (Section 2.1.4.) and adding to it 10 µL of loading solution followed by mixing by pipetting. The prepared sample (40 µL) was then loaded into the loading well, after first removing 40 µL of buffer from the loading well. The cassette allows five samples to be size selected simultaneously.

The selection of the protocol and the correct cassette was made in the set-up menu on the Pippin Prep linked monitor (3% marker P with [use internal standards]). The reference

lane was set to match the lane number, and the protocol editor was set to [broad range] with a range of 124-174 bp. The [START] button was clicked, and the machine underwent automated calibration prior to a run of approximately 80 minutes.

After the run, each sample was checked via the monitor screen visual to ensure they had successfully gone through separation and elution. The samples were then collected by removing the sticker from each elution well and carefully removing (using a pipette and filter tip) a volume of 40 µL which was transferred into a prelabelled Eppendorf tube (as requested by the Sequencing Facility).

Finally, the Pippin Prep was washed using deionised water in a dummy cassette for 2 minutes at room temperature, after which the dummy cassette was removed and discarded.

2.1.4.2. Sample Concentration and Salt Removal Methods

The Sequencing Facility used, Novogene Co., Ltd (Cambridge, UK - referred to as Novogene from this point on), had a specified minimum required concentration of 500 pg/ µL for sequencing a prepared next generation sequencing (NGS) library. To fulfil this requirement, a concentration step had to be performed in some instances.

The concentration step was conducted using the Monarch® PCR & DNA clean up kit (New England Biolabs [NEB], USA). First the sample was diluted with DNA Cleanup Binding Buffer in a ratio of 7:1 (as per NEBnext Multiplex Small RNA library Prep Set for Illumina protocol recommendation) to a volume of 30-40 µL (recommended guidelines 20 - 200 µL).

The sample was mixed well by pipetting up and down or alternatively by flicking the tube – vortexing was not recommended.

Next the sample was loaded onto a spin column inserted into a collection tube. The sample was centrifuged for 1 minute at 16,000 g in a microcentrifuge (Labnet Prism R microcentrifuge) and the flow-through discarded. Two hundred µL of DNA Wash Buffer was added to the column which was then spun as before for 1 minute with the resulting flow-through discarded. The column was then transferred to a clean 1.5 mL microfuge tube. At all times great care was taken whenever removing the column from the tube that the column tip did not come into contact with the flow-through thereby ensuring traces of salt and/or ethanol were not carried over to the next step in the procedure.

Twelve µL of DNA Elution Buffer were added to the spin column and the sample was left to incubate for 1 minute at room temperature. The sample was then again centrifuged for 1 minute at 16,000g to elute the size selected cDNA library sample. Typical elution volumes were 10-12 µL, and the elution buffer was carefully added onto the matrix and not the wall of the column to maximize elution efficiency.

Before further processing or storage at -20°C, confirmation of concentration of the sample was determined by running an aliquot (1 µL) on the Agilent 2100 Bioanalyzer (Section 2.1.3.3.). The sample was of sufficient quality when the concentration met the 500 pg/ µL requirement set out by the Sequencing Facility (Novogene), and a clear peak was identified between 140-165 bp range with no other measurable artefacts.

2.1.5. Sequencing Methodology

The bioinformatics tools and pipelines developed for mRNA or total RNA sequencing data are often not be optimized for the unique characteristics of miRNA sequences. For example, standard alignment algorithms may struggle to accurately map the short miRNA reads to the reference genome or transcriptome, leading to ambiguous or non-specific alignments. This can subsequently impact downstream analyses, such as differential expression analysis and target prediction.

To address these challenges, specialized miRNA bioinformatics pipelines invariably must be put together using existing packages across multiple platforms or in a unique workflow. These approaches often incorporate modifications to the normalisation method or other statistical analysis based around the different limitations and understanding of the differences between miRNAs vs mRNA sequencing analysis. As will be shown later in this Chapter (Figure 2.10. onwards), often the analysis is based on an understanding of first principles rather than a reliance on packages and software outcomes.

In the case of this thesis, while pipelines were analysed for suitability, additional packages were integrated into the pipeline to optimise the results towards limiting false positive results or false negative results. Part of this was the awareness of mRNAs being from a significantly larger potential pool, whereas miRNAs (although isomers are understood to exist and are being researched) the core pool of miRNAs are likely fewer than 3,000.

Sequencing of prepared libraries was via Novogene and the optimum choice of sequencing was 150 bp paired-end reads although for the purpose of analysis only the forward reads

were required. This means that the sequencing machine will count and identify 150 base pairs along each read before stopping and repeating for the reverse. Ideally, only a 50 bp single-end read was required, however, due to costs it was more economical to run a 150 bp paired-end run.

The preparation of the libraries in this case was by multiplexing. Samples are given base pair “bar codes” to enable them to be identified once the sequencing has been completed. These are the adapters. There is an adapter “bar code” to enable the Sequencing Facility to identify reads from the submitted samples and save them to a file for the researcher. This enables multiple researchers to contribute different samples to the sequencer flow cell to all be sequenced simultaneously under the same conditions. For the researcher, it is important to be able to identify individual samples. As such, an index (another base pair adapter “bar code”) is added (Table 2.1.). Both “bar codes” are important and certain manufacturers make compatible indices and adapters.

All details about the samples index sequences used in preparing the libraries therefore are provided to the Sequencing Facility as well as information about the adapters used and the kits, so that the Facility (in this case Novogene) can identify compatible libraries for sequencing at the same time. It is common for up to 75 or so libraries to be run on the same flow cell. This is the standard business model for Novogene. The selection of compatible libraries is often species specific per run followed by risk categories. The researchers submitting libraries, however, have no say in what other libraries are run together on the flow-cell and this is down to the discretion of the Sequencing Facility used.

Pooled sample libraries from this study when received by Novogene underwent internal QC steps (results of which are always shared with the client) involving a tapestation analysis to identify fragment sizes and concentration. The library was then pooled with other high-risk samples (See Chapter 1, Section 1.2.2.) prior to loading onto the sequencer (NovaSeq) flow cell.

Figure 2.8. shows the process of next generation sequencing that occurs on the Illumina sequencing platforms. The Illumina flow cell is loaded onto the NovaSeq instrument, which uses a series of four-color reversible terminators to sequence the DNA. The first read of 150 cycles is performed from one end of the DNA fragment. The process, by which the sequencer identifies what the sequence the base pairs are in, involves cluster generation. The flow cell contains billions of spots each of which is coated with complementary oligonucleotides to the adapters on the DNA fragments. The fragments are then hybridized to the spots and clonally amplified by bridge PCR to form clusters. Once the clusters have been amplified the flow cell is washed, and fluorescently labelled nucleotides are added. Once more, polymerases are used to create complimentary strands of the bound library clusters. Each nucleotide gives off a light signal as it binds that is registered by software within the sequencer. The clusters assist in providing a clear signal for a specific nucleotide at a specific location along the strand. The process is then repeated and the second read of 150 cycles is performed from the other end of the strand, resulting in paired-end reads of 150 bp each. Raw read data is then collected from the machine isolating the sample indexes that were used for each contributing researcher and the data is then distributed to researchers on a hard drive for analysis.

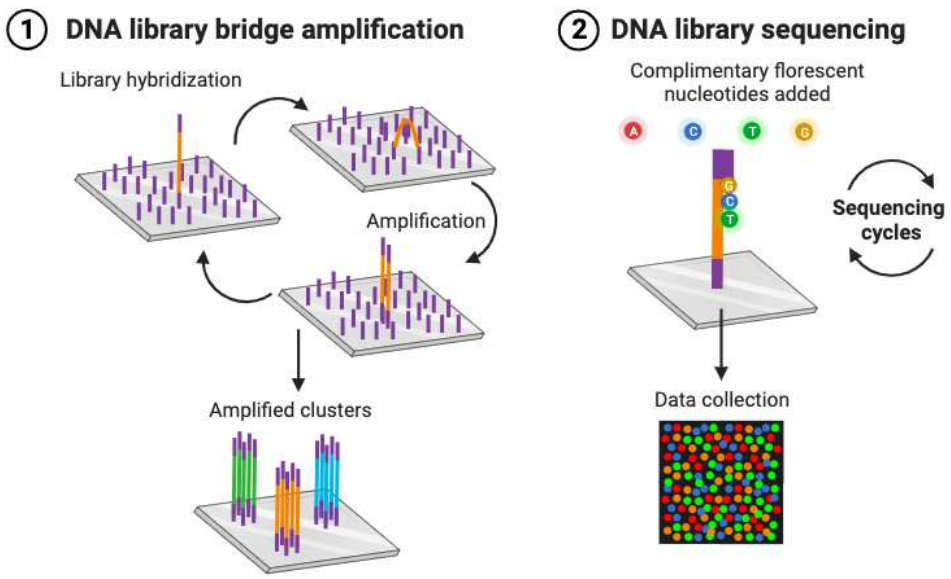


Figure 2.8. Illumina Sequencing Process. *1. DNA bridge amplification: The pooled library was added to a lane on the flow cell and adapters bind to complimentary oligomers bound to the base of the flow cell. Through a process of amplification, whereby complimentary oligomers bind to the other side of the bound library strand forming a bridge, then complimentary strands are created using polymerases followed by denaturing to repeat the process again until clusters are formed. 2. Once the clusters have been amplified the flow cell is washed, and fluorescently labelled nucleotides are added. Once more, polymerases are used to create complimentary strands of the bound library clusters. Each nucleotide gives of a light signal as it binds that is registered by software within the sequencer. The clusters assist in providing a clear signal for a specific nucleotide at a specific location along the strand. Diagram created in Biorender.*

2.1.5.1. Data Processing Methods

Illumina sequencing files (whether generated in house or by a Sequencing Facility) are in FASTQ format. Novogene, as well as providing the raw data, also provided details on the sample distribution on the sequencing runs that were performed. Forward and reverse reads were generated but only the forward reads were used which is the norm as per other published miRNA studies as insufficient evidence at the time indicated that the use of the reverse read would yield any significant benefit (mostly due to a lack of supporting experimental results with which to align the findings). Figure 2.9. below shows a flowchart of how the raw reads received from Novagene were processed and analysed. Data processing (for all four results Chapters 3 to 6) was conducted using the Genboree Workbench within which, the following steps were conducted using the ExceRpt small RNA tool (Rozowsky *et al.*, 2019).

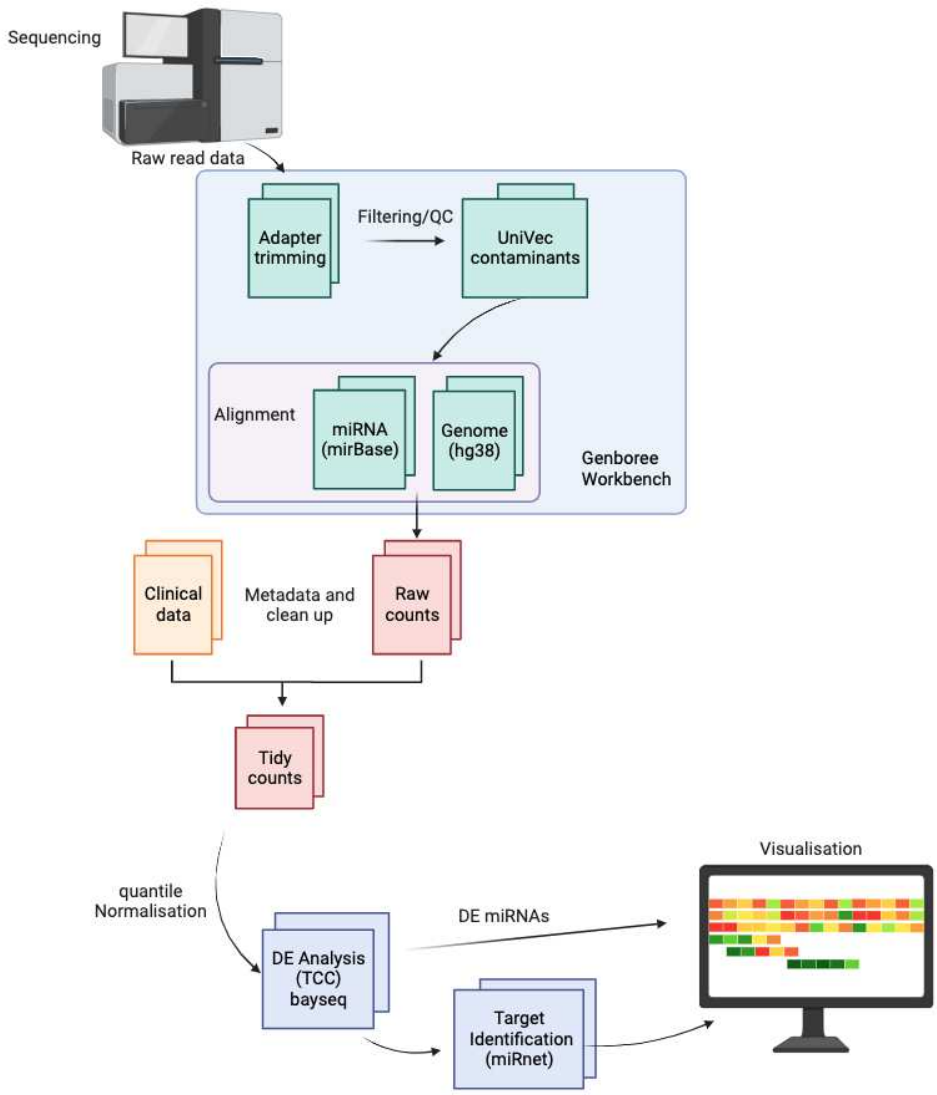


Figure 2.9. Schematic of Data Processing from Raw Sequencing Reads to Visualization of Final Differentially Expressed (DE) miRNAs. *Raw read data was passed through Genboree workbench tools to trim adapters and carry out filtration of vector contaminants and low/poor quality reads. Alignment (using a STAR aligner) miRBase and the human genome then follows. As all reads are complete miRNA sequences there was no collating needed. The result of these first steps was the generation of raw reads for each sample annotated with the number of each miRNA and counts. The raw data was then joined with the metadata (study group) for normalisation, followed by DE analysis conducted with the package TCC* (Tag Count Comparison) *in R. BaySeq was the selected model. Results of DE analysis was considered for visualisation. DE-miRNAs were then processed with software miRNet to identify potential targets. This used targetscan but enabled filtering of experiment type. Only experimental methods such as QPCR were used for inclusion of a target mRNA gene. All experimental methods that used computer prediction or sequencing were excluded. Results were then considered for visualisation. Diagram created in Biorender.*

An initial look at the raw reads was conducted on FASTQC (Andrews, 2010). FASTQC is a bioinformatics tool to assess the quality of sequencing data. All reads from a single sample are run through FASTQC one sample at a time, not all samples will be run unless there is an issue such as significant poor quality from a FASTQC or if the run was complicated in some way. FASTQC assesses eight aspects of sequence quality: an overall Quality Control Summary, Per-Base Sequence Quality, Per-Sequence Quality Scores, Sequence Length Distribution, Sequence Content, Sequence Duplication Levels, Adapter Content, and Overrepresented Sequences. While these are quite commonplace in sequencing preliminary analysis, for miRNA research there are some aspects that may need additional explanation.

For the Quality Control Summary, FastQC generates an overall summary report highlighting any potential issues or areas of concern in the sequencing result. Figure 2.10. below shows an example of a miRNA FASTQC sequencing result. The summary provides a quick overview of the data quality and facilitates decision-making for subsequent data processing steps. FASTQC uses a traffic light system to highlight a pass (green) a warning (amber) and a concern (red). In the case of miRNA sequencing, with the methods used for library preparation (Section 2.1.4.), there are some expected amber and red warnings and these will be detailed and explained in the following sections below.

The basic sequence statistics gives details of the sample name (file name), the FASTQC identifies the encoding and the file type from the input file. A count of the total number of sequences in that file follows with details on the number of sequences flagged as poor quality and the average length. Lastly, the summary highlights the percentage of GC making up the reads.

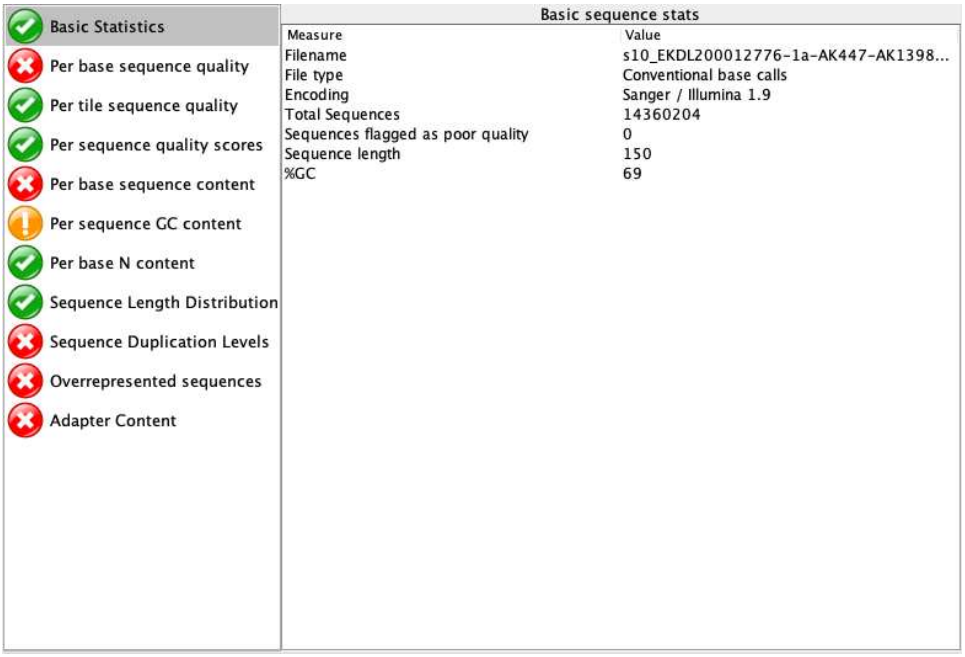


Figure 2.10. FASTQC Quality Control Summary. *The ribbon to the left highlights a general overview of all the areas the FASTQC analyses whilst the ribbon to the right shows the basic sequence statistics. The file name is the individual sample library's raw read file, and FASTQC identifies the file type and the encoding. Then it identifies the total number of reads for that sample, and out of those which are poor quality, and the average length of the reads. Lastly, GC content, for this library there was an amber flag due to the GC content straying from the mean GC content.*

The Per-base Sequence Quality is a graph showing the distribution of quality scores for each base position in the sequencing reads. It helps identify regions with low-quality scores, which may indicate the need for trimming or filtering. An example of a Per-base Sequence Quality from a miRNA sequencing run sample is shown below (Figure 2.11.). It has been annotated for quick reference. The red dashes denote the median value, the yellow boxes are the inter-quartile range (25%-75%) and the upper and lower whiskers represent the 10% and 90% points in the data. The blue line represents the mean quality of the raw reads. Added and not present on FASTQC file results are the representative percentage accuracy of reads landing in that region of the graph. The y axis corresponds to

Phred score, which gives the probability that the corresponding base call is incorrect. A Phred score of 30 is a probability that 1 in 1,000 base calls are incorrect equating to 99.9% accuracy. While a Phred Quality Score of 10 has a 1 in 10 probability of the base call being incorrect, a 90% accuracy. The green bar represents high quality reads with a high confidence of accuracy (99.9% confidence or greater), the orange bar represents medium quality (with between 99%-99.8% confidence), and the red bar represents reads with less than 99% confidence in reads.

The graph (Figure 2.11.) generated by FASTQC is important to assess as the accuracy of sequencing will impact downstream steps. The accuracy of reads is known to fall as a sequencing run proceeds. Along the x axis an additional overlay has been added to show the predicted range of base pair location of the samples that are generated from the methods detailed in this Chapter. The adapters are shown in blue, the “i” in the orange box represents the index primer, and the green box with “miRNA” represents the target sequence (which will not always be miRNA but that is the target).

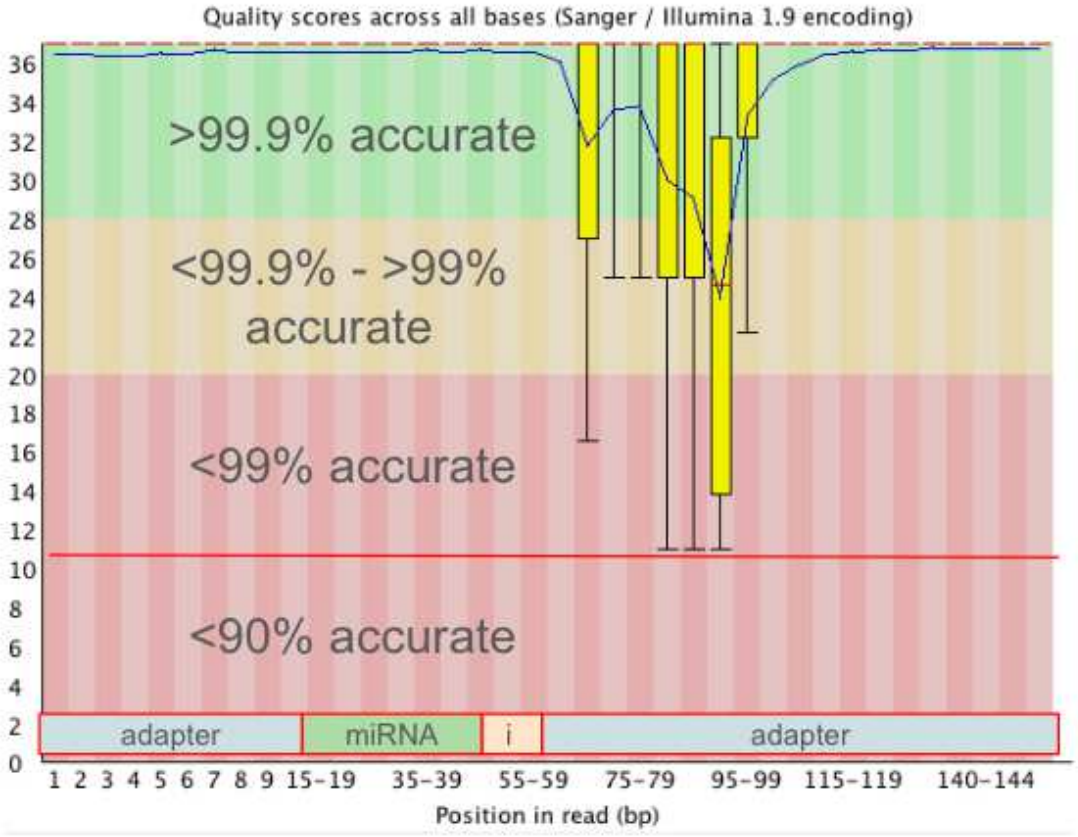


Figure 2.11. FASTQC: An Example of Per-base Sequence Quality with Text Overlay. *Shown here is an average of all the reads for a single miRNA sequenced sample. The y axis corresponds to the Phred quality score, where 30 corresponds to a 99.9% that the base call is accurate (a 1/1,000 chance of the base call being untrue). The green highlights good quality reads (it shows the base call quality is Phred approx. >29), the orange is acceptable quality reads (corresponds to a Phred score of >20 which is 99% that the base call is correct (1/100 chance of being incorrect)), and the red highlights poor quality reads that are likely unreliable or contain a high proportion of errors or repeats (a Phred score of 10 corresponds with an base call accuracy of 90%). In this case we see that the quality is high up until approximately 70-74 bases. Following along the x axis, shown are the specific expected location along the read of features such as adapters in blue, the desired sequence range in green (miRNAs), and the index (i) in orange.*

Per-sequence Quality Scores: This analysis assesses the overall quality score distribution for all sequences in the dataset. It helps identify whether the majority of sequences have consistent high-quality scores or if there are outliers with significantly lower scores. Figure 2.12. below shows an example of a miRNA sequencing Per-Sequence Quality Score. The x axis details the Phred Quality Score whilst the y axis is the number of read sequences.

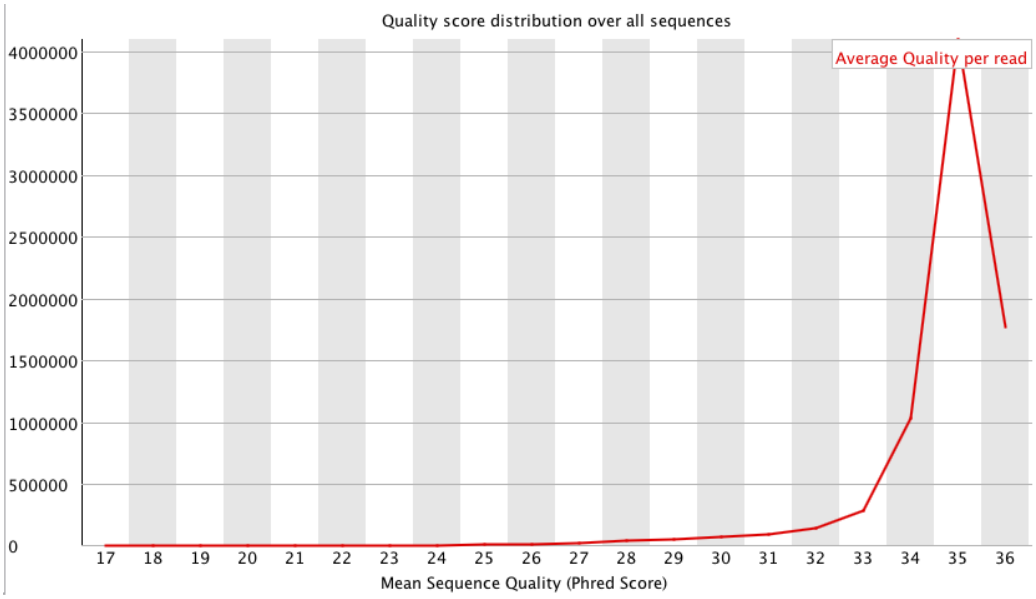


Figure 2.12. FASTQC: An Example of Per-sequence Quality Scores. *Here is shown an average of all the reads for a single miRNA sequenced sample. The x axis corresponds to the Phred quality score, and the y axis the number of sequences. The average quality score shown here is 35 corresponding to a read accuracy score of >99.97%.*

Next is Sequence Length Distribution. FastQC provides a histogram showing the distribution of sequence lengths in the dataset. It helps identify any anomalies or unexpected peaks in the distribution that may require investigation. As the libraries are size selected to be 150 bp in length they should all appear as per the below Figure 2.13.

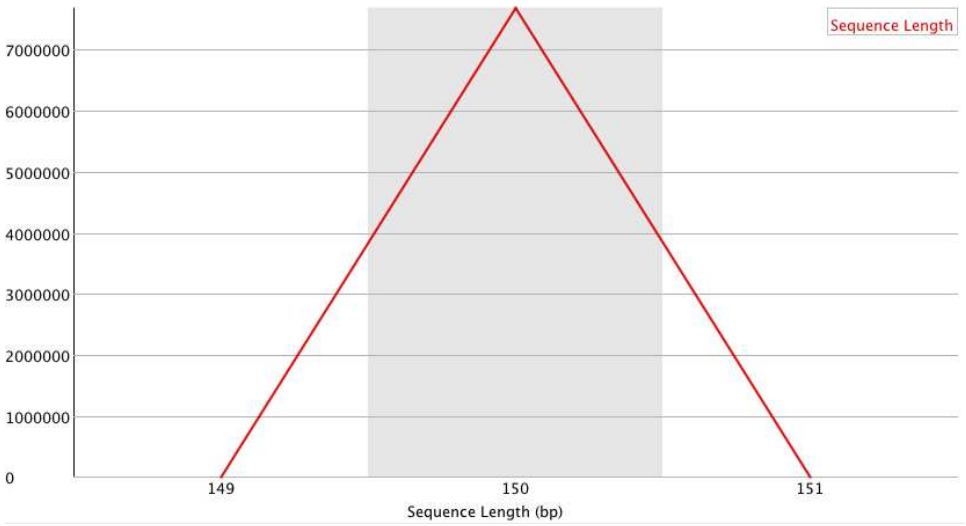


Figure 2.13. FASTQC: Distribution of Sequence Length Across All Sequences. *Here is shown an average of all the reads for a single miRNA sequenced sample. The x axis corresponds to the number of bases across all sequences. The y axis the total number of reads. All reads fall within the 150 bp length range.*

Next FASTQC considers Sequence Content. This analysis checks for the presence of overrepresented sequences or biases in nucleotide composition. It helps identify potential contaminants, adapter sequences, or other artifacts present in the dataset. There are a couple of aspects that result in a flag for this section. Firstly, the entirety of the library is sequenced including the adapters and index sequences which would not be sequenced where the cDNA library size longer than the sequencing coverage. This means that there is an uneven distribution of certain bases as for most of the reads the same base will be identified at the same place, where it would be expected that across the genome all bases are evenly distributed (to some degree). An example of a miRNA base pair content is shown below (Figure 2.14.). As can be seen there is a highly uneven distribution of bases before a moderate attenuation at approximately the 15-19th position in the read. Then a further uneven distribution at the 90th position in the read. miRNAs sequences are not as evenly

distributed in bases as the rest of the genome as there is a smaller pool. As well as other small RNA species contributing to a less than equal distribution (tRNA and fragments).

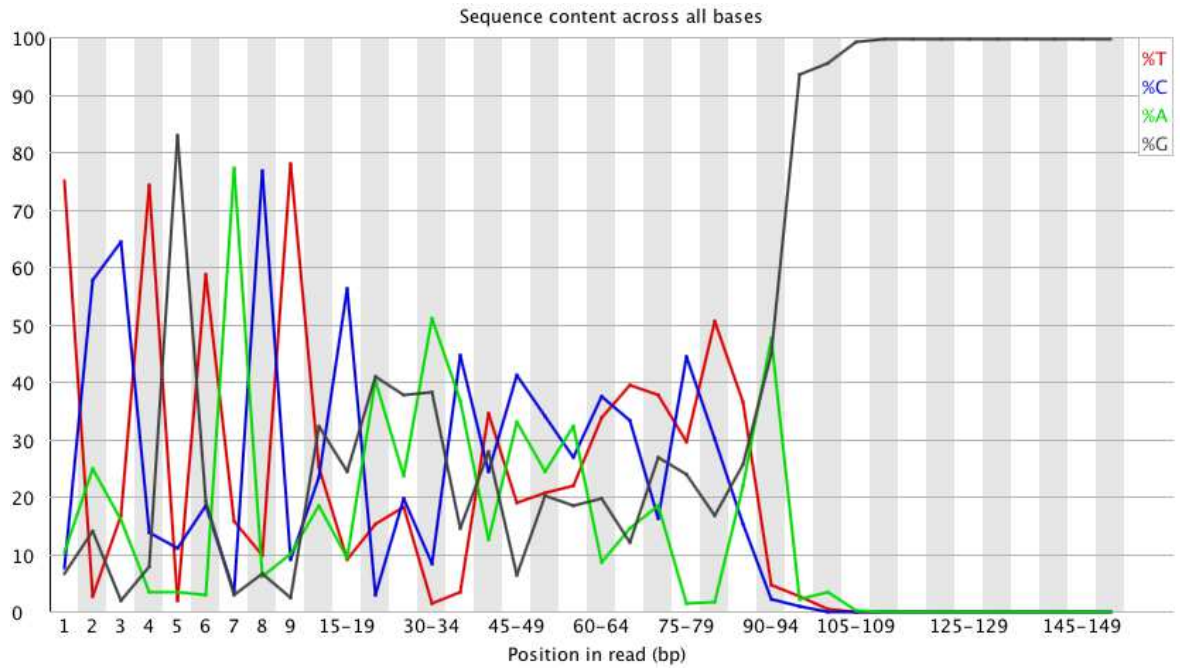


Figure 2.14. FASTQC: Sequence Content Across All Bases. *Example miRNA sequencing FASTQC of base content along reads. Highly unevenly distributed base content seen in the reads before 15-19 bases, before the distribution improves to more even before the 90-94 base position and becomes heavily G based.*

GC distribution is then analysed. GC distribution within a genome roughly corresponds to the GC content of the underlying genome. In the case of size selected specific libraries where the adapter is included in the length of the sequencing run, it would however be expected that the GC count per read would be higher than the theoretical distribution. Figure 2.15. below shows an example of uneven GC distribution.

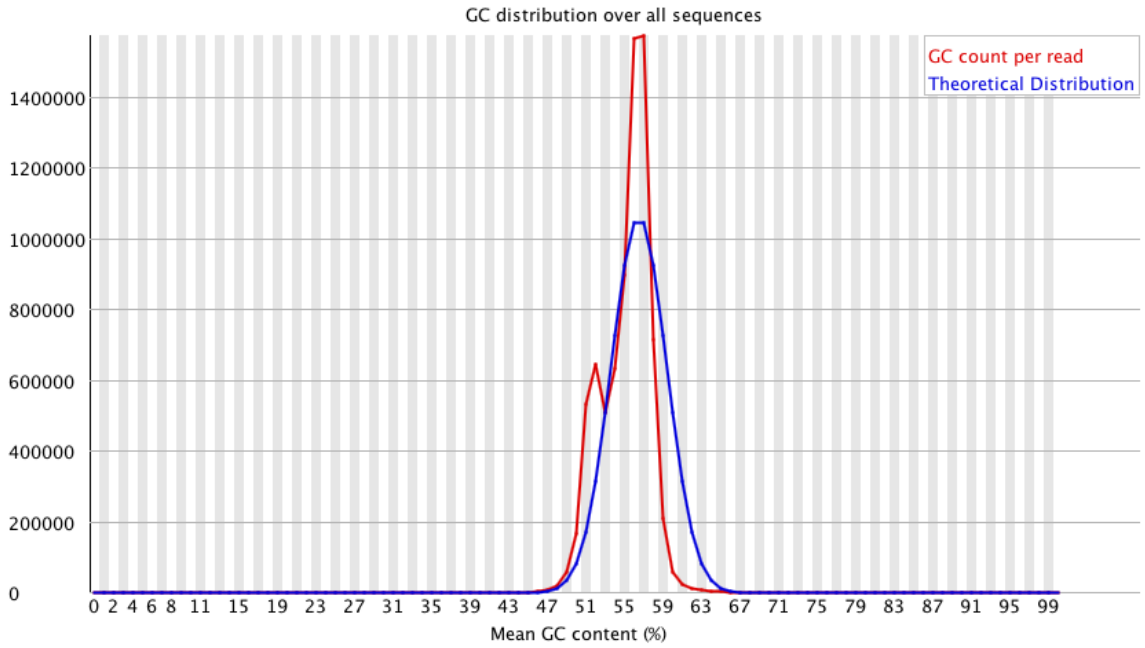


Figure 2.15. An Example of FASTQC GC Distribution Histogram. *Here is shown an average of all the reads for a single miRNA sequencing sample. The y axis corresponds to the number of reads, and the x axis the percentage of reads. This sample's miRNA sequencing data shows that the GC count per read is higher than theoretical distribution across standard RNA sequencing.*

FastQC evaluates the level of sequence duplication levels within the dataset, which can indicate PCR amplification bias or library preparation issues. Figure 2.16. is a histogram showing the high percentage (78%) of duplicated sequences across the dataset as a percentage of total sequences. Duplication rates between 50-90% are expected with miRNA, compared to the much lower duplication rates of 10-20% for mRNA. In addition, a high number of duplicated sequences was expected as the adapter sequence was present on all reads and would be included in this number. The level of variance in miRNAs is not as high as mRNA and therefore would also add to this score.

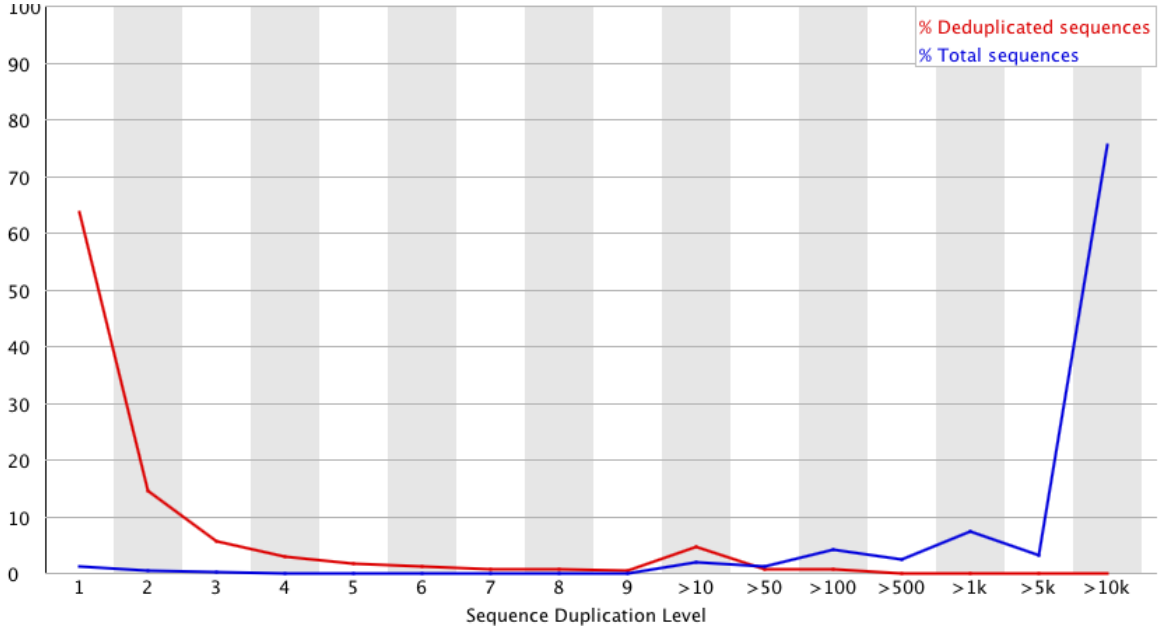


Figure 2.16. An Example of FASTQC Sequence Duplication Histogram. *Here is shown an average of all the reads for a single miRNA sequenced sample. The y axis corresponds to the percentage of reads, and the x axis the percentage of sequence duplication. This sample's miRNA sequencing data shows that there is a high level of sequence duplication.*

FASTQC next checks Adapter Content. It detects the presence of adapter sequences within the reads. In the case shown below, Figure 2.17., the raw data file has been included with the adapter intact.

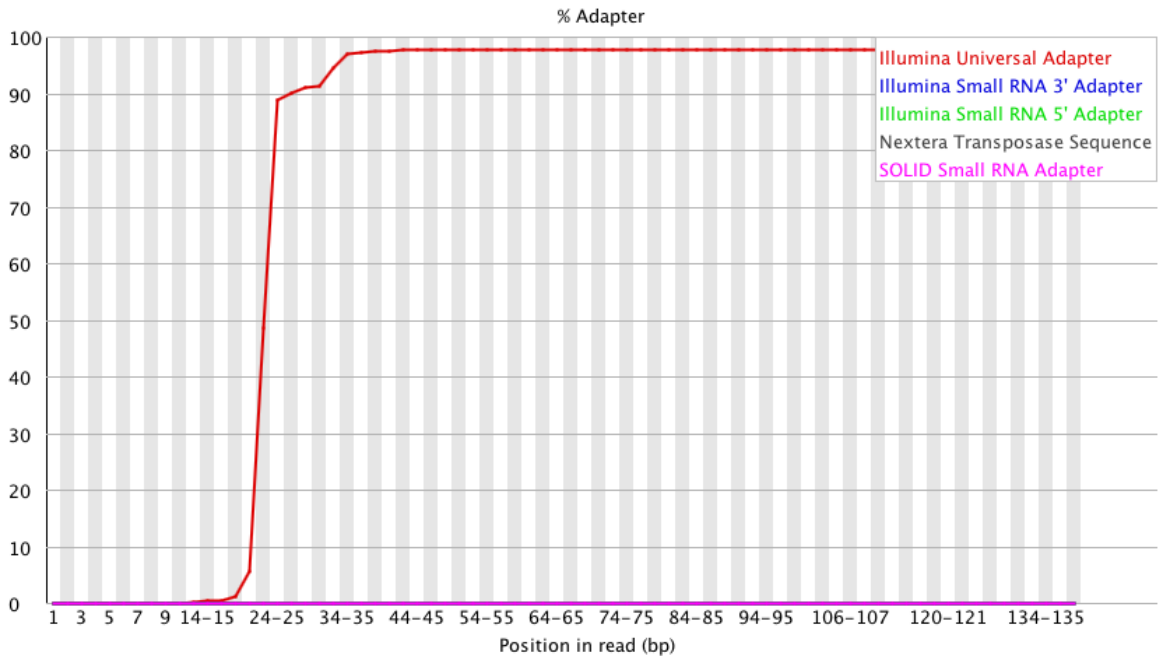


Figure 2.17. An Example of FASTQC GC Distribution Histogram. *Here is shown an average of all the reads for a single miRNA sequenced sample. Adapters tested for are shown in the box at the top right, while only one adapter was present in the sample. The y axis corresponds to the percentage of reads where the adapter was identified, and the x axis the position of where the adapter is identified. In this case the Illumina Universal adapter was identified from position 25 in 90% of reads, and over 95% by position 34-35 in the reads.*

Overrepresented Sequences: FastQC flags sequences that are significantly overrepresented compared to the expected distribution as this can indicate contamination or other anomalies. In the case of the example analysis shown (Figure 2.18., below), the Illumina multiplexing primer has been identified.

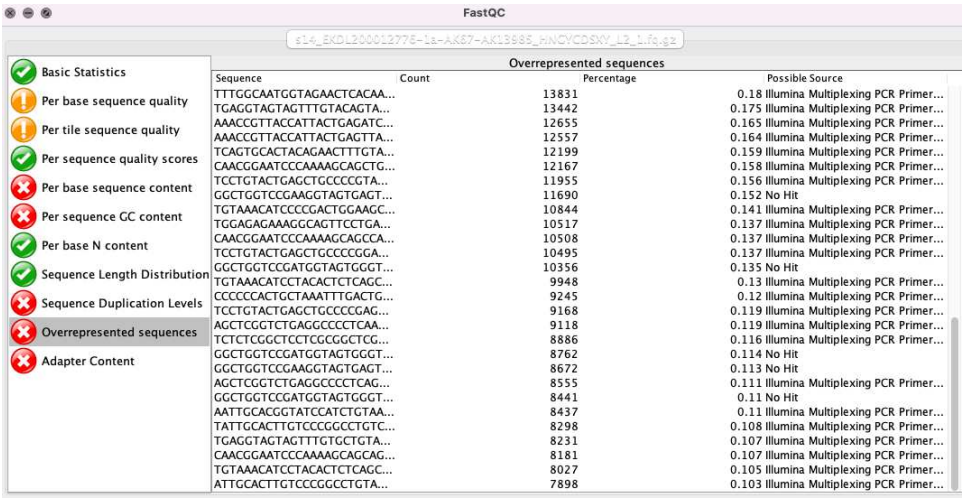


Figure 2.18. An Example of FASTQC Overrepresented Sequence Table. *The sequence that has been identified as overrepresented is in the far-left column, followed by the counts identified, the percentage of reads, and the possible source of the overrepresentation. In this case, the multiplexing primer is identified as it was included in the sequencing length.*

Having run the quality checks using FASTQC, the sequencing reads were pooled into a file on the Genboree workbench and put into a directory for analysis as an experimental set.

Next the sequencing reads were processed and filtered using several criteria. Adapter sequences were clipped and then a quality filter was applied. Homopolymer sequences were removed and any Univec contaminants were filtered out. Finally, any rRNA sequences were filtered out. Next, alignment was performed with the remaining reads this follows the ExceRpt pipeline detailed in Figure 2.19.

Reads were aligned to the human genome Hg38, and miRBase using STAR aligners (with settings allowing no more than one mismatch and requiring a minimum of 18 bp per read), and for the experiments conducted in this thesis raw counts were selected (for full details of sources and links see Table 2.2.). Any reads that did not align to the human genome and the miRbase where then checked to see if they aligned to: tRNA sense, tRNA antisense,

piRNA sense, piRNA antisense, Gencode sense, Gencode antisense, circularRNA sense, or circularRNA antisense. Table 2.2. shows the process in order with details provided for each step conducted.

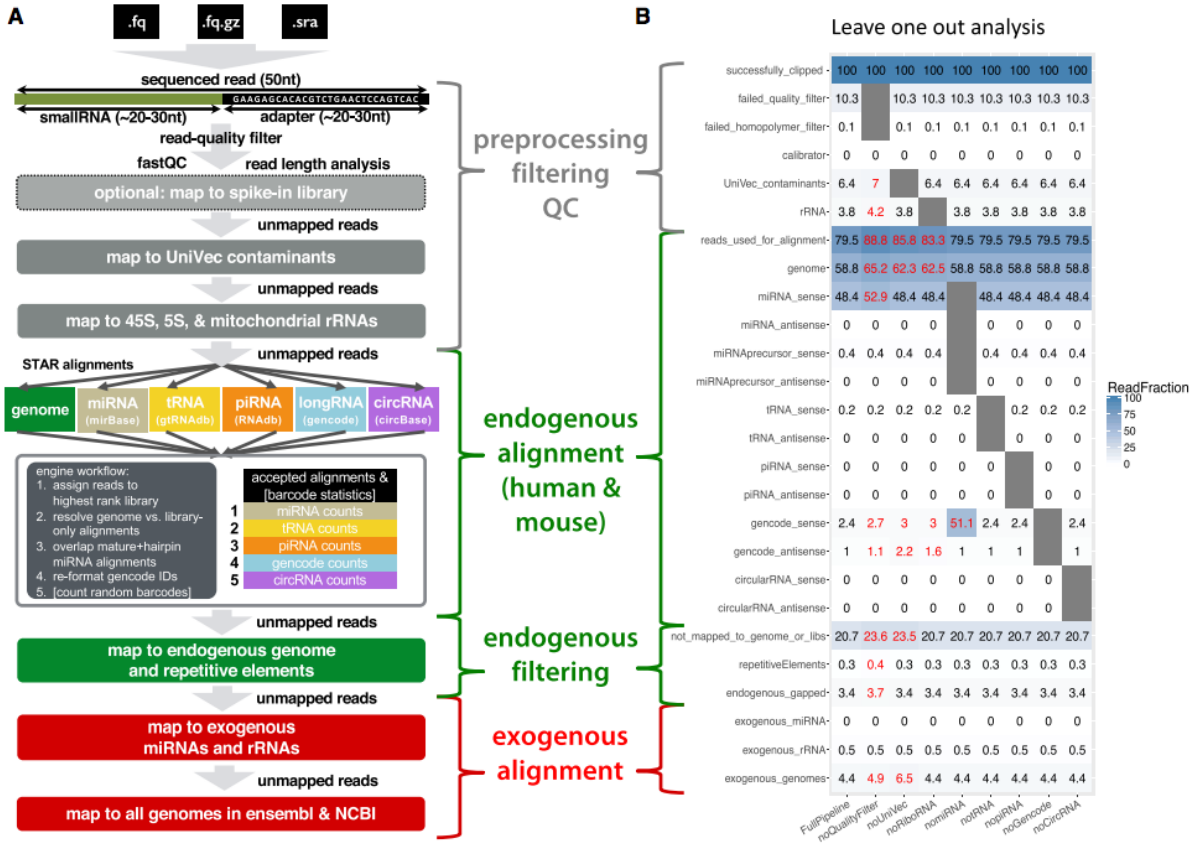


Figure 2.19. ***ExceRpt Pipeline****. taken (with permission) from Rozowsky, J. et al., 2019. ExceRpt workflow (left), and example of result read percentage and alignments.*

Table 2.2. Genboree Workbench ExceRpt Filtering and Alignment Details.

Category	Feature name	Description of Feature
	Successfully clipped	Adapter was successfully clipped from read. ExceRpt automatically identifies 3' adapters.
Quality filters	Failed quality filter	Read fraction quality = 80 (80% of reads must have acceptable quality scores), Minimum quality = 20 (1% error rate)
	Failed homopolymer filter	FASTX toolkit, identifies homopolymer repeats and removes these reads.
	Univec contaminant filter	aligned to sequences in the NCBI's UnivecVec screen database are removed.
	rRNA	Reads are aligned to all primary endogenous ribosomal RNAs (5S,5.8S,18S,28S, and45S) and removed.
Alignments	Reads used for alignments	First pass alignment using STAR aligner with max 1 mismatch per read, minimum read length 18 bases.
	Genome	Hg38.
	MiRNA (sense and antisense)	MiRBase v.22.
	MiRNA (precursor sense and antisense)	MiRBase v.22.
	tRNA (sense and antisense)	tRNAscan database.
	piRNA (sense and antisense)	piRNA annotations.
	Gencode (sense and antisense)	Aligned to GENCODE transcripts (snRNAs, snoRNAs, miscRNAs, protein coding genes and lncRNAs).
	Circular RNA (sense and antisense)	Elected to be aligned using the circRNA database.
Not mapped to the above (and because exogenous mapping was selected)	Not mapped to genome or libs	Second alignment using STAR aligner. Allowing only perfect matches.
	Repetitive elements	Quality control.
	Endogenous gapped	Enables better understanding of reads as a percentage of reads will be identified that were not in earlier alignment steps.
	Exogenous miRNA (input to, and)	Quality control. ("input to" reads that were input and those that successfully aligned). Looking for potential sampling/processing contaminants. Or spike in miRNA.
	Exogenous rRNA (input to, and)	Quality control. Looking for potential contaminants.

Repetitive elements can lead to issues in mapping reads accurately to a reference sequence or the genome. A negative result in alignment to repetitive elements suggests that the main alignment was successful and there is no potential issue with findings. Filtering out these reads may also help in the accuracy in downstream analyses. As mentioned above, mapping accuracy can also be improved by mapping to reads to repetitive elements that have multiple valid mapping locations. In some contexts, repetitive elements have biological significance, specifically as small RNAs (siRNA and piRNAs) can target repetitive elements and regulate activity (Piatek and Werner, 2014; Wu *et al.*, 2019). In the case of this study the rationale for inclusion of filtering for repetitive elements was for quality control.

Sources and references for the databases and software used for the data processing and analysis are given in Table 2.3.

Table 2.3. Key resources for Data Processing and Analysis of miRNA Sequencing Data.

data/software used	**Source**	**reference**
FASTQC v.0.12.0	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	(Andrews and Others, 2010)
Human genome hg38	https://www.ncbi.nlm.nih.gov/grc/human	http://genome.ucsc.edu.
miRbase version 22	http://www.mirbase.org/	(Kozomara and Griffiths-Jones, 2014)
exceRpt version 4.6.2	http://genboree.org/theCommons/projects/ exrna-tools	(Rozowsky *et al.*, 2019)
TCC (R package)	Bioconductor	(Sun *et al.*, 2013)
miRNet version 2.0	https://www.miRNet.ca/	(Chang *et al.*, 2020)
MiRTarBase version 8.0	https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/download.php	(Huang *et al.*, 2022)

2.1.5.2. Data Analysis Methods

Differential expression analysis was performed within the R package TCC (an acronym for Tag Count Comparison). Normalisation of the data, however, was conducted separately as some methods of normalisation that have been identified as being optimal for miRNA analysis are not included in the R package TCC. Consequently normalisation was conducted with "preprocessCore" package in R using the "quantiles" method (Tang *et al.*, 2015) before performing differential expression analysis of the miRNA reads within the TCC package (Di *et al.*, 2011; Sun *et al.*, 2013; Tang *et al.*, 2015). For the latter, the groups used in the comparison are detailed in each results Chapter (Chapters 4, 5 and 6). Differential expression (DE) quantification was performed using the BaySeq method as it has been shown to be effective for the number of reads and the number of samples being analysed for miRNA research (Cordero *et al.*, 2012; Tam, Tsao and McPherson, 2015). The results of the analysis were considered statistically significant if the FDR was less than 0.05 and the fold change (FC) was > 0.5. The FDR was selected as an adjusted *P*-value to account for the effects of multiple testing.

Further correlation analysis was conducted on the selected (based on significance) DE-miRNAs. DE-miRNAs raw counts for each sample were used in the correlation that considered the interaction with clinical variables of interest. Values were converted into factors ("male" or "female" etc), where needed, prior to being input into a Spearman correlation analysis correct to 2 decimal places. NA values were excluded from analysis. Outcomes were given correlation ranges of; 0.0 - 0.19 = very weak, 0.2-0.39 = weak, 0.4-0.59 = moderate, 0.6+ = strong. This was for both positive and negative correlations.

2.1.5.3. miRNet mRNA Target Analysis

DE-miRNAs were processed with the software miRNet to identify potential targets. The software uses Targetscan but has the option to filter based on experiment type (see later).

miRNet is available as a graphical user interface (GUI) on the internet (Chang et al. 2020). There is the capacity to upload raw read data or upload lists (DE-miRNAs in the case of this research project). Figure 2.20. shows the selections made before input of DE-miRNA lists. Lists must be uploaded in one of two types of formats: mirbase ID (eg. hsa-miR-123) or accession number.

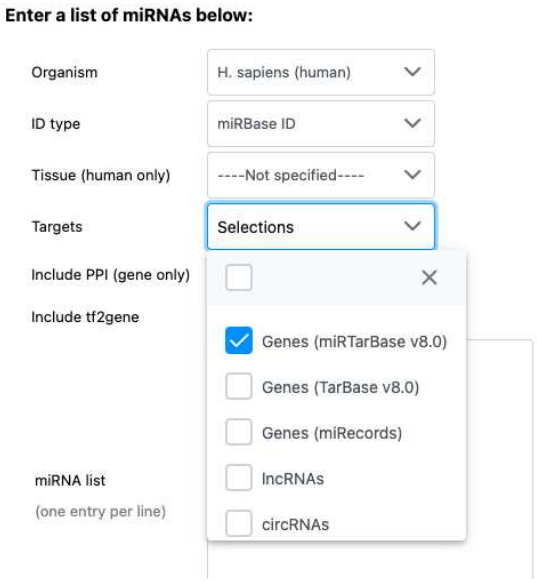


Figure 2.20. miRNet Selection and Input Screen*. Selection of homo sapiens and mirbase ID, and tissue was not specified, miRTarBase v8.0 was selected. Then miRNAs are added as a comma separated list with each miRNA on a separate line.*

Once input a selection screen gives further options prior to network viewer mode. Figure 2.21. shows how users can select from Interaction tables, Networks, and Network tools for filtering. From the interaction tables, the types of experiments from which the targets were

identified can be seen and can be filtered out. For example, in this thesis all experimental targets that were identified through PCR or arrays were included, whilst a default of excluding computational or other sequencing analysis was initially applied (e.g. any that contained CHIP, CLASH, or Sequencing). For the example (Figure 2.21.) the top five miRNAs from a previously published list were input (Lee *et al.*, 2023). Any changes to these baseline criteria will be discussed and rationalle explained on a case by case basis.

Once the target genes meeting the criteria of experimental identification method were selected for, the results were further filtered to analyse only those with the strongest associations. For this thesis, targets were selected for degree and then betweeness. Filtering for degree enables the identification of important network nodes that have the most connections, whilst filtering for betweenness measures the shortest paths through a node and identifies bottlenecks in a network.

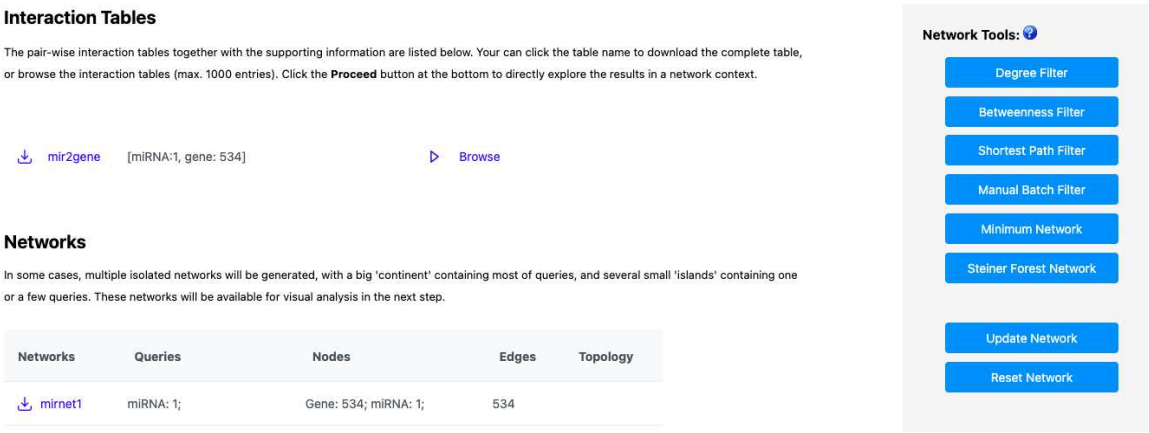


Figure 2.21. miRNet Network Analysis Filtering Screen. *Once the list of miRNAs has been input, options regarding how networks are to be filtered (if needed) before going into viewing mode.*

Then, the final selection of gene targets was input into Gene Ontology analysis. Three aspects of Gene Ontology were considered, Biological Process, Molecular Function, and Cellular Components. This way a broad understanding of the potential impact on the organism through the effects that might occur in Biological Process, and what molecular mechanisms might be involved, as well as an understanding of what cell machinery may be involved in this process. Not all gene targets will be identified in all tests therefore a broader understanding of the processes can be informative in seeking targets of interest to validate and research further. The last element for consideration with Gene Ontology analysis, with respect to miRNA research, are the number of targets a specific DE-miRNA has and the total number of DE-miRNAs that target a single gene as this can potentially guide researchers to understand associated processes that might be key to understanding the underlying processes involved in the pathology studied.

When selecting from the Function Explorer in the Network View pane (Figure 2.22.), possible choices are: all genes, highlighted genes, all miRNAs, highlighted miRNAs. The Algorithm options are: Hypergeometric test or Empirical Sampling followed by choices of: KEGG, GO:BP, GO:MP, GO:CC. In this thesis, all genes were selected using the Empirical sampling, and all GO options.

While Hypergeometric tests (one sided Fisher's exact tests) are often conducted for Gene Ontology analysis, it assumes that genes are selected uniformly at random from a finite population. This may not be the case given the target genes were identified (in this thesis) as being linked to the input miRNA. In some cases Hypergeometric tests did not generate a

result and Empirical sampling was therefore chosen (as it was considered the less conservative method).

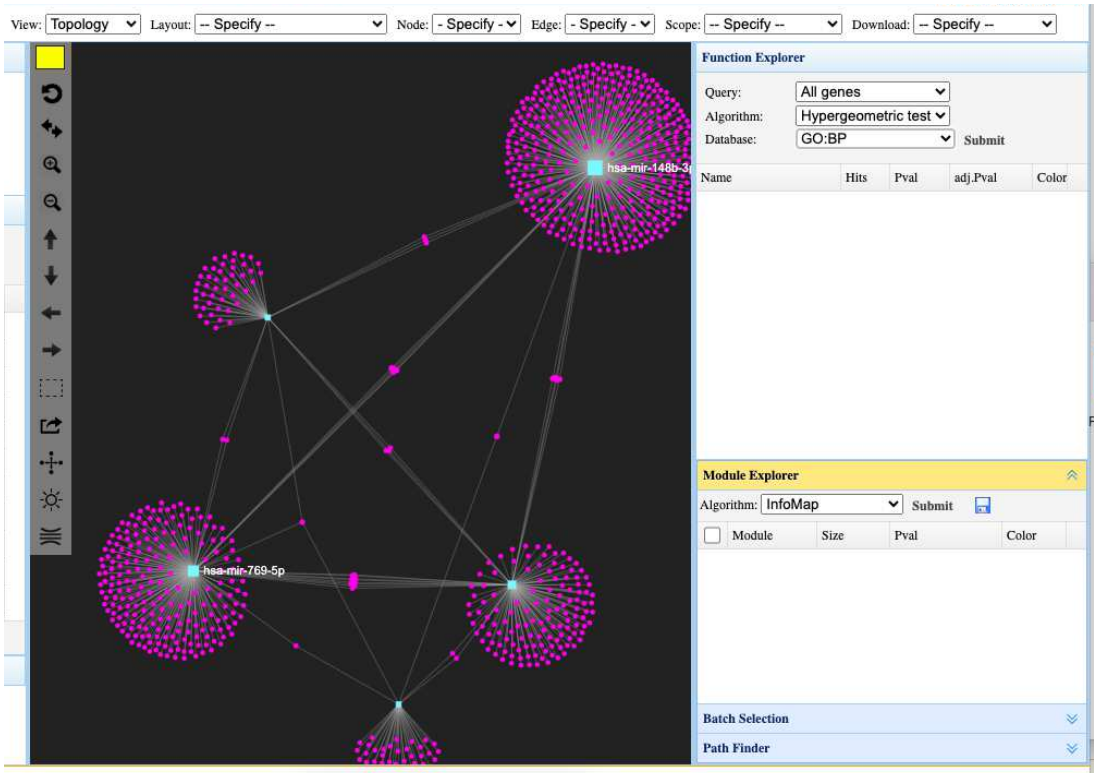


Figure 2.22. miRNet Network Viewer Mode. *Interaction network in left panel showing the five input miRNAs, and on the right options for exploring function links with the identified mRNA targets (highlighted pink in the left panel). Shown here are five miRNAs (blue squares) with target genes and linked associations with other miRNAs (pink circles). Grey lines identify links between network connections.*

Once generated, the GO analysis lists detail the name of the function, the number of hits (as a list that can be expanded to identify individual miRNAs/genes), and lastly, the *P* and adjusted *P* values.

2.1.5.4. Literature Review

A literature review was conducted for comparison of miRNAs with results in Chapter 3, Chapter 4, and Chapter 5 with the review focusing on research papers where miRNAs were sequenced for asthma (adult or childhood) from any human sample type and between the years 2016-2023. Studies where patients were treated with novel compounds or with oral corticosteroids as part of the study or as part of the inclusion criteria of the study were excluded.

2.1.5.5. Reference miRNA Analysis

As discussed (Chapter 1, Section 1.3.2.), identification of stable reference genes is crucial for experimental validation of findings. Several computational programs have been designed to identify stably expressed genes across a range of test and control samples for a given experimental context.

The GeNorm software determines the *M* value, or stability score, according to the average pairwise variation between a specific gene and all other candidate genes (Vandesompele *et al.*, 2002). The gene with the highest variation in expression based on the paired expression values in all the studied samples is eliminated, and the best pair of reference genes is determined. The genes with the most stable expression have the lowest *M* values (Vandesompele *et al.*, 2002).

Using a standard deviation (SD) and coefficient of variance (CV) of candidate gene expression, BestKeeper (Pfaffl *et al.*, 2004) calculates, and identifies, the lowest CV and SD

which indicate the most stable reference gene. The software also estimates the inter-gene relations of all possible candidate gene pairs by applying the Pearson correlation coefficient and the probability of significance of the pairs (Pfaffl *et al.*, 2004).

The comparative ΔCt method is based on a comparison of the relative expressions of pairs of possible reference genes within each sample. The stability of the candidate genes is ranked according to the reproducibility of the gene expression differences in the studied samples (Silver *et al.*, 2006).

Normfinder is an algorithm based tool that ranks gene expression based on a given sample set and experimental design to identify genes with the most stable expression. It is novel as it runs estimates of stability of expression between subgroups within a given population (Andersen, Jensen and Ørntoft, 2004).

RefFinder integrates GeNorm, NormFinder, BestKeeper, and the comparative ΔCt method. It assigns an appropriate weight to an individual gene, calculates the geometric mean of these weights, and gives the rank of the best candidate reference genes (Xie, Wang and Zhang, 2023).

A comparison was conducted between GeNorm alone and RefFinder, which includes GeNorm, to identify how differently miRNAs were weighted by the combinatorial analysis.

While there is little published on minimum numbers for reference gene identification from sequencing, recommendations from Sigma-Aldrich is 5 samples per condition or group (*qPCR Reference Gene Selection Protocol*, no date). As such, analysis results with fewer than 5 samples will not be considered in this thesis.

After filtering out miRNAs with fewer than one read, raw count data was input for analysis. For GeNorm, this was conducted with the "genorm" package in R, and for RefFinder the online Graphical user interface (GUI) was used. Results were then passed into R to generate figures, or figures were generated by the software. *M* values, or a stability score below 1.5 were considered stably expressed and suitable as reference miRNAs.

3. Chapter 3: Method Development for Nasosorption sampling for miRNA sequencing

3.1. Introduction

Asthma is a chronic inflammatory disorder of the airways, affecting over 300 million people worldwide, as discussed in Chapter 1 (Section 1.1.1.). Accurate diagnosis, monitoring, and treatment of asthma remain challenging due to the heterogeneous nature of the disease. Current diagnostic methods, such as spirometry and symptom assessment, have limitations and may therefore not accurately reflect the underlying pathophysiology. Consequently, many patients experience suboptimal disease control leading to frequent exacerbations, decreased quality of life, and increased healthcare costs. The development of novel, minimally invasive diagnostic tools that can provide insights into the molecular mechanisms driving asthma could revolutionize disease management and improve patient outcomes.

MicroRNAs (miRNAs) are small non-coding RNA molecules that play crucial roles in regulating gene expression (Chapter 1, Section 1.3.1.). Dysregulation of miRNAs has been implicated in various diseases including respiratory conditions such as asthma. The study

of miRNA expression in the respiratory tract has, however, been challenging due to the difficulties in obtaining relevant samples. As discussed in Chapter 1 (Section 1.4) traditional methods, such as bronchoscopy, are invasive and not clinically viable for ongoing assessment of the condition.

Nasosorption is a novel, minimally invasive technique that allows for the sampling of the nasal mucosal lining fluid. This method involves the gentle absorption of the nasal fluid onto a synthetic absorptive matrix (SAM) which is then processed for downstream analysis. Nasosorption has the potential to provide insights into the molecular mechanisms underlying respiratory diseases by enabling the analysis of miRNA expression in the nasal mucosa, the latter being a site of interest for conditions like asthma.

While nasosorption has been previously used for the analysis of proteins and other biomolecules (Chapter 1, Section 1.4.), its application in miRNA profiling has not been extensively explored. This research project aimed to develop and optimize a protocol for miRNA sequencing from Nasosorption samples with the long-term goal of identifying differentially expressed miRNAs (potential biomarkers) in asthma and other respiratory diseases.

The specific objectives of this Chapter were:

To optimize the protocol for RNA extraction from Nasosorption samples, ensuring high yields and RNA quality suitable for miRNA sequencing.

To develop and optimize a library preparation and sequencing pipeline for miRNA analysis from Nasosorption samples.

To perform a pilot study comparing miRNA expression profiles in Nasosorption samples from healthy individuals and individuals with asthma, with the aim of identifying potential miRNA biomarkers.

To validate the identified miRNA biomarkers and investigate their functional roles in the pathogenesis of asthma.

3.2. Methods

3.2.1. Study Participants

Ethical approval was given by the West London and GTAC Research Ethics committee (REC reference number 15/LO/0444). The ethics allowed both asthma patients and healthy control subjects to be sampled. Work conducted for this Chapter was focused on designing and optimising the miRNA extraction method. This work was therefore conducted on Nasosorption samples from healthy subjects only.

3.2.2. Nasosorption Collection and RNA Isolation

Nasosorption (NS) was conducted as previously described (Chapter 2, Section 2.11.) using CE-marked NS medical Synthetic Absorptive Matrix (SAMs) (Hunt Developments UK Ltd) (Thwaites *et al.*, 2018) (Chapter 2, Figure 2.1.). In brief, Nasosorption samples were collected via insertion of the SAM into the nostril under direct vision and placed against the

inferior turbinate. The subject then pressed their nostril closed with their index finger. The SAM remained in place for a timed 60 seconds. The SAM was then gently removed and replaced within its holder. Storage of samples collected prior to elution and extraction varied as the Processing Protocol evolved and was optimised and is therefore detailed appropriately below. From each healthy subject, a Nasosorption sample was collected per nostril (i.e. both left and right nostrils were sampled) and then pooled together for subsequent processing.RNA isolation was conducted using either the Monarch or the Zymo total RNA (including small RNA) isolation kits. Although the kits provide all the reagents once the RNA is bound to the spin column matrix, the method of ensuring preparation of the sample for optimum RNA binding to the spin column had not previously been identified for this sample type. A number of adjustments to the steps of the method were therefore taken, to attain an optimised protocol. Both kits were used for Processing Protocol 1 only. From then on, only the Zymo kit was used.

For the initial protocol trialled, the Nanodrop and the Qubit 4 were used for RNA quantification (Chapter 2, Section 2.1.3.2.). Subsequently, the most sensitive method of quantification available was used: the Agilent Bioanalyzer (Chapter 2, Section 2.1.3.3.) in combination with the small RNA kit.

3.2.2. Library Preparation and Sequencing

For each SAM extracted RNA sample, cDNA library preparation was conducted using NEBnext small RNA library synthesis, compatible with Illumina (Chapter 2, Section 2.1.4.). In brief, the 3' adapter sequence was ligated to the RNA to enable attachment of the reverse transcription primer. Next the 5' primer was ligated to the other end of the RNA strand. Reverse transcription then followed, enabling compatible nucleotide transcription up to and including the 5' adapter. A unique index sequence was then ligated to the end of the sequence at the complementary 3' end (different for every SAM extracted RNA sample to enable multiplexing), as well as a final adapter sequence to both the 3' and the 5' end of the strand. PCR enrichment followed, amplifying all species of RNA present in the sample. The cDNA library was then stored at -80°C until quantification and size selection.

Quantification of the generated cDNA library was performed using the Agilent Bioanalyzer and the small RNA Bioanalyzer kit as detailed in Chapter 2 (Section 2.1.3.3). Due to the narrow range between the desired library and adapter-dimers, a highly precise size selection method was used: the Sage Science Pippin Prep. Full details of the Pippin Prep size selection can be found in Chapter 2 (Section 2.1.4.1.). In brief, proprietary Pippin Prep 3% Agarose dye free cassettes and associated reagents were used alongside the Pippin Prep machine. The cassette wells were filled with buffer and tested and calibrated, and then the size selection protocols were selected on the computer software. The sample was prepared and loaded into the cassette and run on the Pippin Prep. Post the run and using a Gilson pipette, the sample (40 μL) was collected and transferred into a pre-labelled Eppendorf tube and sent to Novogene (Cambridge, UK) for sequencing (Chapter 2, Section 1.5.).

3.2.3. Data Analysis

Detailed description of the methods of data analysis performed on the sequencing data generated are given in Chapter 2, Section 2.1.5.1. Briefly raw FASTQ files were passed through FASTQC to look at the quality of the reads and any flags in the data. Once this was passed, FASTQ files were uploaded into the Genboree workbench for analysis using ExceRpt (a small RNA alignment workflow). Reads were trimmed of adapter sequence and aligned to both miRNA and genomic sequences as well as filtered for UniVec contaminants and repeat sequences. Raw counts were then produced as a text file. All alignment information was recorded in an additional text file for observation of counts that had passed through the different filtering and alignment steps. As this Chapter involved optimisation of the samples processing method, ultimately only one sample was sequenced and therefore the raw counts were migrated into R to allow comparison with the published literature (Chapter 2, Section 2.1.5.1. and Section 3.2.5. below). The comparison comprised of filtering miRNAs with more than 100 reads and identifying matches based on miRNA names and numbers. Some data manipulation of the literature sources was required to ensure miRNAs were in the same format (e.g. miR and not mir, and miR-456 not 456).

3.2.4. Literature Review

A literature review was conducted for comparison of miRNAs identified from sequencing of the single Nasosorption sample as no internal comparison was possible. Full details of the process of selecting matches with this study are detailed in Chapter 2 (Section 2.1.5.3). There were no studies identified where miRNA expression was researched in solely healthy

individuals. To enable comparison, a search of the literature was conducted where healthy controls were compared with asthmatics (therefore miRNA expression of compared miRNAs would have to be present in sufficient quantity for comparison). As the goal is to develop this methodology for comparing healthy controls to asthmatics it was viewed that there might be comparable expression of miRNAs between the nasosorption (NS) sample of this study and a study comparing healthy controls and asthmatics.

3.3. Results

3.3.1. RNA Isolation

The two goals of the experiments performed in this Chapter were interlinked. The first was to optimise a protocol for the successful isolation of miRNAs from nasal samples that were collected by NS using Synthetic Absorptive Matrix (SAMs). The next goal, as at the date of this experimental work miRNA profiling had been array based (Chapter 1, Section 1.2.2.), was to determine if the RNA obtained was of sufficient quantity and quality for miRNA profiling by next generation sequencing. If this proved possible and miRNAs present were identifiable and at a sufficient level this would support that exploration with larger sample numbers (subsequent experiments) was worthwhile. The focus of the said experiments being to identify differentially expressed (DE) nasal miRNAs between healthy control samples versus those from patients with asthma.

The next sections outline the various modification and results as the processing protocol was developed and optimised. Only for the first protocol (Processing Protocol 1)

experiments were conducted with both kits – Monarch total RNA (including small RNA) isolation kit and the equivalent kit manufactured by Zymo (Chapter 2, Section 2.1.2.1.). For subsequent optimisation (Processing Protocols 2, 3 and 3a) the Zymo kit was adopted.

3.3.1.1. Processing Protocol 1.

Figure 3.1. details the initial protocol that was conducted regarding sample handling and storage, as well as the practical usage of the RNA extraction kits. There was some evidence in the literature to suggest that miRNAs are more stable than mRNA and could be placed on ice without risk of degradation (Glinge *et al.*, 2017). As sampling would need to be done in a clinical setting, the first processing protocol (Figure 3.1.) attempted to allow the nasosorption SAMs (one SAM from each nostril of a subject) to be placed on ice after sampling for up to an hour prior to the subsequent steps being carried out within the laboratory. Chapter 2 (Section 2.1.1.1.) provides details of the sampling methods and equipment used.

Post collection, and using sterile autoclaved scissors, each nasosorption SAM was detached from its holder (device) and the two SAMs (one from each nostril of each subject) were placed together in a Costar spin tube. These tubes have a removable inner casing (basket) with a sieve-like plastic mesh filter at the base to allow fluids to pass through whilst retaining the SAMs. Two hundred μL of RNAlater (Sigma-Aldrich, Darmstadt, Germany) were added by pipetting with incubation of the sample at room temperature for 1 minute. RNAlater is an agent that ensure immediate RNase inactivation hence maximising integrity

of RNA that can be obtained through extraction. The sample was then centrifuged for 20 minutes at 1,600 g (Labnet Prism microcentrifuge).

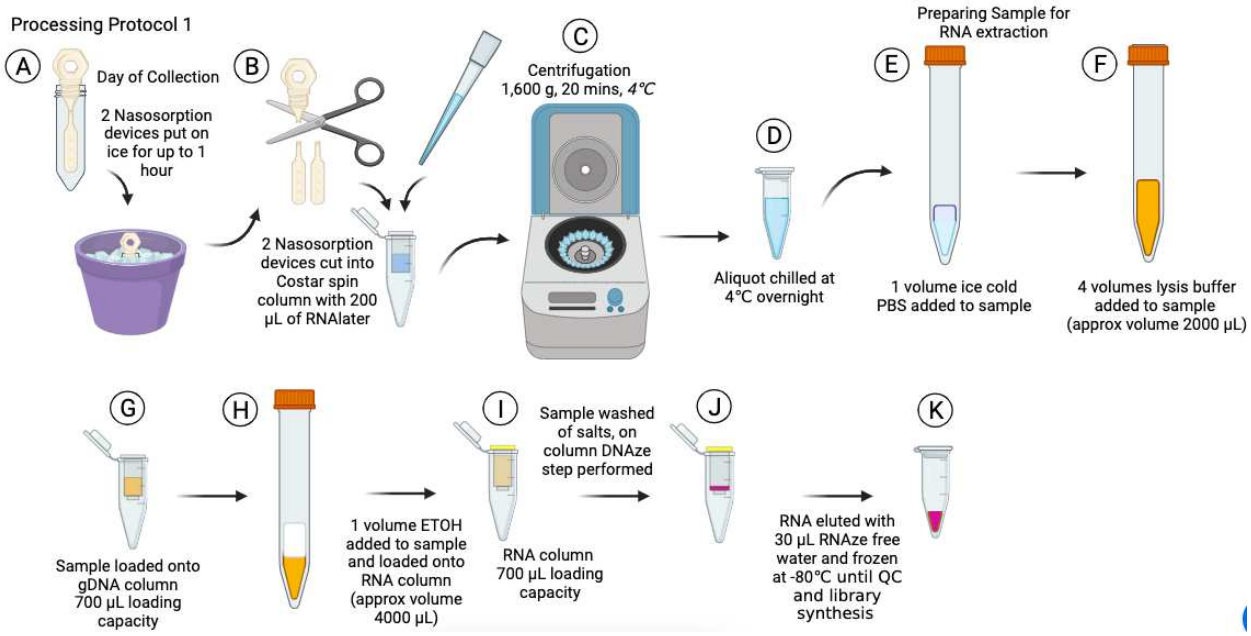


Figure 3.1. ***Outline of Sample Processing Protocol 1. A.*** *Samples were placed on ice after collection for up to one hour.* ***B.*** *The nasosorption devices were trimmed from the holders with sterile scissors and the sample placed into a Costar spin column (mesh, without filter). Two hundred µL of RNAlater was added to the nasosorption filters.* ***C.*** *The samples were centrifuged at 1,600 g for 20 minutes at 4℃.* ***D.*** *Samples were then stored overnight at 4℃.* ***E.*** *Note the transfer of sample to a 5 mL falcon tube as multiple volumes were added to the sample in RNAlater. The sample was then prepared for RNA extraction by the addition of one volume of ice-cold PBS,* ***F.*** *followed by four volumes of lysis buffer.* ***G.*** *The sample and buffer mix were then loaded onto a gDNA column 700 µL at a time. The gDNA column was then centrifuged for 1 minute at 20,000 g.* ***H.*** *The flow through was transferred to a 5 mL falcon tube.* ***I.*** *Further loadings of the sample buffer mix were then loaded onto the gDNA column before being centrifuged until all the sample-buffer mix had been passed through the column. One volume of ethanol was then added to the captured gDNA flow through sample. The loading and centrifugation process was then repeated with an RNA column, this time the flow through was discarded as the RNA was captured onto the silica matrix filter of the column.* ***J.*** *The sample was then washed to remove salts and other impurities (using proprietary reagents provided within the kit)* ***K.*** *The sample was then eluted using 30 µL RNAse free water and frozen at -80℃ in cryotubes prior to QC and downstream experiments.*

The spin column and SAMs were discarded, and the resulting sample stored overnight at 4°C. The next day one volume of ice-cold PBS was added followed by 4 volumes of lysis buffer (provided in the kit and following the manufacturer's kit instructions). This resulted in the volume of each sample in buffer mix to be between 2,500-3,000 µL (2.5 to 3 mL). With the further addition of one volume of ethanol it resulted in a volume between 5,000 to 6,000 µL (5 to 6 mL). This meant that post discarding the spin column and SAMs the sample was transferred to either a 5 or 15 mL falcon tubes for these additions. It also meant that for the subsequent extraction steps multiple loadings onto the gDNA and RNA columns had to be performed as the maximum loading volume of the columns was 700 µL.

Firstly, different total RNA extraction kits (Monarch and Zymo) were compared for reliability and highest yield. Total RNA extracted from Nasosorption samples collected using Processing Protocol 1 was quantified by both the Nanodrop (Chapter 2, Section 2.1.3.1.) and the Qubit 4 (Chapter 2, Section 2.1.3.2.). The Qubit 4 detected no RNA in all tested samples (below minimum range) and the Nanodrop (Figure 3.2.) very low concentrations. The NanoDrop spectrophotometer is a commonly used instrument for quantifying RNA but it can overestimate concentrations due to the presence of contaminants like phenol, guanidine, or protein that absorb at 260 nm. On the other hand, the Qubit 2.0 fluorometer uses fluorescent dyes that bind specifically to RNA providing more accurate quantification with less interference from contaminants. The Qubit 2.0 however has a lower sensitivity (although higher detection) limit compared to the NanoDrop. This means it is less suitable for quantifying low concentrations of RNA, hence at very low concentrations the Qubit was unable to quantify the samples shown in Figure 3.2 below.

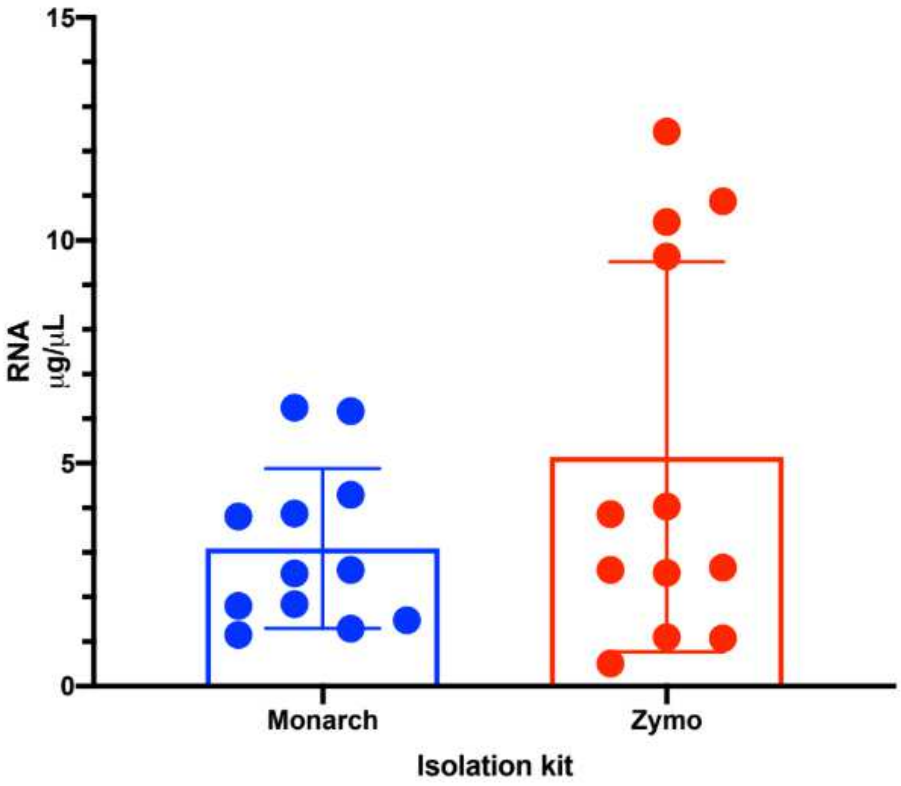


Figure 3.2. ***Total RNA Extractions Conducted from Nasosorption Samples Using two Different Total RNA extraction Kits*** *– one manufactured by Monarch the other by Zymo. Nanodrop quantification of total RNA isolated with Processing Protocol 1, N = 12 individuals (healthy) (biological repeats) 2 NS SAMs combined, not significant in Wilcoxon signed rank test (P=0.15).*

3.3.1.2. Processing Protocol 2.

MiRNA makes up approximately 1% of total RNA. While kits for preparing libraries for next generation sequencing are optimised for low inputs, Processing Protocol 1 had proved to be unsuccessful and had not yielded sufficient RNA for sequencing. Processing Protocol 1 had assumed that the sample, once obtained, would be stable until processing one hour after collection (Figure 3.1.). Although small RNAs may or may not be more stable than larger

species, such as mRNA, as total RNA was the input measure for later steps (library synthesis) it seemed key to preserve as much total RNA as possible. As such, an adaptation of Protocol 1 was made whereby RNAlater (Sigma-Aldrich, Darmstadt, Germany), that permeates tissues (in this case the SAM) and can inactivate RNases protecting RNA present in the sample, was added earlier as soon as the Nasosorption sample was collected (Figure 3.3.).

Therefore upon removal of the SAM from the inferior turbinate, the SAM was immediately cut from its holder (as was the SAM taken from the other nostril) and then both were placed together into a 1.5mL Eppendorf containing 200 μL of RNAlater. As sampling was done in a clinical sampling setting this required appropriate risk assessment and supply in advance of eppendorfs containing the RNAlater. Post transfer of the SAMs to the 1.5 mL Eppendorf, this was then stored at room temperature until elution following the rest of Protocol 1.

In addition to the change at the point of sampling, it was decided to quantify the RNA extracted from the Nasosorption samples using the Bioanalyzer (small RNA chip) (Chapter 2, Section 2.3.1.1.). Use of the Bioanalyzer would also enable quality-and-size-assessment of the extracted RNA samples and therefore the concentration of miRNA present within each sample to be determined.

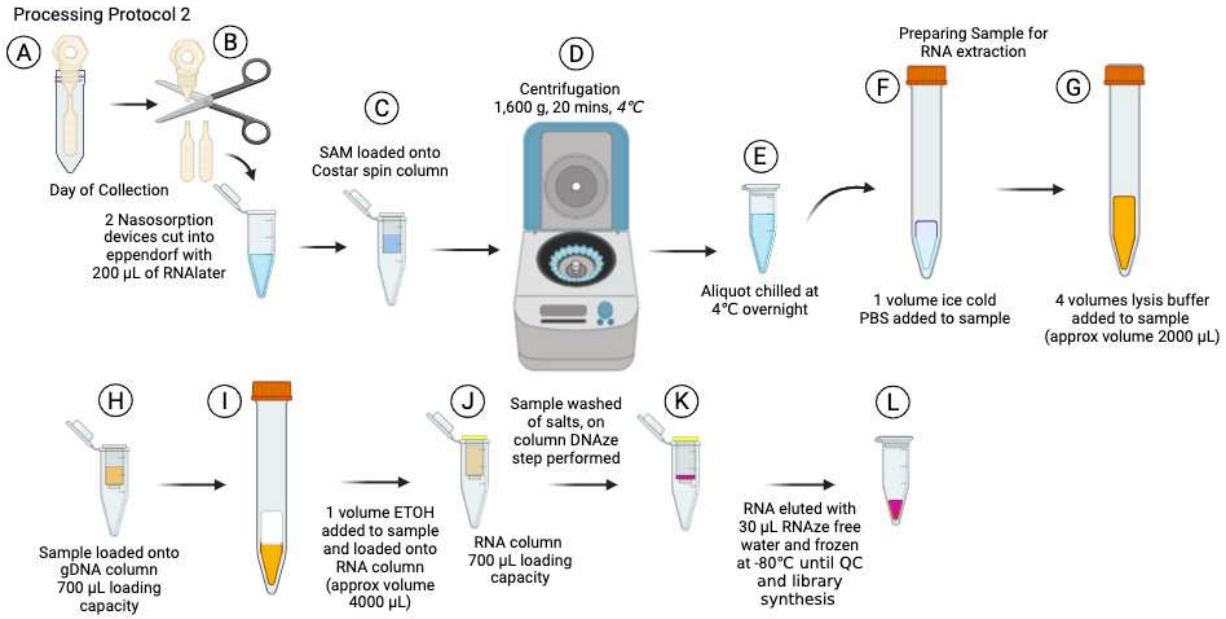


Figure 3.3. Outline of Sample Processing Protocol 2. A*. Nasosorption samples were added to RNAlater immediately after sampling was performed, and processing continued as Protocol 1.* ***B****. This involved the nasosorption devices (SAMs) being trimmed from their holders with sterile scissors.* ***C****. Then SAMS and any liquid were transferred into a Costar spin column (mesh, without filter).* ***D****. The samples were centrifuged at 1,600 g for 20 minutes at 4℃.* ***E.*** *The sample was then stored overnight at 4℃.* ***F.*** *Next the sample was prepared for RNA extraction by the addition of one volume of ice-cold PBS.* ***G****. followed by four volumes of lysis buffer.* ***H****. The sample and buffer mix were then loaded onto a gDNA column 700 µL at a time. The gDNA column was then centrifuged for 1 minute at 20,000 g and the flow through transferred to a 5 mL falcon tube. Further loadings of the sample buffer mix were then loaded onto the gDNA column before being centrifuged until all the sample-buffer mix had been passed through the column.* ***I****. One volume of ethanol was then added to the captured gDNA flow through sample.* ***J****. The loading and centrifugation process was then repeated with an RNA column, this time the flow through was discarded as the RNA was captured onto the silica matrix filter of the column.* ***K****. The sample was then washed of salts and other impurities (using proprietary reagents provided within the kit).* ***L****. The sample was then eluted using 30 µL RNAse free water and frozen at -80°C in cryotubes prior to QC and downstream experiments.*

Despite the modification at the sampling stage, a low concentration (and therefore yield – elution volume 30 µL) of miRNA was obtained from the Nasosorption samples. An advantage of running the samples on the Bioanalyzer was that size selection of fragment size was possible - ~150 bp in the case of miRNA – and exact concentration of miRNA

species determined as well as overall total RNA yield. Figure 3.4. gives an example of the trace obtained for one sample and as can be seen concentration was less than 1ng/ µL for the miRNA peak.

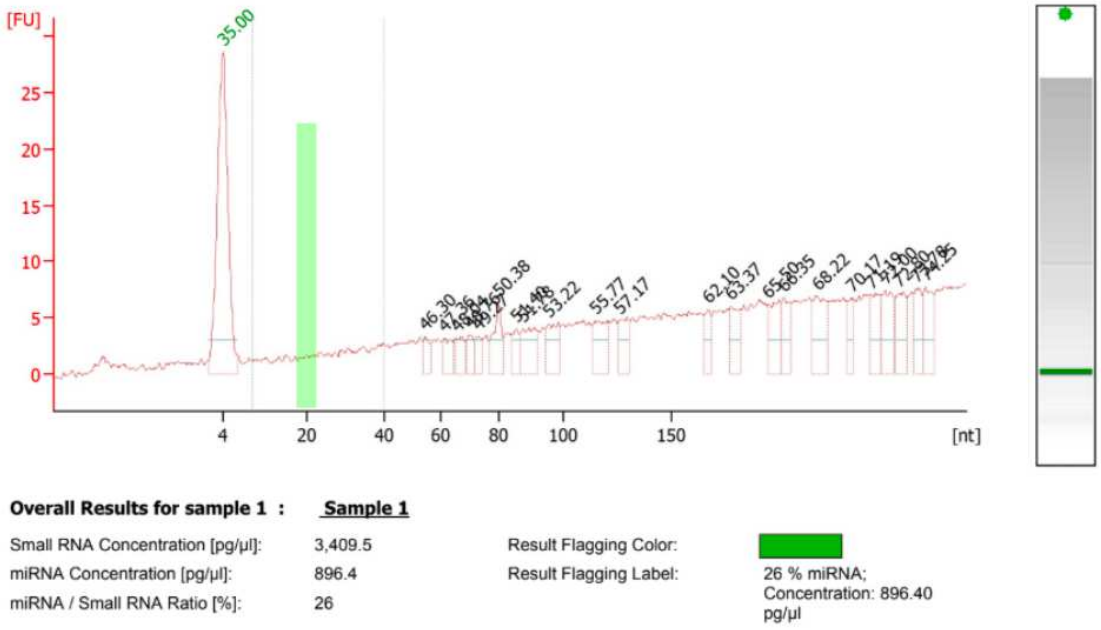


Figure 3.4. Small RNA Bioanalyzer Trace of Extracted Total RNA Sample from Protocol 2. *Although miRNA constitutes 26% of the total RNA as can be seen it is of very low concentration - less than 1 ng / µL miRNA - and therefore yield. The large peak is the marker, the blue lines identify the selected range of measurement, and the green bar indicates the range that miRNA would be expected to be observed.*

3.3.1.3. Processing Protocol 3.

Discussions (verbal) with the manufacturer of the RNA extraction kit (Zymo) identified that a gDNA column could significantly reduce RNA isolation in a low RNA sample, and that the loading of any spin column was ideally between one and two loadings. Due to the volumes,

as indicated earlier, for Protocols 1 and 2 the number of loadings had been between 5 and 6. To therefore reduce the number of loadings onto the RNA column, RNAlater was substituted with an alternative stabilising agent RNA/DNA shield (Zymo Research, California, USA). In combination with the Zymo extraction kit smaller volumes of reagents would therefore be required thereby minimising the number of loadings onto spin columns but nonetheless providing comparable RNA protection (Soliman, 2018). Lastly, it was noted that irrespective of which protectant fluid was used (RNAlater or RNA/DNA shield) due to their viscosity and the volume added the fluid tended to settle into the tip of the bottom of the Eppendorf tube and complete absorption into the nasosorption SAM device was potentially not occurring. Therefore, for Protocol 3 round-bottom Eppendorfs, that enabled the SAM to contact the tip of the Eppendorf, were used. Figure 3.5. shows the adapted Processing Protocol 3 with the aforementioned changes.

So in summary, the sampling step for Protocol 3 was: upon removal of the SAMs from the inferior turbinate, the two SAMs (one from each nostril of the subject) were detached from their devices and placed into a 2 mL round bottom Eppendorf tube containing 200 μL RNA/DNA Shield (Figure 3.5.). The Eppendorf was then flicked to ensure complete saturation of the SAMs. The sample was then kept at room temperature until processing (<24 hours).

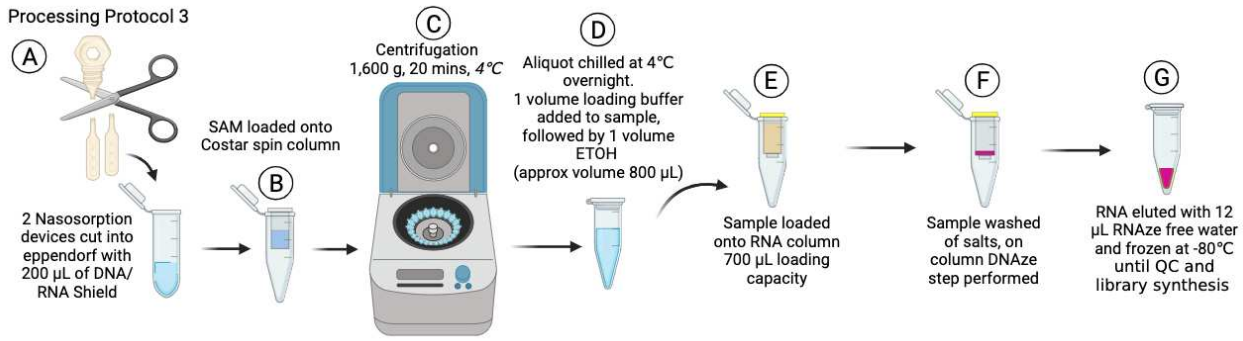


Figure 3.5. Outline of Sample Processing Protocol 3. A. *Featuring the round bottom Eppendorf tube for collection, with the RNA/DNA Shield, and the absence of the gDNA column step. Immediately post sampling, the nasosorption devices were trimmed from the holder with sterile scissors into a 1.5 mL Eppendorf containing 200 µL of RNA/DNA Shield. The stabilised sample was stored at room temperature until processing (< 24hr).* ***B.*** *The sample was then transferred into a Costar spin column (mesh, without filter), any liquid remaining in the original collection eppendorf was transferred into the spin column with mesh. The sample was then centrifuged at 1,600 g for 20 minutes at 4℃.* ***D.*** *The samples was then stored overnight at 4℃. Next the sample was prepared for RNA extraction by the addition of one volume of loading buffer followed by one volume of ethanol.* ***E****. The sample and buffer mix were then loaded onto an RNA column with a capacity of 700 µL. The RNA column was then centrifuged for 1 minute at 20,000 g and the flow through discarded as the RNA was captured onto the silica matrix filter of the column.* ***F****. The sample was then washed of salts and other impurities (using proprietary reagents provided within the kit)* ***G****. The sample was then eluted using 12 µL RNAse free water and frozen at -80℃ in cryotubes prior to QC and downstream experiments. Figure created using Biorender.*

The changes implemented in Processing Protocol 3 had a notable impact with substantial increase in RNA yield. Figure 3.6. shows an example Bioanalyzer trace showing a peak at 20 nucleotides of a concentration of ~26 ng/ µL. Yields ranged from 17 ng/ µL to 28 ng/ µL (mean ± SEM = 24 ± 2.4) for miRNA specific peaks across 4 samples.

Table 3.1. Procesing Protocol 3 Concentrations at 20 Nucleotides from Small RNA Bioanalyzer Analysis

Sample Number	miRNA Peak Yield
1	26
2	17
3	28
4	25

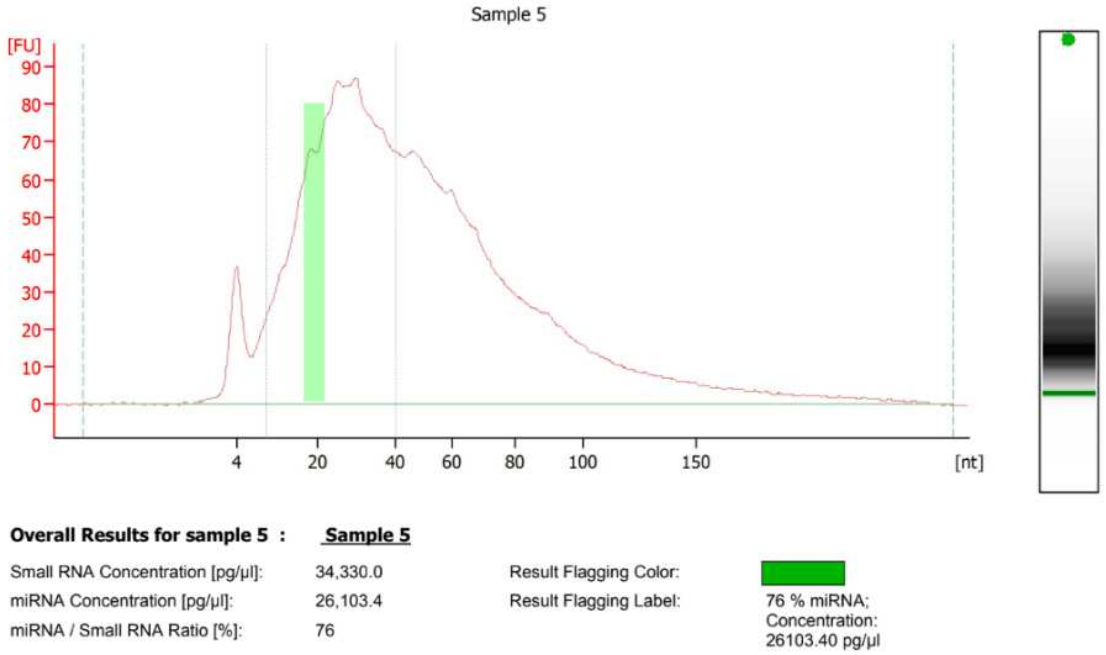


Figure 3.6. Small RNA Bioanalyzer Trace of Nasosorption Sample. *RNA isolated using Sampling Protocol 3. miRNA concentration identified at 26 ng / µL. Example trace of one of four samples processed using Processing Protocol 3.*

3.3.1.4. Processing Protocol 3a

One further modification was made to Processing Protocol 3. Substitution of the V- bottom Eppendorf tube with a round-bottom Eppendorf had enabled more protectant fluid to be used more efficiently meaning less was required. Protocol 3 was therefore further amended reducing the volume of RNA/DNA Shield to 100 µL, as well as chilling samples in RNA/DNA Shield at 4°C for 24 hours followed by freezing at -80°C until RNA extraction (Figure 3.7.).

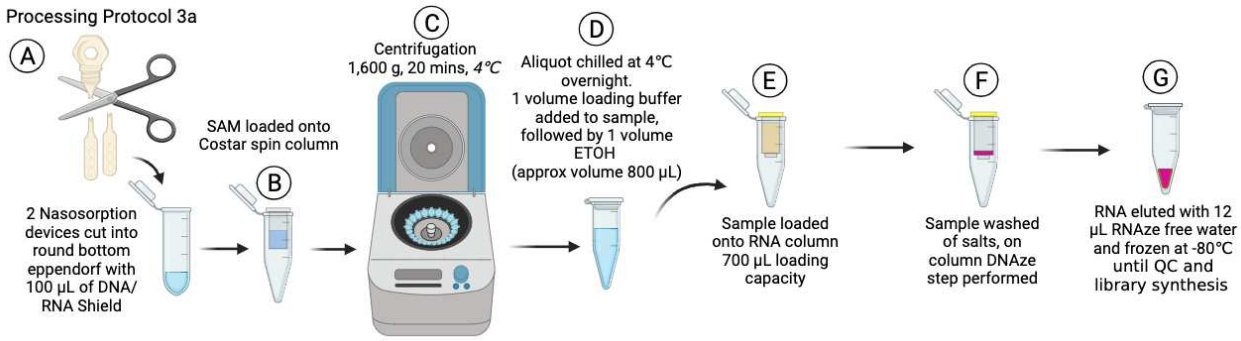


Figure 3.7. Outline of Sample Processing Protocol 3a. A. *Immediately after sampling the Nasosorption strips were trimmed from the device and added to 100 µL of DNA/RNA Shield in a round bottom 200 µL Eppendorf.* ***B.*** *The sample SAM and liquid transferred into a Costar spin column (mesh, without filter) and centrifuged at 1,600 g for 20 minutes at 4℃.* ***D.*** *Samples were then stored overnight at 4℃. Samples were then prepared for RNA extraction by adding one volume of loading buffer to the sample, followed by one volume of ethanol.* ***E.*** *The sample and buffer mix were then loaded onto an RNA column with a capacity of 700 µL. The RNA column was then centrifuged for 1 minute at 20,000 g and the flow through discarded as the RNA was captured onto the silica matrix filter of the column.* ***F.*** *The sample was then washed of salts and other impurities (using proprietary reagents provided within the kit)* ***G.*** *The sample was then eluted using 12 µL RNAse free water and frozen at -80℃ in cryotubes prior to QC and downstream experiments. Figure created using Biorender.*

Total RNA quantification using Qubit (Figure 3.8.) revealed RNA yields from Nasosorption samples using Processing Protocol 3a of between 90 ng/ µL and 200 ng / µL (3 out of the 4 samples being ≥ 100 ng/ µL. As the Illumina minimum library synthesis requirements are 100 ng in 6 µL Processing Protocol 3a was used for the final RNA isolation of the sample that was to undergo the pilot sequencing.

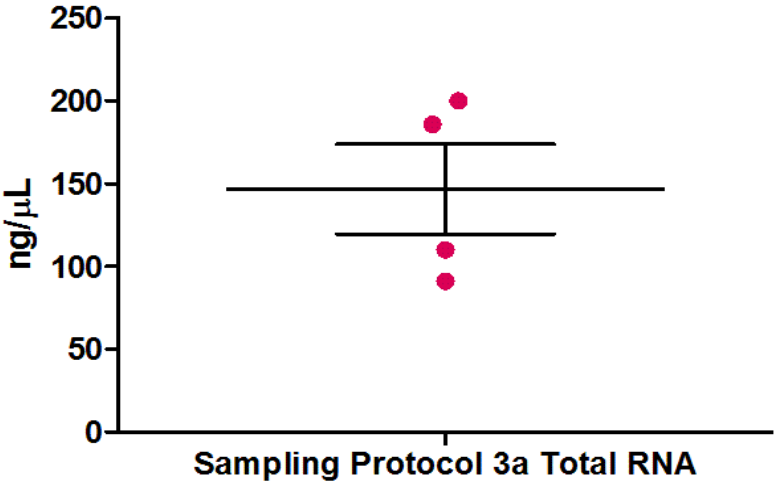


Figure 3.8. Summary RNA Concentrations of Samples from Sampling Protocol 3a. *Measured with Qubit (N = 4) biological repeats. Total RNA quantities fall between 90 ng/ µL and 200 ng/ µL.*

A schematic to highlight the changes in protocol sample handling and sample stabilizing development as described above is shown in Figure 3.9. The focus of the optimisation protocol was sample handling prior to starting the manufacturer optimised protocols for the RNA isolation kits. This handling was crucial as it impacted yield, going from insufficient yield to sufficient total RNA for downstream small RNA library synthesis for next generation sequencing. The gDNA column was retained as it was optimal for the expected yield.

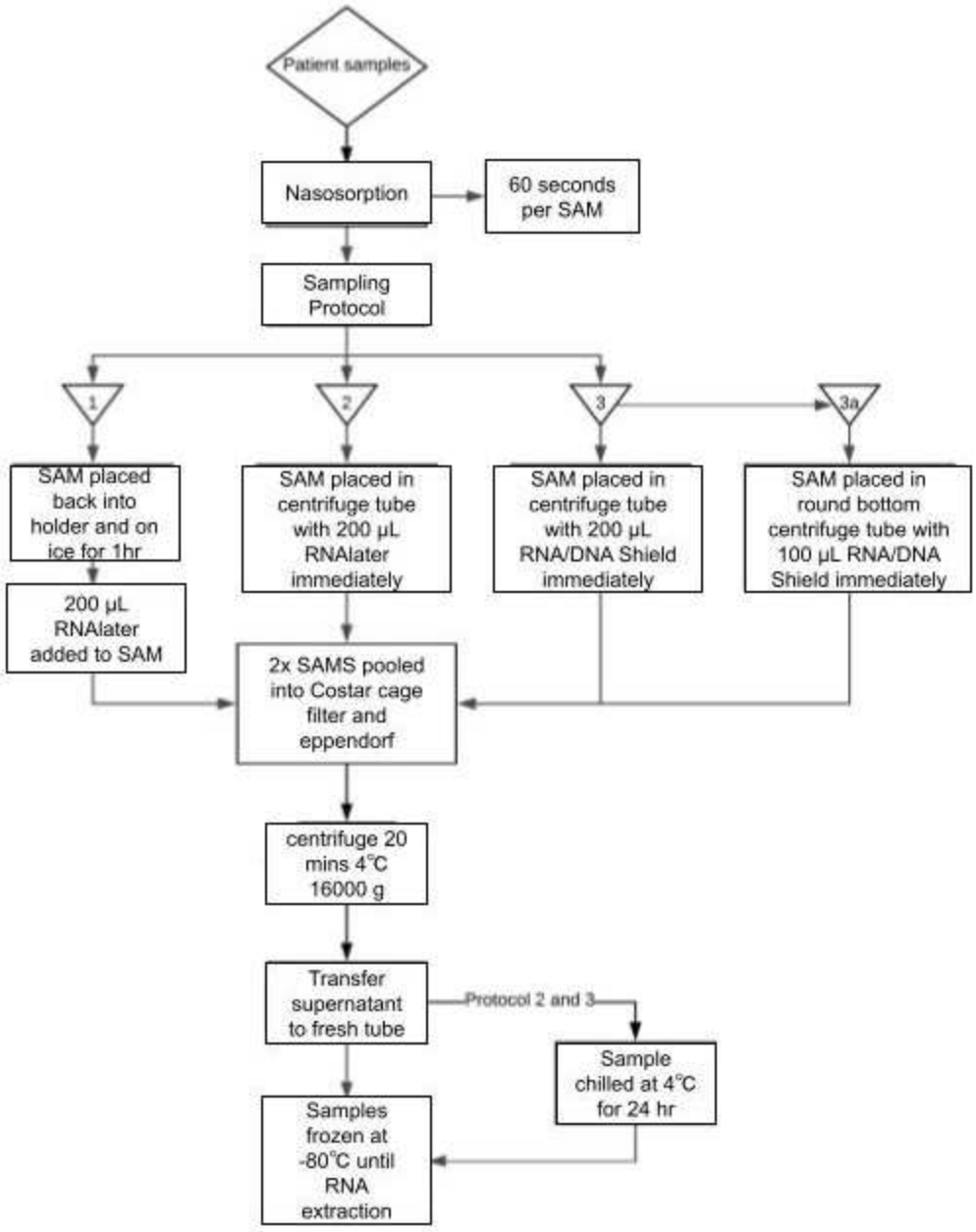


Figure 3.9. Schematic of Sampling Handling and Stability Adaptations. *Protocols numbers are within triangles denoting changes in protocol around the recommendations from the manufacturer of the Zymo total RNA isolation kit. Nasosorptions were conducted with the Synthetic Absorptive Matrix (SAM) device.*

3.3.2. Library Synthesis

Following on from optimisation of the sample processing protocol to the end point of extracted RNA, the next step was to synthesise a cDNA library and size select for miRNA species (that were reverse transcribed into cDNA with the corresponding adapters and indices). Full details of the library synthesis process are given in Chapter 2, Section 2.1.4. with a brief summary in Section 3.4 above. The process the sample needed to go through prior to sequencing, was conversion into cDNA with the addition of adapters and an index that would enable the sample to be identified at the sequencing facility post sequence generation. As the focus was specifically miRNA and due to the size of adapter dimers being so close in size to the miRNA fragments, size selection was needed. Then depending on the concentration of the library a further concentration step can be required to meet the minimum sequencing concentration as set out by the Sequencing Facility (Nanogene in this instance). QC steps on the Agilent Bioanalyzer (Chapter 2, Section 2.1.3.3.) were used to identify if the mentioned aspects above were successfully achieved.

As NS samples were not specifically covered in the manufacturer guidance it was considered that optimisation might improve the chance of success. Libraries synthesised for performing optimisation experiments were from air liquid interface (ALI) total RNA the latter extracted by a colleague within the Genomic Medicine Group at NHLI (Ms Verdiana Martello). Once an initial protocol was optimised, it was then tested on several Nasosorption healthy subject samples to validate that a clean sequencing compatible library was produced. The rationale for this was two-fold. First, to establish using the same kits and equipment what number of cycles of PCR amplification was optimum for the 100

ng total input of RNA. Second to establish, within the method of small RNA library focus, what ranges were optimum for size selection (see Section 3.3.3. below) to isolate the reverse transcribed miRNA from other RNA species that may be present.

3.3.3. Size Selection

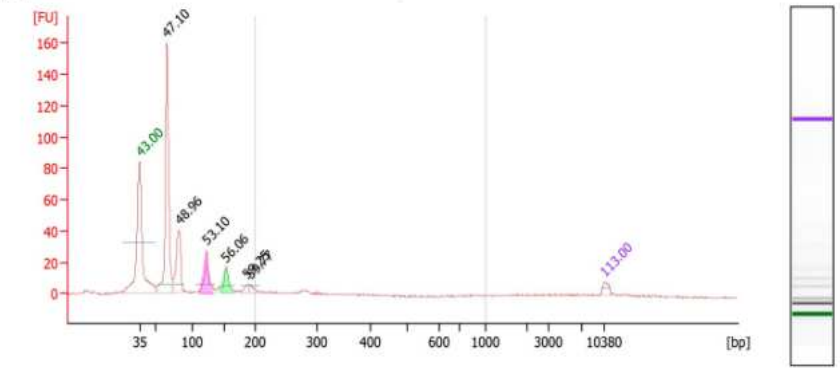
Small RNA, such as miRNA, and library synthesis carry unique complications. The adapter sequences used to enable binding to the Illumina flow cell are of a dictated length. The length of the adapters alone is 120 bp. The length of miRNAs are between 18 and 25 nucleotides. In addition to the index sequence of 6 nucleotides, the complete length of an Illumina-ready miRNA sequence is around 150 base pairs. This gives only a 30 base pair difference between success and failure of the run and the importance of the size selection step. Due to such a fine threshold for success, the Pippin Prep size selector (Sage Science Ltd) was used to isolate the cDNA as it could be set to enable a (claimed) 97% accuracy in relation to fragment size.

Sage Science Ltd, the manufacturer of the Pippin Prep, (following consultation with their technical support) gave some initial recommendations of range for the library and therefore for optimisation a range broader than the standard recommendation (125-174 bp) was tested, all the way to 112-186. Figure 3.10. A. below shows a typical cDNA raw library trace showing the markers with green and purple labels at 35 and 10,380 bp and peaks in between denoting the concentrations and sizes of specific cDNAs present in the sample. The pink highlighted peak shows the adapter dimers and the green highlighted peak shows the miRNA cDNA range. The insignificant size difference between the two is

one of the key challenges of miRNA sequencing as adapter dimers must be removed completely for successful sequencing. Therefore, the goal of optimisation was to identify the best range for this specific method regarding two aspects of the size selection, firstly to acquire the highest yield possible, and secondly to ensure no adapters or other sizes of cDNA strands made it through at that range. Figure 3.10. B. below shows the ranges tested. At the recommended range the expected yield from the input NS RNA library was 20%.

Due to the limited quantities of NS RNA available, this optimisation, as mentioned above, was conducted using ALI sourced RNA that had been converted into cDNA (using the same method as that for the NS sample). By using the ALI sourced RNA there was also the option to use a higher quantity of input cDNA to better assess the removal of unwanted DNA fragments from the size selection process. The actual yield when the size range was increased (denoted in Figure 3.10. B.) resulted in a decrease to 26.6% at the broadest range, where adapter dimers were also identified. At the range narrower than this (116-182), while adapters were not drawn through into the elution well with the sample library, the yield, somewhat counterintuitively, was only 28.9%. The original range of 125-174 bp was therefore adopted for the library clean-up. Figure 3.10. C., below, shows the bioanalyzer trace of the result of size selection applying that range.

A.



B.

Lane	Start	Stop	Yield %	Adapter detected
1. (current setting)	124	174	46	no
2.	120	178	29.5	no
3.	116	182	28.9	no
4.	112	186	26.6	yes

C.

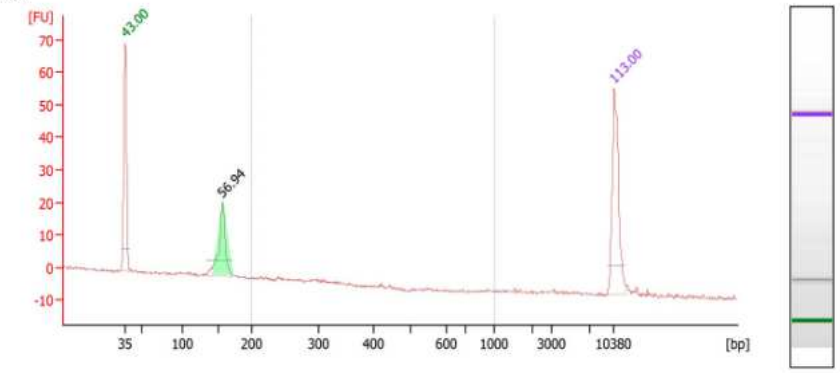


Figure 3.10. cDNA Library Isolation Optimisation A. *High sensitivity DNA Bioanalyzer trace of Air-Liquid Interface culture sourced total RNA converted to cDNA. Markers at 35 bp and 1380 bp shown denoting the maximum and minimum measurable range of DNA fragment length (please note a peak marker error occurred on this run and the peaks are incorrectly labelled). Raw library trace, 122 bp peak (pink) (adapter dimers) 672 pg / μL and 149 bp peak (green) (miRNA) 412 pg / μL.* ***B.*** *Table showing results of Pippin Prep optimisation of range parameters. 149 bp peak 107 pg / μL. N = 1 per range.* ***C.*** *HS DNA Bioanalyzer trace of ALI library after Pippin Prep following 124bp start -174 bp stop range.*

With the final range optimised, RNA isolated from Nasosorption samples were converted into libraries and size selected. Example Bioanalyzer traces (pre and post size selection) of the result from a NS cDNA library are shown in Figure 3.11. As can be seen the majority of the adapter dimer peak, highlighted in pink in Figure 3.11. A. has been removed through the size selection with successful elution of the miRNA peak (in green – Figure 3.11. A. and B.).

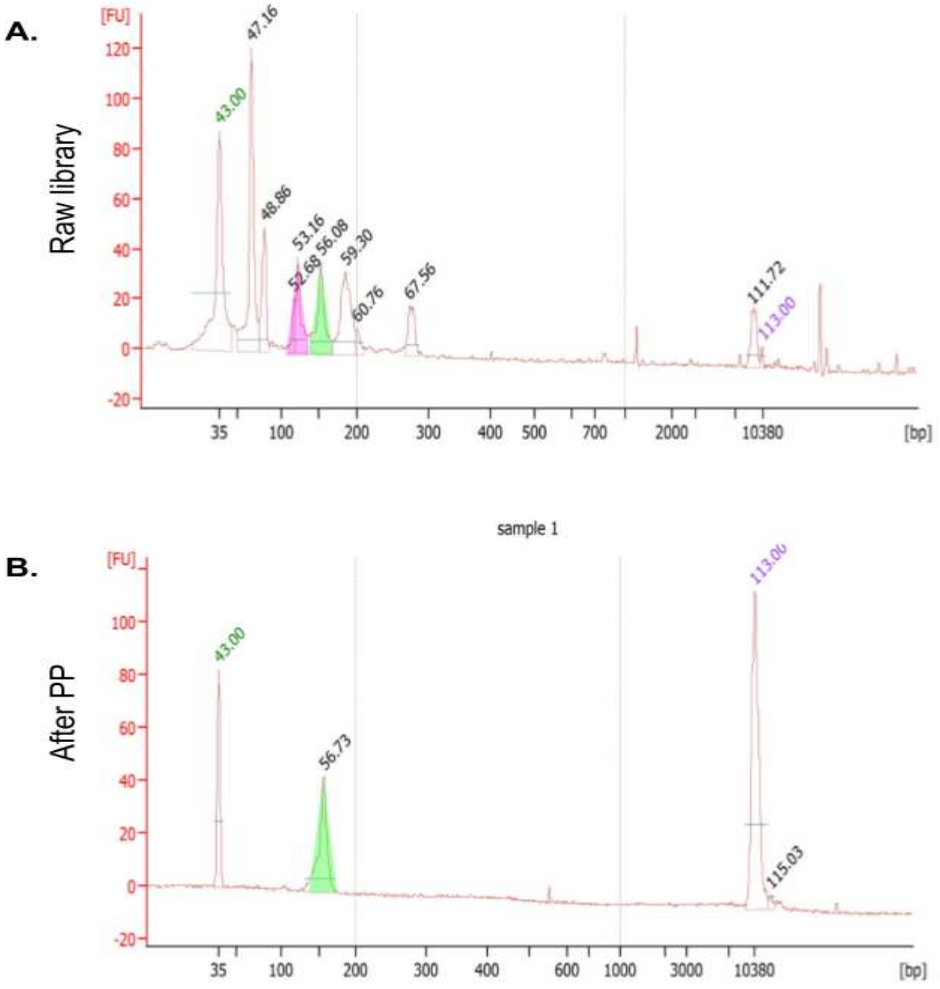


Figure 3.11. High Sensitivity DNA Bioanalyzer Trace of cDNA Library from Nasosorption Samples. A. *HS DNA Bioanalyzer trace of NS sample, species of interest (miRNA library) highlighted green, and adapter dimers highlighted pink.* ***B.*** *HS DNA Bioanalyzer trace after size selection using the Pippin Prep showing adapter dimers were effectively removed leaving only sample cDNA library (of miRNAs).*

3.3.3.1. Library Concentration for Sequencing

The process of size selection enabled isolation of a specific range of cDNA fragment length to meet the expected length of strands that would match miRNA species including the attached adapters necessary for sequencing. The size selection also successfully removed the adapter dimers that if left would severely impact the success of the sequencing run.

The electrophoresis process within the lane of the Pippin Prep also means, however, that there can be a significant loss in yield as the sample has the opportunity to dissipate through the gel. Sage Science notes that the Pippin Prep recovery is approximately 20% (Munafo *et al.*, 2013). The volume is also affected since the elution volume is larger than the input volume. Collectively this means that the output yield (concentration) of samples that have undergone size selection is low.

Based on optimisation tests using the ALI sourced RNA input at the same concentration as that available from Nasosorption RNA isolation (100 ng/ μL), it was identified that the size selected cDNA libraries would be too low a concentration to meet the minimum concentration input set by the Sequencing Facility (500 pg/μL Novogene). As the volume needed by Novogene was well below the Pippin prep output it was possible to remedy this issue by concentrating the sample by reducing the overall volume of the library using a DNA concentration kit.

One concern, however, if implementing a concentration step was the potential further loss of sample and impact on overall yield. Therefore, first the DNA concentration kit (NEB Monarch® PCR DNA clean up kit) was tested to determine potential sample loss. A 50bp

ladder was used to identify an accurate percentage of the concentration that was eluted from the clean-up process as it had known input concentrations for a range of fragment lengths. The latter also enabled it to be determined whether yields of fragments of particular size were being impacted by the concentration step.

Figure 3.12. shows the range and concentration of each size on the ladder, as well as Bioanalyzer traces of the result of each experiment and the accompanying calculation identifying the percentage recovered from the initial input. The focus was at 150 bp as the final size selected library was that range. The experiment was conducted once only due to supply chain issues (because of the COVID pandemic). Ideally, an average of three repeats would be conducted to identify a more accurate recovery yield. The experiment identified that nearly all of the sample (passed through the DNA clean up and concentration kit [NEB]) was recovered (99.7%) and 55% of the concentrated sample was then recovered from the Pippin Prep size selection at the selected range, well above the Sage Science reported recovery yield of 20%.

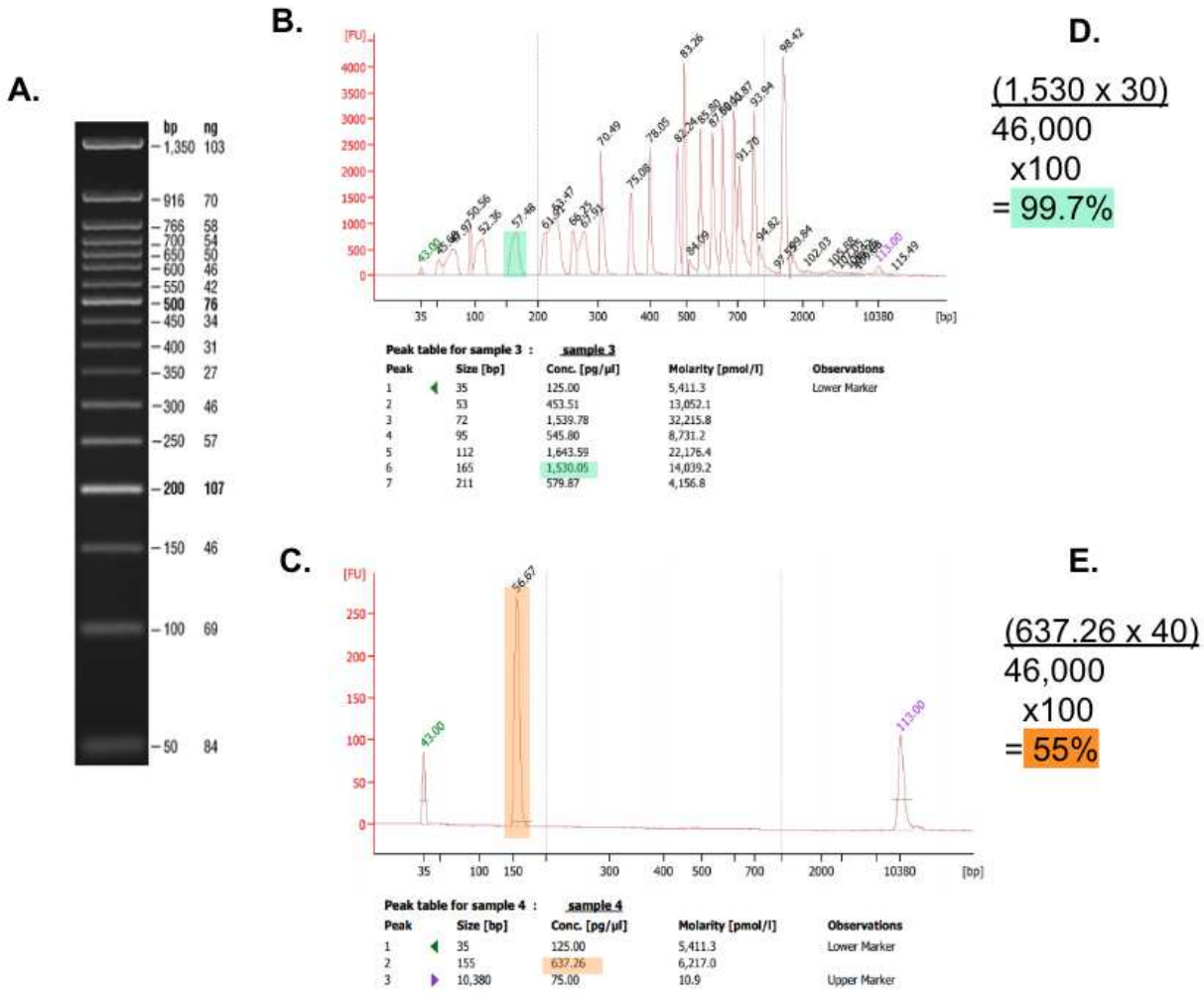
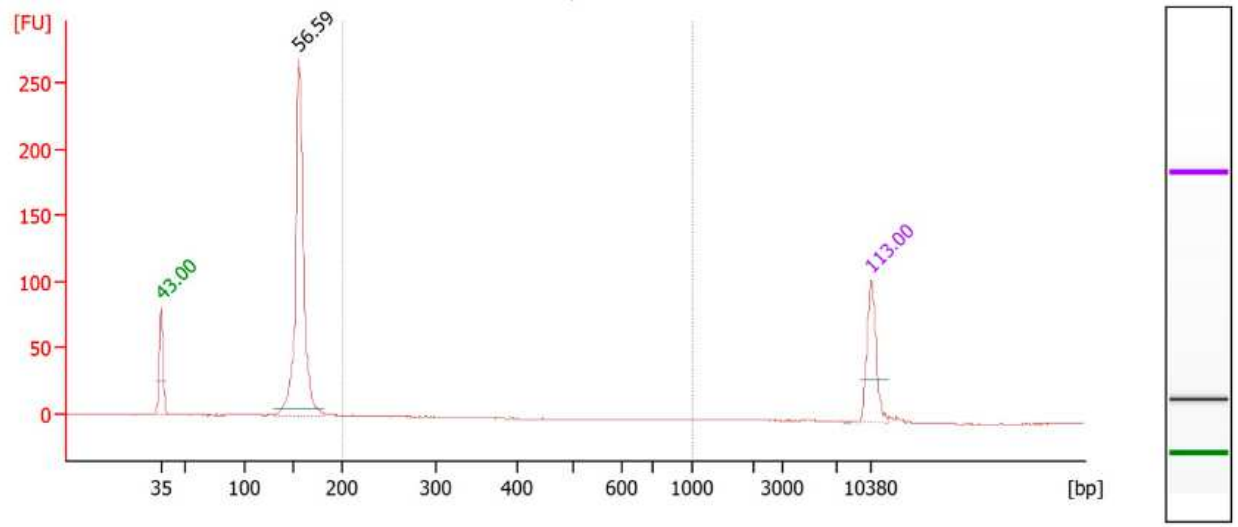


Figure 3.12. cDNA Library Optimisation Clean-up and Size Selection Yields. *Sample yields were assessed for loss of sample through different procedures, Size selection was to be used to finely select the range of cDNA for miRNA and eliminate adapters.* ***A.*** *Manufacturer image (NEB) of 50 bp ladder and input concentrations in 1 µL (as used).* ***B.*** *High sensitivity DNA Bioanalyzer trace showing results of passing 50 bp ladder through NEB DNA clean up kit. Green highlight Identifies range of interest (165 bp on trace) at 1,530 pg / µL.* ***C.*** *HS DNA Bioanalyzer trace of 50 bp ladder after Pippin Prep following 124bp start -174 bp stop range. One hundred and 55 bp concentration 637 pg / µL.* ***D.*** *and* ***E.*** *Yield calculation (concentration identified from trace output multiplied by output volume, divided by input concentration.*

Four RNA samples extracted from Nasosoprtion samples obtained from healthy volunteers that met the minimum input requirements of the library synthesis protocol (100 ng/ μL) were passed through into cDNA libraries. Of the four, one stood out concentration wise being over 600 pg/ μL. Figure 3.13. shows the Bioanalzyer trace of this sample (which was submitted to Novogene for sequencing) with a clean peak at 156 bp. The remaining three samples processed had between 200-400 pg/ μL.



Overall Results for sample 2 : **sample 2**

Number of peaks found:	1	Corr. Area 1:	15.1
Noise:	0.4		

Peak table for sample 2 : **sample 2**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	156	650.38	6,315.3	
3	10,380	75.00	10.9	Upper Marker

Figure 3.13. cDNA library of Sample Submitted for Sequencing. *Sufficient library concentration of sample after size selection was acquired for this sample (>500 pg/ μL) which was submitted for sequencing.*

3.3.4. Sequencing Analysis

The chosen library was sequenced on the Illumina Novaseq platform (sequencing conducted by Novogene. The sequencing was successful with the desired number of reads (5M) achieved, and the sequencing data passed the desired quality control measures (FASTQC) (per sequence quality scores) (Chapter 2, Section 2.1.5.1.). Due to adapters making up a high percentage of the sequenced length, as anticipated there were flags on read diversity where the adapters had been sequenced.

Figure 3.14. below shows an example trace of the expected sequencing read composition. Adapters are comprised of repeated sequences that will be the same for every read, therefore each read will contain the exact same codon in the same location, which comes up as a red flag on FASTQC as this impacts the expected equal distribution of all 4 codons.

Typically, for mRNA sequencing, the sequencer will read fragments of the strand length as the majority are > 150 bp in length. In the case of miRNA sequencing, on a sequencer set to cover 150 bp, the entirety of the strand length will be covered by the sequencer due to the size of the miRNA and its adapters (Chapter 2, Section 2.1.5.). This means that the apparent uneven distribution of bases is even more pronounced. Each of the sequencing reads is comprised of a short sequence of adapter (approximately 13 bases in length). The sample test sequence then follows (this can be miRNA but also other RNA species). Then the adapter, which is then followed by the index (identifying the sample) followed by further adapters (for details see Chapter 2, Section 2.1.5.1.). Figure 3.14. shows a FASTQC trace of an average of all the reads across the single sample sequenced. If the first 13 bases are the

adapter and the following 18-25 bases are the miRNA, this means that the section of interest lies between the 14th and 39th position in the read. While the quality being affected was highlighted in the FASTQC, it was not of concern as the quality score remained high up until the 70th base pair position in the read bases (details on interpretation and percentage scores are in Chapter 2, Section 2.1.5.1).

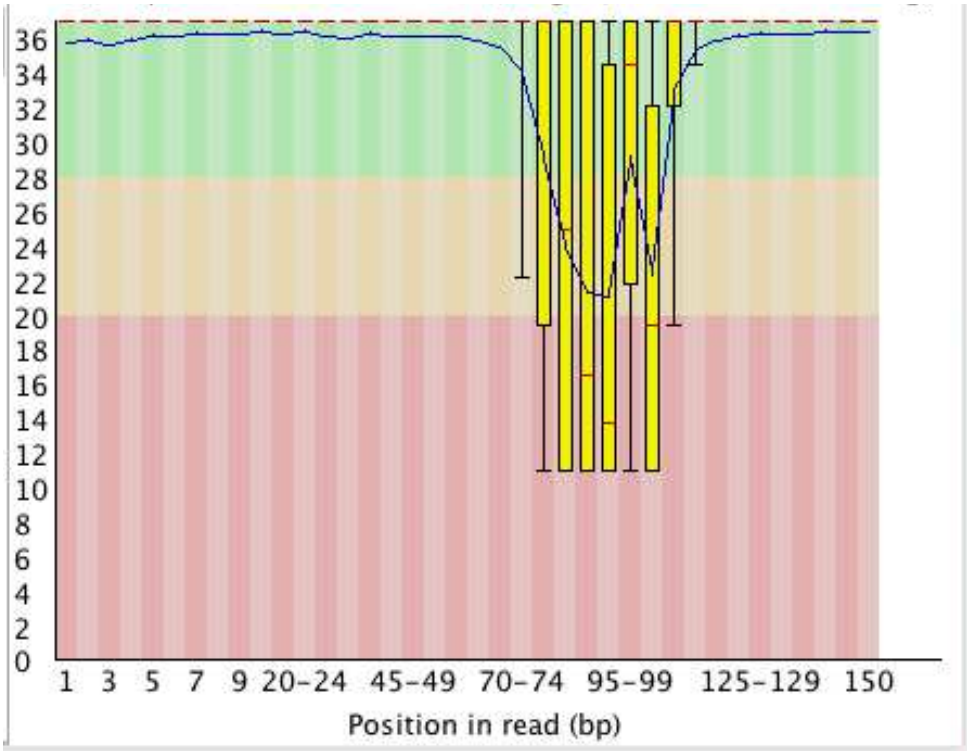


Figure 3.14. FASTQC Result for Per Base Sequence Quality. *The x axis represents the position in base pairs along the read, while the y axis represents the quality score (Q). The central red line is the median value. The yellow box represents the inter-quartile range (25-75%). The upper and lower whiskers represent the 10% and 90% points. The blue line represents the mean quality. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). High read quality was identified up until between 70-74 bases. Adapter sequence at the 5' end (start) of the read is 13 bases, while the miRNA sequences comprise between 18-25 bases, giving sufficient coverage to ensure all reads were high quality.*

The sequencing read files were passed into the alignment software on the Genboree workbench and processed (Chapter 2, Section 2.1.5.1.). An outline of the alignment results are shown in Figures 3.15. and 3.16. A high percentage of reads were successfully clipped of their adapters (98%). A low percentage failed the subsequent quality filter (3%), homopolymer filter, and UniVec contaminant filter. There was a small percentage of reads (6%) that were also removed due to matching ribosomal RNA (rRNA). Nonetheless a high percentage of the reads were successfully used for alignment (89%).

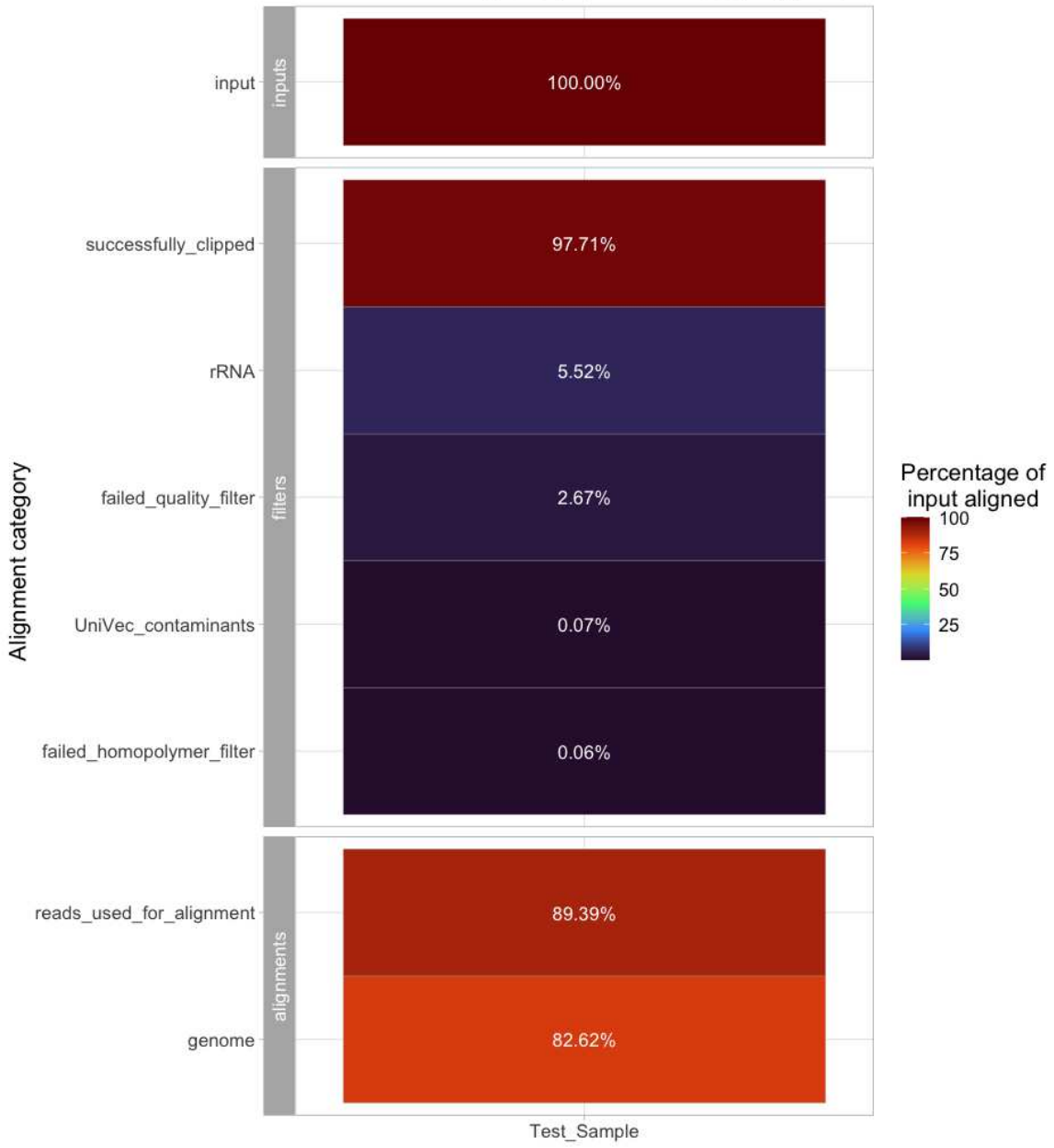


Figure 3.15. Scaled Percentage Filtered Alignment of Raw reads from Healthy Control Nasosorption Sample. *Percentage numbering shown for >1%. Reads were aligned along a pathway of aligners including filtering for UniVec contaminants and repetitive elements. Starting from the input raw reads (100%) those that were successfully clipped (98%) were then passed on through the remaining filters, of which 6% aligned to rRNA, 3% failed quality filter, and less than 0.1% aligned to Univec contaminants and failed the homopolymer filters. With those reads removed, a high percentage of reads were used for alignment (89%). Of these 83% aligned to the human genome.*

Looking at the composition of the reads that aligned to the genome (Figure 3.15.), 235,355 (2%) of the sequencing reads aligned with miRNAs, with tRNA making up the bulk (68%) of the remaining 7,793,019 reads and other RNA species fragments and endogenous genomic transcripts making up just over 12%. A further 7% was found to align to transcripts matching exogenous species and 1.2% of these were exogenous miRNA matches. While miRNAs were as best possible size selected, it was expected that miRNAs might make up a greater portion of the total. tRNAs are between 70-100 nucleotides in length, which would be filtered out during size selection, however, tRNA fragments (tRFs) are between 13-35 nucleotides long and a biologically active regulator in their own right that are derived from tRNAs and this most likely explains the high incidence of them. Due to the size of tRFs, tRNA matches are always conducted prior to miRNAs to eliminate false positive matches to miRNAs and this will be discussed further in the discussion section of this Chapter.

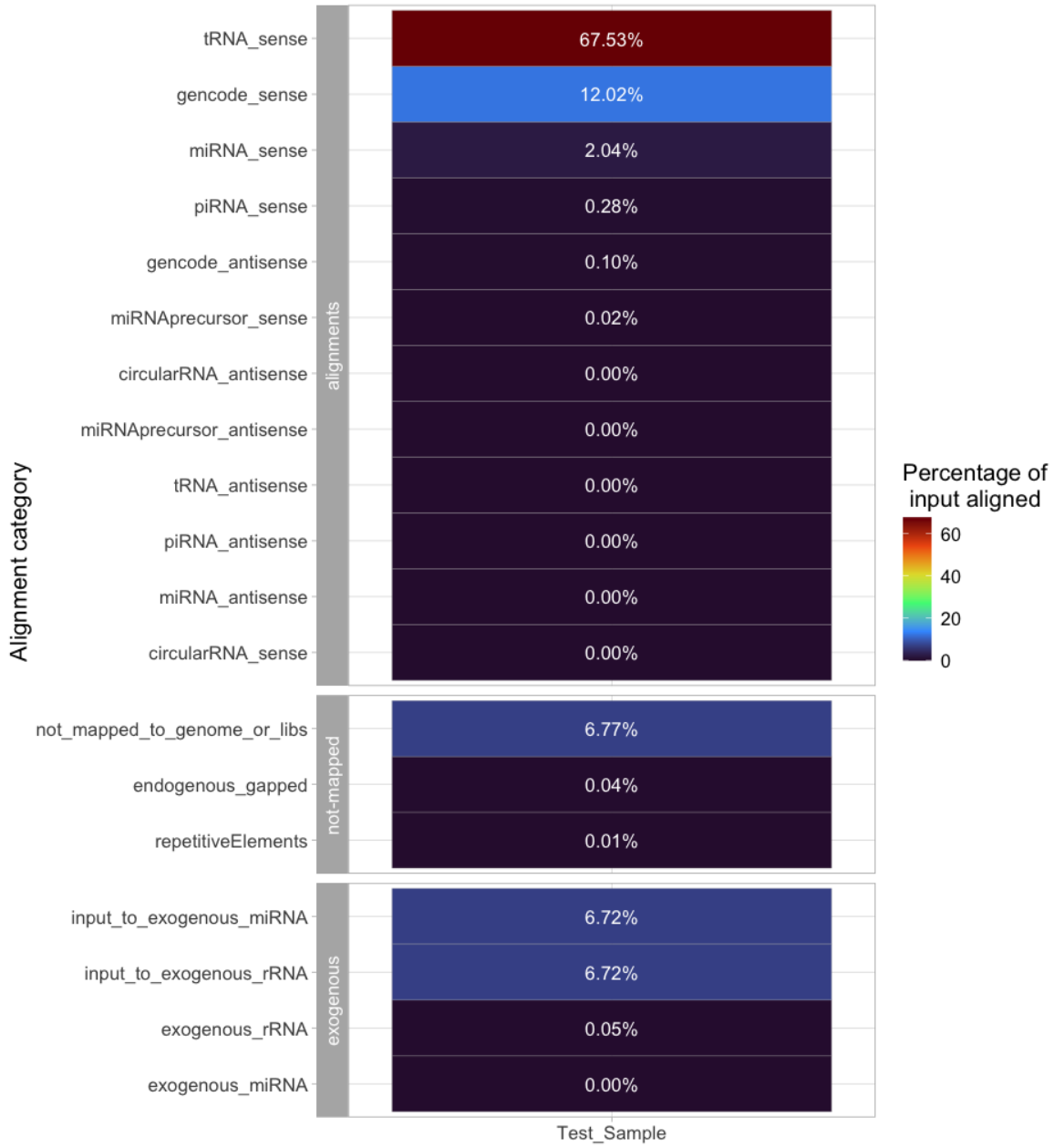


Figure 3.16. Scaled Read Alignment of Healthy NS sample. *Fractions of aligned reads (normalised by total reads acquired from sequencing (not included in final figure)). Sense miRNAs make up 2.04% of total reads identified. Alignments shown in order of percentage by category of alignment. Sense and antisense alignments were tested, all antisense alignment had 0%. Each descending category receives the reads that did not align with the above category, but percentages are based on total reads input into analysis (including reads used for filtering shown in Figure 3.15.).*

3.3.4.1. Top Expressing miRNAs and Literature Review

Despite the low % of reads aligned to miRNAs for the NS sample sequenced, a total of 645 unique miRNAs were identified with total read counts per miRNA ranging between 1 and 21,244 reads. After filtering out miRNAs with low read count (<10 reads), as would be done if performing differential expression (DE) analysis, 286 miRNAs remained. This number was reassuring and would be sufficient input if considering and planning comparative miRNA studies using NS sampling.

For visualisation and further analysis, the results were further pruned to those miRNAs (N = 49) having >1,000 reads (Figure 3.17.). miR-223-3p was found to be the most highly expressed miRNA having over 20,000 reads.

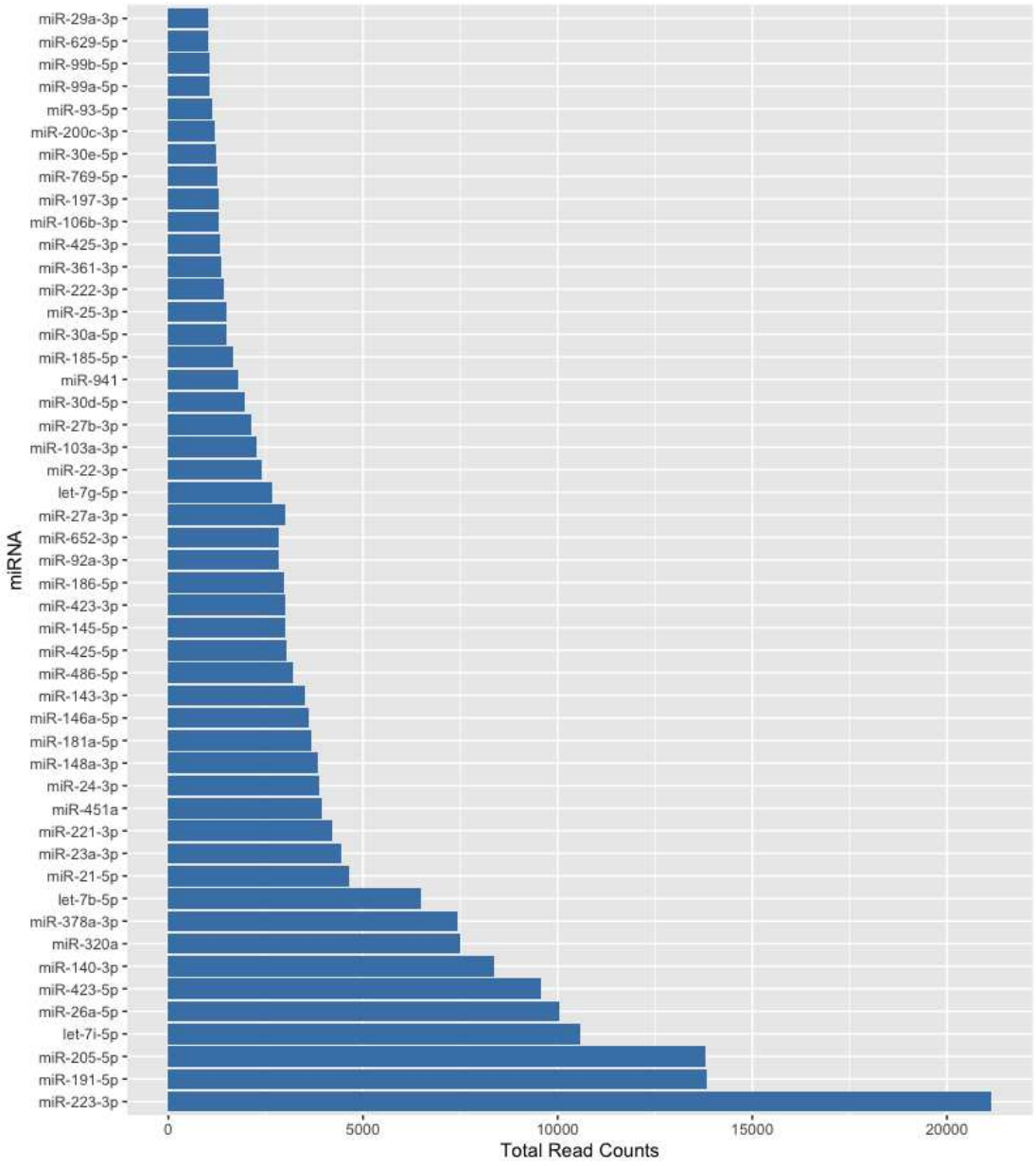


Figure 3.17. Bar Chart of miRNAs with Total Read Counts Greater than 1,000 from Sequencing of a Nasosorption Sample. *Forty-nine unique miRNAs had a sequencing read count greater than 1,000 reads with number of reads ranging from 1,000 to 21,144 with miR-223-3p having the most reads.*

A comparison with the literature was conducted to identify if any of the identified miRNAs had been observed in healthy control subjects in previous published respiratory studies. As one of the intentions was to understand the difference between miRNA expression in the nasal airway compared to the lung, a study of miRNA expression in bronchial biopsy samples (in healthy controls and asthmatics) was used (Roffel *et al.*, 2021). The Roffel *et al.* study comprised miRNAs from bronchial biopsy samples from 79 severe asthmatics and 82 healthy controls. The miRNAs identified by Roffel *et al.* as having different expression levels between healthy controls and asthmatics can be used to identify a subset of miRNAs that were expressed at measurable levels in healthy controls alone. The following comparison was conducted retrospectively, as at the time the experiments were conducted there were no sequencing miRNA expression studies with biopsy samples available for comparison.

Table 3.1 shows the miRNAs from the study and the miRNAs identified in this present study that were used for comparison as well as the matches identified. In the first two rows of the table all identified miRNAs found to be differentially expressed in the study from the literature (43 miRNAs) or from this study the 31 miRNAs with greater than 1,000 reads per miRNA, with the following corresponding columns giving details of the sample size, type of sample, and details of the comparison, with the last column giving the reference. The final row in the table highlights the specific 18 miRNAs that were present between the Roffel *et al.* study and those identified in this study with more than 1,000 reads and this is also shown in the Venn Diagram Figure 3.18.

Table 3.2. miRNAs identified in Literature compared to miRNA hits from the Nasosorption sample Sequencing.

miRNAs identified	**Sample size**	**Sample**	**details**	**Reference**
miR-99b-5p, miR-181b-5p, miR-141-3p, miR-374b-5p, miR-92a-3p, miR-203a-3p, miR-23b-3p, miR-100-5p, miR-150-5p, miR-199b-5p, miR-139-5p, miR-320a, miR-200c-3p, miR-92b-3p, let-7a-5p, miR-374a-5p, miR-214-3p, miR-1247-5p, let-7d-3p, miR-574-3p,miR-26a-5p, miR-190b, miR-223-3p, miR-200a-3p, let-7b-5p, miR-30d-5p, miR-30a-3p, miR-19b-3p, miR-103a-3p,miR-99a-5p, miR-200b-3p, miR-183-5p, miR-148b-3p, miR-21-5p, miR-195-5p, miR-34c-3p, miR-361-5p, miR-142-5p, let-7e-5p, miR-199b-3p, miR-182-5p,miR-34c-5p, miR-10a-5p, miR-193a-5p, let-7c-5p, miR-10b-5p, miR-652-3p, miR125a-5p, miR-429, miR-20a-5p, miR-486-5p, miR-25-3p, miR-425-5p, miR-185-5p, miR-106b-5p, miR-125b-5p, miR-15b-5p, miR-449a, miR-93-5p, miR-451a, miR-16-5p	79 asthma, 82 Healthy Controls	Bronchial biopsy	DE-miRNAs identified in comparison	Roffel *et al.*
miR-223-3p, miR-191-5p, miR-205-5p, let-7i-5p, miR-26a-5p, miR-423-5p, miR-140-3p, miR-320a, miR-378a-3p, let-7b-5p, miR-21-5p, miR-23a-3p, miR-221-3p, miR-451a, miR-24-3p, miR-148a-3p, miR-181a-5p, miR-146a-5p, miR-143-3p, miR-486-5p, miR-425-5p, miR-145-5p, miR-423-3p, miR-186-5p, miR-92a-3p, miR-652-3p, miR-27a-3p, let-7g-5p, miR-22-3p, miR-103a-3p, miR-27b-3p, miR-30d-5p, miR-941, miR-185-5p, miR-30a-5p, miR-25-3p, miR-222-3p, miR-361-3p, miR-425-3p, miR-106b-3p, miR-197-3p, miR-769-5p, miR-30e-5p, miR-200c-3p, miR-93-5p, miR-99a-5p, miR-99b-5p, miR-629-5p, miR-29a-3p.	Healthy Control N = 1	Nasosorption	miRNAs identified with counts greater than 1,000	This Thesis
miR-223-3p, miR-26a-5p, miR-320a, let-7b-5p, miR-21-5p, miR-451a, miR-486-5p, miR-425-5p, miR-92a-3p, miR-652-3p, miR-103a-3p, miR-30d-5p, miR-185-5p, miR-25-3p, miR-200c-3p, miR-93-5p, miR-99a-5p, miR-99b-5p.			Matching miRNAs between the Roffel *et al.* study and this study.	

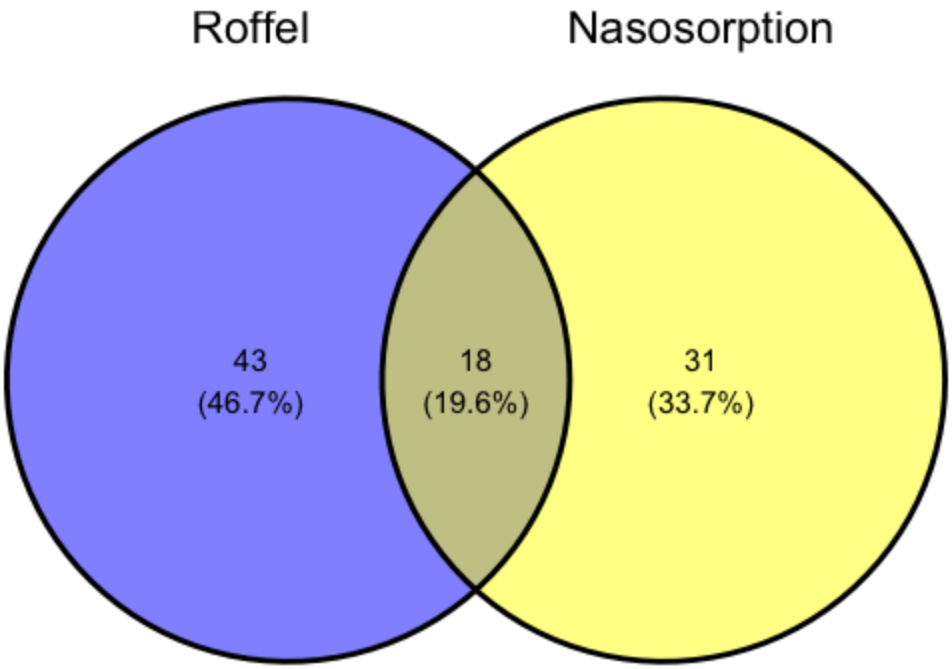


Figure 3.18. Venn Diagram of miRNAs Comparison. *miRNAs identified in this study with total read counts greater than 1,000 compared to DE-miRNAs in published findings in (Roffel et al., 2021) bronchial biopsy comparison of Severe asthma and Healthy Controls. There were 18 shared miRNAs identified in the comparison.*

3.4. Discussion

Research into mucosal sampling of the distal airway has been developing at an accelerated rate in recent years; likely because of advances made in RNA biology over the COVID-19 pandemic. The understanding of miRNA expression in different tissues in health and disease, however, is still in its infancy. The present study provides insights into the feasibility of miRNA next generation sequencing from nasal mucosal samples for the purpose of potential differential expression (DE) analysis. The work has been focused upon and achieved the optimisation of a methodology for the preservation of nasal mucosal samples for the purpose of RNA extraction, as well as the optimisation of retaining cDNA library yield after library synthesis for the purpose of next generation sequencing. From the work conducted the following conclusions can be drawn.

Nasosorption is a viable method of sampling and can provide sufficient total RNA for miRNA sequencing through optimisation of sample processing protocols, and careful removal of adapter dimers from cDNA libraries.

Total RNA yield met the minimum requirements for next generation library synthesis (Figure 3.8.). Post this present study being completed (the work described in this Chapter was completed shortly before the COVID-19 pandemic and lockdown in the United Kingdom), Farr *et al.* in 2022 published a study of upper respiratory tract miRNA profiling, using next generation sequencing and nasal swabs, for detection of SARS-CoV2 infection (Farr *et al.*, 2022). Nasosorption may be a more precise method of mucosal sampling as it gently absorbs, rather than by friction, the mucosal lining fluid of the epithelial surface.

Sampling using the more rigid standard classical cotton swab with circular movements (creating friction) may initiate a localised inflammatory response, which may in turn lead to changes in gene expression. The present study has shown an NGS library generated from a Nasosorption sample can be successfully sequenced using the optimised methodology developed in this Chapter.

This present study has also highlighted that reagent volumes are an important factor when considering sample processing as they can impact downstream processing yield. There was a clear impact on yield depending on the number of times a silica matrix column was loaded with RNA suspended in a buffer (Figure 3.5.) (Farrell, 2009). The present study also established that the shape of containers in relation to the shape of the SAM can also impact the efficacy of the protectant fluid, quantity required, and consequently the RNA yield obtained. The differences in Bionalyzer trace yields (seen between Figure 3.4. and Figure 3.6.) highlighted the impact of such a small change on required reagent volumes in relation to assisting in sample stabilisation.

The Pippin Prep is a standard piece of equipment used by many sequencing facilities for the size selection of next generation sequencing libraries prior to loading on the sequencer. The Pippin Prep is made by Sage Science who report that a sample library yield recovered is in the region of ~20% and therefore is comparable with gel size selection yields (Munafo *et al.*, 2013). In this thesis the use of the Pippin Prep, while essential for removing unwanted adapter dimer sequences, *had a considerable impact on sequencing library yield with the output being approximately 50% less than the input concentration (based on tests with input 50 bp ladder and recovered yield of the 150 bp fragment).* A reduction in the library yield

may also impact the total miRNA variation which could skew downstream analysis as the true representation of miRNAs might no longer be present in the sample (Buschmann *et al.*, 2016) It could be argued that size selection is an essential part of library synthesis preparation and, as long as all samples undergo the step, the dataset remains comparable. The issue then lies with differences that might be seen between datasets that have been prepared in different ways: gel electrophoresis vs. magnetic beads. This has already been shown when comparing library synthesis methodologies alone (Heinicke, Zhong, Zucknick, Breidenbach, Sundaram, T Flåm, Leithaug, Dalland, Farmer, *et al.*, 2020). As an alternative option, as the input volume of the Pippin Prep is fixed to 30 μL, a sample could be concentrated to maximise the total load per well of 5 μg per lane. Concentrations between 100 ng/ μL to 200 ng/ μL (Figure 3.8.) of cDNA were input into the Pippin prep in the experiments detailed in this present study. In light of these concentrations, it would therefore be possible to concentrate the entire volume of the cDNA library synthesis (100 μL) into the fixed load of the Pippin Prep i.e. 30 μL without going above the 5 μg total per lane. The significant risk, however, would then be that should the lane fail for any reason, the entire sample would be lost.

Concentration of samples does not appear to affect the yield of the library (Figure 3.12. B.); however, there may be an impact on the miRNA read variation. Concentration kits, as used in this Chapter, involved a silica column and buffers to load the sample onto a matrix column; wash the sample of salts; and re-elute cDNA at a desirable concentration (when input at a known concentration) or volume. Quality control steps in library synthesis and clean-up were able to be utilised to assess the concentration of the library. Further analysis however is needed to compare sequenced libraries (derived from the same original sample)

that have and have not been concentrated in this manner to understand if there is an impact on the total number and which miRNAs are identified (miRNA profile obtained).

Sequencing of small RNA libraries, such as those expected to be predominantly miRNA sequences, requires a minimum recommended read depth of 50 bp. The use of a 150 bp read depth was quite effective practically and, in this case, cost-effective. The entirety of the fragment was sequenced and as the tail end of the read often declines in quality, the desirable section of the read was 'protected' as it occurred earlier in the sequence. The FASTQC results indicated that read quality (Figure 3.14.) was very good for the reads of the NS sample sequenced.

The protocol developed in this Chapter for Nasosorption sample next generation sequencing was successful based on a high percentage of reads passing quality filters and successfully being aligned to miRNAs and the genome (Figures 3.15. and 3.16.). Sample alignment of small RNA sequences through this method showed a highly successful sequencing run, based on the less than 9% of reads that did not pass the quality filters and between 89% of reads that made it to alignment. There was, however, an apparent low percentage of aligned reads to miRNA sequences at 2% although as this was at the time the first such study this may be the standard outcome.

Despite the methods of library synthesis attempting to select for miRNA sequences, through small-RNA-inclusive-total-RNA-isolation and size selection of cDNA libraries specific to the expected fragment size of a miRNA library with appropriate adapters attached (Figure 3.13.), as just mentioned sense miRNA reads made up only 2% of total reads (Figure 3.16.). The majority of the reads (69%) were tRNA sequence fragments. This

has been found previously by (Van Goethem *et al.*, 2016) with methods of filtering out tRNA sequences being described. In the study Van Goethem *et al.* identified that 5' tRNA halves, a tRNA derived fragment, made up an abundant number of reads when total RNA (high input) extracted from mouse serum was sequenced. The Van Goethem *et al.* study suggested that 5' tRNAs can make up 88% of reads with 5% of the remaining reads being miRNA sequences. To date this work has not been replicated by other researchers nor attempted with human samples but nonetheless it is interesting that the findings of this study reflect that of the Van Goethem *et al*. study.

Whilst a high percentage of reads in the Nasosorption sample were comprised of tRNA fragments, nonetheless a sufficient number of miRNAs were identified for comparison with the literature. A total of 31 miRNAs were identified with read counts greater than 1,000 with 286 miRNAs meeting the standard for differential expression analysis (DE). This supports that the optimised protocol of nasosorption-sample miRNA library-preparation and next generation sequencing would be a plausible method for analysis of Nasosorption samples to understand potential differences in miRNA expression between asthma patients versus healthy controls.

Post the initial data analysis and the COVID-19 pandemic and change of direction of work of the thesis, that led to the work conducted for the next Chapters, a retrospective analysis was conducted using the Roffel *et al., (*2020) study. It would have been ideal to compare miRNAs identified in healthy controls from this study, sadly this data was not made available or included in the publication. As the goal however was to compare miRNA expression between healthy controls and asthmatics it seemed relevant to compare the

miRNAs identified in the DE analysis from this study. To ensure a conservative comparison a high threshold of miRNA counts was applied, significantly higher than those filtered for when comparing in DE analysis. This identified 18 miRNAs shared between the two studies. This suggests that the nasosorption-sampling method developed in this Chapter to allow small RNA next generation sequencing analysis might be a viable method for the comparison of nasal miRNA expression between asthma and healthy controls. While no other sequencing study, to the best of the author's knowledge, has published expression of healthy controls in the airway, one study of QPCR of healthy biopsy samples has been published and 11 of the highly expressed miRNAs in the airway matched those found highly expressed in this nasosorption study (miR-223, miR-191, miR-205, Let-7i, miR-26, miR-140, miR-21, miR-30d, miR-30a, miR-200c, and miR-93) (Williams *et al.*, 2009). This supports the findings in this thesis that Nasosorption sampled miRNA expression could be representative of expression in the lower airway.

Objectives 1 and 2 of this Chapter were achieved, however the COVID-19 pandemic occurring prevented objectives 3 and 4 from being started and completed. Therefore, research pivoted towards identifying if circulating miRNA expression could be used for the identification of a signature of asthma between healthy controls and asthmatics. Nevertheless, objectives 1 and 2 were met and despite this study consisting of an N of 1, it is the first to consider the application of Nasosorption for isolation of miRNAs for next generation sequencing. These findings will inform the design of future experiments in the identification of differences in nasal miRNA expression in respiratory conditions and diseases such as asthma.

4. Chapter 4: miRNA Expression Comparison Between Severe Asthmatic and Healthy Control Whole Blood Samples

4.1. Introduction

As highlighted in prior Chapters, athma remains a prevalent chronic respiratory condition, affecting millions worldwide and imposing substantial burdens on individuals and healthcare systems (Chapter 1, Section 1.1.1.1.). Despite advances in treatment and management strategies, the precise molecular mechanisms underlying asthma pathogenesis are not fully understood. Severe asthma presents with symptoms that significantly impair daily life, including frequent and acute asthma attacks, persistent shortness of breath, wheezing, coughing, and chest tightness (Chapter 1, Section 1.1.1.4.). Despite adherence to conventional asthma treatments like inhaled corticosteroids and bronchodilators, individuals with severe asthma often experience inadequate symptom control.

Despite new therapies emerging including the use of biologic medications targeting specific pathways of inflammation, such as monoclonal antibodies against IgE or interleukin-5 (Chapter 1, Section 1.1.5.), due to the heterogeneity of the condition positive responses to

these medications are unpredictable. This results in various attempts with different medications to manage symptoms causing additional distress to the patient. Furthermore, patients with severe asthma typically still experience exacerbations despite treatment, leading to hospitalizations and reduced quality of life (Papiris *et al.*, 2002). Moreover managing severe asthma often requires close monitoring by healthcare providers, regular check-ups, and lifestyle adjustments to minimize triggers, all of which have significant financial and economic impacts (Chaplin, 2018).

Expanding our current understanding of the molecular signatures associated with asthma is essential for the development of targeted therapies (Alipoor *et al.*, 2016). It also affords the opportunity for personalized-medicine approaches to relieve patients of the process of iteratively trialling various medications and optimizing responses.

In recent years, microRNAs (miRNAs) have emerged as crucial regulators of gene expression and therefore important players involved in both health and disease (Gulyaeva and Kushlinskiy, 2016). In the asthma field, research to date has implicated dysregulation of certain miRNAs being associated with modified pathways related to inflammation, airway remodelling, and immune regulation (Chapter 1, Section 1.4.1.). These small non-coding RNAs have the potential to serve as diagnostic biomarkers and therapeutic targets for asthma management. Despite these promising findings, however, the precise roles of miRNAs in asthma pathophysiology remain incompletely understood. Further research is essential to elucidate the intricate interactions between miRNAs and their target genes, as well as the influence of environmental factors and genetic variability.

Accessibility of samples from the lung, however, is not trivial (Chapter 1, Section 1.3.) but whole blood represents a valuable resource for studying systemic molecular alterations in asthma, given the ease of access to it and that it captures the circulating immune response. Investigating differential miRNA expression in whole blood between asthmatics and healthy controls therefore has the potential to unveil novel insights into the pathophysiology of asthma and identify potential biomarkers for diagnosis and prognosis.

The overall aim of this Chapter was to conduct an initial pilot study applying the approach of next generation sequencing to profile and understand to what extent miRNA expression is altered in severe asthmatics compared to healthy controls and establish if specific miRNAs play a role in the phenotypic presentation of asthma. Due to having access to samples of the Celtic Fire Study cohort (Appendix: Figure 3), it was possible to select subjects for the pilot study with careful matching of cases and healthy control subjects to, as best as possible, control for known risk factors associated with asthma.

The specific objectives of this Chapter were to:

1. Investigate differential miRNA expression profiles in whole blood samples from Severe asthmatic individuals compared to Healthy Controls.
2. Explore potential functional roles and pathways enriched among dysregulated miRNAs in asthma.

4.2. Materials and Methods

4.2.1. Study Participants: The Celtic Fire Study

The blood samples used for the initial pilot study had been biobanked from participants of a larger research study called Celtic Fire. Ethical approval for the Celtic Fire Study was granted by the London-Stanmore Research Ethics Committee (reference 14/LO/2063). Celtic Fire was a cross-sectional study with recruitment of subjects from three centres (the Royal Brompton Hospital, London, UK, Connolly Hospital, Dublin, Ireland, and Swansea University Medical School, Swansea) with a range of biological samples obtained from healthy controls and asthmatics of differing severities. No adverse reactions were noted from the sampling. All subjects were non-smokers or ex-smokers for a minimum of 12 months prior to sampling with a $\leq$5 pack-year history. All subjects were free of other significant diseases. Healthy controls did not have a history of asthma or rhinitis and had to have a pre-bronchodilator FEV1 of $\geq$ 80% predicted to be eligible for the study.

Asthmatic patients of all categories required specialist physician diagnosis of asthma plus one of the following recorded within the last 2-year period:

1. A decrease in Forced Expiratory Volume in one second (FEV1) > 12 % and > 200 mL within 24-48 hours after discontinuing treatment of ICS (Inhaled Corticosteroids), LABA (Long-acting beta2-agonist), and SABA (Short-acting beta2-agonist).*

2. An improvement in FEV1 ≥ 12 % or 200 mL predicted after inhalation of 200 µg salbutamol.*

3. A positive airway hyperresponsiveness to methacholine or histamine (PC20 < 80 mg / mL).*

*As part of preparation for a bronchodilator response test, 24 hours (ideally 48 hours) discontinuation of ICS+LABA combination, or 6 hours discontinuation of SABA inhaler was necessary.

Recruitment to the Celtic Fire Study was irrespective of atopic status and both non-severe and severe asthmatics were recruited. For this pilot study, only severe asthmatics were studied with severity defined as:

Dependence on systemic corticosteroid therapy for disease control.

In addition to the range of biological samples collected for downstream experiments, each participant completing a questionnaire (Mc Brien, 2020) (Cuthbertson *et al*. 2022). Samples taken included whole blood for serum isolation (for subsequent Immunoglobulin E [IgE] measurements), and whole blood collected for DNA extraction. For obtaining peripheral blood for subsequent RNA extraction, whole blood samples were collected into PAXgene tubes (Chapter 2, Section 3.1.3.2.).

4.2.2. Sample Processing

Prior to this present study, the questionnaire data and IgE measurements for all subjects recruited to the Celtic Fire Study (N = 110) had been compiled into a metadata file. In addition, for some of the study participants RNA had already been extracted from the PAXgene blood tubes (extraction performed by a former member of the Genomic Medicine Group, NHLI, Dr S. Darby). Using the metadata file, and considering participants who already had RNA extracted (as at the time there were supply issues in relation to RNA extraction kits due to the COVID pandemic), a subset of individuals for the initial miRNA profiling pilot study were selected and 9 subjects per group were used for the pilot study. Figure 4.1. shows the sample processing flowchart detailing steps from sample selection to sequencing and data analysis.

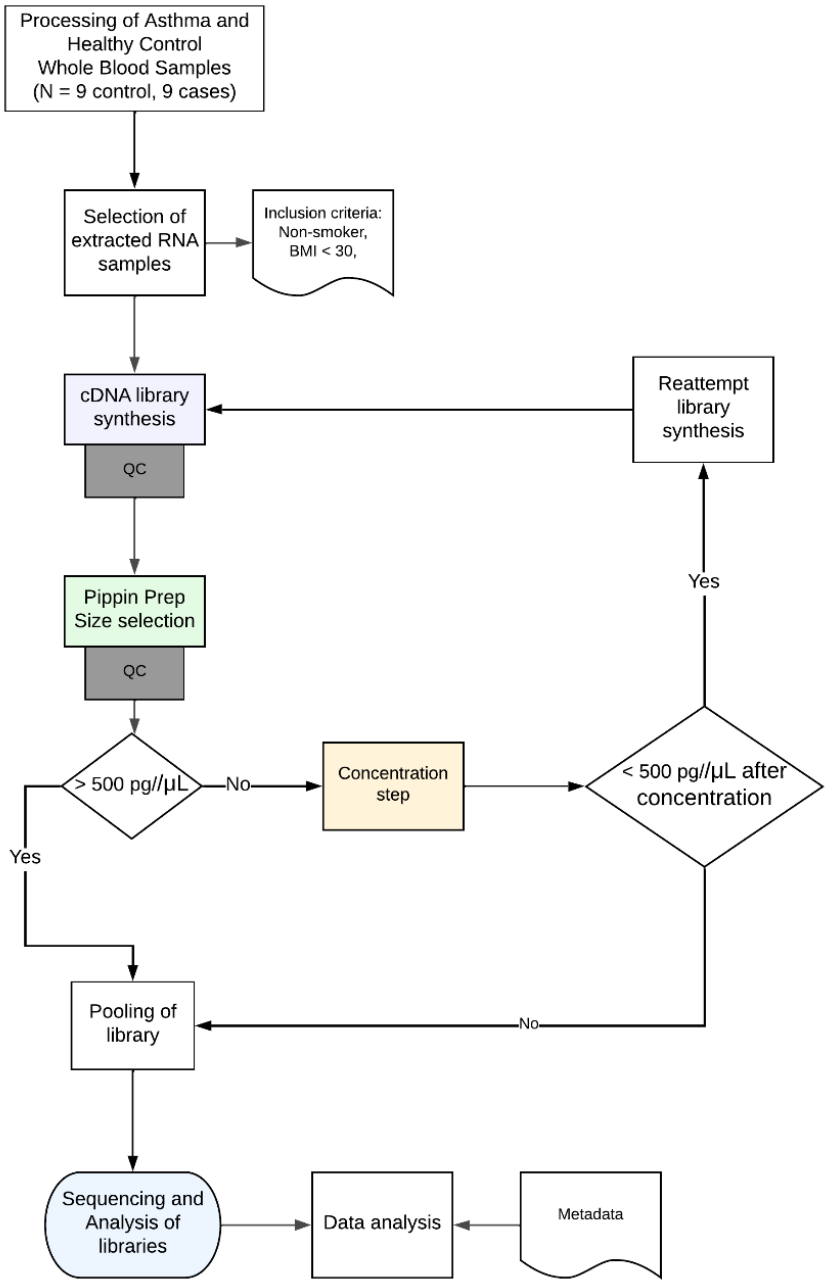


Figure 4.1: Sample Processing Flowchart. *Steps detailed from sample selection to sequencing including quality control (QC). QC was conducted on Agilent Bioanalyzer. Library synthesis was conducted using NEB small RNA library synthesis kit. Pippin Prep was the method of library sample size selection. Concentration of samples was conducted using NEB Monarch® PCR & DNA cleanup kit. Two attempts of library synthesis were made for samples that failed the initial library synthesis*

The extracted blood RNAs for the 18 selected individuals were first quantified (Chapter 2, Section 2.1.1.2.) and then underwent conversion into indexed libraries using the NEBnext® Multiplex Small RNA Library Prep Set for Illumina® (Set 1) following the recommended Illumina protocol (Chapter 2, Section 2.3.). For the initial optimisation experiments, only one index was used as samples did not undergo library pooling prior to sequencing. A selection of index sequences were selected from the 44 available for optimum compatibility as recommended by the manufacturer (NEB) in order to allow pooling of libraries generated from the blood RNA samples.

Indexed cDNA library samples were stored individually at -20°C prior to quality checking.

Quality checking involved samples being run on the Agilent 2100 High sensitivity DNA Bioanalyzer microfluidic chip (Agilent® Technologies, Inc.) to determine both library quality and size (Chapter 2, Section 2.3.3.). The expected size range of the Illumina-adapter-linked cDNAs generated is between 147-157 bp (NEB. 2020).

4.2.3. Sequencing and Analysis

The sequencing protocol was conducted as described previously (Chapter 2, Section 2.1.4.). Likewise, the analytical pipeline was performed as described in Chapter 2, Section 2.1.5. In brief, sequencing reads from samples were passed through FASTQC, followed by Genboree workbench small RNA pipelines where trimming and alignment was conducted. Differential analysis was then conducted using Tag Count Comparison (TCC) in R. The pilot study groups were Healthy Controls and Severe asthmatics. Significance was given to DE-miRNAs with a FDR (q-value) <0.05 and FC >0.05. DE-miRNAs were then input into miRNet for mRNA target analysis and Gene Ontology.

4.3. Results

4.3.1. Subject Demographics and Clinical Correlations

Eighteen subjects were selected from the larger Celtic Fire Study cohort for the pilot study – nine healthy controls and nine severe asthmatics (Section 4.2.1.). All individuals were current non-smokers with a BMI below 30. Clinical factors such as age, gender and BMI have been previously reported to impact asthma severity (Chapter 1, Section 1.1.1.3.). To therefore try and reduce the impact of these variables, the two groups were matched as best as possible for age, gender and BMI. This was done successfully as no significant differences between the groups were seen for the three traits (Table 4.1.). Other clinical

features that have been key to asthma diagnosis and disease monitoring are IgE, blood Eosinophil count, as well as the ratio between forced expiratory volume in 1 minute and forced vital capacity (FEV1/FVC ratio). Serum IgE levels between the groups showed a small difference (P = 0.04) Wilcox signed-rank test). There was no difference in blood eosinophil counts between the groups, but, as expected, there was a very significant difference between the groups regarding FEV1/FVC ratio (P< 0.0001).

Table 4.1. Demographics for Case and Control Groups*. Table shows the distribution of subjects across the two groups and key factors considered in the study. Patients were selected to match as best as possible for age, gender and BMI. There were no significant differences between the groups regarding age, gender, blood eosinophil count, or body mass index (BMI). Total Immunoglobulin E (IgE) and the ratio of forced expiratory volume in 1 minute and forced vital capacity (FEV1/FVC ratio), were identified as having statistically significant differences in a Kruskal-Wallis test. P values generated through a Wilcox test (ordinal variables) or a Fisher's exact test (categorical variables). SD = standard deviation. BMI (Body mass index), FEV1FVCratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

	Control (N = 9)	Case – Severe asthma (N = 9)
Age: *P* = 0.06		
Min	22	25
Max	59	71
mean (SD)	32.78 ± 14.02	47.67 ± 12.82
Gender: *P* = 1		
Female	4 (44)	5 (56)
Male	5 (56)	4 (44)
Total IgE kU/L: *P* = 0.04		
Min	2.4	16.9
Max	258.5	404
mean (SD)	7; 54.56 ± 90.71	125.01 ± 112.63
Data unavailable	2	0
Blood Eosinophils: *P* = 0.14		
Min	0.1	0.1
Max	0.7	0.69
mean (SD)	0.20 ± 0.21	0.30 ± 0.19
Data unavailable	1	1
FEV1/FVC ratio: *P* = 4e-05		
Min	82	57
Max	89	77
mean (SD)	84.62 ± 2.61	68.69 ± 7.75
BMI Class: *P* = 1		
Normal	6 (67)	5 (56)
Overweight	3 (33)	4 (44)

An analysis was conducted to assess potential correlations between clinical variables of interest (Figure 4.2.) (outcome ranges of correlations applied are detailed in Chapter 2, Section 2.1.5.2.). Age, BMI and Gender have been associated with affecting asthma disease severity and were therefore included to assess the contribution that might affect downstream miRNA analysis. Other factors that are often co-associated, such as serum IgE and Eosinophil counts, were also included (Chapter 1, Section 1.1.4.) A strong negative correlation was identified between patient disease Severity and FEV1/FVC ratio. Total serum IgE also showed a moderate negative correlation with BMI and a moderate positive correlation with Severity. Eosinophil count and IgE showed a weak positive correlation.

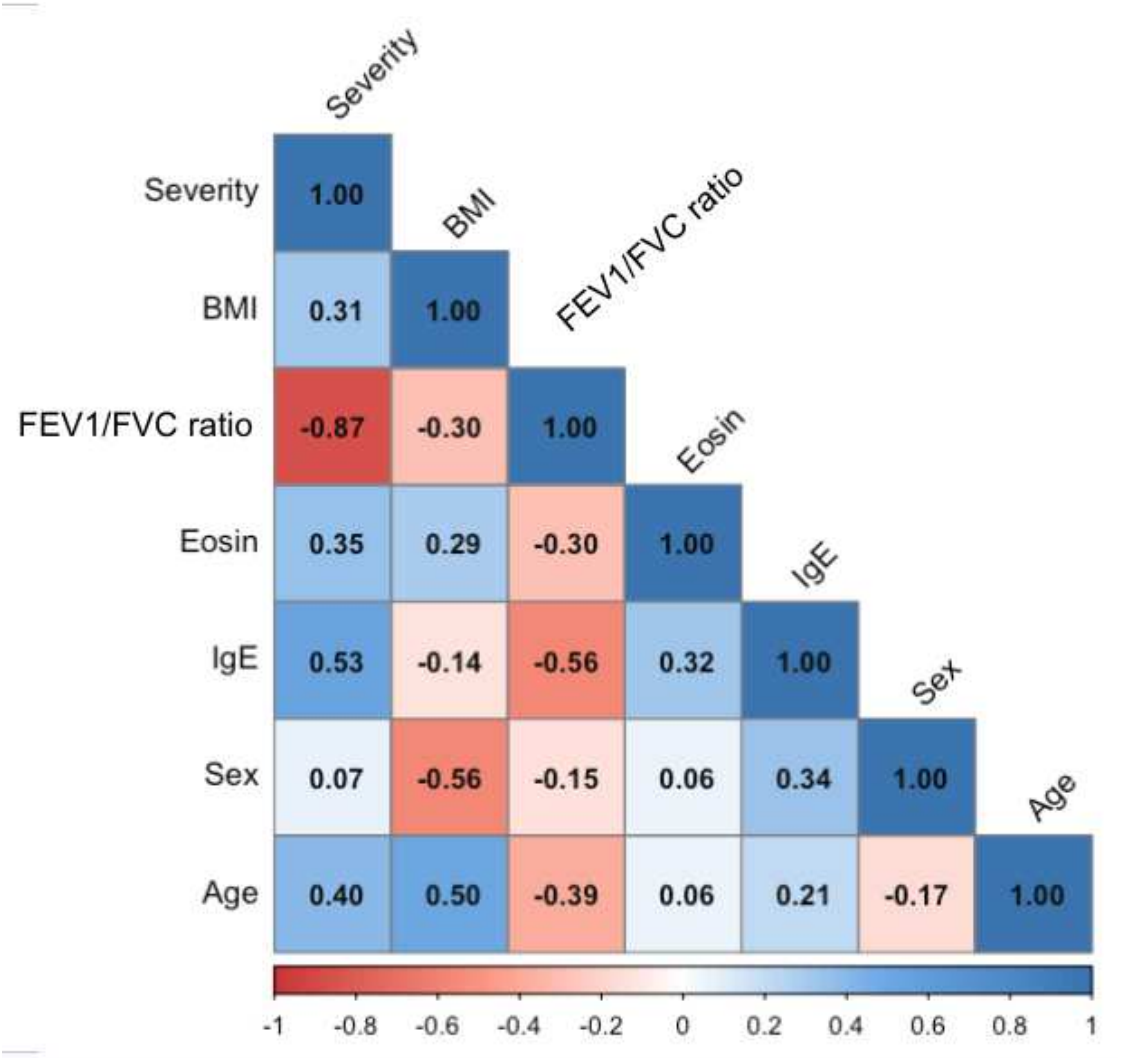


Figure 4.2. Clinical Correlation Analysis of Key Clinical Variables*. Severity represents: Severe Asthmatics and Healthy Controls. Severity (Control, Severe), BMI (Body mass index), FEV1FVCratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years). As per scale, red indicates negative correlation and blue positive correlation. Strong negative correlation was identified between patient disease Severity and FEV1/FVC ratio. IgE (total serum immunoglobulin E) also showed a moderate negative correlation with BMI (Body Mass Index), and a moderate positive correlation with Severity. Eosin and IgE showed a weak positive correlation.*

4.3.2. cDNA Library Processing

A quantity of 100ng of total RNA in a volume up to 6μL was required for input into the library synthesis protocol and the total RNA (extracted from blood) samples for all 18 subjects met this requirement.

Each total RNA sample was converted into a cDNA library using the NEBnext small RNA library synthesis kit, with a unique index barcode added for later identification and to enable pooling of all samples prior to sequencing. As much as possible, samples were processed in batches ensuring even distribution of status (case or control) and gender between batches.

After library synthesis, cDNA libraries were size-selected to filter out all other species of RNA other than the target miRNAs (approximately 150 bp long including adaptors). Sample dropout is always a risk since size selection using the Pippin Prep, like other electrophoresis methods, results in some loss of concentration (see Chapter 3, Section 3.3.3.1. and Figure 3.12.). The minimum input requirement for sequencing (Novogene Sequencing Facility) was 500ng/ μL (within 1%).

Post size selection, five of the sample libraries attained the required minimum input concentration (Figure 4.3.). The remaining 13 samples had to undergo a concentration step using the NEB Monarch PCR & DNA cleanup kit (Chapter 2, Section 2.3.6) to meet the minimum input concentration required (+/- 1%) for pooling and sequencing. Figure 4.4. shows the libraries of the 13 samples pre and post the concentration step.

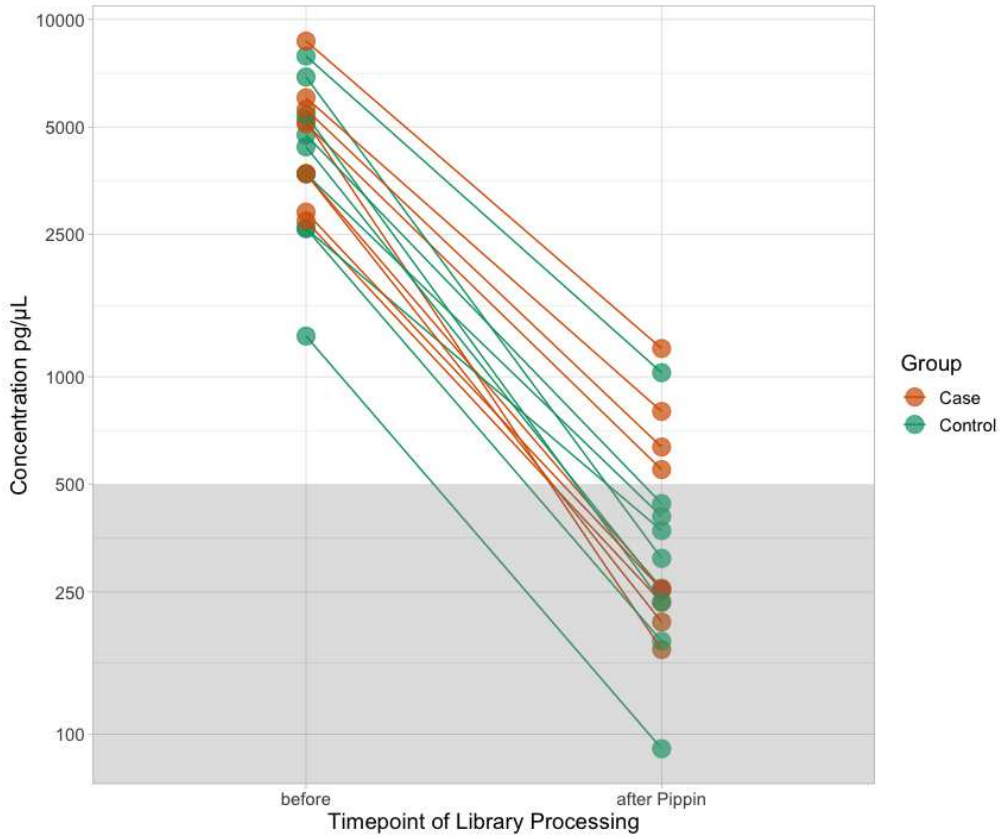


Figure 4.3. cDNA Library Concentration after Pippin Prep Size Selection. *N = 18. After library synthesis size selection is conducted to isolate miRNA cDNA species (with attached adapters). Grey section highlights sample concentrations that do not meet the minimum input concentration for library sequencing (500 ng/ µL) Novogene Sequencing Facility). As can be seen from the figure, five samples were above the minimum concentration threshold with the remaining thirteen samples requiring further concentration prior to sequencing.*

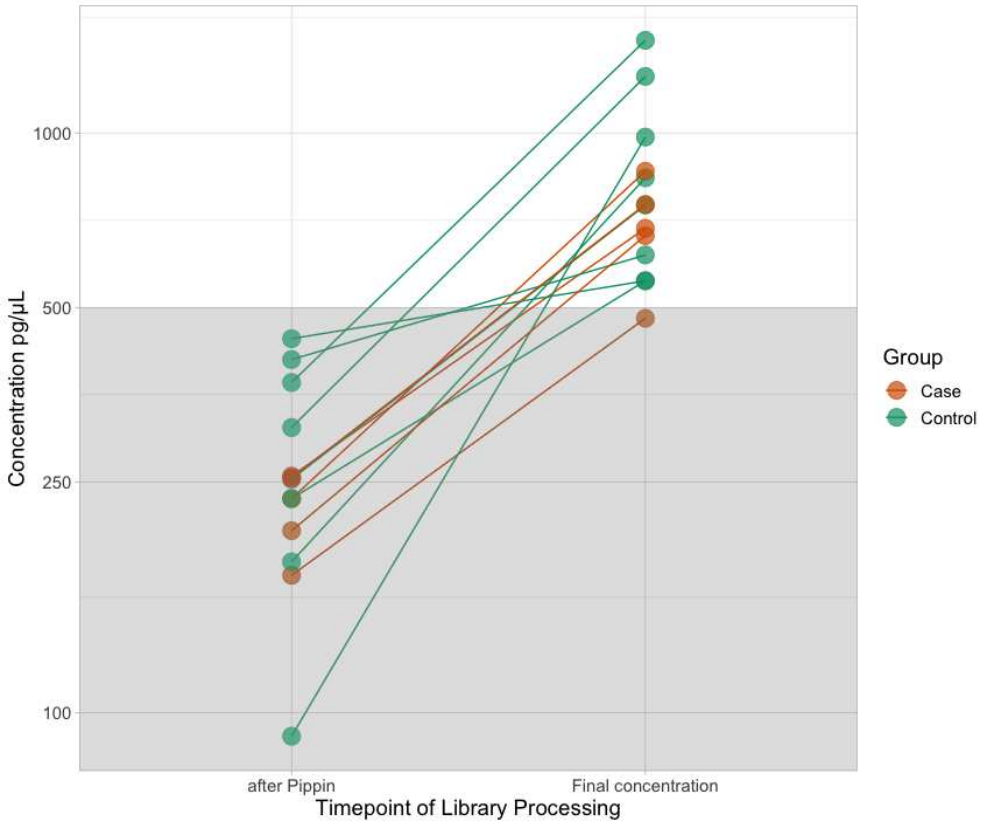


Figure 4.4. cDNA Library Concentration after Concentration Clean-up. *N = 13. Of the 18 total libraries prepared, thirteen cDNA libraries required further processing to the meet minimum concentration input requirements for sequencing (500 ng/µL). All samples shown here were processed through the NEBnext PCR & DNA clean-up columns and eluted in 6-10 µL (variable depending on the amount of concentration required). As can be seen all samples met the concentration input requirement (within 1%) post-clean up.*

4.3.3. Sequencing Results

Raw reads from Novogene arrived on a hard drive in FASTQ file format. The raw reads were then assessed via FASTQC. FASTQC results returned flags related to adapter sequence contributing to some homogeneity among approximately the first half of the read length. This was expected in this experiment as the sequence length was greater than the unique miRNA sequence and included the adapter and barcode sequences (Chapter 1, Section 1.2.2.).

Of the 18 samples, two yielded fewer than the planned 5M reads with 2.9M and 4.6M respectively; 14 samples yielded between 5M and 9.8M reads; and 3 samples had total reads between 10.8M and 11.6M. As the planned number of total reads per sample (on average) was 5M reads, this indicated that the sequencing overall was successful. Figure 4.5. below shows a summary of total reads between the two subject groups – Severe asthmatics versus Healthy Controls.

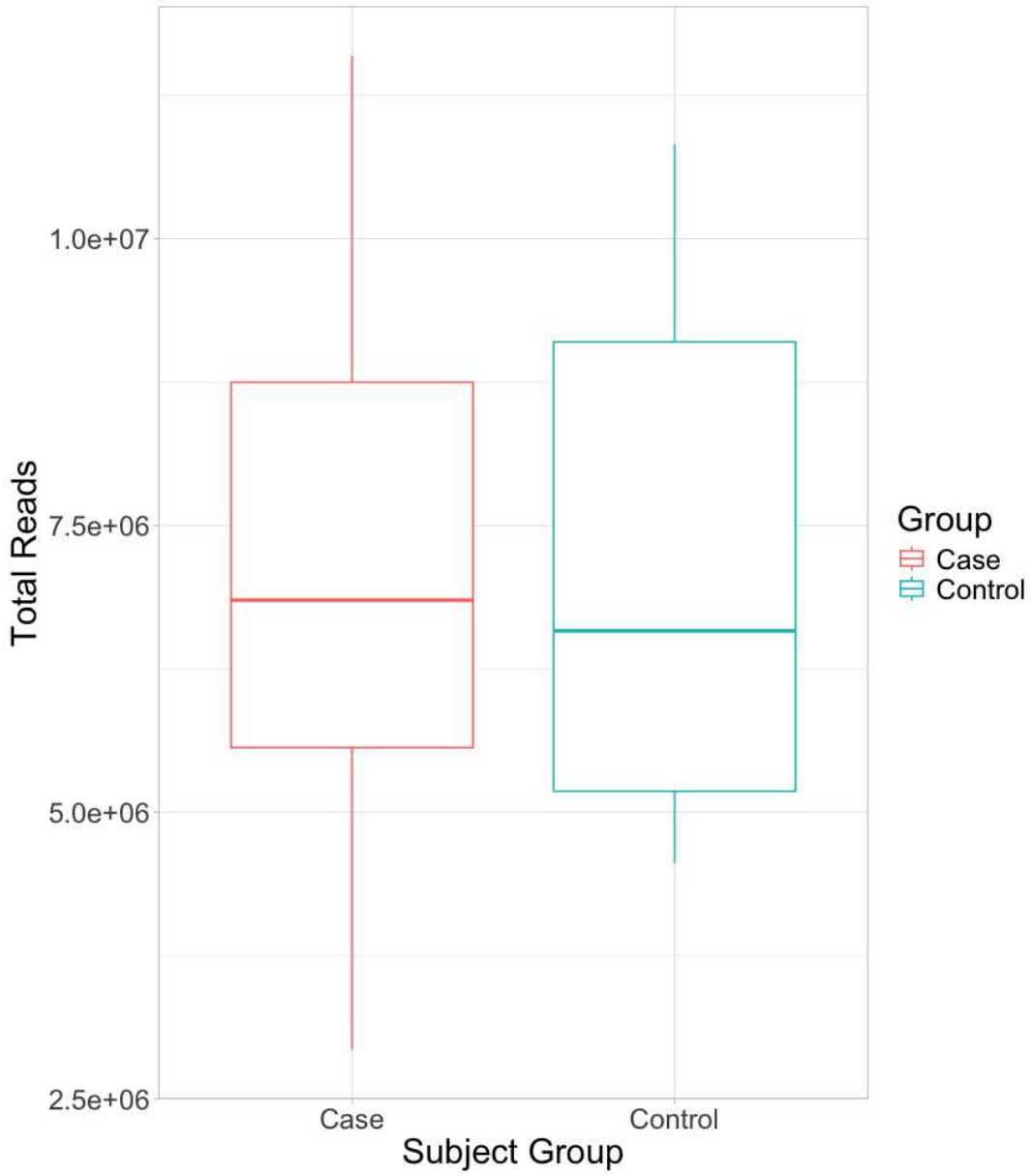


Figure 4.5. Boxplot Summary of Raw Reads for Control and Severe Asthma (Case) Sample Sequencing. *(N = 9 per group). Boxplot denotes median with the box denoting the interquartile range (IQR), and the whiskers 1.5 times the IQR. There were no identified outliers in this case. The average reads acquired for each group was above the 5M reads.*

Quality of reads was high as the per-base sequence quality (the confidence that the base pair read was accurate at that position) was of high confidence beyond the read length that would cover the adapter and the miRNA. Sample quality was further analysed through a range of different alignment factors including, for example, miRNA precursors, as well as miRNA antisense (as a negative control). Table 4.2. below details the factors involved at each step in greater depth. These come as part of the package provided on the Genboree workbench (Chapter 2 Section 2.1.5.1.).

Table 4.2. Alignment and Read Processing Filters in ExeRpt Package. *Categories, Features and Descriptions of alignment focus.*

Category	Feature name	Description of Feature
	Successfully clipped	Adapter is successfully clipped from read. ExceRpt automatically identifies 3' adapters.
Quality filters	Failed quality filter	Read fraction quality = 80 (80% of reads must have acceptable quality scores), Minimum quality = 20 (1% error rate)
	Failed homopolymer filter	FASTX toolkit, identifies homopolymer repeats and removes these reads.
	Univec contaminant filter	aligned to sequences in the NCBI's UnivecVec screen database are removed.
	rRNA	Reads are aligned to all primary endogenous ribosomal RNAs (5S,5.8S,18S,28S, and45S) and removed
Alignments	Reads used for alignments	First pass alignment using STAR aligner with max 1 mismatch per read, minimum read length 18.
	Genome	Hg38
	MiRNA (sense and antisense)	MiRBase v.22
	MiRNA (precursor sense and antisense)	MiRBase v.22
	tRNA (sense and antisense)	tRNAscan database
	piRNA (sense and antisense)	piRNA annotations
	Gencode (sense and antisense)	Aligned to GENCODE transcripts (snRNAs, snoRNAs, miscRNAs, protein coding genes and lncRNAs).
	Circular RNA (sense and antisense)	Elected to be aligned using the circRNA database.
Not mapped to the above (and because exogenous mapping is selected)	Not mapped to genome or libs	Second alignment using STAR aligner. Allowing only perfect matches.
	Repetitive elements	Quality control
	Endogenous gapped	Enables better understanding of reads as a percentage of reads will be identified that were not in earlier alignment steps.
	Exogenous miRNA (input to, and)	Quality control. ("input to" reads that were input and those that successfully aligned).Looking for potential sampling/processing contaminants. Or spike in miRNA.
	Exogenous rRNA (input to, and)	Quality control. Looking for potential contaminants.

Most of the sequencing reads were successfully clipped of adapter (between 97-99%), and few failed the quality filters (1%) (Figure 4.6.). There were negligible-or-no UniVec contaminants (<1%) and low rRNAs identified in the run (between 1-3%). Between 92-97% of reads were used for alignment and 92-95% aligned to the human genome and 74-87% to miRNA sense databases (as only the forward read was required for further steps). To allay concerns that possible contaminants had been introduced from the environment during the preparation of samples, factors that would align the reads to exogenous RNA species were also considered (Table 4.2). There was a negligible alignment to exogenous RNA species (<1%). A total of 1,364 unique miRNAs were identified, with between 667 and 928 miRNAs identified in each sample.

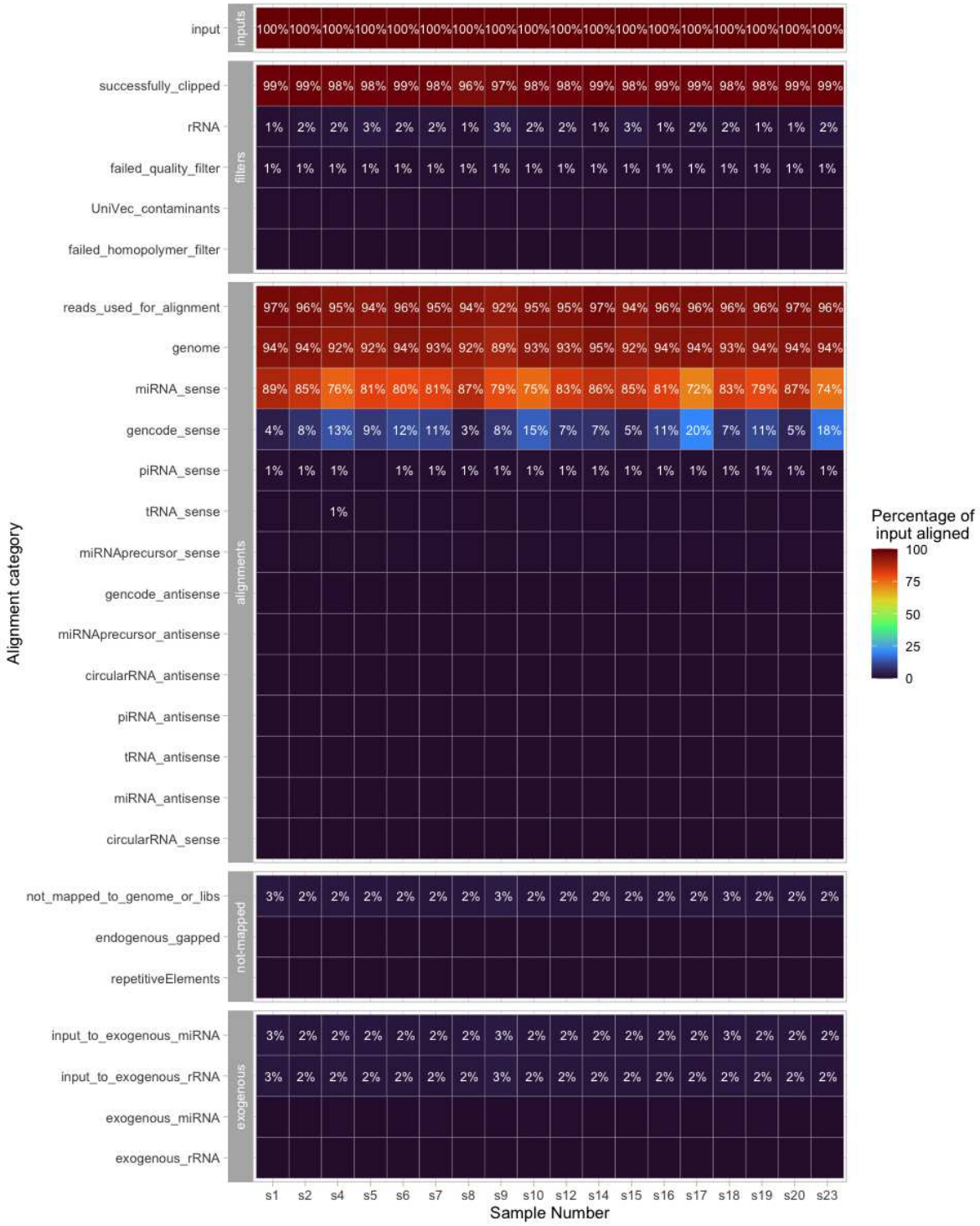


Figure 4.6. Control vs Severe Raw Read Alignment. *N = 18. Raw sequencing reads pass through multiple quality control and alignment steps, the count of reads that pass through each step are accounted for in the figure for each sample. Percentages shown are between 1-100%, below 1% are left blank. A high percentage of reads (>=94%) were successful through filters and aligned to the genome (>= 89%) and to sense miRNA (>= 72%). Very low/no reads aligned to contaminants or failed the quality filters. There was no indication of contaminants introduced during the library preparation based on low percent of reads aligning to exogenous genomes or other exogenous RNA species (<1%).*

As some samples required concentrating to meet the minimum input requirements set by the Novogene Sequencing Facility, a check was run to ensure that miRNA diversity had not been unduly affected by the concentration procedure (as to date this has not been discussed in the literature in any detail). Comparison of miRNA diversity between the samples that were and were not concentrated (non-concentrated) was conducted. Figure 4.7. shows the plot generated for miRNA diversity in each sample separated by library treatment and defined by patient group. A Wilcoxon signed-rank test between the two groups did not identify any significant difference in miRNA diversity, although there appears to be a trend that the samples that were not concentrated have a broader range with regard to number of unique miRNAs per sample.

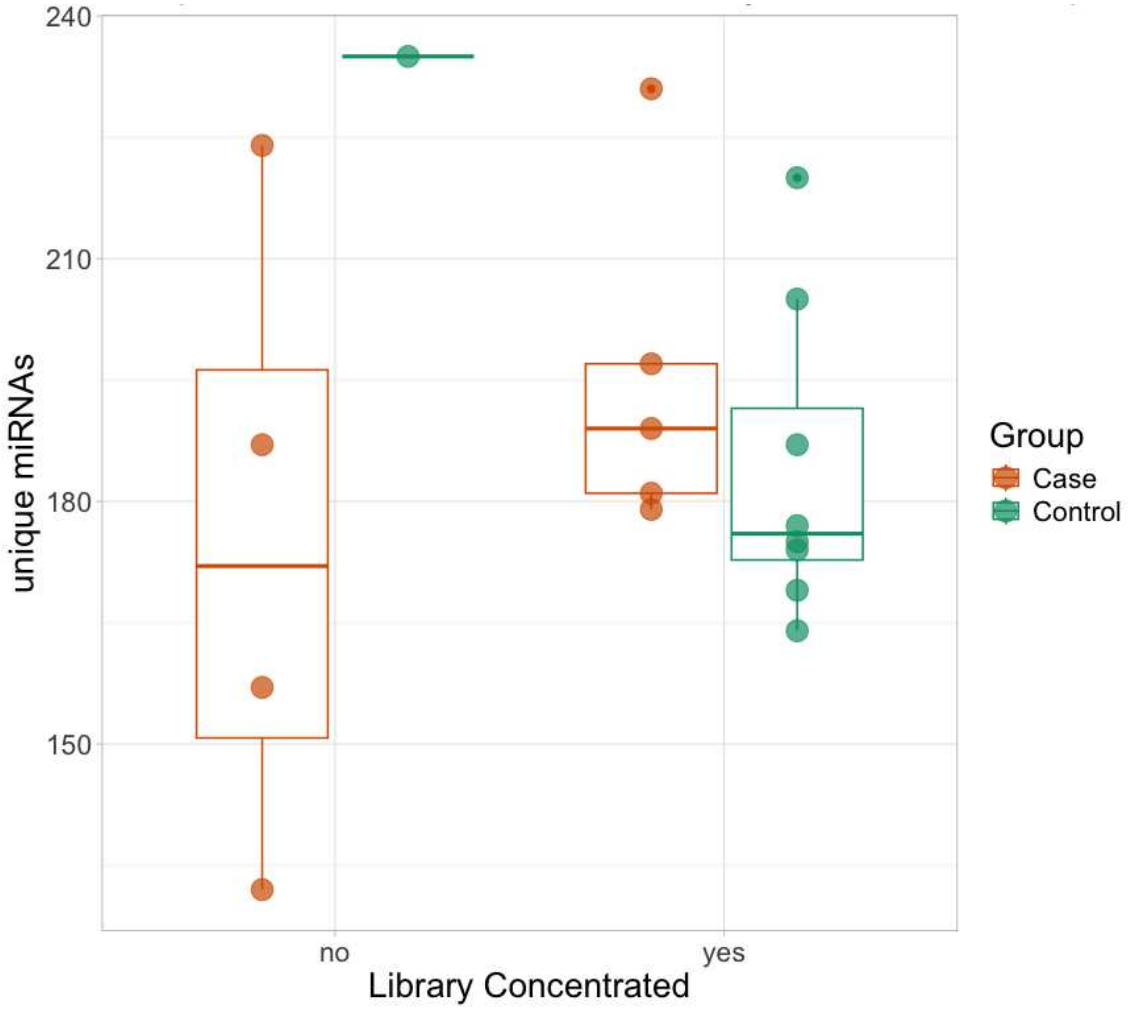


Figure 4.7. miRNA Diversity between Concentrated and Non-concentrated Samples *(N = 18), (P > 0.05 Wilcox signed-rank test). There appeared to be a trend towards wider distribution of number of unique miRNAs in samples that did not undergo library concentration compared to those that did undergo the process.*

Whole blood can contain a variety of sources of RNAs so there was a possibility that the miRNA diversity might have arisen from expected differences in white cell populations between the clinical groups that could contribute to the overall abundance of miRNA. As white cell counts were done at the time of blood sampling and this data was available for all subjects, a Wilcoxon signed-rank test was conducted but was found not to be significant ($P = 0.98$) (Figure 4.8.).

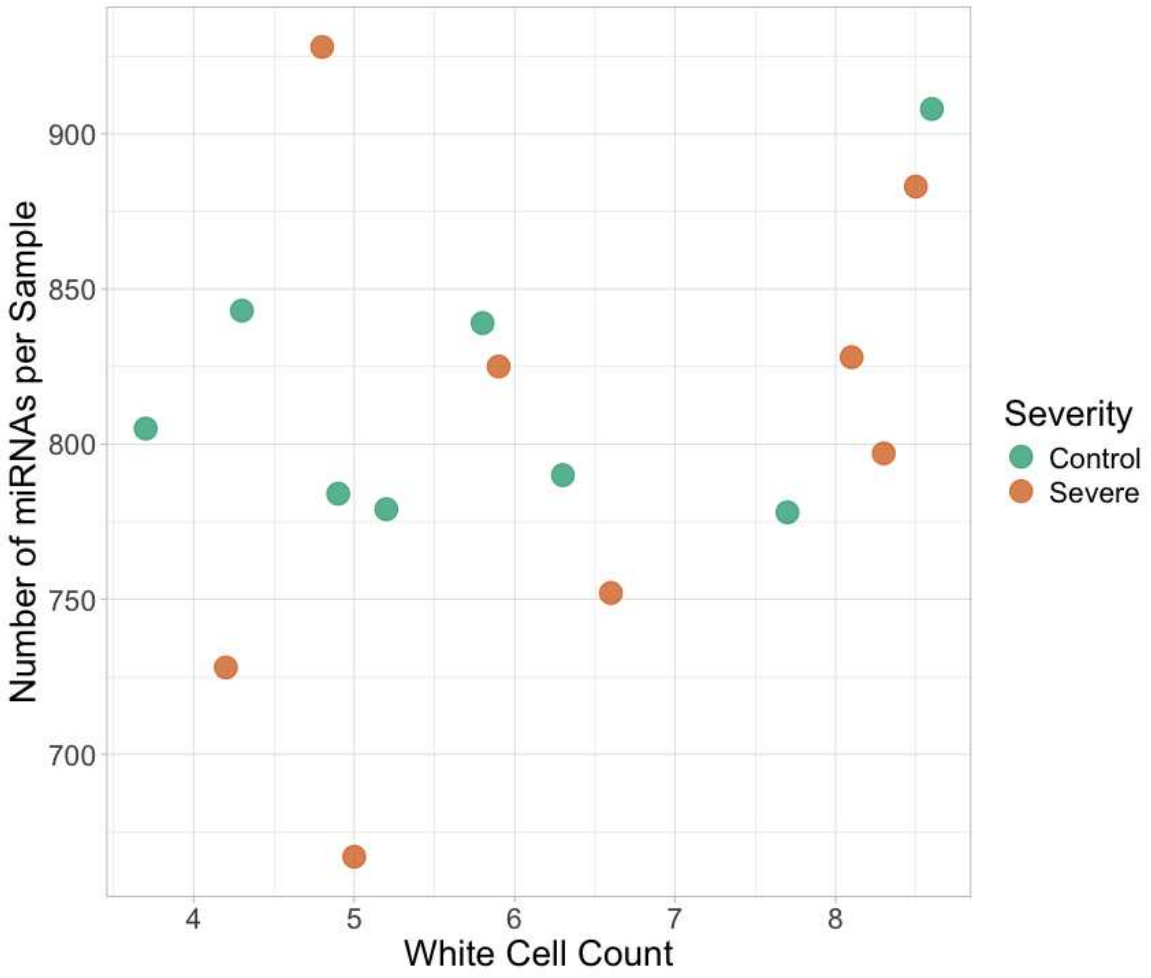


Figure 4.8. Pilot Study Sequencing Results miRNAs per Sample vs White Cell Count*. Total white cell count vs total number of miRNAs points as individual samples categorised by patient Severity. No significant difference in Wilcoxon signed-rank test (P > 0.05).*

A DE analysis was conducted on the aligned miRNA sense reads (1,364 in total) comparing between the two groups (Controls (healthy) and Cases (Severe Asthma)). A total of 12 miRNAs met the requirements of having an FDR < 0.05, while only 10 miRNAs had a fold change of greater than 0.5 and an FDR < 0.05. This group of 10 from this point forward will be referred to as DE-miRNAs. Figure 4.9. shows the results of this analysis in the form of a Volcano plot. All 10 DE-miRNAs (miR-1304-3p, miR-4482-3p, miR-374a-3p, miR-378c, miR-378d, miR-1307-5p, miR-32-5p, miR-224-5p, miR-17-3p, miR-24-2-5p) showed reduced expression in severe asthma compared to healthy controls.

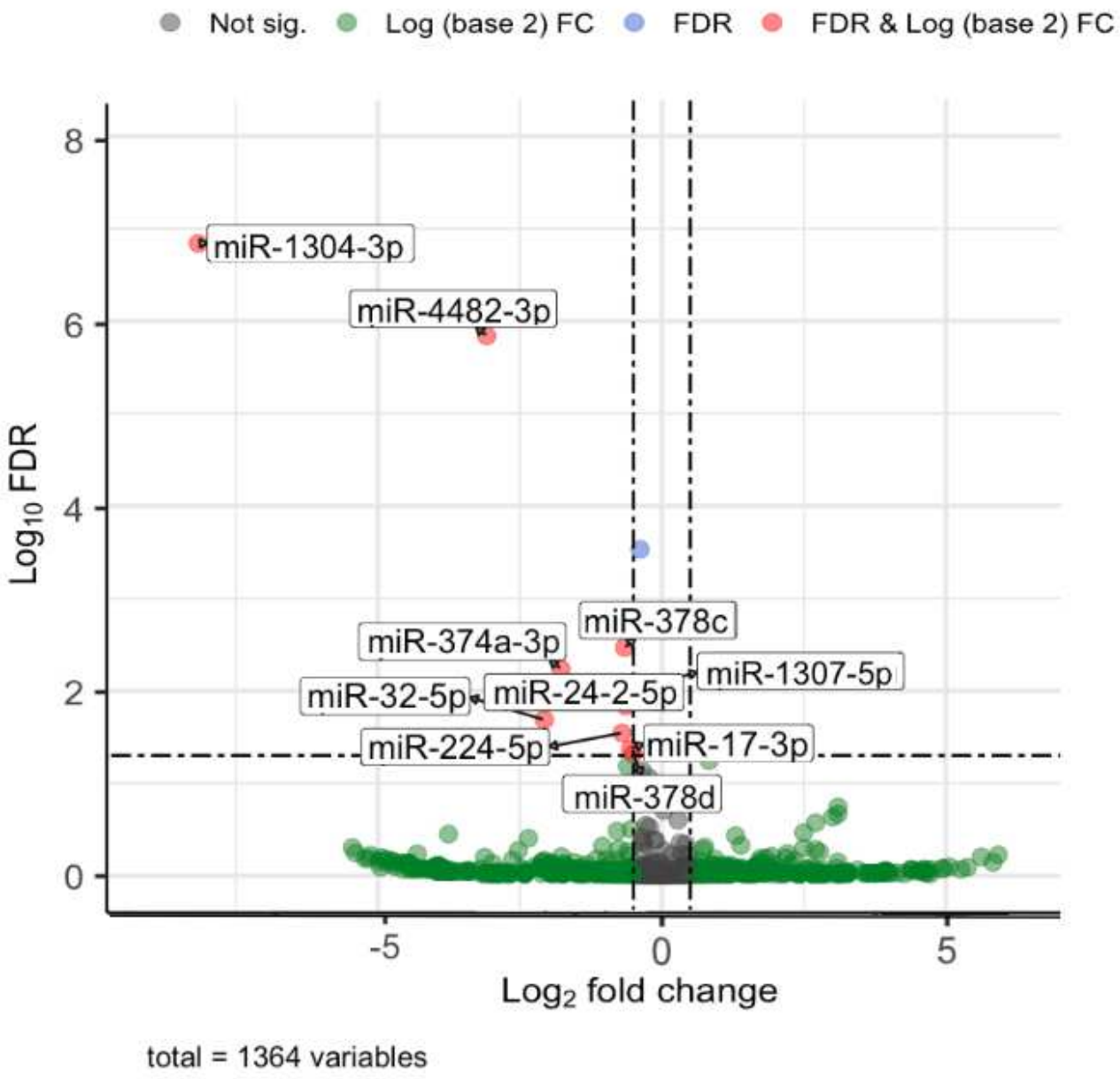


Figure 4.9. Volcano Plot of DE Analysis Result of miRNA Reads Compared by Severity *(Healthy Controls vs Severe Asthma) by FDR and Fold change (log2). Analysis was conducted using data normalised by quantiles and using the BaySeq method of DE analysis. Named miRNAs show 10 miRNAs that meet both the FDR < 0.05 and fold change > 0.5 requirements by the Volcano plot input (software limits the number of miRNAs labelled), 2 additional miRNAs (blue dots) met the FDR requirements for significance (one is obscured by labels). All DE-miRNAs showed reduced expression in Severe asthma compared to Healthy Controls.*

A correlation analysis was conducted to identify any potential associations between DE-miRNAs and six key clinical variables of interest: Age, Sex, BMI, FEV1/FVC ratio, Eosinophil counts and total IgE (Figure 4.10.). Patient age, gender, BMI, and total IgE showed no significant correlations with any DE-miRNAs.

Asthma severity, however, correlated with clinical features (as expected and previously noted Section 4.3.1.) as well as with miR-4482-3p. FEV1/FVC ratio showed a significant correlation with miR-32-5p. Neither Eosinophil count (Eosin) nor total serum IgE (IgE) shared a correlation with any of the DE-miRNAs. Eosinophil count, however, did have significant correlations with miR-378c and miR-378d. miR-1304-3p, miR-1307-5p, miR-224-5p, and miR-24-2-5p showed no significant correlations with clinical outcomes but each of them did correlate with one or more other miRNAs. Of the interactions between miRNAs, miR-32-5p showed the strongest association with miR-17-3p: a positive correlation of 0.79.

miR-1304-3p showed no correlation with clinical measures but did have weak positive correlation (of 0.17) with miR-4482-3p. miR-1307-5p showed no correlation with clinical measures but had a moderate positive correlation (0.6) with miR-24-2-5p. miR-224-5p and miR-24-2-5p showed no significant correlations with clinical outcomes but did correlate with one or more other miRNAs.

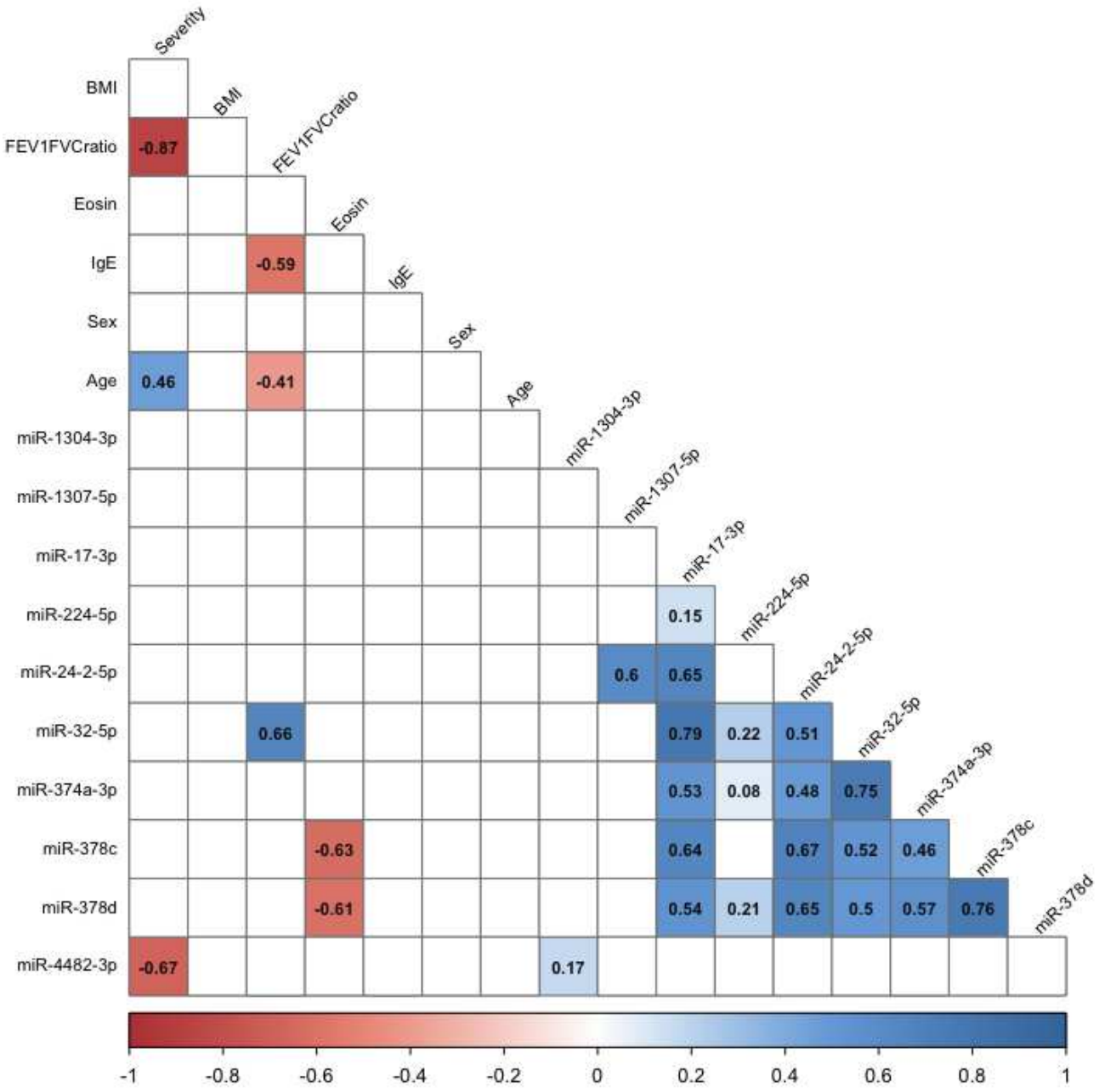


Figure 4.10. Correlation Heatmap of DE-miRNAs and Clinical Features of Interest*. A Spearman correlation with a 95% confidence level was used with a P value > 0.05 was considered significant. Only significant features (FDR < 0.05, and FC > 0.5) highlighted with correlation coefficients shown. Non-significant correlations have been left blank for ease of reference. BMI (Body mass index), FEV1/FVC ratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

4.3.4. miRNet mRNA Target Analysis

All 10 DE-miRNAs, with a FDR < 0.05 and fold change > 0.5, were analysed for targets using miRNet (Chang *et al.*, 2020). miRNet enables the user to conduct database target searches based on raw counts data or lists of DE-miRNAs, to identify miRNA-target-interaction networks based on multiple factors such as species, tissue and specific target database of interest. Results can be further filtered by the methodology that generated the target (full details of the methods and selections possible are given in Chapter 2, Section 2.1.5.3.). In the case of this experiment, only experimentally validated mRNA targets of the input miRNAs were used for the target identification and all those identified by sequencing-related target identification were excluded.

The results of the miRNet analysis were then selected within the platform and put through a series of Gene Ontology (GO) analyses. These link the identified gene targets to associated wider Biological Processes. GO analysis can give insights, for the condition being studied (in this case severe asthma) as to what Biological Processes (BP), Molecular Functions (MF) and Cellular Components (CC) are being impacted. This information can then be used to make selections of miRNAs or genes of interest for future work, identify patterns in miRNA target downstream pathways, or identify miRNAs with shared networks. During selection, all genes were analysed using Hypergeometric sampling for the three analyses (MF, CC, and BP) (Ju, 2011). The statistical methodology is discussed in Chapter 1. Section 1.2.2, and Chapter 2. Section 2.1.5.3). Graphs of each aspect of Gene Ontology were made with the genes identified as targets from the miRNet target analysis (Figures 4.11. to 4.13.).

Six of the 10 DE-miRNAs (miR-224-5p, miR-24-2-5p, miR-17-3p, miR-32-5p, miR-374a-3p, miR-4482-3p) were found to be associated with targets (N = 56) in the Biological Processes (BP)

Gene ontology analysis. As 100 categories were identified as being significant, only the top 10 are shown (Figure 4.11.). Please see Supplementary (Chapter 4: [CF1]) for tables of gene targets identified for all 100 categories with a $P < 0.05$.

The target category with the largest number of gene targets was 'positive regulation of metabolic process' N = 32 (adjusted [adj] $P < 0.001$) (Figure 4.11.) and the gene category with target genes with the highest significance was in 'regulation of phosphorylation' N = 22 (adj $P < 0.001$) (Figure 4.11.).

The top 10 categories were made up of 49 of gene targets and distribution of the latter across the categories is shown in Figure 4.11. There was one gene Homeodomain-interacting protein kinase 2 (*HIPK2*) that appeared in 8 of the top 10 categories. miR-224-5p has the highest total number of gene targets (N = 22), it also shares targets with miR-24-2-5p (target gene: *BCL2*), and miR-374a-3p (*RAC1*), and miR-17-3p (*NCOA3*). miR-17-3p also shares gene targets, with miR-32-5p (*ICAM1, PTEN*), and miR-4482-3p (*SOD2*).

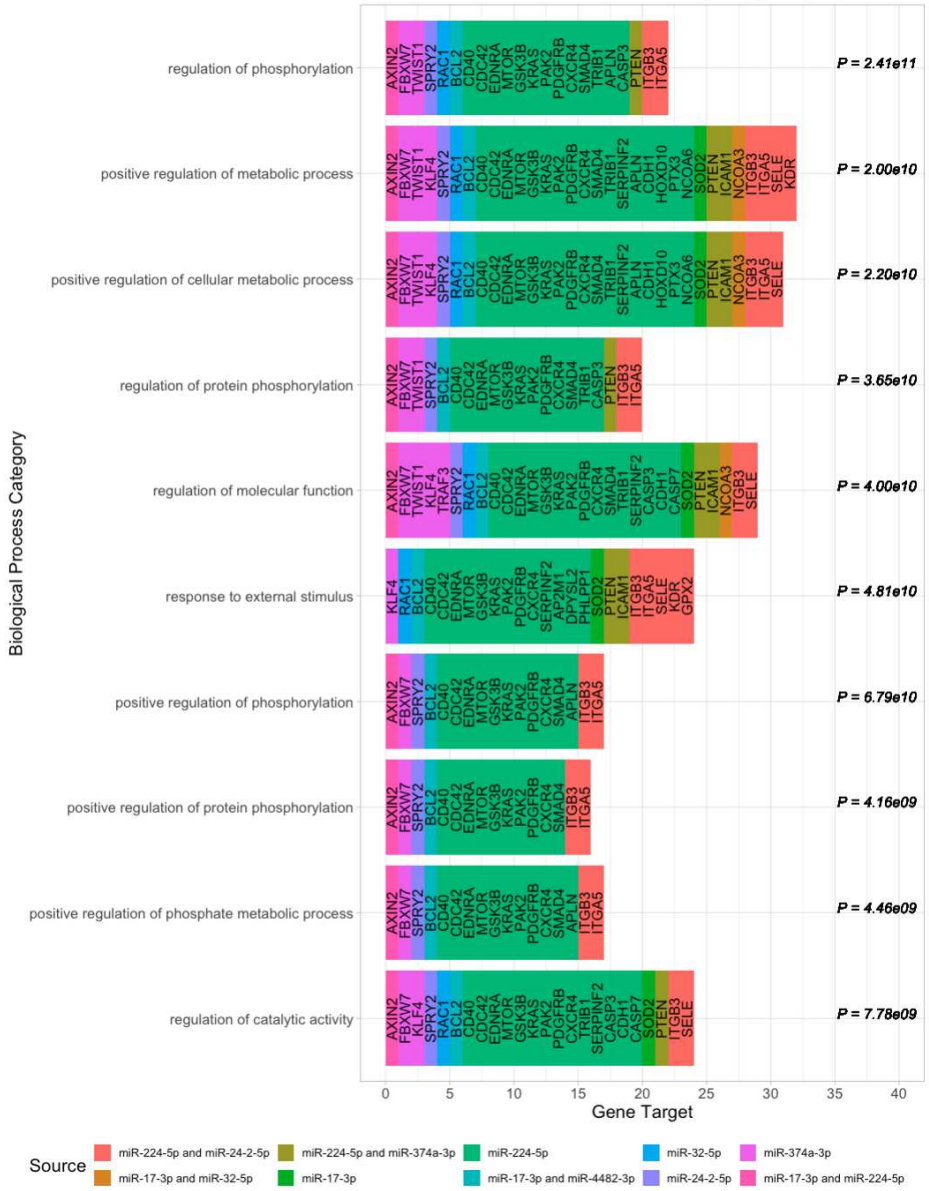


Figure 4.11. Partial Gene Ontology Classification (Biological Process) features and number of Target genes for differentially expressed miRNAs between Severe and Healthy Controls. *All Differentially expressed miRNAs between Severe and Healthy Controls were assessed for targets in miRTarBase 8.0 and the resulting targets analysed through GO classification Biological Process. The bars represent the number of target genes in each Biological Process category, ordered from most significant (top) to least (bottom) adjusted P-value within each category. The categories have been grouped on broad terms into facets to show any trends in gene targets. The bars are coloured to identify the source miRNA associated with the target gene, where more than one miRNA targets the same gene a different colour is allocated and noted in the legend 'Source'.*

Six of the 10 DE-miRNAs (miR-224-5p, miR-24-2-5p, miR-17-3p, miR-32-5p, miR-374a-3p, miR-4482-3p) were found to be associated with 55 unique gene targets in the Molecular Function (MF) Gene ontology analysis. The targets fell into 100 MF categories, the top 10 were selected for further analysis. Once filtered, there were 28 unique gene targets across all 10 of the top categories. The category with the greatest number of target genes was 'enzyme binding' N = 17 (adj $P < 0.001$) (Figure 4.12.), and the gene category with target genes with the highest significance were in 'protein kinase binding' N = 10 (adj $P < 0.001$) (Figure 4.12.) category. Full tables of the analysis results are available in Supplementary (Chapter 4: [CF1]).

In this analysis, miR-224-5p again had the highest number of gene targets (N = 19); it also shares targets with miR-24-2-5p (*BCL2*), miR-374a-3p (*RAC1*) and miR-17-3p (*NCOA3*), interactions that were also identified in the Biological Process Gene Ontology analysis above (Figure 4.11.). The same interactions between miR-17-3p gene targets, (with miR-32-5p (*ICAM1, PTEN*) and miR-4482-3p (*SOD2*) were also observed in the Biological Process analysis. Unique to the Molecular Function analysis were the shared targets between miR-224-5p and miR-32-5p (*MBD2*, and *PHLPP2*). Figure 4.12. below highlights the targets and shared interactions for each gene and category. Full results tables of the Molecular Function analysis are available in Supplementary (Chapter 4: [CF1]).

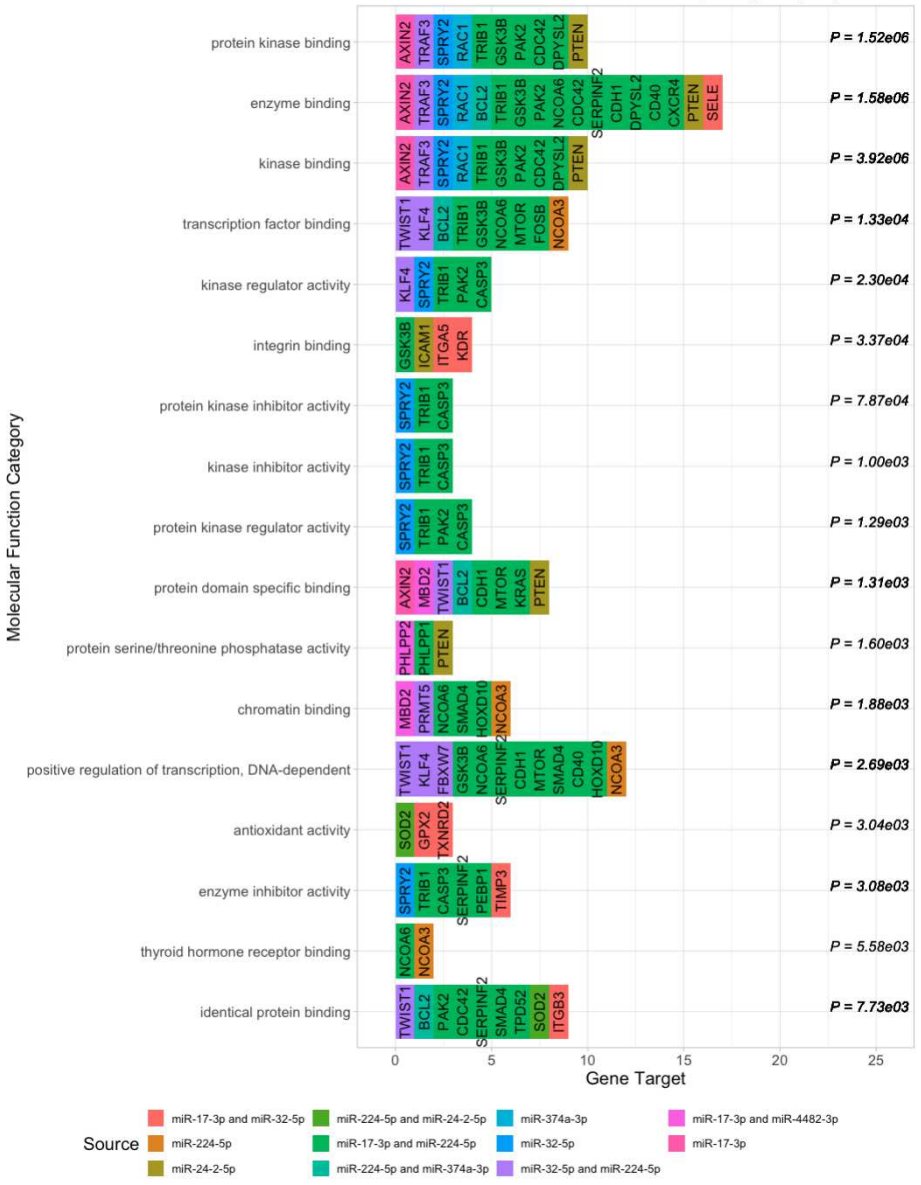


Figure 4.12. Partial Gene Ontology Classification (Molecular Function) features and number of target genes. *For differentially expressed miRNAs between Severe and Healthy Controls. All Differentially expressed miRNAs between Severe and Healthy Controls were assessed for targets in miRTarBase 8.0 and the resulting targets analysed through GO classification Molecular Function. The bars represent the number of target genes in each Molecular Function category, ordered from most significant (top) to least (bottom) adjusted P-value within each category. The categories have been grouped on broad terms into facets to show any trends in gene targets. The bars are coloured to identify the source miRNA associated with the target gene, where more than one miRNA targets the same gene a different colour is allocated and noted in the legend 'Source'.*

The final step in the GO analysis was to consider Cellular Components (CC) which revealed five of the 10 DE-miRNAs (miR-224-5p, miR-24-2-5p, miR-17-3p, miR-32-5p, miR-374a-3p) w to be associated with gene targets (N = 52) across 100 categories. These were filtered for the top 10 categories for further analysis and visualisation resulting in 24 gene targets. The description (category) with the greatest number of gene hits was the 'plasma membrane part' N = 15 ($P < 0.001$), and the gene target category most significantly associated was 'cell projection part' N = 5 ($P < 0.001$) (Figure 4.13.).

miR-224-5p had the highest number of gene targets (N =15). This miR also shares targets with miR-374a-3p (*RAC1*), as seen in both the Biological Process and Molecular Function GO analyses, as well as sharing another target with miR-32-5p (*PHLPP2*), seen in the Molecular Function GO analysis. Also, as seen in the previous MF analysis, the same interactions between miR-17-3p gene targets, with miR-32-5p (*ICAM1, PTEN*) were observed.

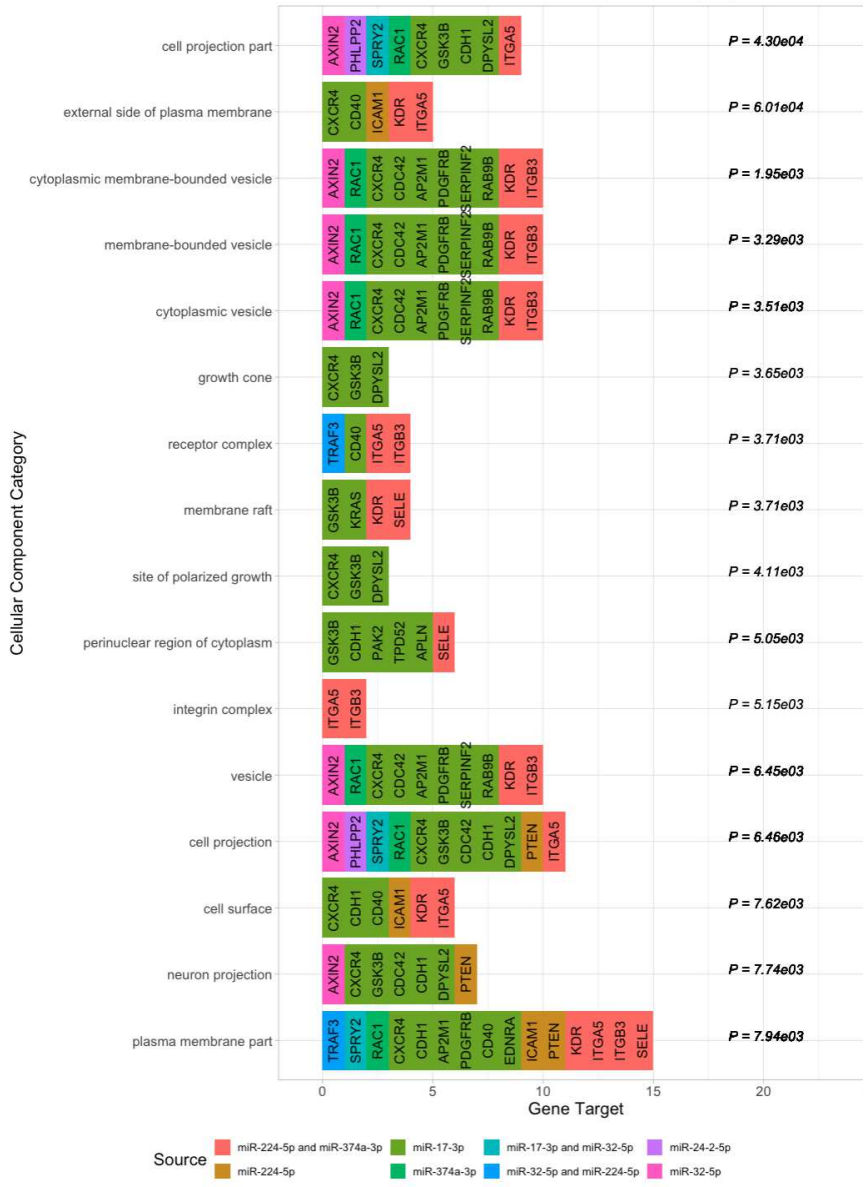


Figure 4.13. Partial Gene Ontology Classification (Cellular Components) Features and Number of Target Genes for Differentially Expressed miRNAs Between Severe and Healthy Controls. *All Differentially expressed miRNAs between Severe and Healthy Controls were assessed for targets in miRTarBase 8.0 and the resulting targets analysed through GO classification Cellular Components. The bars represent the number of target genes in each Cellular Components category, ordered from most significant (top) to least (bottom) P-value within each category. The categories have been grouped on broad terms into facets to show any trends in gene targets. The bars are coloured to identify the source miRNA associated with the target gene, where more than one miRNA targets the same gene a different colour is allocated and noted in the legend 'Source'.*

4.3.4.1. Comparison of Gene Target Identification among GO Analysis Categories for Control vs Severe Asthma Pilot

Next the results of the GO analyses were cross-validated, to identify significant gene targets across different biological contexts. This was to filter the genes for those that might have the highest chance of being biologically significant or a higher functional consistency. miRNAs that regulate such genes are likely critical in maintaining cellular homeostasis and function. Genes that are targets in multiple GO categories might be central nodes in biological networks. These central nodes (often referred to as hubs) are crucial for the stability and functionality of biological networks, indicating that miRNA interactions with these genes are essential for network integrity. The following results put together the GO gene targets from all analyses and relate back to the original DE-miRNAs identified.

From the results of the GO analysis from the DE-miRNAs with identified target associations (N =6), a Venn diagram was made to show crossover of gene targets among the GO-tested categories (Biological Process [BP], Molecular Function [MF], Cellular Components [CC]) top 10 most significant result categories (Figure 4.14.). Just under 40% of the genes identified were shared among all tests (17 genes). Details of the genes within each section of the Venn diagram are given in Table 4.3.

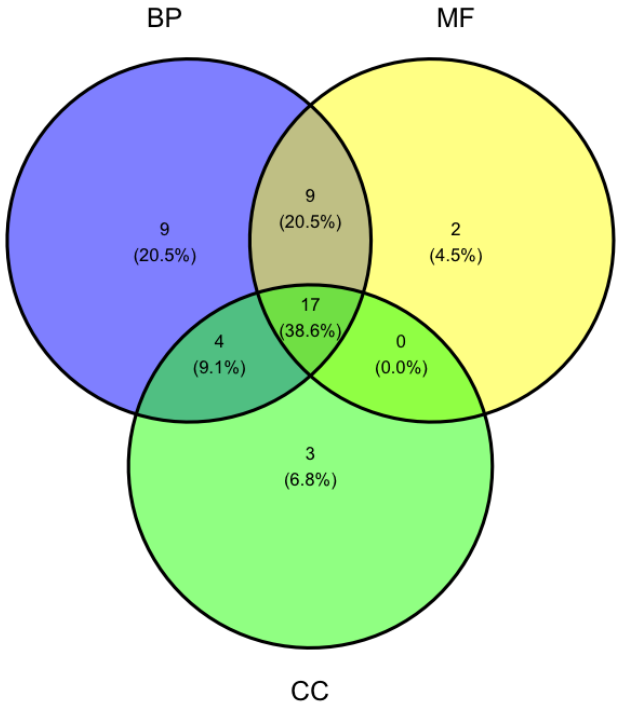


Figure 4.14. Venn Diagram of Target Genes Control vs Severe Asthma. *Genes identified as Significant in the Top 10 categories of each GO analysis Category in Healthy Control vs Severe asthma study blood sample miRNA expression. DE-miRNAs (N = 10) were input into miRNet database and passed through GO analysis categories (Biological Process (BP), Molecular Function (MF), and Cellular Components (CC)). Resulting gene targets were then compared among the three categories.*

Table 4.3. Target Genes from GO Analysis: Top 10 Most Significant (P < 0.05) Categories Compared Against One Another. *DE-miRNAs (N = 10) identified from blood sample miRNA expression between Healthy Controls vs Severe asthmatics. Unique genes and genes identified among other GO analysis categories.*

Interaction (analysis)	***Gene names***
Unique to Biological Process	*PTX3, CASP7, PHLPP1, SOD2, SMAD4, EDNRA, HOXD10, GPX2, FBXW7*
Unique to Molecular Function	*MBD2, FOSB*
Unique to Cellular Components	*RAB9B, PHLPP2, TPD52*
All GO analysis categories	*ITGA5, CD40, CDC42, GSK3B, KRAS, PAK2, RAC1, CXCR4, SPRY2, AXIN2, ICAM1, KDR, SELE, CDH1, SERPINF2, TRAF3, DPYSL2.*
BP and MF shared	*PTEN, TRIB1, BCL2, NCOA3, NCOA6, TWIST1, KLF4, MTOR, CASP3*
MF and CC shared	
BP and CC shared	*AP2M1, PDGFRB, ITGB3, APLN*

To identify the interaction between the miRNAs input and the gene targets identified as significant within specific categories of each GO analysis, the DE-miRNAs were aligned to their corresponding associated target genes. The following results are those associations; the goal being to identify (from the pool of DE-miRNAs) any specific target processes or pathways that are more targeted.

With the goal of filtering for the gene targets and source miRNAs that target those genes, the GO results were pooled and analysed. Gene targets were filtered for those with at least 1 source miRNAs and a stacked bar graph was created to show both the interactions and how the gene targets were shared among the GO categories (Figure 4.15).

Of the 10 miRNAs, that were identified as DE between healthy controls and severe asthmatics with an FDR < 0.05 and a FC > 0.5, there were 5 source miRNAs identified among all GO analysis categories with at least one gene target or more. While all shared interactions are considered below with the 6 source miRNAs, the focus was on those shared in all categories and was therefore limited by the results of the CC analysis (where only 5 source miRNAs had associations with gene targets) (outlined in red, Figure 4.15.). There were 17 genes shared among all three GO analysis categories. Two genes in the shared group had two source miRNAs (*ICAM1* targeted by: miR-32-5p and miR-17-3p, *RAC1* targeted by: miR-374a-3p and miR-224-5p). The predominant source miRNA across all targets was miR-224-5p. The remainder had only a single source miRNA.

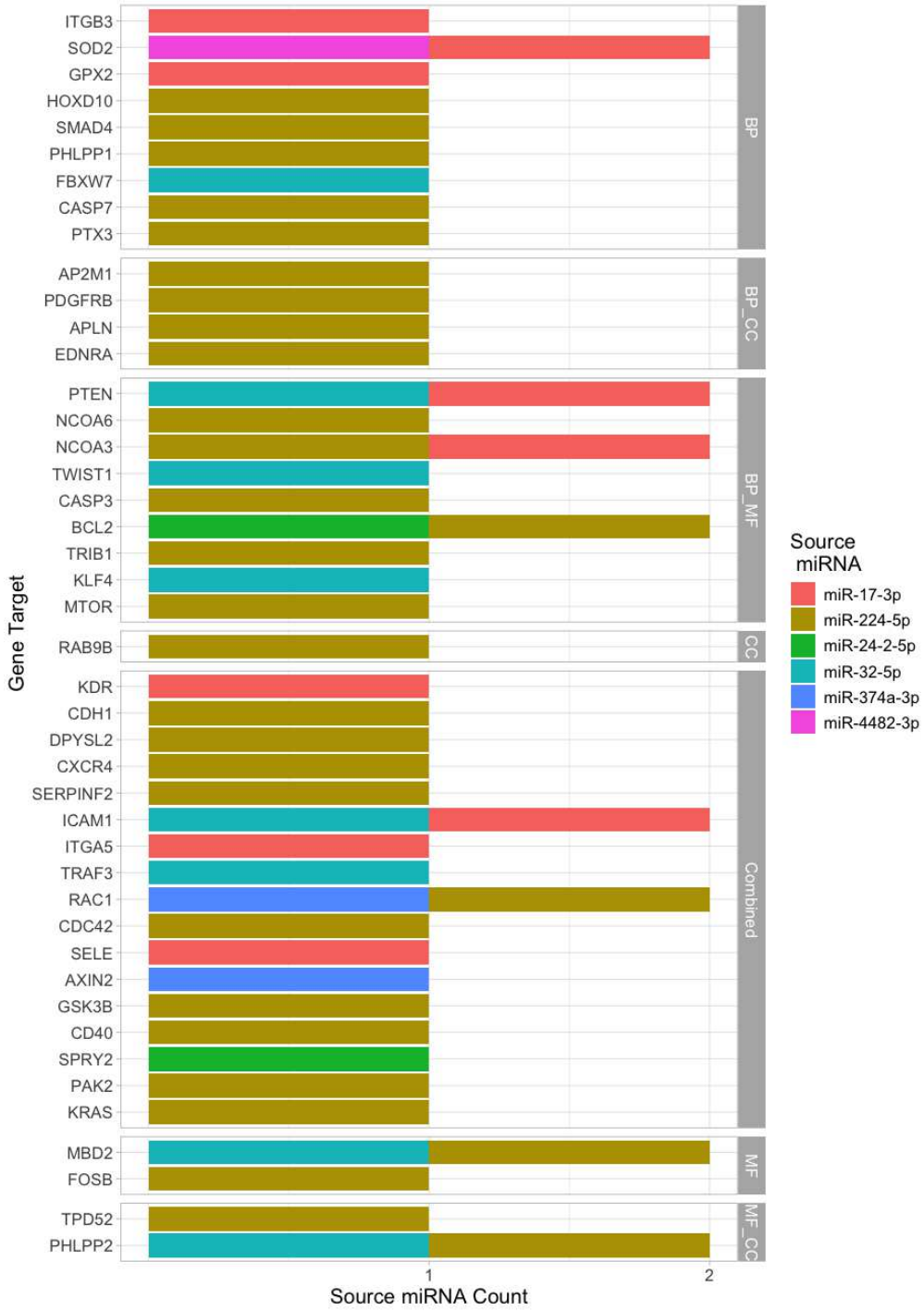


Figure 4.15. Gene Targets and Source miRNAs from GO Analysis Faceted by Analysis Interaction. *DE-miRNAs (N = 10) from blood sample analysis between Healthy Controls and Severe Asthma (N = 18) were input into multiple GO analyses, 6 of the DE-miRNAs were identified as having published associations with genes. The x-axis shows the gene names, and the fill shows the miRNA source, while the y-axis shows a count of how many miRNAs are sources for that gene target. Grey bar denotes the GO categories (Biological Process (BP), Molecular Function (MF), and Cellular Components (CC), as well as genes found between categories, and those found in all three analyses (Combined). Among all GO analyses, two genes were targeted by two miRNAs (ICAM1 and RAC1).*

4.3.5. Literature Comparison

A review of the literature was conducted to compare the identified DE-miRNAs with those found in sequencing experiments. The search looked at any sequencing studies that included miRNA and where all DE-miRNAs were published or made available (either through supplementary materials provided within publications or direct communication with the authors).

There was no specific tissue or sample-type required for inclusion other than being from human patients or volunteers. For inclusion into the comparison, the research also needed to be isolated to the comparison of asthma to healthy controls and no treatment groups.

Of the research published between 2015 and 2023, two studies published DE-miRNA findings regarding differences between asthmatics and healthy controls. The Roffel *et al*. (2021) study was conducted on bronchial biopsy samples, and features such as smoking and ICS-use were controlled for. The Francisco-Garcia *et al*. (2023) study, not published at the time of the Chapter 3 Nasosorption analysis, was conducted on bronchoalveolar lavage from healthy controls and severe asthmatics. Table 4.4. shows the DE-miRNA findings from the present study and DE-miRNAs details from the two papers identified by the literature search.

Table 4.4. DE-miRNAs Identified in Literature (and the Results of This Study)*. Comparison findings of two published studies comparing asthma and healthy Control miRNA expression with this current study. Details on DE-miRNAs, patient type and numbers, as well as sample type.*

DE-miRNAs identified	Sample size	Sample	PMID	Reference
miR-99b-5p, miR-181b-5p, miR-141-3p, miR-374b-5p, miR-92a-3p, miR-203a-3p, miR-23b-3p, miR-100-5p, miR-150-5p, miR-199b-5p, miR-139-5p, miR-320a, miR-200c-3p, miR-92b-3p, let-7a-5p, miR-374a-5p, miR-214-3p, miR-1247-5p, let-7d-3p, miR-574-3p,miR-26a-5p, miR-190b, miR-223-3p, miR-200a-3p, let-7b-5p, miR-30d-5p, miR-30a-3p, miR-19b-3p, miR-103a-3p,miR-99a-5p, miR-200b-3p, miR-183-5p, miR-148b-3p, miR-21-5p, miR-195-5p, miR-34c-3p, miR-361-5p, miR-142-5p, let-7e-5p, miR-199b-3p, miR-182-5p,miR-34c-5p, miR-10a-5p, miR-193a-5p, let-7c-5p, miR-10b-5p, miR-652-3p, miR125a-5p, miR-429, miR-20a-5p, miR-486-5p, miR-25-3p, miR-425-5p, miR-185-5p, miR-106b-5p, miR-125b-5p, miR-15b-5p, miR-449a, miR-93-5p, miR-451a, miR-16-5p	79 asthma, 82 Healthy Controls	Bronchial biopsy	34446467	Roffel *et al.*
miR-9-5p, miR-10b-5p, miR151a-5p, miR-151a-3p, miR-202-5p, miR-202-3p, miR-224-5p, miR-568, miR-581, miR-615-3p, miR-625-3p	8 Healthy Controls, 12 (Severe) asthma	Broncho alveolar Lavage	31684064	Francisco-Garcia *et al.*
miR-1304-3p, miR-4482-3p, miR-181a-5p, miR-378c, miR-374a-3p, miR-1307-5p, miR-24-2-5p, miR-32-5p, miR-224-5p, miR-191-5p , miR-17-3p, miR-378d	9 Healthy Controls, 9 Severe asthma	Whole blood	This thesis	Headley

To aid interpretation of the data in Table 4.4., a Venn diagram was created to visualise the overlap in terms of identical miRNAs between the three studies, the two published studies and the present study (Severe study) (Figure 4.15.). Between the published papers, only one DE-miRNA was shared (miR-10b-5p). No DE-miRNAs were seen in common between the present study (Severe study) and the Roffel *et al.* study but one DE-miRNA miR-24-2-5p was in common between the present study and the Francisco-Garcia *et al.* paper.

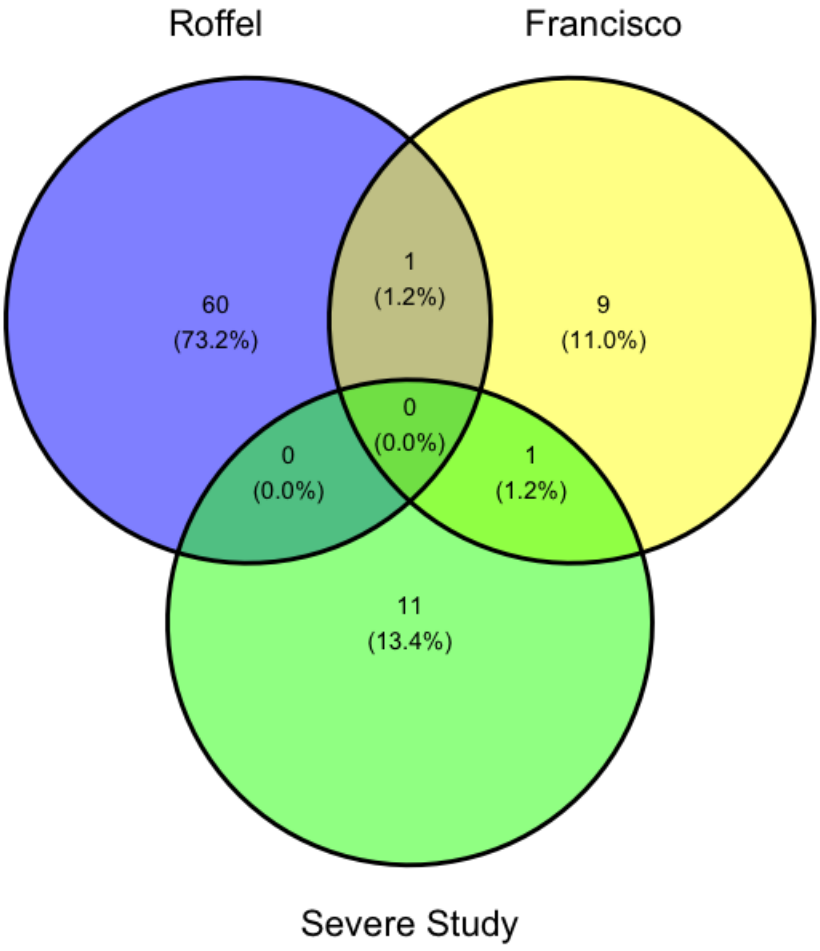


Figure 4.16. Venn diagram of comparison of DE-miRNAs with literature and Control vs Severe*. Two publications comparing miRNA expression between asthmatics and healthy controls, and the results of the present study comparing Severe asthma miRNA expression to Healthy Controls. Of the 12 DE-miRNAs identified in the present study, only one was identified as a DE-miRNA in the Francisco (Francisco-Garcia et al. 2023) study. Between the Francisco and Roffel (Roffel et al. 2021) studies only one miRNA, miR-10b-5p, was identified in both.*

4.4. Discussion

This Chapter has described a pilot study conducted to identify if differences exist between the expression of circulating (blood) miRNAs of healthy Control subjects versus Severe asthmatics through applying an unbiased sequencing approach. Blood samples of study subjects were obtained via the Celtic Fire Study (Section 4.2.1.). The size of the latter study, plus its recruitment criteria, meant that subjects for the present pilot study could be selected to ensure the exclusion of smokers and those with a BMI over 30 - both being factors recognised to be associated with asthma (Beasley, Semprini and Mitchell, 2015; Wong, Forno and Celedón, 2022). Gender and age (Chowdhury *et al.*, 2021) were also taken into consideration and balanced as much as possible with no significant differences seen between the two groups (Healthy Controls and Severe asthmatics). Comparison of several other clinical features revealed, as expected, differences between groups for the ratio between forced expiratory volume in one second and forced expiratory volume (FEV1/FVC ratio), and total Immunoglobulin E (IgE) serum levels.

RNA isolation and library synthesis, using a total RNA extraction kit (that included isolation of small RNAs) followed by a small RNA library synthesis kit and size selection method using the Pippin Prep, were successfully conducted. A concentration kit was required to concentrate a subset of the samples following size-selection on the Pippin Prep. The Pippin prep is restricted by small input volumes and the experimental methods in this thesis had the limitation of low input concentrations. Although experimental results from this thesis revealed a sample yield of approximately 40% (surpassing the manufacturer's quoted recovery of 20%), this created additional challenges, on top of low input concentrations, towards ensuring there was sufficient concentration for subsequent sequencing steps. The HT Pippin (a high throughput version of the Pippin Prep) does have a more efficient size-selection over the Pippin Prep, purely as the

electrophoresis gel wells are smaller. This project did not have access to a HT Pippin but for future studies its use would be worth considering.

Prepared libraries were of suitable quality for sequencing as assessed from the Agilent Bioanalyzer output and this was confirmed by the Sequencing Facility (Novogene - in-house quality control using Tape station). Except for two samples (that generated 2.9M and 4.6M reads respectively), all other RNA samples generated the minimum goal of 5M reads. Quality of reads, based on FASTQC and read alignments, was high. Read alignment methods using the Genboree workbench proved effective, enabling a selection of quality filters and negative control filters.

There was a high percentage of reads for alignment to the genome and mature miRNA sense, indicating that the samples (the libraries) and the sequencing were of good quality. Alignments to exogenous and other types of RNA (piRNA, tRNA) were low and confirmed that the protocols employed for the miRNA sequencing were robust.

There was a concern around the miRNA variance regarding variance between groups due to the source of the miRNA being from whole blood. As asthma is an inflammatory condition associated with elevated eosinophils, differences in white cell populations may have had an impact on miRNA abundance between groups. This was tested for and white cell counts were not significantly different between the two groups.

Taking the aligned miRNA data, a differential expression analysis was conducted comparing the Healthy Controls group versus the Severe asthma group. Ten miRNAs were identified as differentially expressed. All DE-miRNAs were found to be more highly expressed in healthy controls compared to severe asthma. The DE-miRNAs all had significant (P <0.05) fold changes greater than 0.5 and these were (in order of fold change given in brackets): miR-1304 (8.2),

miR-4482-3p (3.1), miR-32-5p (2.1), miR-374a-3p (1.8), miR-224-5p (0.7), miR-378c (0.7), miR-24-2-5p (0.6), miR-1307-5p (0.6), miR-17-3p (0.6), miR-378d (0.5).

Correlation analyses were conducted, revealing a number of associations with particular clinical measures and other DE-miRNAs. While few of the DE-miRNAs were identified as having a significant correlation with clinical measures, all had at least one correlation with another DE-miRNA. Each of the 10 DE-miRNAs will now be discussed in more detail.

Of the 10 DE-miRNAs, miR-1304-3p had the most significant and largest fold change (8.2) higher in healthy controls compared to severe asthmatics. There is very little in the published literature about miR-1304-3p and its contribution to asthma. One *in vitro* lung cancer cell line study reported tumour-suppressor targeting heme-oxygenase 1 (*HO-1*) as a target of miR-1304-3p (C.-G. Li *et al*., 2017). One aspect of the research that was supportive of the complicated nature of miRNAs is that HO-1 expression could be recovered through a different mediator despite miR-1304-3p expression, in a synthetic environment. Therefore, understanding interacting suppressing and supporting pathways could give a clearer indication as to what miRNAs are crucial to the key physiological responses seen in asthma. By contrast, miR-1304-3p has been identified as being upregulated in pulmonary hypertension as well as in coronary heart disease (Chen *et al*., 2023; Onuoha *et al*., 2022). While a promising idea in that miR-1304-3p is seen as upregulated in miRNA dysregulated conditions in cancers, the impact on increasing regulation on multiple systems with the awareness of which have functional compensation may limit the practicality of a pharmacological intervention harnessing the miR-1304-3p pathway. The correlation analysis conducted in this present study identified no association with clinical measures, but some correlations with one other DE-miRNA within the study i.e. miR-4482-3p.

MiR-4482-3p was found to have a significant fold change of 3.1 in this present study – being higher in Healthy Controls compared to Severe Asthmatics. In a study, published in 2023 (X. Ma *et al.*, 2023), miR-4482-3p was identified as part of a group of non-coding RNAs (two miRNAs and two circRNAs) that were significantly downregulated (>2 FC) and collectively a potential marker for non-small cell lung cancer. The network of the four non-coding RNAs attenuated expression was able to identify between the stages of cancer within patients as well as the degree of metastasis (X. Ma *et al.*, 2023). The authors of the study also highlighted the existence of a single nucleotide polymorphism of the miR-4482-3p miRNA within Chinese populations. This polymorphism increases the susceptibility of individuals to suffering from rheumatoid arthritis, through a suggested effect around the alteration in targets because of the polymorphism (a gain of 1,677 targets and a loss of 1,456), which could significantly alter an inflammatory response. A link between miR-4482-3p with inflammation was also reported earlier in a 2020 study by Rotival *et al.* (Rotival *et al.*, 2020). The study identified differences in the expression of miRNAs in CD4+ cells isolated from Europeans and African participants treated with various immune activating compounds (including LPS and a strain of influenza). One of the miRNAs identified was miR-4482-3p, which was found to be more highly expressed in individuals with African ancestry compared to those with European ancestry. This could limit the potential of miR-4482-3p as a diagnostic marker in isolation, and highlights the importance of understanding miRNA expression across ethnicity as a potential confounding factor. Conducting network and pathway analysis could identify interactions between miRNAs and target genes and what pathways are enriched providing insight into the functional effects of a tested intervention method.

Asthma is understood to be a heterogeneous condition with individuals being affected with unique sensitivities and responses. CD4+ are so-called helper T cells (Chapter 1, Section 1.1.2.), important for fighting infection and asthma exacerbations can arise from both viral and bacterial

infections. The variation in expression of miRNAs in response to an immune challenge could go some way to explaining differences in asthma development as well as progression of the condition, or even the response to pharmacological interventions. In this present study, miR-4482-3p was negatively correlated with patient severity as all DE-miRNAs had lower expression in severe asthmatics than healthy controls.

miR-32-5p was also found to have a significant fold change of 2.1, in this present study- being higher in Healthy Controls compared to Severe Asthmatics. There is little presently in the literature around the involvement of miR-32-5p in asthma. miR-32-5p has been identified as having a potential role in non-small cell lung cancer and is often found to be downregulated with an impact on the SMAD3 family of genes (Gao *et al.*, 2016). An interaction between the SMAD genes (1/3/6) and TGF-β in allergen challenge has been identified in asthmatics (Panek *et al.*, 2022). This may be a link between the expression difference identified between the two groups in the present study and the gene target effect.

Although a literature search revealed no publications in relation to miR-374a-3p, more broadly, miR-374a (the 5p variant) has been reported to be associated with lung function in children (Kho *et al*., 2016). One further study identified a reduction in miR-374a expression when there was an elevated level of circulating eosinophils in allergic asthmatics (Weidner *et al*., 2020). In this present study, no clinical measures (including eosinophil counts) were correlated with miR-374a-3p.

Downregulated in the severe asthmatics in this present study, miR-24-2-5-p, has also been found to be downregulated in a house dust mite (HDM) model of asthma (Cay *et al*., 2022).

While miR-17-3p has not been identified in the literature to date to have a relationship to asthma, it has been reported to be upregulated in respiratory *Candida* infections (Muhammad *et al.*,

2015). This suggests that miR-17-3p may have an inflammatory role in response to fungal agents that may be present as part of or contributing to the severity of the asthma (Rick *et al.*, 2020). It has been proposed that fungal sensitisation may be linked to asthma severity (Denning *et al*., 2006). miR-17-3p was not found to be correlated to any of the clinical measures in this present study and no fungal measures were available for the study subjects

That miR-224-5p was downregulated in severe asthmatics in this present study, agrees with the findings from a study (granted of a different tissue type but respiratory in nature) of nasal biopsy samples from asthmatics and healthy controls. The study identified miR-224 expression to be downregulated in asthmatics with allergic rhinitis (Suojalehto *et al*., 2014).

In this present study several of the DE-miRNAs showed significant correlations with other miRNAs (miR-374a-3p, miR-17-3p, miR-224-5p), and whilst some showed no correlation with clinical measures these three (brackets above) miRNAs exhibited consistent alterations in expression levels between asthmatics and controls (Figure 4.10.), suggesting a shared regulatory response to the underlying biological stimuli. Moreover, analyses reveal significant correlations among these DE-miRNAs, indicating potential cooperativity or coordination in their regulatory functions, and as such, merit note.

The 10 DE-miRNAs identified were next analysed for gene targets and Gene Ontology networks using a strict experimentally validated inclusion criteria of interaction table information. The analysis considered three aspects of gene function: Biological Processes, Molecular Function, and Cellular Components. Only experimentally validated mRNA targets of the input miRNAs were used for the target identification and all sequencing-related target identification was excluded. This is particularly important as miRNA research is not as well-known and conservative associations may be the most helpful in drawing conclusions regarding possible pathways and for

further exploration. It has already been suggested that one of the challenges of identifying the functional significance of miRNA is the lack of genome wide expression data combining miRNA results. Genome wide transcriptome sequencing still, to some extent, focusses on mRNA (Barbato *et al.*, 2009). This explains why, out of the 10 DE-miRNAs that were identified, only 6 were acknowledged on the database as having associated gene targets (BP and MF), and only 5 miRNAs for the CC analysis.

The results of the miRNet analysis then underwent the Gene Ontology (GO) analysis. The latter allows one to link the identified gene targets to associated wider Biological Processes. GO analysis can give insights, for the condition being studied (in this case severe asthma), as to what Biological Processes (BP), Molecular Functions (MF) and Cellular Components (CC) are being impacted.

In order to focus on the key most significant pathways of each analysis, the top ten categories were selected and the gene targets and miRNA sources were used for the analysis that followed, enabling a focused view of the key pathways and processes that were the most targeted. Since few experiments have been done to date and there is little information on unbiased miRNA expression in asthma, application of this approach was viewed as an efficient way of identifying unique pathways relating to the processes of severe asthma compared to non-asthmatics.

Some of the gene targets shared among the three categories (BP, MF, CC) will be discussed as well as any that link to that grouping. For Biological Process, GO analysis revealed regulation of phosphorylation (adj $P < 0.001$) to be the category with the highest significance. Several other categories, within the top 10 most significant, also notably involved cell turnover regulation i.e. regulation of cell proliferation, cell proliferation, and regulation of programmed cell death (See Supplementary for full results tables for GO analysis conducted in this study). Considering the

different cell types involved in asthma pathogenesis, Liu *et al.* (2023) highlighted an identified impairment of eosinophil apoptosis resulting in increased inflammation with a cause unknown. It has been suggested that regulation of this apoptotic resistance could be a potential treatment for asthma inflammation (Ilmarinen and Kankaanranta, 2014).

It is interesting to note that in the case of miR-378c, Li *et al.*, (2019) identified the miRNA to be involved in the promotion of apoptosis resistance and proliferation of airway smooth muscle cells from asthmatic children in cell culture experiments. While this interaction has been noted in the literature, it was not one of the 6 source miRNAs with experimentally validated interactions in the Gene Ontology analysis, which may give some insight as to the nascent understanding of miRNA involvement in human biology. Whilst four of the top 10 most significant categories were involved in cell turnover interestingly, a further three were to do with metabolism (Figure 4.11.). It is also interesting that dysregulated metabolic pathways have likewise been identified in asthma (Ho *et al.*, 2013).

Considering Molecular Functions: enzymatic binding and inhibition made up several of the most significant categories identified. While the targets (*FOSB*, *MBD2* - Table 4.3.), that were found exclusively in the Molecular Function category with the 10 DE-miRNAs input, encode for enzyme linked proteins, several of those found in all the GO categories were linked with the most significant function categories made up of kinase and kinase-binding or enzymatic-binding (Figure 4.11.). For example: *CD40* and *CDH1* (Cadherin 1) whilst not directly encoding enzymes are related to signalling pathways and cellular adhesion molecules that regulate enzymatic activities indirectly. CD40 expression is increased in bronchial epithelial cells, smooth muscle cells, and eosinophils in asthmatic patients (Park *et al.*, 2007), has been targeted through

monoclonal treatment in human asthma trials (Samant, Ziemniak and Paolini, 2023) providing potential alternate methods of targeting this mechanism with miRNA treatments.

Another target found in all GO categories was *GSK3B* (Glycogen synthase kinase 3 beta) which encodes an enzyme involved in various signalling pathways thereby regulating numerous cellular processes. Smooth muscle hypertrophy in patients with asthma have been identified as linked to this protein (Deng *et al.*, 2008). No human studies have been conducted to identify the mechanism although several murine studies suggest that *GSK3B* may be a contributor of airway smooth muscle hypertrophy (Bao *et al.*, 2007).

Others, not found in all categories but in at least two included *PHLPP1* and *PHLPP2* (PH domain and leucine-rich repeat protein phosphatase 1 and 2) that encode enzymes involved in the regulation of various signalling pathways. *MTOR, PHLPP1* and *PHLPP2* are linked through interactions and dependencies in inactivating *AKT* which is an oncogenic kinase that, although not previously noted to be linked to asthma, is linked to several cancers including lung cancer (Newton and Trotman, 2014).

While *MTOR* was not found as a target between the combined GO analysis it was identified as a target between the BP and MF groups. *MTOR* (Mechanistic target of rapamycin) encodes a protein kinase the latter being an essential component of the mTORC1 complex involved in cellular metabolism. The *MTOR* pathway signalling is known to be involved in asthma onset and linked with homeostasis between TH17/Treg and TH1/TH2 cytokines (Zhang *et al.*, 2017).

Together, the gene targets and Gene Ontology results provide a breadth of links of how the differentially expressed miRNAs might be altering gene expression through their impact on Biological Processes, Molecular Functions and Cellular Components combined. The results of the combined Gene Ontology analyses suggest a miRNA contribution to known processes and

pathways identified in asthma, such as impaired apoptotic pathways in eosinophils and airway cells, all supporting the miRNAs identified as being contributory factors in asthma pathophysiology.

Lastly, tying back to the DE-miRNAs that showed no correlation with clinical outcomes but significant correlations with other DE-miRNAs, miR-224-5p was the source of many targets across all GO categories tested (Figure 4.14.). While this may mean it is the most researched miRNA of those identified in this study, the correlated expression patterns further suggest potential interactions or regulatory relationships among these miRNAs, where changes in the expression of one miRNA may influence the expression of others within the group. These findings highlight the importance of considering the collective behaviour of miRNAs in deciphering their regulatory roles and understanding the intricate molecular mechanisms underlying the observed biological responses. Further exploration of the functional implications and regulatory networks involving these DE-miRNAs could provide valuable insights into their roles in the context of the experimental conditions and their potential contributions to cellular processes and disease mechanisms.

In conclusion, this pilot study has (using a robust sequencing approach) identified DE-miRNAs that appear to be of biological relevance in relation to severe asthma and show potential, if validated, of use as biomarkers. Only 10% of asthmatics have severe disease and therefore the following Chapter will investigate miRNA expression between mild and severe asthma.

5. Chapter 5: miRNA Expression Comparison Among Severe Asthmatics, Non-Severe Asthmatics and Healthy Control Whole Blood Samples

5.1. Introduction

As highlighted previously (Chapters 1, 3 and 4), asthma is a chronic inflammatory disease of the airways characterised by symptoms such as wheezing, shortness of breath, chest tightness, and coughing. The disease affects people of all ages, with varying degrees of severity ranging from mild to severe. In the prior Chapter (Chapter 4) the role of miRNAs was explored for severe asthma as this poses a significant burden on patients and healthcare systems due to its resistance to standard treatments and the associated morbidity and healthcare costs. Severe asthma, however, only affects approximately 10% of asthma patients. Therefore, an understanding of the underlying processes of differing severities of asthma could be helpful in more effective treatment strategies and the development of novel treatments.

As detailed in Chapter 4, the Celtic Fire Study not only recruited severe asthmatics and healthy controls but also patients suffering from mild/moderate (Non-Severe) asthma. Therefore, in this Chapter as well as considering severe asthma the study was extended to include non-severe asthmatics, enabling a more robust statistical analysis (due to increasing size of subject groups) and potentially uncovering novel miRNA signatures associated with asthma severity. As before

(Chapter 4), in this Chapter the samples have been analysed in relation to miRNA expression, correlation with clinical outcomes, and Gene Ontology.

The specific objectives of this Chapter were:

1. To identify differentially expressed (DE) miRNAs in whole-blood samples across the three groups: Healthy Controls, Non-Severe asthmatics, and Severe asthmatics.
2. To explore the potential target genes and associated biological pathways of the identified DE-miRNAs thereby providing insights into the molecular mechanisms underlying asthma pathogenesis and severity.
3. To compare and contrast the findings from this larger cohort study with the previous pilot study and other relevant published literature, highlighting differences and similarities in expression targets or patterns.

5.2. Materials and Methods

5.2.1. Study Participants and Study Conduct

The study was conducted applying the same methodology as Chapter 4 but including mild-moderate asthmatics recruited by the Celtic Fire Study. The three groups of subjects therefore were:

Healthy Controls (Non-asthmatics).

Mild asthmatics: *symptoms were controlled with intermittent inhaled SABA only.*

Severe asthmatics: *dependence on systemic corticosteroid therapy for disease control.*

5.2.2. Overview of Sample Processing

An overview of sample processing is shown below in Figure 5.1. Sample processing decisions are highlighted, and quality control (QC) steps are noted in grey.

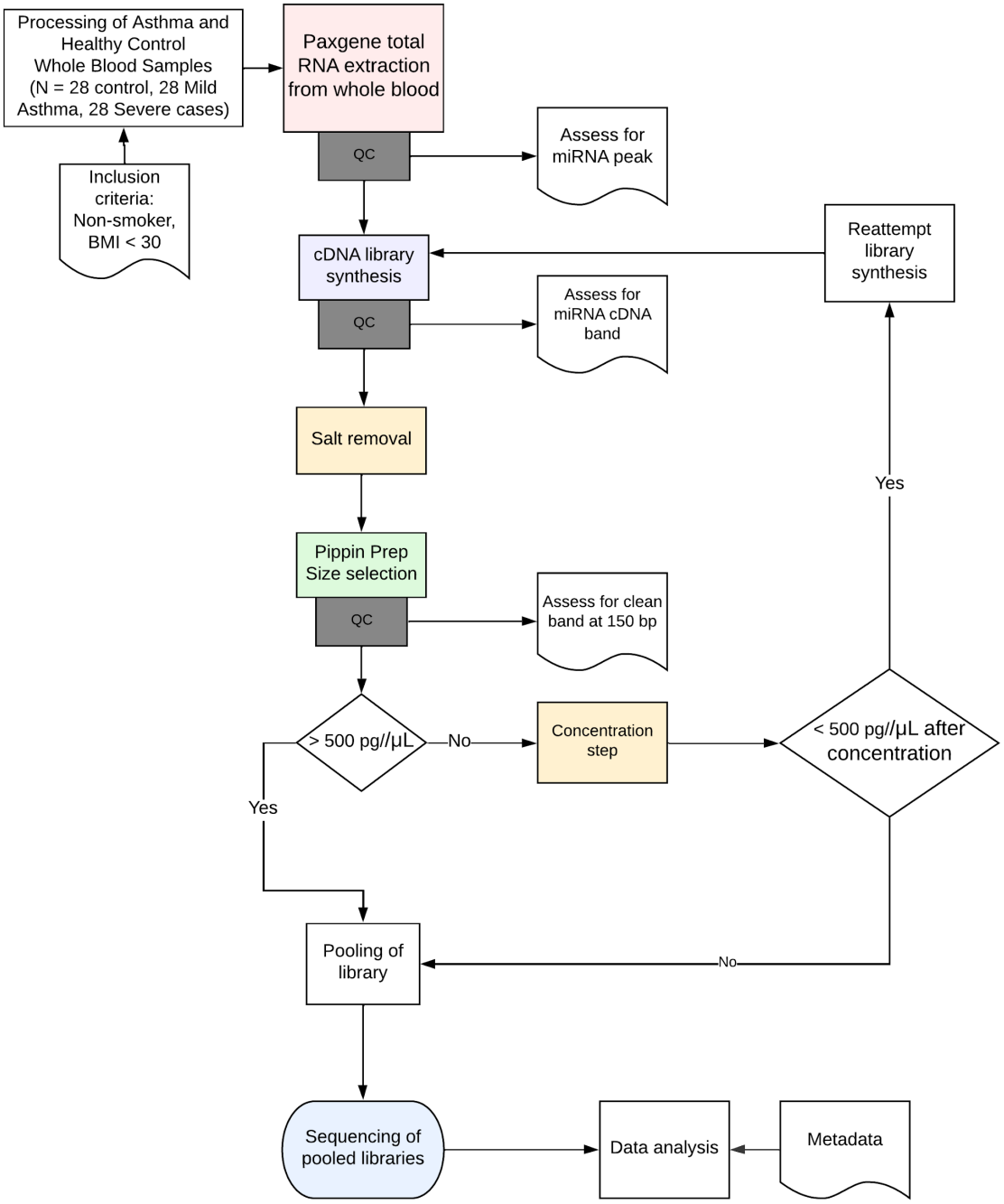


Figure 5.1. Modified Protocol after Pippin Prep Replacement*. Modified protocol of that shown in Chapter 4. Section 4.2. As a result of the Pippin Prep replacement as detailed in Section 5.2.2. below.*

For the pilot study, miRNAs were already extracted for the subjects. For this subsequent larger independent study (so no subjects from pilot study included as the plan was to first compare results) of severe asthmatics (N =17), non-severe asthmatics (N = 15), and healthy controls (N = 24), all blood samples required extraction. This was performed using the PAXgene blood miRNA extraction protocol and the PAXgene blood RNA extraction kit (Chapter 2, Section 2.2.2.). It was considered optimal use of data to compare the results of the pilot with the follow up study. By comparing the results with those from the larger follow-up study, and identify potential biases, or evaluate effect size changes, thereby strengthening the validity of the overall research. There may also be benefits of providing sequencing results from samples prepared in using the same reagents and equipment would minimise the contribution of sampling and processing to the sequencing results over a pilot study and a follow up study.

Library synthesis and Quality Control was conducted as previously described (Chapter 2, Section 2.3.). Quantified samples underwent conversion into indexed libraries using the NEBnext® Multiplex Small RNA Library Prep Set for Illumina® (Set 1) following the recommended Illumina protocol (Figure 2.5.).

Table 5.1. details the index sequences used for the study, and the sequencing lane distribution. There were insufficient indexes for all samples (for which libraries were successfully generated) to be placed on a single lane. As such samples were distributed evenly between two lanes taking into consideration patient severities and gender.

Table 5.1. Index Sequences used for Study 2*: Two pools of sample libraries were created (one pool of 29 samples the other 32 samples) to evenly distribute samples by asthma severity and gender across the two sequencing lanes. There was some drop out during library synthesis resulting in Lane 1 with 24 samples and Lane 2 carrying 32 samples.*

	Lane 1		Lane 2	
Sample number	**Index number**	**Base sequence**	**Index number**	**Base sequence**
1	1	ATCACG	1	ATCACG
2	2	CGATGT	2	CGATGT
3	3	TTAGGC	4	TGACCA
4	4	TGACCA	5	ACAGTG
5	5	ACAGTG	6	GCCAAT
6	6	GCCAAT	7	CAGATC
7	7	CAGATC	8	ACTTGA
8	9	GATCAG	9	GATCAG
9	10	TAGCTT	10	TAGCTT
10	12	CTTGTA	11	GGCTAC
11	13	AGTCAA	23	GAGTGG
12	14	AGTTCC	24	GGTAGC
13	15	ATGTCA	25	ACTGAT
14	16	CCGTCC	26	ATGAGC
15	17	GTAGAG	27	ATTCCT
16	19	GTGAAA	28	CAAAAG
17	21	GTTTCG	29	CAACTA
18	25	ACTGAT	30	CACCGG
19	26	ATGAGC	31	CACGAT
20	28	CAAAAG	32	CACTCA
21	29	CAACTA	33	CAGGCG
22	30	CACCGG	34	CATGGC
23	32	CACTCA	35	CATTTT
24	33	CAGGCG	36	CCAACA
25	35	CATTTT	37	CGGAAT
26	41	GACGAC	38	CTAGCT
27	42	TAATCG	39	CTATAC
28	43	TACAGC	40	CTCAGA
29	44	TATAAT	41	GACGAC
30			42	TAATCG
31			43	TACAGC
32			44	TATAAT

5.2.3. Sequencing and Analysis

Sequencing protocols were conducted as per the pilot study as was the analytical pipeline (Chapter 4, Section 4.2.3.) although for the latter the three subject groups were considered Healthy Controls, Mild (will be referred to as Non-Severe from this point), and Severe asthmatics. miRNAs were considered if they achieved an FDR <0.05 and a FC of > 0.5.

5.3. Results

5.3.1. Subject Demographics and Clinical Correlations

A total of 28 subjects per group were matched for sample processing (total of 84). There was a drop out resulting in a reduction of between 4 and 13 subjects in each group. Despite this decrease there was no significant impact on the matching conducted between the groups. Of the 56 successfully processed samples (for inclusion criteria see Chapter 2, Section 2.2.1.), subjects were either severe asthmatics (N = 17), non-severe asthmatics (N = 15), or healthy controls (N = 24). As highlighted previously (Chapter 4), clinical factors such as age, gender, and BMI have been known to impact asthma severity and to try and reduce the impact of these variables, the groups were matched as best as possible at the start of the study with the samples available. Despite the subsequent decrease in sample numbers, for which data was generated, no significant differences between the three subject groups were seen for age, gender or BMI (Table 5.2.). Other clinical features key to asthma diagnosis and disease monitoring IgE, blood Eosinophil count and FEV1/FVC ratio, serum IgE levels between the groups showed a difference (P = 0.02 – levels highest in Severe group) as did blood eosinophil counts (P = 0.01). There was no statistical difference between groups regarding FEV1/FVC ratio (P = 0.07).

Table 5.2. Patient Demographics of Patients with Samples that Completed Processing and Sequencing. *P values generated through Kruskal-Wallis (ordinal variables) or Fisher's exact test (categorical variables). There were significant differences between groups regarding total IgE, and blood eosinophil counts. All other features did not result in significant differences between the groups. SD = standard deviation. BMI (Body mass index), FEV1FVCratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

	Control (N = 24)	Non-Severe (N = 15)	Severe (N = 17)
Age: P = 0.22			
Min	18	20	28
Max	69	60	70
mean (SD)	42.04 ± 13.18	40.13 ± 13.69	48.47 ± 15.55
Gender: P = 0.78			
Female	9 (38)	5 (33)	8 (47)
Male	15 (62)	10 (67)	9 (53)
Total IgE kU/L: P = 0.02			
Min	1	3.6	8.9
Max	1117	1452	3597
mean (SD)	20; 124.89 ± 248.40	342.05 ± 367.55	467.39 ± 892.62
Data unavailable	4	0	0
Blood Eosinophils: P = 0.01			
Min	0.05	0	0
Max	0.4	0.9	0.6
mean (SD)	0.13 ± 0.09	0.30 ± 0.29	0.32 ± 0.19
Data unavailable	8	6	5
FEV1/FVC ratio: P = 0.07			
Min	69	42	42
Max	87	94	90
mean (SD)	78.50 ± 4.66	71.41 ± 14.38	69.93 ± 13.18
BMI class: P = 0.09			
Normal	9 (38)	5 (33)	3 (18)
Overweight	12 (50)	10 (67)	8 (47)

As before (Chapter 4, Section 4.3.1.2.), an analysis was conducted to assess potential correlations between clinical variables of interest (outcome ranges of correlations applied are detailed in Chapter 2, Section 2.1.5.2.). Patient disease severity showed a weak positive correlation with total IgE (0.26) and BMI (0.35), and a weak-to-moderate positive correlation to eosinophil counts (0.46). Severity also had a negligible negative correlation with FEV1/FVC ratio and patient gender. BMI showed weak positive correlations with patient gender (0.34) and disease severity (0.35). Patient lung function measure (FEV1/FVC ratio) showed weak negative correlation with patient gender (-0.36). Age appeared to show the weakest correlations with other clinical measures with correlations of less than 0.2 (positive and negative).

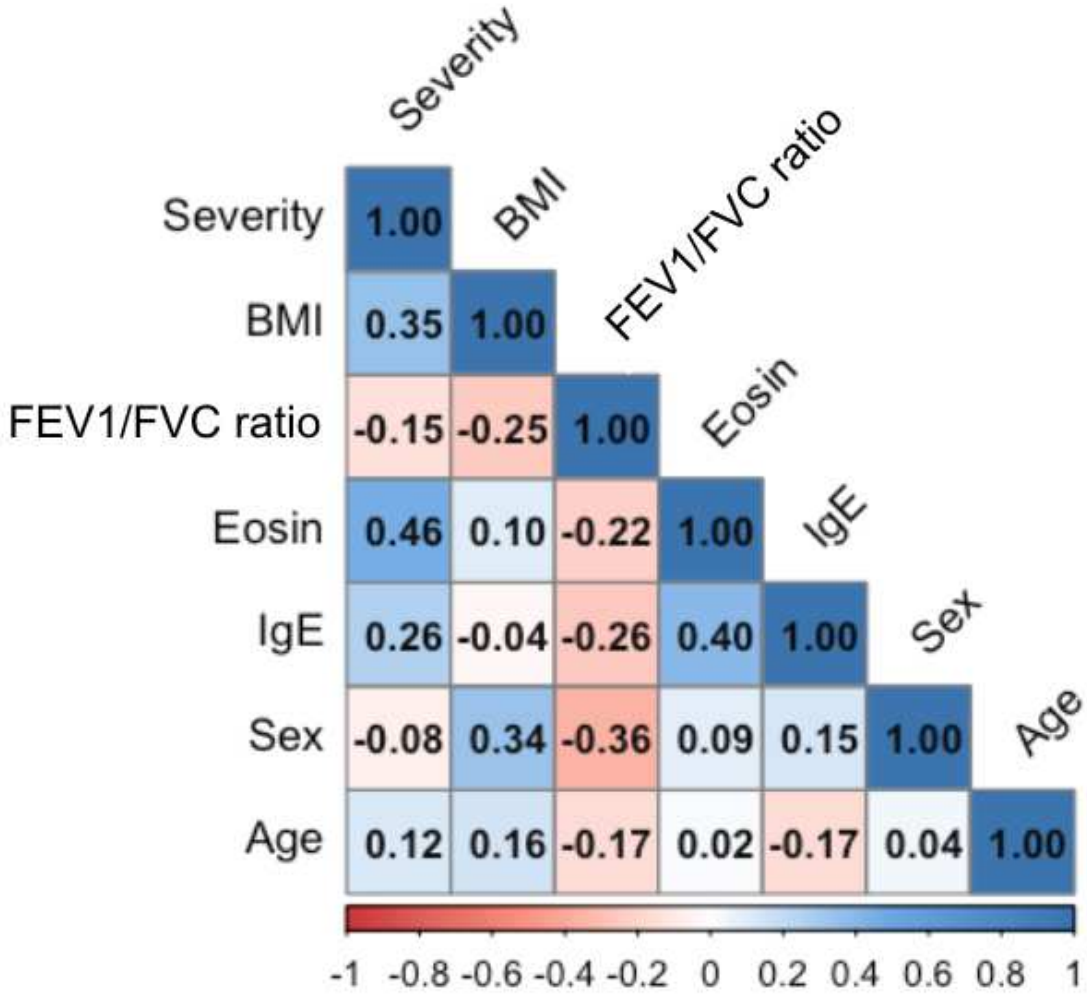


Figure 5.2. Correlation Analysis of Clinical Outcomes of Patient cohort. *N = 56 (Ns = 24 [Controls], 15 [Non-Severe asthmatic], 17 [Severe asthmatic]). Clinical measures were converted to factors where necessary, and a correlation analysis was conducted. Severity (Control, Non-Severe, Severe), BMI (Body mass index), FEV1FVCratio (FEV1/ FVC ratio, measure of lung function), Eosin (Total blood eosinophils), IgE (serum IgE), Sex (patient gender), Age (patient age in years). As per scale, red indicates negative correlation and blue positive correlation. Moderate correlations were identified between disease Severity and BMI, and a moderate positive correlation between Severity and Eosinophil counts. Eosin and Ige also shared a moderate positive correlation.*

A Kruskal-Wallis test was conducted to identify if patient age was significantly different between groups, and was found not to be significant ($P = 0.33$) (Figure not shown). Eosinophil counts and IgE are an important component of asthmatic disease and diagnostics (Chapter 1). Eosinophil counts with serum IgE were therefore compared between the groups and as already shown in

Table 5.2., there was a statistically significant difference between the groups and serum IgE in the Control vs Severe asthma (P = 0.01).

A comparison of the selected patients' BMI categories (Normal = 18- <25, Overweight = 25- <30, Obese >30) and severity was made as BMI has an established connection with asthma severity (Chapter 1, Section 1.1.3.). Although not significant (Kruskal-Wallis test), there was nonetheless a trend suggesting that those with more severe the asthma tend to be distributed within the obese and overweight categories.

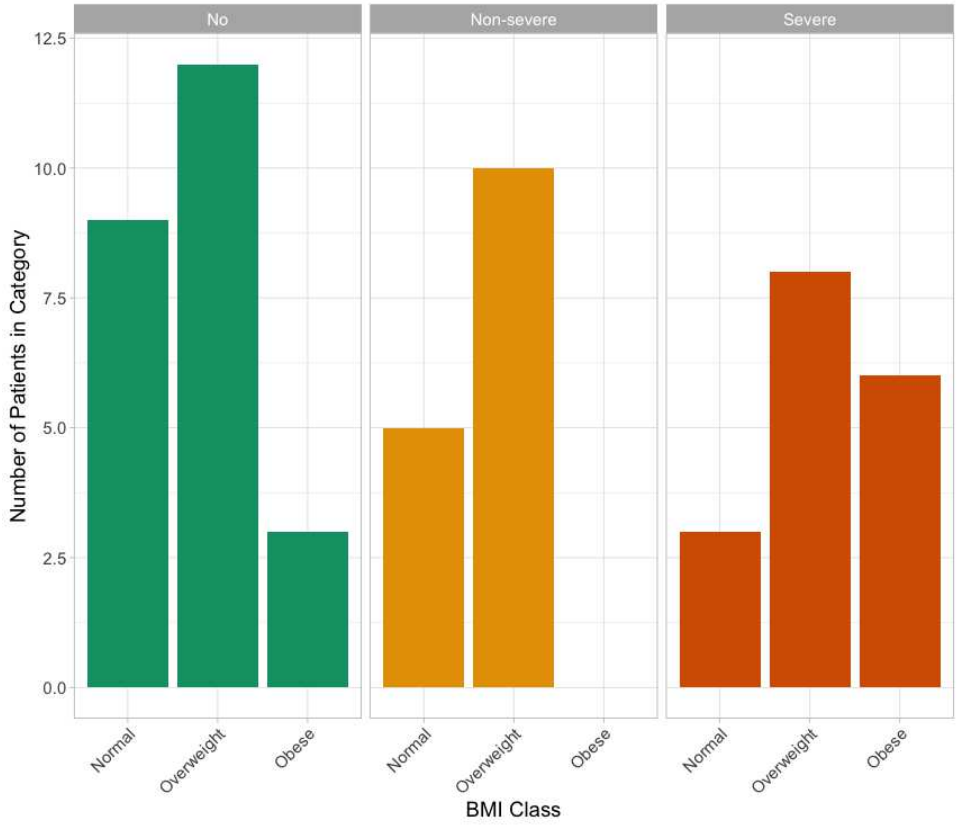


Figure 5.3. Patient BMI Class versus Asthma Severity. *There was no significant difference between the BMI classes and severity groups Kruskal-Wallis test (P = 0.17) although, as can be seen, there is a trend towards higher BMI and asthma Severity. BMI (Normal = 18- <25, Overweight = 25- <30, Obese >30).*

5.3.2. Sample Processing

Eighty-four blood samples (28 Severe, 28 Non-Severe, and 28 Healthy Controls) had RNA extracted from them and were passed through library synthesis and clean-up protocols. Fifty-six libraries met the minimum input concentration for sequencing (500 ng/μL) and underwent sequencing (for further information on sequencing requirements see Chapter 2, Section 2.1.4.).

As highlighted earlier (Section 5.2.2.) due to the number of index barcodes available, two sequencing runs were performed. The first was conducted between lockdowns using the resources already in the laboratory and resources made available at the Genomics Facility of the NIHR Imperial Biomedical Research Centre: https://imperialbrc.nihr.ac.uk/facilities/genomics-facility/) as there was a period where manufacturers were unable to meet delivery deadlines. As such, the second set of samples had total RNA isolated and libraries prepared but size selection of the libraries could not begin due to Pippin Prep kits (that had been backlogged due to COVID) being unavailable.

When the new kits arrived, the Pippin Prep was re-tested. During a run, it was found that across several if not all lanes an appropriate current could not be established. Through a process of elimination, it was ascertained that there was no issue with the reagent mixing nor with the use-by date of the Pippin cassettes; no faults in the cassette preparation; and that reagent temperatures were all optimal. Having carried out this array of tests, the manufacturer finally acknowledged that the fault likely rested with the machine itself and they replaced the faulty machine free of charge.

When the new machine arrived, a cassette (5 samples) was run following the previously established protocol (Chapter 2, Section 2.2.6.) but the subsequent QC, using the Agilent Bioanalyzer (trace shown in Figure 5.5. A.), revealed that size selection had been ineffective at removing unwanted species of cDNA. On further inspection of the Pippin Prep run, lane current artefacts suggestive of salt contamination were identified. The current artefacts are seen as "humps" in the Bioanalyzer trace (Figure 5.5. A.).

A modification to the original protocol was therefore made wherein samples were passed through a clean-up process before size selection on the Pippin Prep. Clean-up and concentration had been conducted in prior experiments with the clean-up step showing no reduction in sample yield (Chapter 3, Section 3.3.3.1.). Figure 5.5. B shows an example Bioanalyzer trace for a library that underwent this clean-up step with a size-selected library of the correct length, and a Pippin trace with no salt "hump". Samples were therefore from this point forward processed in this manner before pooling. The first set of samples sequenced (prior to the Pippin Prep machine being replaced) will be referred to as Run 1 and the second set (with the adapted method detailed above) will be referred to as Run 2 in the following results.

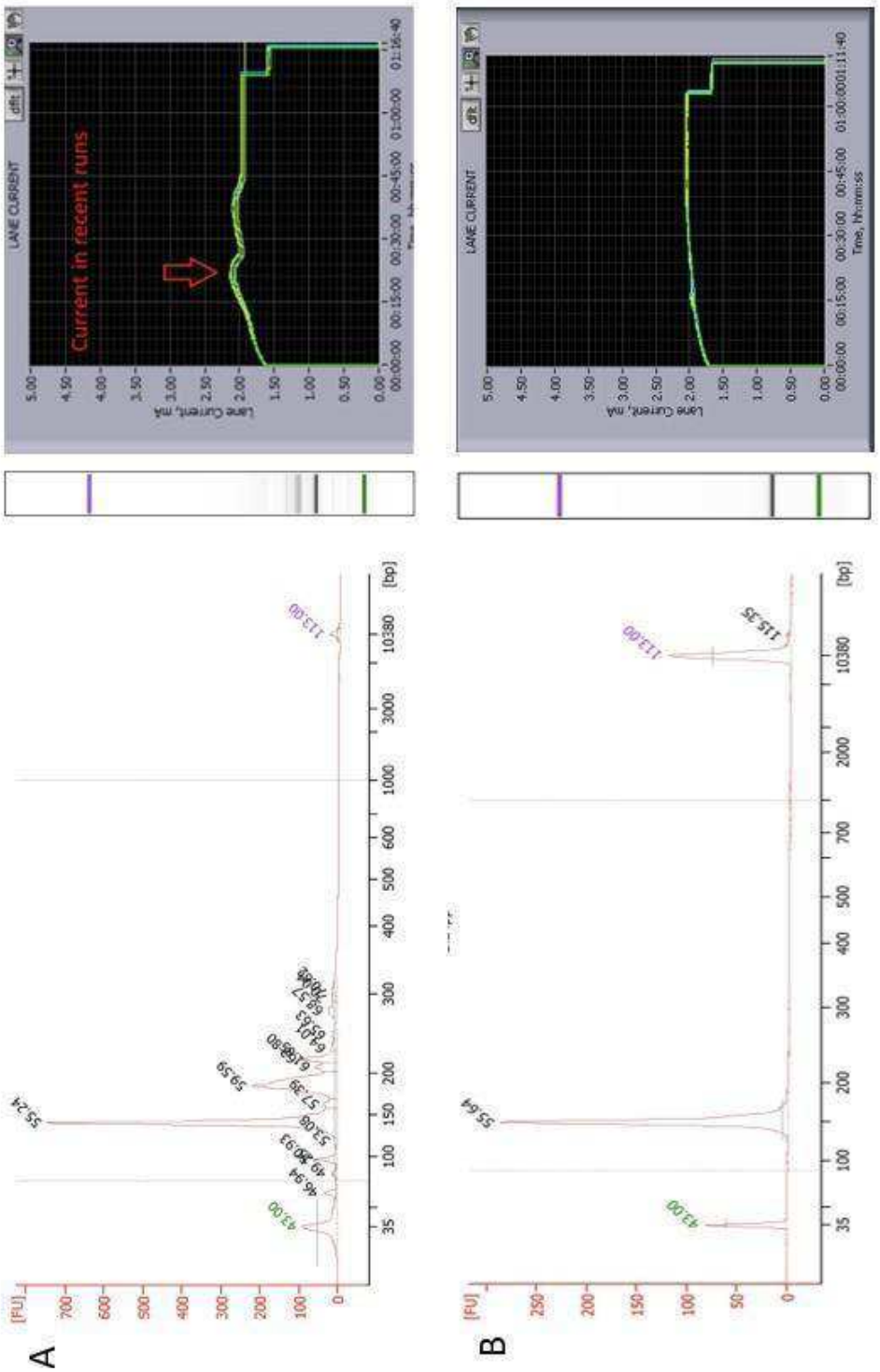


Figure 5.4. Bioanalyzer Trace Result and Corresponding Sample Trace from Pippin Prep Size Selection. ***A***. *Before protocol modification showing cDNA species contaminants in Bioanalyzer trace and salt hump in Pippin trace.* ***B.*** *Bioanalyzer trace showing clean size selected library ready for pooling, and clean Pippin trace.*

5.3.3. Sequencing Results

The data from the two sequencing runs were compared to identify any differences between the raw counts and miRNA variance. This was because the Novogene Sequencing Facility had advised that, although the first run (Run 1) had performed well, for the second run (Run 2) technicians at Novogene had observed orphan samples on the lane and read counts lower than observed in previous runs. More details on the methods of how samples are pooled and how the sequencing facility identifies samples is provided in Chapter 2 (Section 2.1.5.). Novogene provided information about the orphan reads, namely the index sequence by which to identify them. Crucially, the index sequences were not available or present in any of the present study's index kits; therefore, the orphan samples were not a result of this research. There was reassurance from Novogene that the orphan index sequences were not incompatible with the run. There were 71 other libraries (submitted by other customers) combined with the library from this study that underwent sequencing on Run 2. The source of the orphan samples could have come from any one (if not more) of the 71 other customer libraries. No further information whether this was the case was provided by Novogene, and disappointingly as some sequencing data had been generated they were not contractually bound to do a re-run despite clearly there being an issue with the Run 2.

As before, raw reads from Novogene arrived on a hard drive in FASTQ file format. The raw reads were assessed via FASTQC. FASTQC results returned flags related to adapter sequence contributing to some homogeneity among approximately the first half of the read length. This was expected as the sequence length was greater than the unique miRNA sequence and included the adapter and barcode sequences (discussed in Chapter 2, Section 2.1.5.1.). Quality of reads was

high as the per-base sequence quality (the confidence that the base pair read was accurate at that position) was of high confidence beyond the read length that would cover the adapter and the miRNA. As for the pilot study (Chapter 4, Figure 4.6.) sample quality was further analysed through a range of different alignment factors including, for example, miRNA precursors, as well as miRNA antisense (as a negative control) using the Genboree workbench. The requested number of reads for each sample was 5M. Total reads ranged from 456,000 reads to > 15M reads, with the Severe group average reads sitting below the 5M mark, and the Control and Non-Severe groups read average sitting above the 5M mark. Figure 5.6. shows a summary of the total reads grouped by patient group.

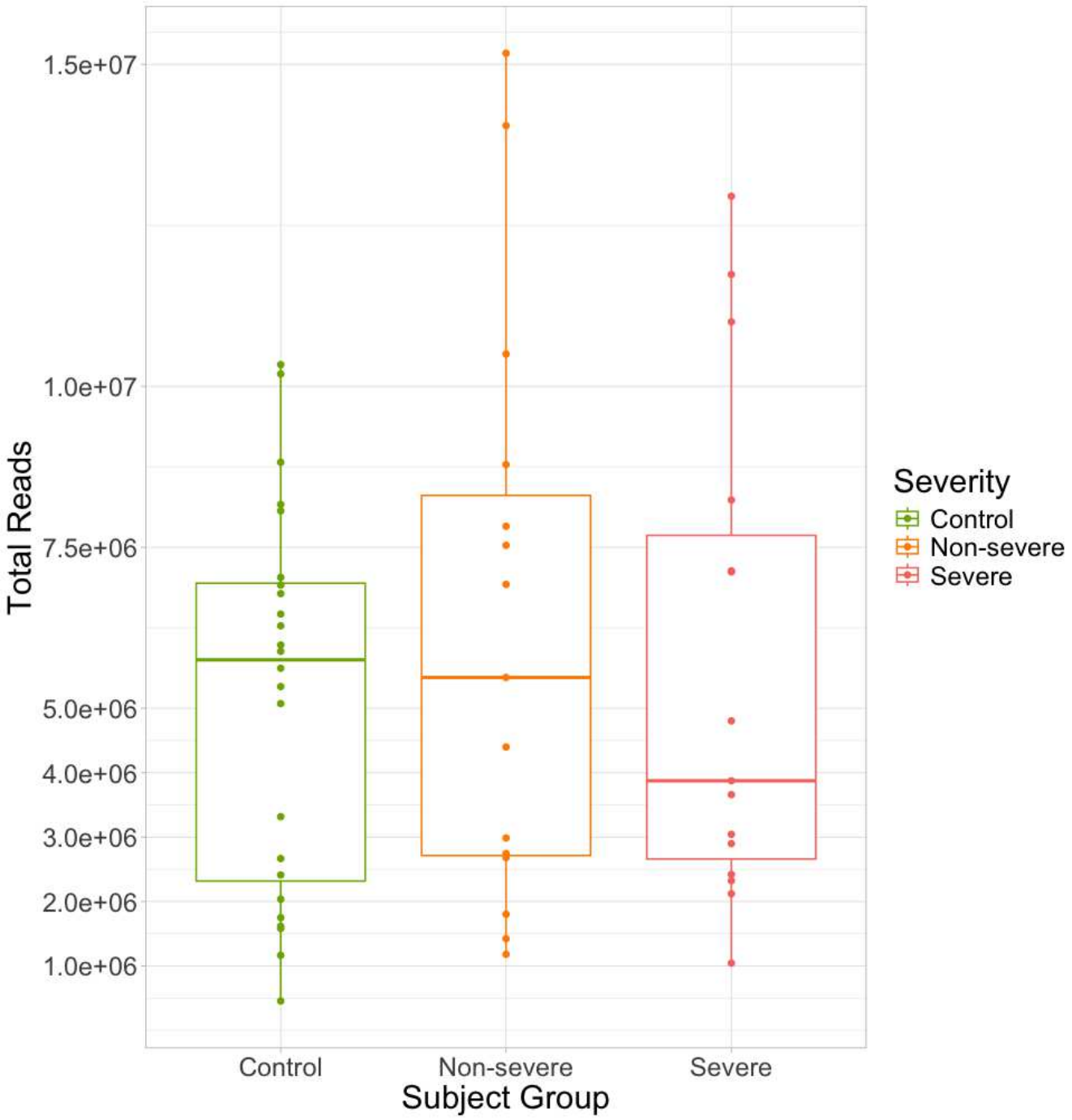


Figure 5.5. Summary Boxplot of Raw Reads input into Alignment for Each Patient Category. *Boxplot denotes median with the box denoting the interquartile range (IQR), and the whiskers 1.5 times the IQR. There were no identified outliers in this case. Of the 5M reads requested per sample for each library the Control (N = 24) and Non-Severe (N = 15) groups average read count exceeded this while the Severe (N = 17) group had an average of just below 4M reads.*

Alignment was then conducted. The majority of sample reads were successfully clipped of adapter (between 91-99%), and very little failed the quality filters (<1-1%). There were negligible-or-no UniVec contaminants (<1%) or rRNA identified in the run (1-5%). Reads were aligned to the human genome and miRNA with high success (82-97% of reads aligned to the genome, and 69-91% of reads aligned to miRNA sense for both lanes). There was also <1-1% alignment to tRNA.

As before (Chapter 4, Section 4.3.3.) factors that would align the reads to exogenous RNA species were also considered. There was a negligible alignment to exogenous RNA species (<2%). Figure 5.7. highlights the alignment percentages faceted by the test lane conducted (Run 1 shown on the left of the Figure and Run 2 on the right). While there was a difference in the total number of reads acquired during sequencing, there was no significant difference between the lane quality (Run 1 and Run 2) as far as alignment and filtering were concerned.

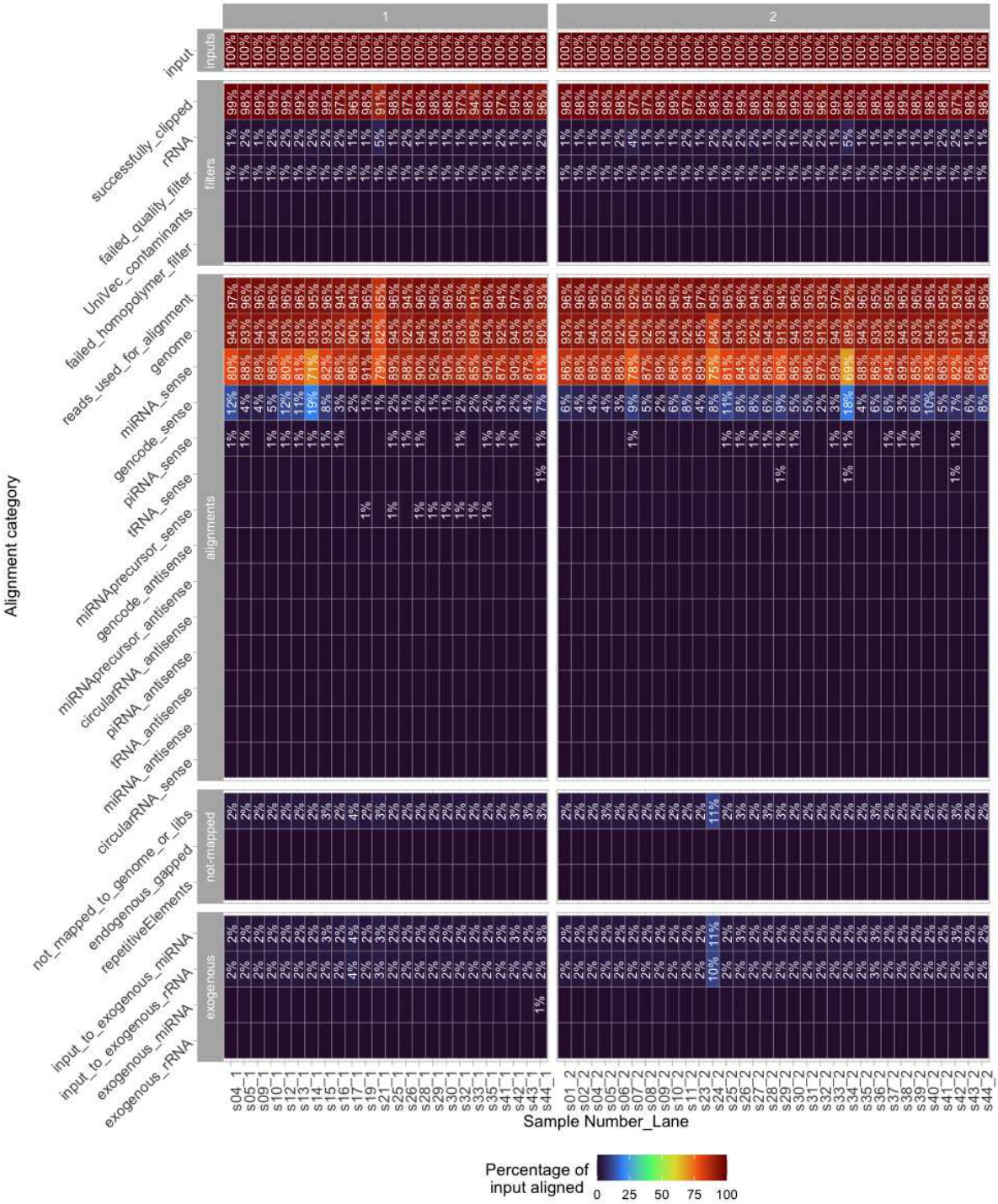


Figure 5.6. Raw Read Alignment. *N = 56. Run 1 shown on left and Run 2 shown on right. Raw sequencing reads pass through multiple quality control and alignment steps, the count of reads that pass through each step are accounted for in the figure for each sample. Percentages shown are between 1-100%, below 1% are left blank. A high percentage of reads (>=85%) were successful through filters and aligned to the genome (>= 82%) and to sense miRNA (>= 69%). Very low/no reads aligned to contaminants or failed the quality filters. There was no indication of contaminants introduced during the library preparation based on low percent of reads aligning to exogenous genomes or other exogenous RNA species (<1%).*

To minimise batch effects, prior to sending the samples for sequencing, it should be noted that it had been ensured that numbers of controls, non-severe asthmatics and asthmatic samples were evenly distributed between the runs. Figure 5.8. shows that comparing the two runs there were significantly fewer total miRNA counts in the second sequencing run (P = 5.57e-13) but this decrease was consistent across all three subject groups and not driven by one particular group.

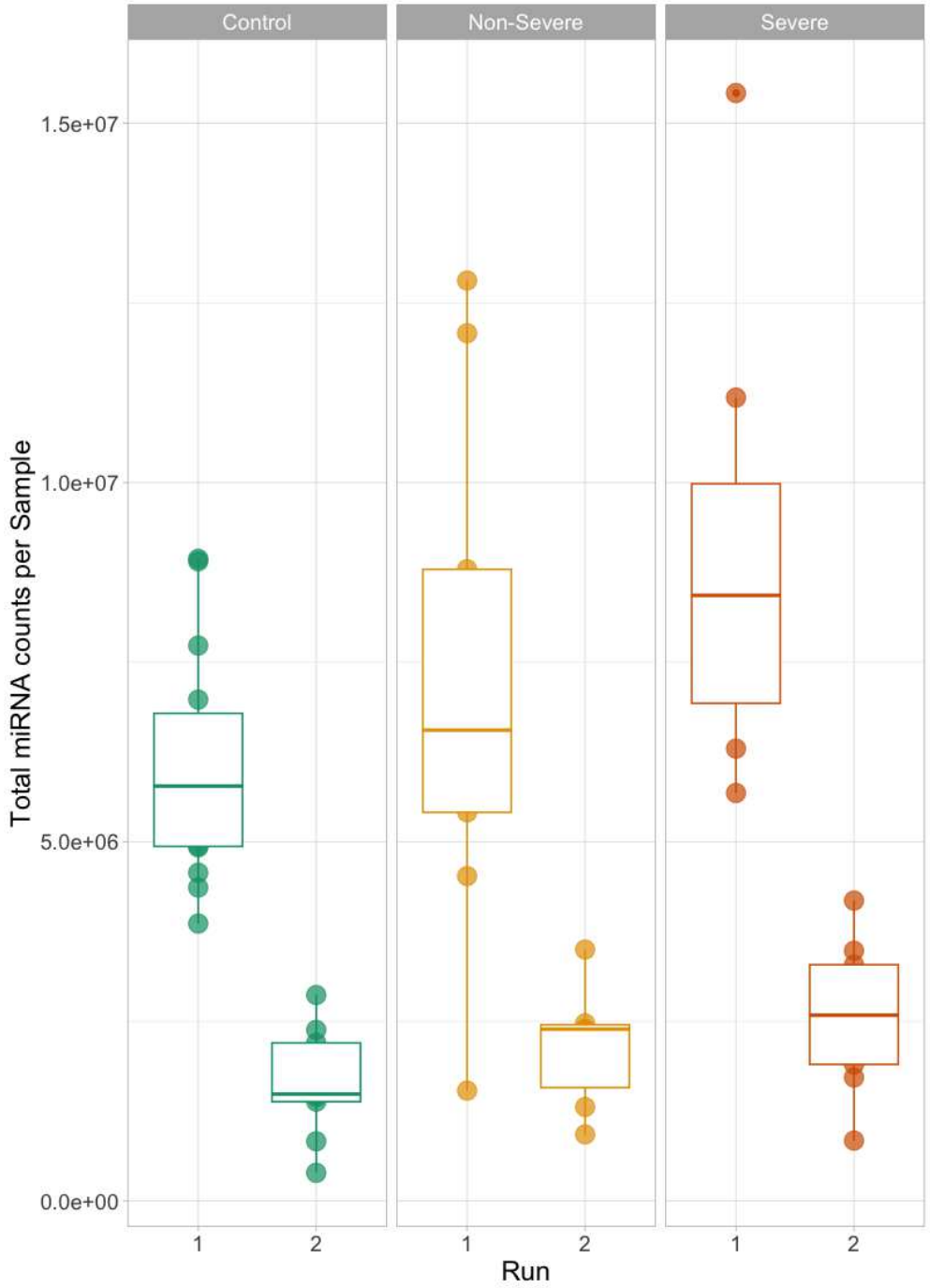


Figure 5.7. miRNA Total Reads by RNA Sequencing Run by Patient Severity. *A Wilcoxon rank sum test found the differences in read counts significantly different between sequencing runs (P < 0.0001).*

With fewer total reads captured, there was a risk that this might impact the number of unique miRNAs identified. This was examined (Figure 5.9.) and it was observed that across all three groups, Run 2 had resulted in fewer being identified compared to Run 1 (*P* = 1.4e-07).

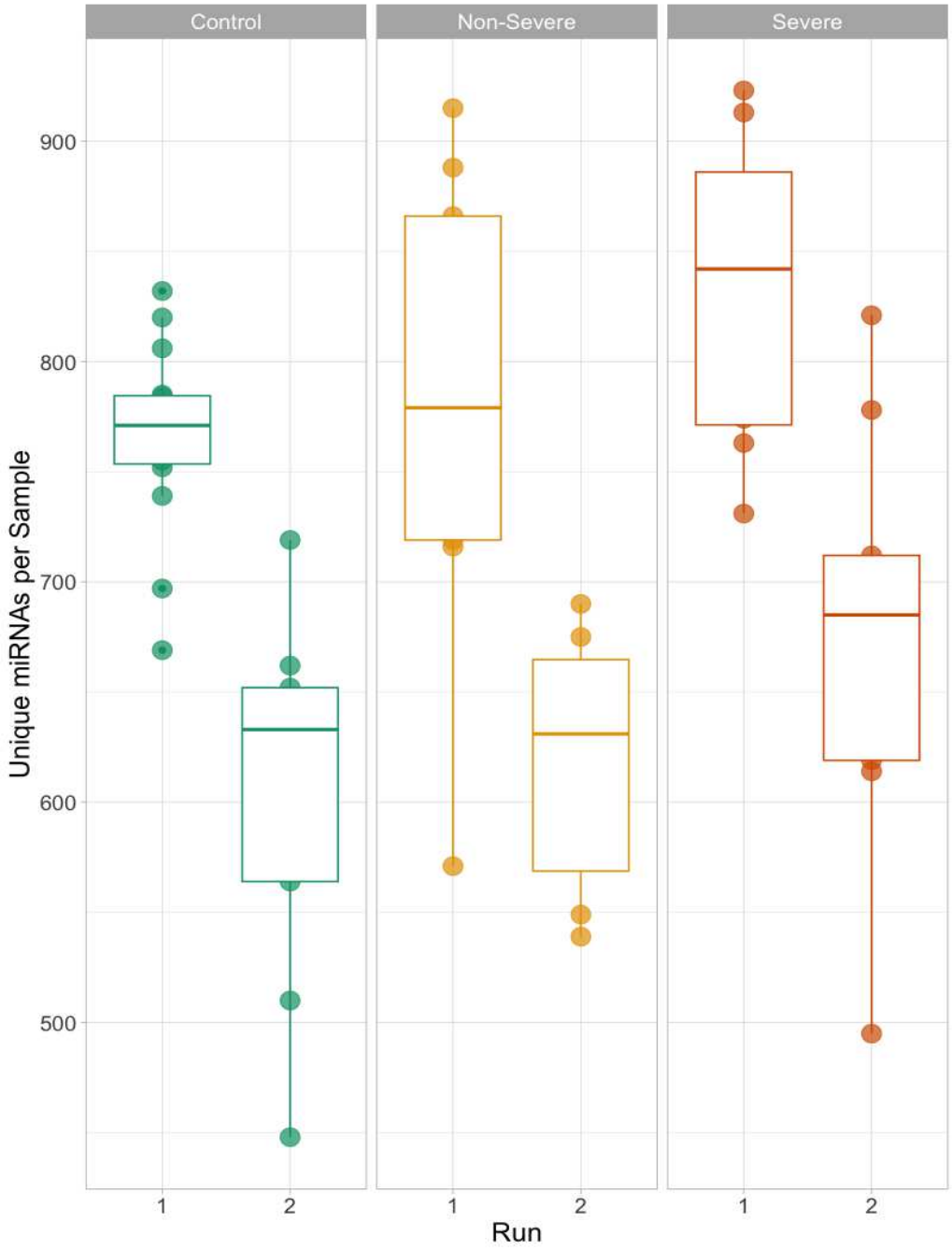


Figure 5.8. Aligned miRNA diversity grouped by sequencing run and patient severity*. A Wilcox test found the differences in read counts by run significantly different (P < 0.0001) between patient severity groups.*

The two graphs (Figure 5.8. and Figure 5.9.) show that there was an impact on the total number of reads based on the sequencing run as well as the total number of unique miRNAs identified. Although there was no identifiable reason (in relation to library preparation), there was a noticeable paucity in the number of reads generated in the second sequencing Run compared to the first. All three groups were similarly affected. As potential variation in Lane and batch effects was considered in the design, results from Run 1 and Run 2 were combined as planned for analysis.

As whole blood can contain a variety of sources of RNAs, beyond the vessels and cells, there was the possibility that the miRNA diversity might arise from expected differences in white cell populations between the clinical groups that could contribute to the overall abundance of miRNA. A Kruskal-Wallis test was conducted to identify if this might be the case and was found that white cell count was not to be a significant contributor to the miRNA abundance ($P = 0.74$) (Figure not shown).

5.3.3.1. Differential Expression Analysis

As dealing with three groups of subjects, miRNAs were analysed for differential expression (DE) using TCC (an acronym for Tag Count Comparison) a package that enables analysis of two or more groups for DE (for more details on the normalisation and the specific packages please see Chapter 2, Section 2.1.5.2.). For multiple comparison results where more than two groups are compared no Fold change (FC) information is generated; as such, only FDR values were used for selection of DE-miRNAs from the analysis result. Within the multiple comparison analysis (Controls vs Non-Severe vs Severe), 52 miRNAs were identified as having an FDR of < 0.05 and were considered significant. miR-92a-3p showed the highest expression in raw counts followed by several of the let-7 family of miRNAs. The majority of the DE-miRNAs appeared to be expressed in all tested groups, apart from miR-5968 (not expressed in Healthy Controls), miR-337-5p, miR-6753-5p and miR-676-3p (not expressed in Non-Severe asthmatics) and miR-5684 (not expressed in Severe asthmatics). Figure 5.9. highlights all DE-miRNAs from most significant to least (right to left), and the average counts for that miRNA within each respective group.

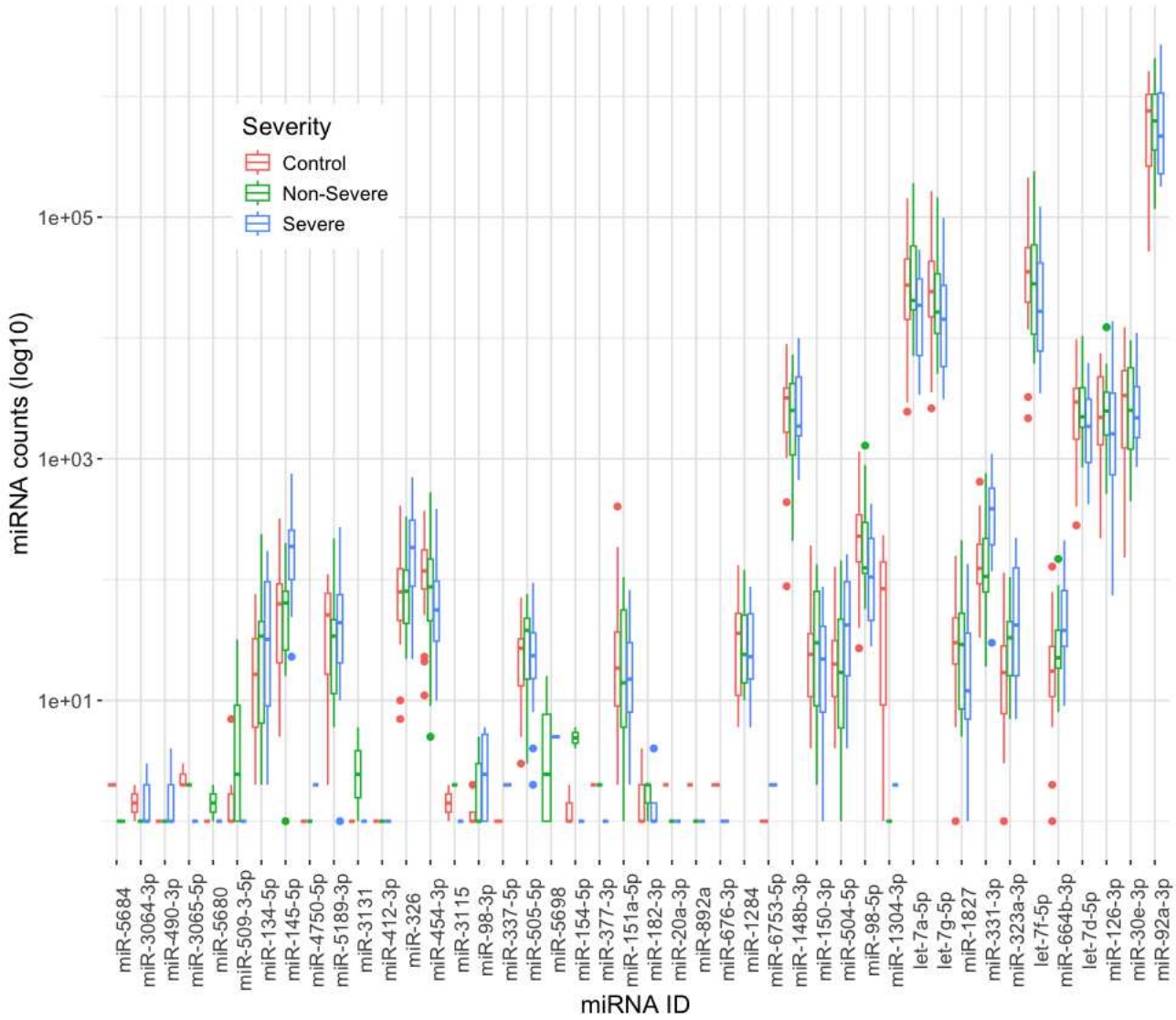


Figure 5.9. DE-miRNAs Identified from Multiple Comparison Analysis Among Control, Non-Severe, and Severe asthma. *DE-miRNAs identified from BaySeq multiple comparison analysis using the package TCC (Tag Count Comparison) in R. DE-miRNAs from most significant to least (right to left), and the normalised average counts (log10) for that miRNA within each respective group.*

Next comparisons were made between groups separately (Controls vs Severe, Controls vs Non-Severe, Non-Severe vs Severe) and this was again conducted in the TCC package using two-group analysis parameters with volcano plots produced using the resulting FDR values and fold change (FC) data. Labels were selected for the top 20 (when present) of the DE-miRNAs meeting both FDR value (< 0.05) and FC requirements (change of > 0.5). The three volcano plots of the results from the three comparisons are shown in Figures 5.10., 5.11., and 5.12.

In the Control vs Severe comparison 35 DE-miRNAs were identified: miR-145-5p, miR-223-3p, miR-1304-3p, miR-98-3p, miR-26b-5p, let-7c-5p, miR-98-5p, miR-6753-5p, miR-4750-5p, miR-452-5p, miR-326, let-7a-5p, miR-5698, miR-664b-3p miR-548t-5p miR-4638-5p miR-6827-3p miR-3064-3p, miR-331-3p, let-7d-5p, miR-4655-5p, miR-4775, miR-532-3p, miR-369-3p, miR-182-3p, miR-144-5p, miR-1278, miR-425-5p, miR-374a-5p, miR-339-5p, miR-15a-5p, miR-32-5p, miR-454-3p, miR-548ap-5p|miR-548j-5p, miR-296-3p.

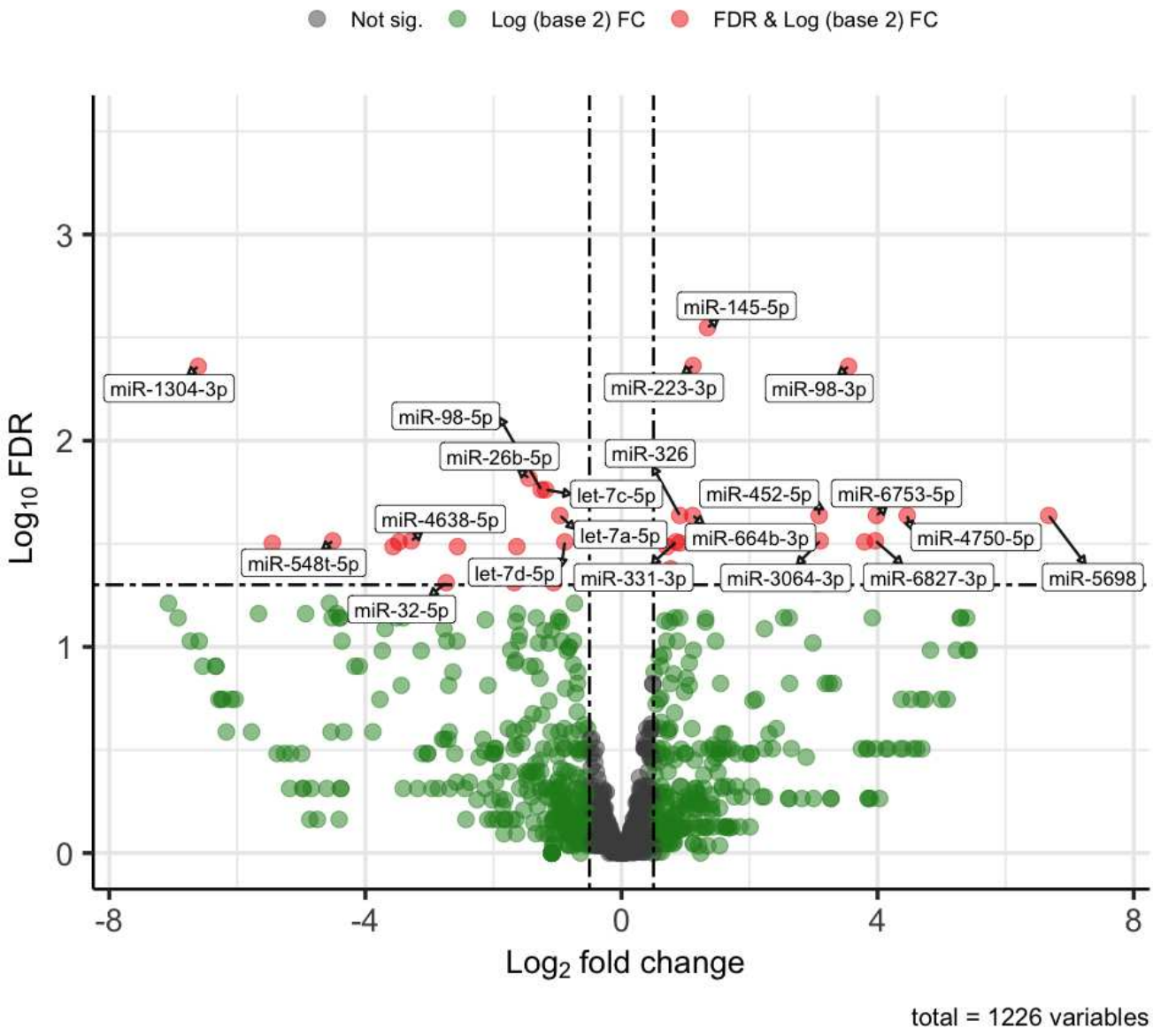


Figure 5.10. Volcano Plot of Controls vs Severe Asthma DE-miRNA Analysis. *FDR (Q value) < 0.05, Fold Change (FC) > 0.5. miRNAs are labelled red (meeting both FDR and FC requirements) or green (miRNAs meeting FC requirements). Top 20 miRNAs from BaySeq DE analysis are labelled. Positive FC is associated with Control expression being higher than Severe. Thirty five miRNAs identified meeting both FDR and FC requirements for significance. The label for miR-32-5p is obscured by let-7a-5p.*

A comparison was conducted between the Pilot study (Chapter 4, Control vs Severe) DE-miRNAs (N = 10) and these thirty five Control vs Severe miRNAs. Two miRNAs were in common between the two studies: miR-32-5p, and miR-1304-3p with both expressed more highly in healthy controls compared to severe asthmatics.

Next considering the Non-Severe versus Severe groups within this Chapter, 5 DE-miRNAs were identified (Figure 5.11.). One miRNA, miR-145-5p, showing higher expression in non-severe asthma compared to severe, while the remaining 4 DE-miRNAs showed higher expression in severe asthma (let-7a-5p, let-7c-5p, let-7d-5p, and miR-486-3p).

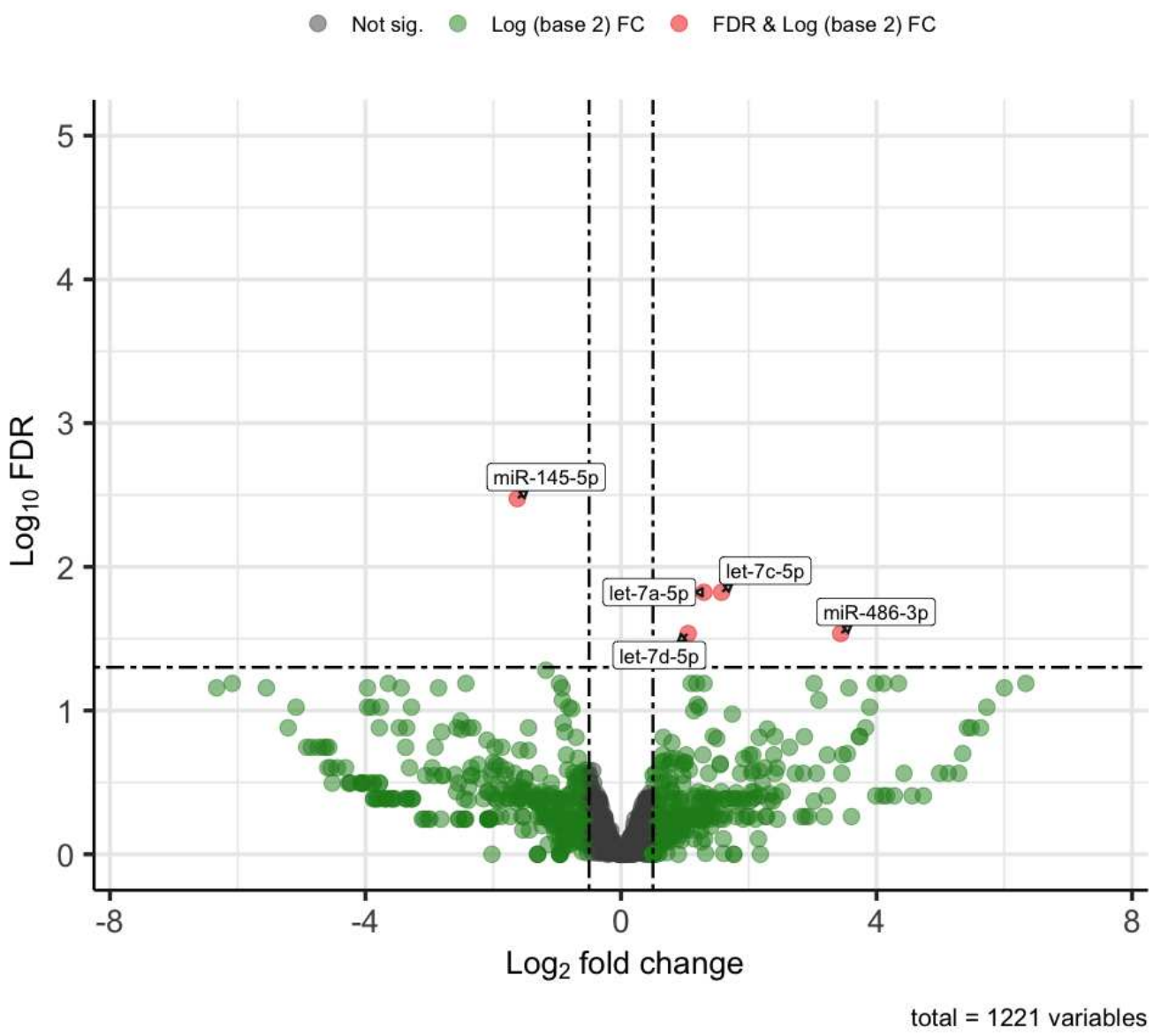


Figure 5.11. Volcano Plot of Non-Severe vs Severe Asthma DE-miRNA Analysis*. FDR (Q value) < 0.05, Fold Change (FC) > 0.5. miRNAs are labelled red (meeting both FDR and FC requirements) or green (miRNAs meeting FC requirements). Positive FC is associated with Non-Severe expression being higher than Severe. Five miRNAs met the FDR and FC requirements for significance.*

The final comparison Controls versus Non-Severe asthma groups identified 2 DE-miRNAs. One miRNA, miR-1304-3p, showed higher expression in controls compared to non-severe asthma, while miR-5698 showed higher expression in non-severe asthma.

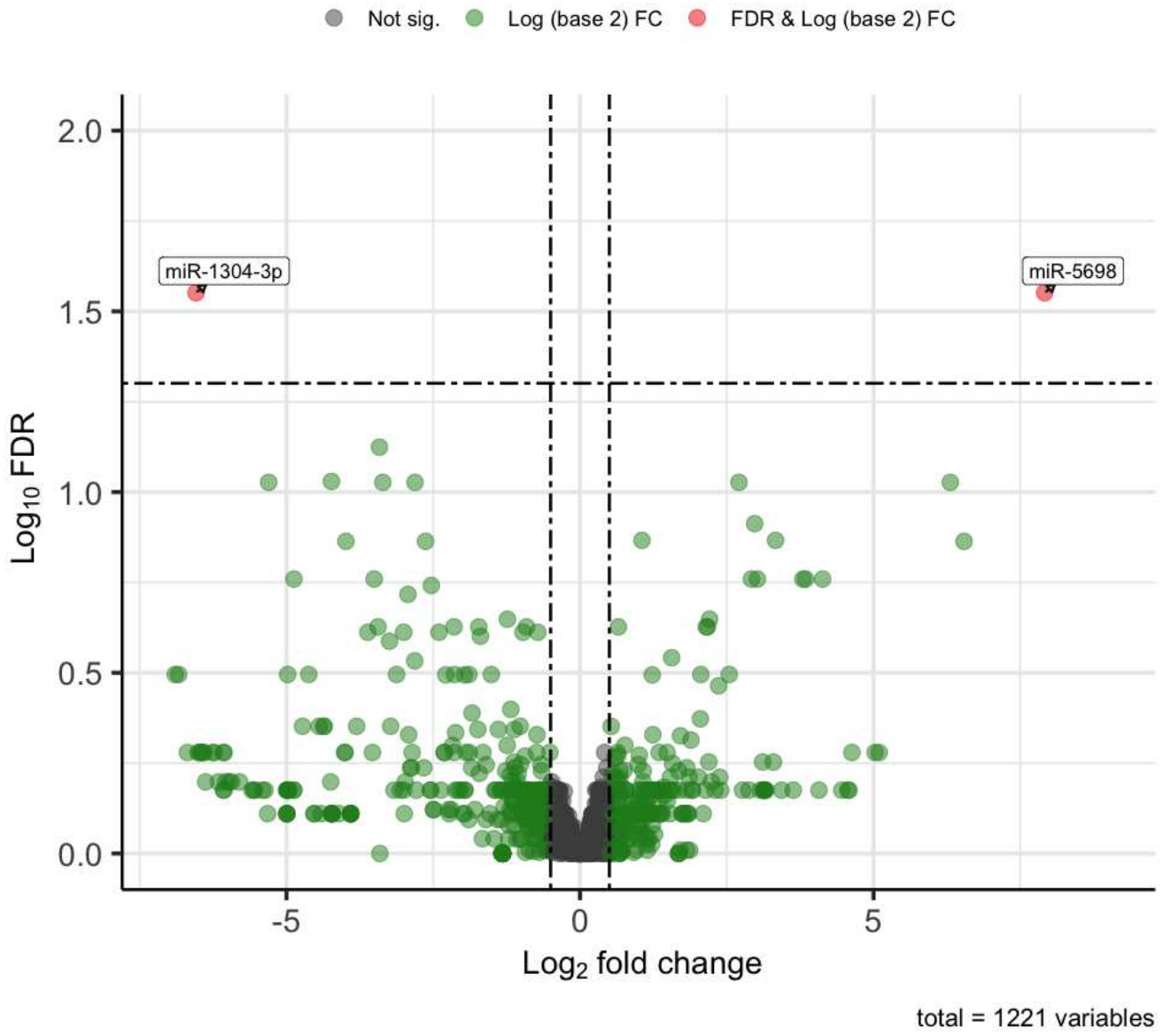


Figure 5.12. Volcano Plot of Control vs Non-Severe Asthma DE-miRNA Analysis*. FDR (Q value) < 0.05, Fold Change (FC) > 0.5. miRNAs are labelled red (meeting both FDR and FC requirements) or green (miRNAs meeting FC requirements). Positive FC is associated with Control expression being higher than Severe. Two miRNAs met the FDR and FC requirements for significance.*

A Venn diagram was made to consider the number of unique miRNAs identified between analysis pairings tested and those shared among the pairings (Figure 5.13.). No DE-miRNAs were shared between all three groups or Control vs Non-Severe and Non-Severe vs Severe and similarly among all three tested pairings there were no shared DE-miRNAs. Between Control vs Severe and Non-Severe vs Severe there were 4 DE-miRNAs in common (miR-145-5p, let-7c-5p, let-7a-5p, let-7d-5p). Both DE-miRNAs identified between Control vs Non-Severe were shared in the Control vs Severe analysis (miR-1304-3p, miR-5698). There were no shared DE-miRNAs between Non-Severe vs Severe and Control vs Non-Severe (Table 5.3.).

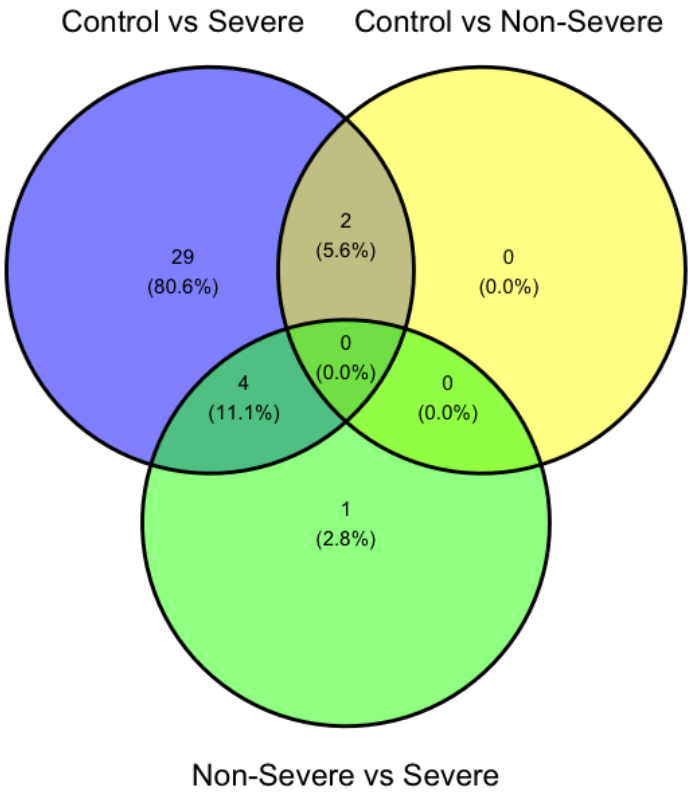


Figure 5.13. Venn Diagram of DE-miRNAs Identified in DE Comparisons Between Groups. *DE-miRNAs identified met FDR < 0.05 and fold change of greater than 0.5. No DE-miRNAs were shared between all tests.No DE-miRNAs were shared between Control vs Non-Severe and Non-Severe vs Severe. Control vs Severe identified 35 DE-miRNAs, Non-Severe vs Severe identified 5 DE-miRNAs, and Control vs Non-Severe identified 2 DE-miRNAs. Between Non-Severe vs Severe and Control vs Severe, 4 miRNAs were shared. Between Control vs Non-Severe and Control vs Severe two miRNAs were shared. DE-miRNAs identified through BaySeq analysis between test groups using package TCC (Tag Count Comparison) in R.*

Table 5.3. Table of DE-miRNAs found Among Analysis Pairs*. DE-miRNAs identified met FDR < 0.05 and fold change of greater than 0.5. DE-miRNAs identified through BaySeq analysis between test groups using package TCC (Tag Count Comparison) in R.*

Analysis	DE-miRNAs
Control vs Severe only	miR-145-5p, miR-223-3p, miR-1304-3p, miR-98-3p, miR-26b-5p, let-7c-5p, miR-98-5p, miR-6753-5p, miR-4750-5p, miR-452-5p, miR-326, let-7a-5p, miR-5698, miR-664b-3p miR-548t-5p miR-4638-5p miR-6827-3p miR-3064-3p, miR-331-3p, let-7d-5p, miR-4655-5p, miR-4775, miR-532-3p, miR-369-3p, miR-182-3p, miR-144-5p, miR-1278, miR-425-5p, miR-374a-5p, miR-339-5p miR-15a-5p miR-32-5p miR-454-3p, miR-548ap-5p|miR-548j-5p, miR-296-3p
Control vs Severe and Non-Severe vs Severe	miR-145-5p, let-7c-5p, let-7a-5p, let-7d-5p
Non-Severe vs Severe only	miR-486-3p
Non-Severe vs Severe and Non-Severe vs Control	
Control vs Non-Severe only	
Control vs Non-Severe and Control vs Severe	miR-1304-3p, miR-5698

5.3.3.2. DE-miRNAs and Clinical Outcome Correlation

A correlation analysis was carried out to compare all DE-miRNAs with the clinical outcomes of each of the comparison groups (Control vs Severe [DE miRNAs N = 35], Non-Severe vs Severe [DE miRNAs N = 5], and Control vs Non-Severe [DE miRNAs N = 2]) to identify any possible correlations of miRNA expression and physiological disease presentation (FEV1/FVC ratio, IgE, Eosinophil counts), or factors that have been shown to affect disease presentation (Age, Gender, BMI). Figures 5.14., 5.15., and 5.16., show the significant findings from the correlation analyses. MiRNAs with at least one correlation with a clinical outcome or other DE-miRNA are shown.

For the Control vs Severe correlation analysis 18 DE-miRNAs were identified (out of the 35) with significant correlations with other DE-miRNAs or clinical outcomes (Figure 5.14.). Patient Severity showed a weak negative correlation with three DE-miRNAs (miR-1278 [-0.38], miR-4775 [-0.28], and miR-548t-5p [-0.37]). There were also correlations identified among DE-miRNAs. The correlations were predominantly positive between miRNAs. None of the other clinical measures showed correlations with the DE-miRNAs (BMI, FEV1/FVC ratio, Age, Gender, IgE and Eosinophil counts).

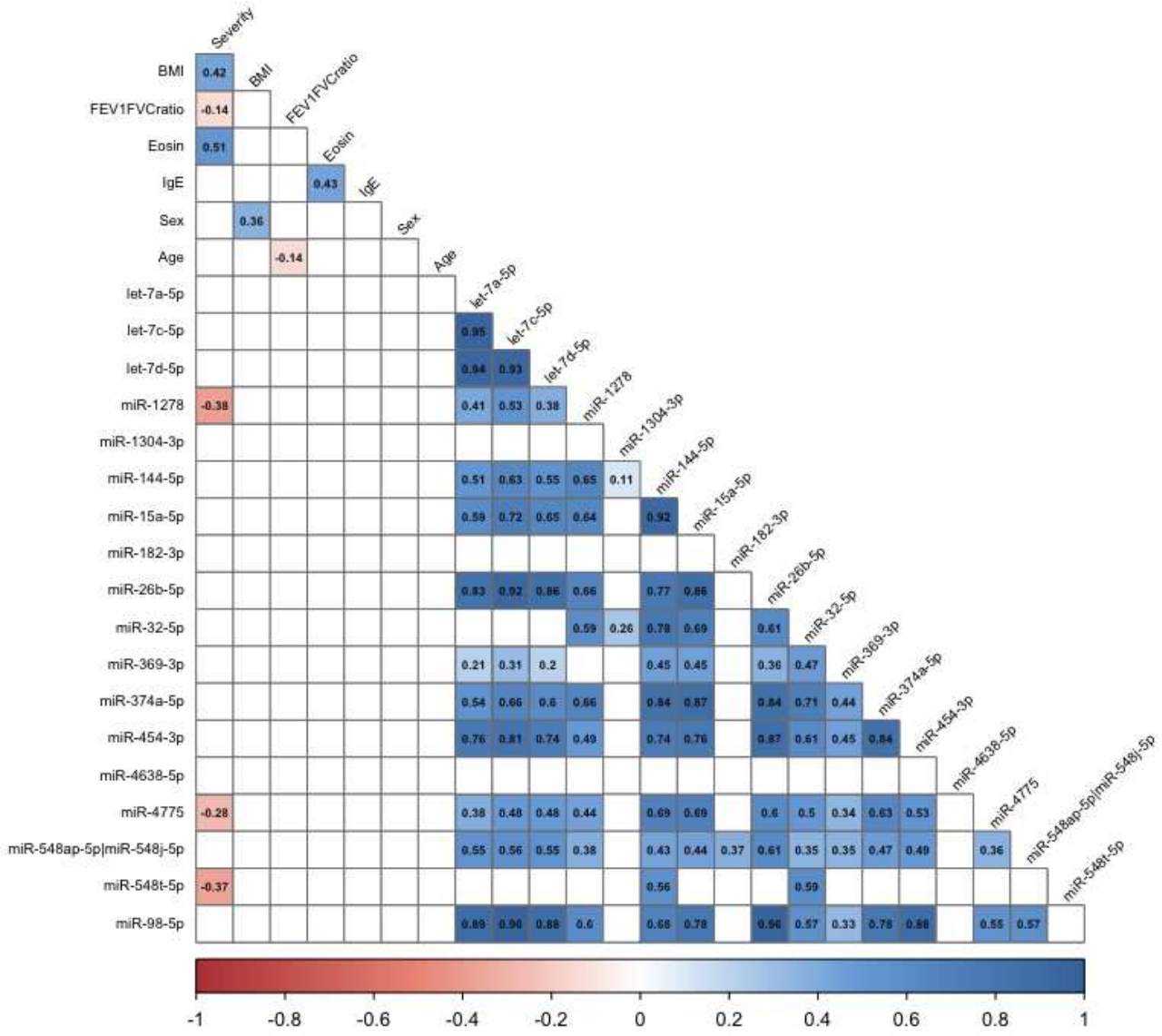


Figure 5.14. Correlation Analysis of Clinical Variables vs DE-miRNAs from Control vs Severe*. A Spearman correlation with a 95% confidence level was used, a P value > 0.05 was considered significant. Only significant correlations highlighted. Non-significant results are blank. DE-miRNAs identified through BaySeq DE analysis, where FDR < 0.05 and fold change > 0.5. BMI (Body mass index), FEV1FVCratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

For the Non-Severe vs Severe groups correlation (Figure 5.15), patient severity showed a moderate positive correlation with miR-145-5p (0.49). The latter also showed a weak to moderate correlation with BMI (0.38). The Let-7 miRNAs identified as DE-miRNAs (let-7a-5p, let-7b-5p, and let-7d-5p), showed positive correlations among themselves but no correlations to other clinical measures or severity. The remaining DE-miRNA, miR-486-3p, showed no correlation with the other miRNAs or clinical measures.

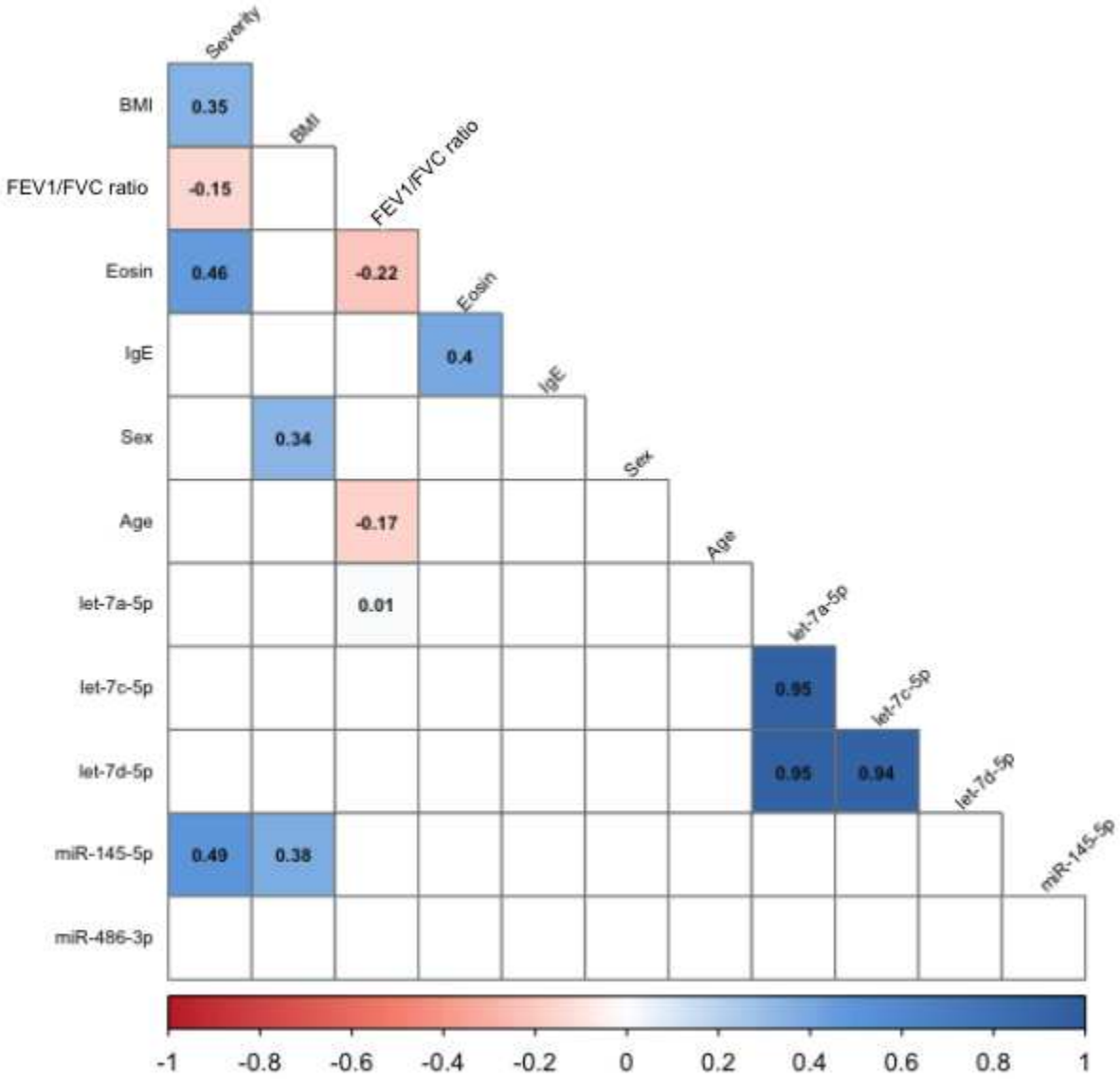


Figure 5.15. Correlation Analysis of Clinical Variables vs DE-miRNAs from Non-Severe vs Severe*. A Spearman correlation with a 95% confidence level was used, a P value > 0.05 was considered significant. Only significant correlations highlighted (P < 0.05). Non-significant results are blank. DE-miRNAs identified through BaySeq DE analysis, where FDR < 0.05 and fold change > 0.5. BMI (Body mass index), FEV1/FVC ratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

For the Control vs Non-Severe groups correlation (Figure 5.16), only BMI was identified as having a correlation with one (miR-1304-3p [-0.34]) of the two DE-miRNAs identified.

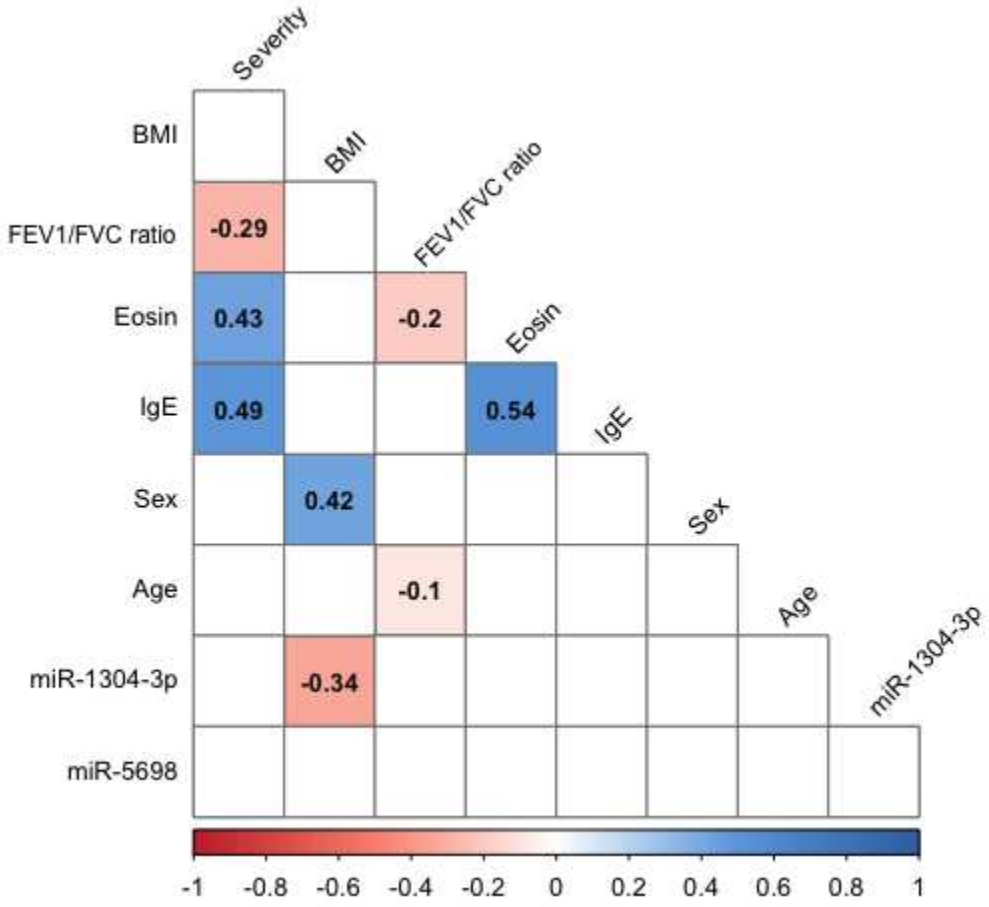


Figure 5.16. Correlation analysis of clinical variables vs DE-miRNAs from Control vs Non-Severe*. A Spearman correlation with a 95% confidence level was used, a P value > 0.05 was considered significant. As per scale, red indicates negative correlation and blue positive correlation. Only significant correlations highlighted (P < 0.05). Non-significant results are blank. DE-miRNAs identified through BaySeq DE analysis, where FDR < 0.05 and fold change > 0.5. BMI (Body mass index), FEV1/ FVC ratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

5.3.4. miRNet mRNA Target Analysis

As per the pilot Celtic Fire study (Chapter 4, Section 4.3.4.) Gene Ontology classification analysis was conducted to look at potential gene targets and functional aspects that might be affected by the identified DE-miRNAs.

For each analysis, DE-miRNAs with a FDR <0.05 and a FC > 0.5 were analysed for targets using miRNet (Chang *et al*. 2020) (for Control vs Severe N = 35, for Non-Severe vs Severe N = 5, and for Control vs Non-Severe N = 2). As before (Chapter 4, Section 4.3.4.), the results of the miRNet analysis were then selected within the platform and put through a Gene Ontology analysis to identify numbers of genes associated with Biological Process (BP), Molecular Function (MF), and Cellular Components (CC). During selection, all genes were analysed using hypergeometric sampling for the three analyses (MF, CC, and BP). Graphs of each aspect of Gene Ontology were made with the genes identified as targets from the miRNet target analysis (Figures 5.18. to 5.22.).

As previously in Chapter 4, the top 10 categories will be considered for each of the domains.

The figures (Figures 5.18. to 5.22.) show bar charts of the associated genes and functional categories identified from the input DE-miRNAs, as well as the number of genes within each category. The same gene targets may appear in multiple categories of the Gene Ontology analysis; for each analysis, a total number of unique genes that made up the top 10 categories have been given. For full tables of the analysis data for each section (adj $P < 0.05$), see Supplementary Materials: “Chapter 5 (CF2): Supplementary Materials”; “CF2_GO-sig_table_ControlvsSevere".

5.3.4.1. Control vs Severe Target Analysis

Only thirty-four of the 35 DE-miRNAs (Table 5.3.) identified in the BaySeq analysis of Control versus Severe asthmatics were input into the miRNet database for analysis as one miR was a duplicate sequence (with a different name (miR-548ap-5p|miR-548j-5p) hence it was not included.

Six thousand and 258 gene targets were identified. After filtering to remove interactions from sequencing and computational methods of target identification, and as there were > 4,785 targets remaining, degree filtering (Chapter 2, Section 2.1.5.3.) was also performed resulting in 2,372 targets for GO analysis. All gene targets identified, for all GO categories, and significance are available in the Supplementary file (CF2_GO_sig_table_ControlvsSevere).

Compared with the pilot study (Chapter 4 Control versus Severe asthmatics) there were some shared targets between that study and this present study and these were: *ITGB3*, *PTEN*, *TWIST1*, *FBXW7*, *BCL2*, *CASP3*, *KRAS*, *SMAD4*, *TRIB1*, *KLF4*, *TRAF3*.

Unlike in the pilot study, where no more than two source miRNAs were targeting the same gene (Chapter 4, Section 4.3.4. and Figure 4.11.), in this study there was greater complexity in the number of miRNAs with shared targets with instances of up to 10 miRNAs sharing the same gene target. To give an idea of the increased complexity of the interactions Figures 5.17. and 5.18. show the combination of source miRNAs and the specific genes that are targeted as identified in the Biological Process Gene Ontology analysis. A complex interaction was seen with 5 source miRNAs (let-7a-5p, let-7d-5p, let-7d-5p, miR-26b-5p, and miR-98-5p) that together all target 4 genes (*PDGFB*, *INTS7*, *HERPUD1*, *DVL3*) all within the most significant BP GO category (Regulation of cellular metabolic process).

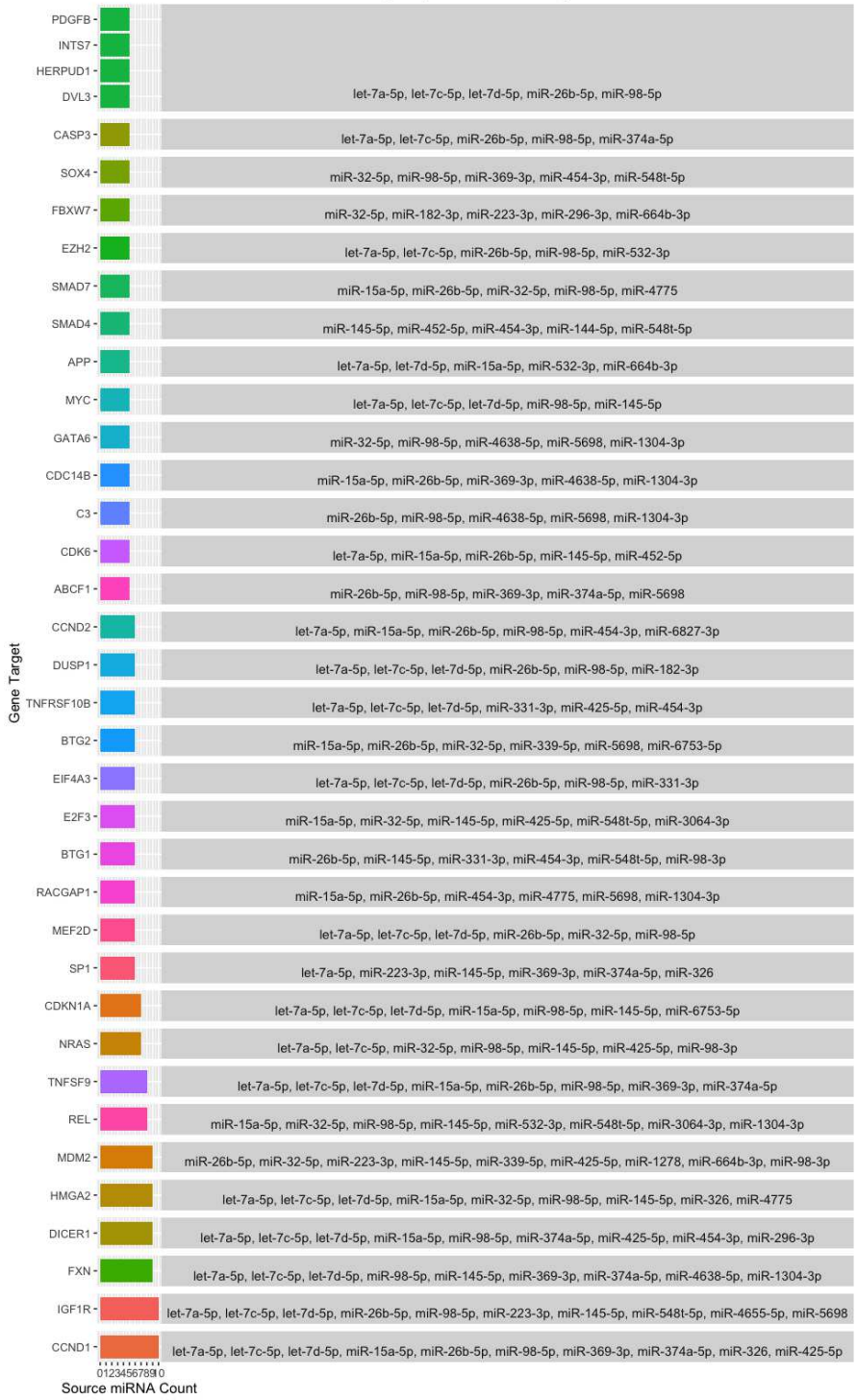


Figure 5.17. Source miRNAs with Shared Targets. *Genes targets are shown on the y axis, with a count of the source miRNAs in the coloured bars, the specific miRNAs are named. This chart colour corresponds with the colours of Figure 5.19 for cross-reference. Only gene targets with more than or equal to 5 source miRNAs have been shown.*

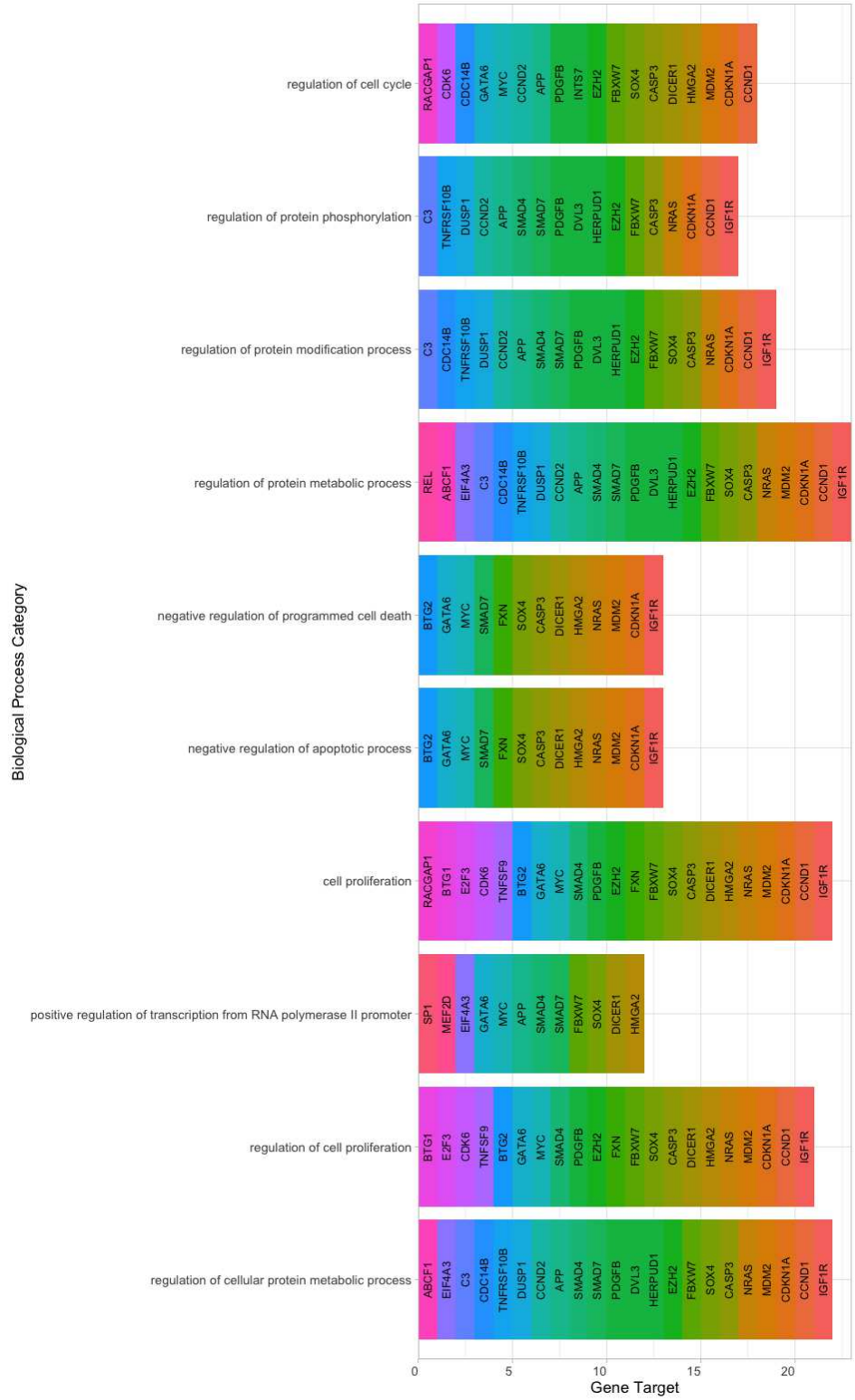


Figure 5.18. Gene Targets and top 10 Biological Process Category from DE-miRNA input from Control vs Severe Asthma for genes with 5 or more source miRNAs*. Gene targets are coloured based on source miRNAs combinations shown in Figure 5.18. above.*

The Biological Process category with the highest significance was in 'regulation of cell cycle' N = 138, $P < 0.0001$ (genes listed in Table 5.4. at the end of this section). Of the top 10 categories (all $P < 0.0001$), the category with the highest number of unique gene targets was in 'cell proliferation' N = 154. There was a total of 272 genes that made up the top 10 GO Biological Process target categories generated from the 34 DE-miRNAs input. There were 5 genes that appeared in all top 10 categories (*IGF1*, *JUN*, *SMAD3*, *TP53*, and *WNT5A*). Figure 5.19. shows the distribution of numbers of gene targets across the top 10 categories identified by their adjusted *P* values.

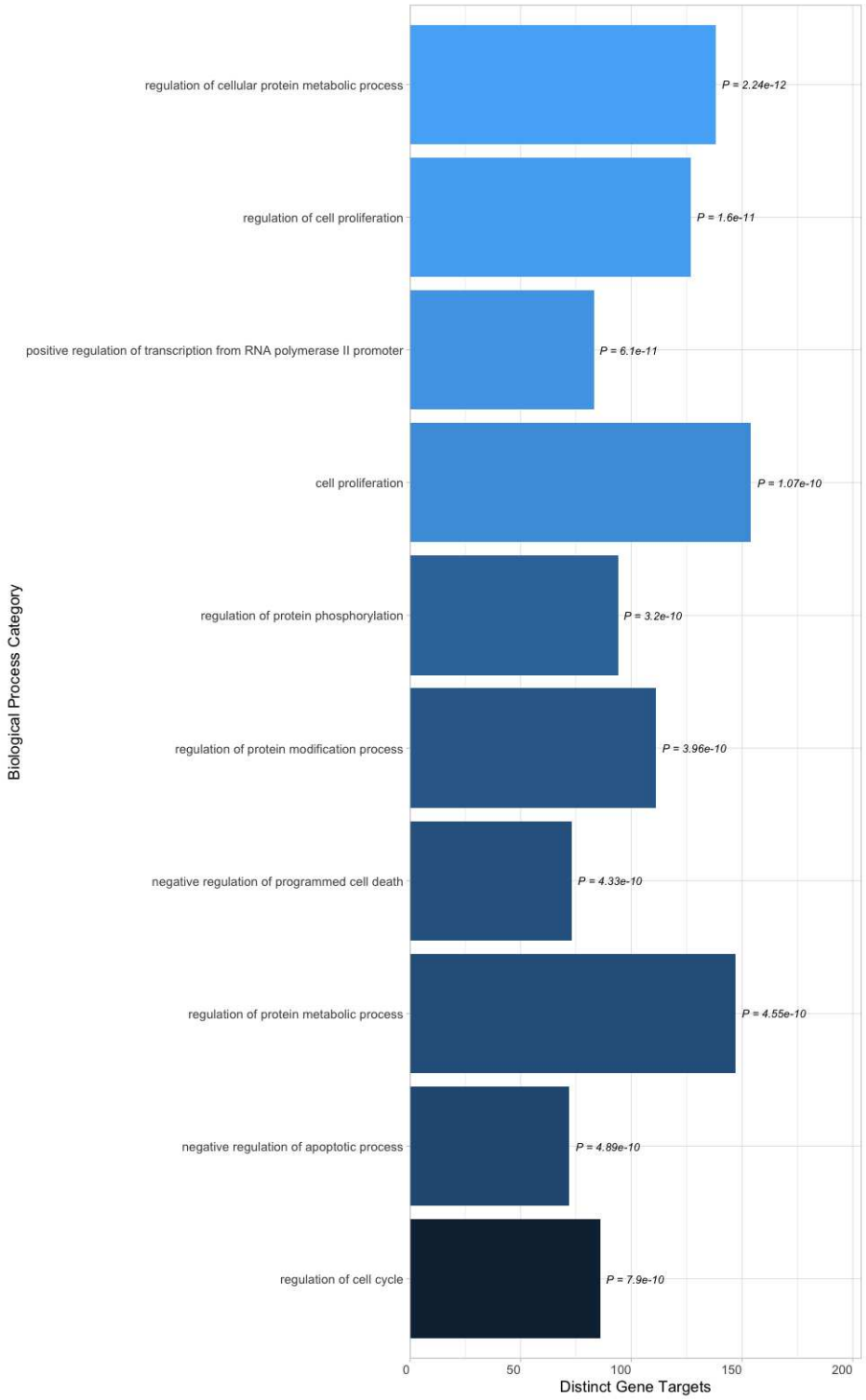


Figure 5.19. Gene Ontology Classification (Biological Process) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls. *The thirty-four differentially expressed miRNAs between Controls vs Severe and were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. Only the top 10 categories are shown here. For full details of all GO results including lists of the genes see Supplementary (Chapter 5: CF2). The bars represent the number of target genes in each Biological Process category, ordered (top to bottom) from most significant to least significant P-values. These results provide insight into the Biological Processes underlying asthma.*

There was a total of 350 genes that made up the top 10 GO Molecular Function target categories (Figure 5.20.) generated from the input 34 DE-miRNAs. The Molecular Function category with the highest significance was in 'positive regulation of transcription (DNA dependent)' N = 105, $P < 0.0001$. Of the top 10 categories (all with a $P < 0.001$) the category with the highest number of hits was in 'nucleotide binding' N = 159. There was one gene that appeared in 9 of the top 10 categories (*TP53*).

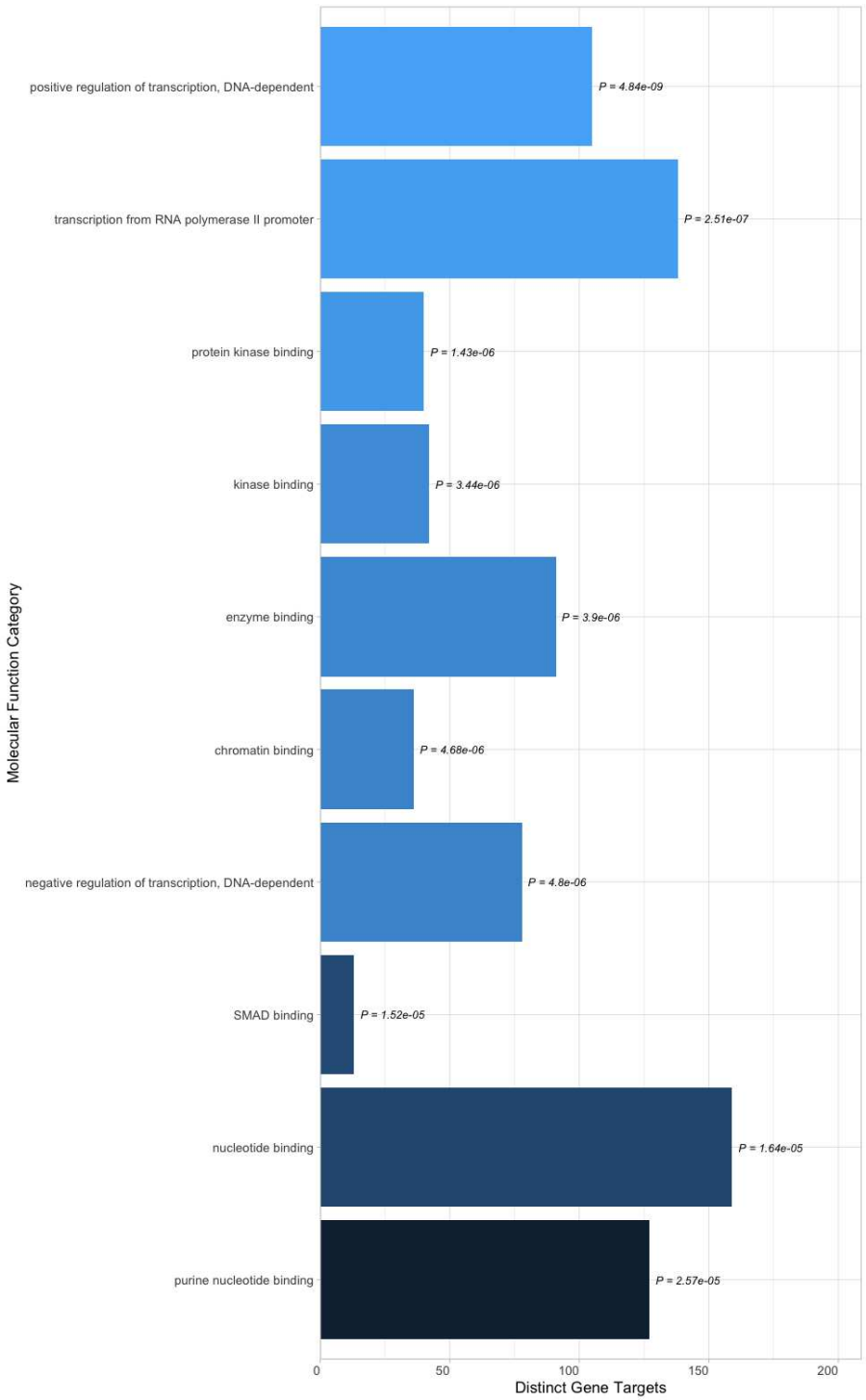


Figure 5.20. Gene Ontology Classification (Molecular Function) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls and the Significance of Each Category*. The thirty-four differentially expressed miRNAs between Controls vs Severe were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. Only the top 10 categories are shown here. The bars represent the number of target genes in each Biological Process category, ordered (top to bottom) from most significant to least significant P-values.*

There were a total 347 genes that made up the top 10 GO Cellular Component target categories generated from the input 34 DE-miRNAs identified between controls and severe asthmatics. The Cellular Components category with the highest significance was in 'nucleoplasm' N = 153, $P < 0.0001$. Of the top 10 categories (all with a $P < 0.0001$), the category with the highest number of hits was in 'membrane-enclosed lumen' N = 216. Figure 5.21. shows the top 10 gene targets from the GO Cellular Component analysis from the DE-miRNAs identified between severe asthma and healthy controls from blood samples. Of note, the gene *TP53* appeared in 9 of the top 10 categories.

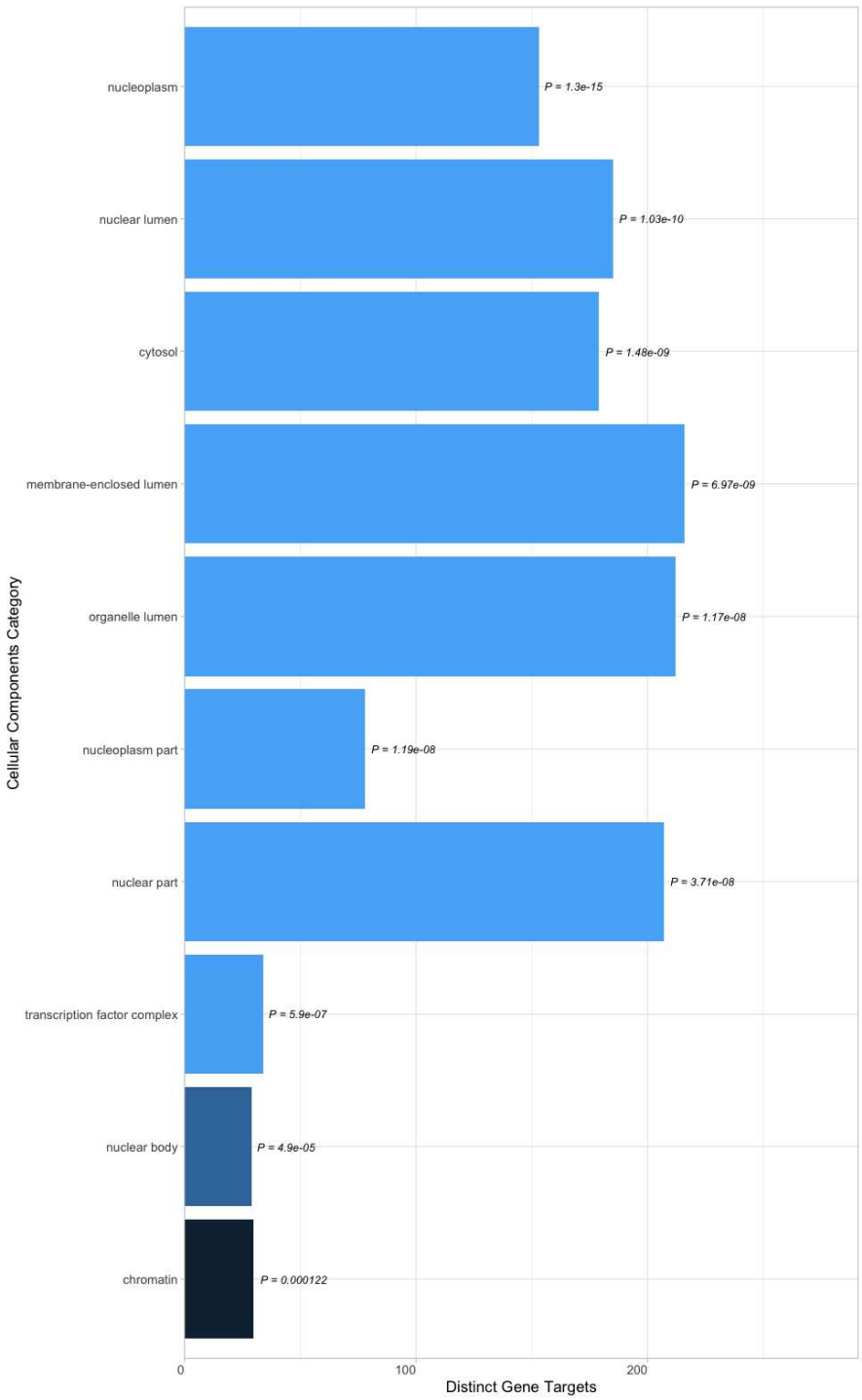


Figure 5.21. Gene Ontology Classification (Cellular Component) Features and Number of Target Genes for Differentially Expressed miRNAs between Severe and Healthy Controls and the Significance of Each Category. *The thirty-four differentially expressed miRNAs between Severe and Healthy Controls were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. Only the top 10 categories are shown here. The bars represent the number of target genes in each Biological Process category, ordered (top to bottom) from most significant to least significant P-values.*

The Venn diagram in Figure 5.22. shows the cross over of gene targets among the GO analysis domains (Biological Process [BP], Molecular Function [MF], Cellular Components [CC]) top 10 most significant result categories for the thirty-four DE-miRNAs. Just over 25% of the genes identified were shared among all tests (135 genes). Cellular Component analysis had the highest number of unique gene targets (99 genes). Table 5.4 details the specific gene targets for each GO category.

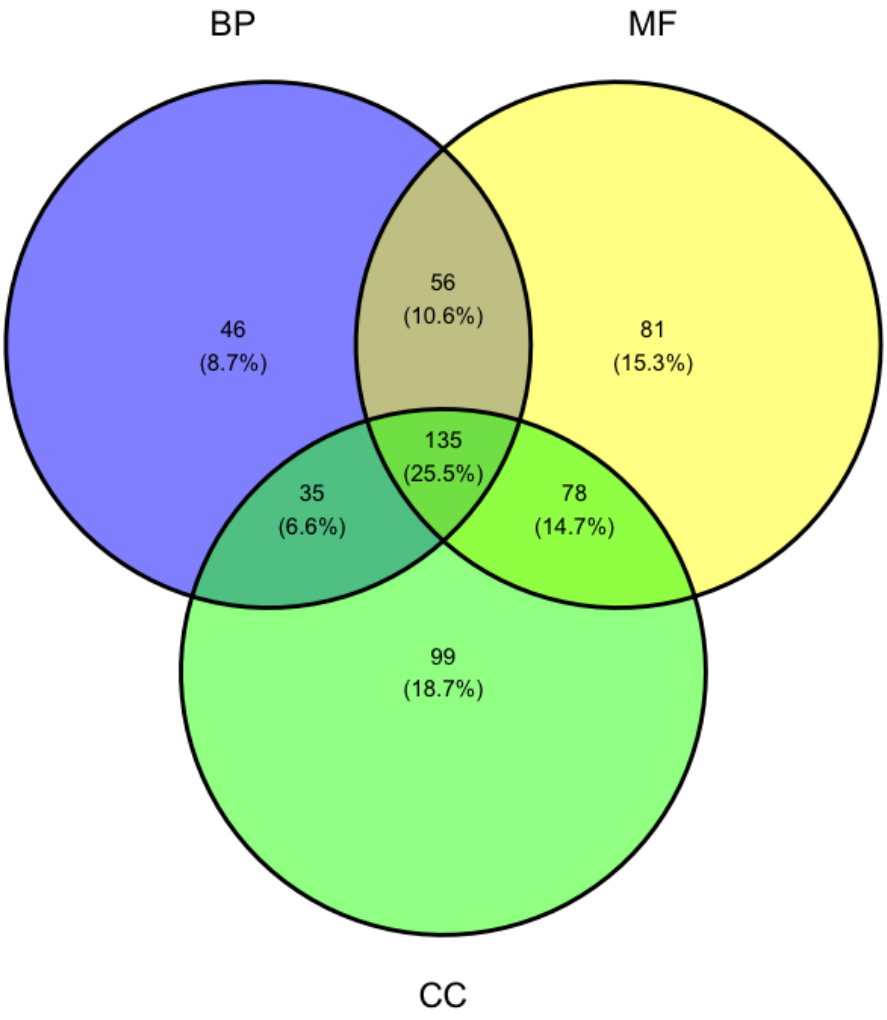


Figure 5.22. Venn Diagram of Target Genes Identified as Significant in the Top 10 Categories of each GO Analysis Category in Control vs Severe Asthma Study Blood Sample miRNA Expression. *DE-miRNAs (N = 34) were input into miRNet database and passed through GO analysis categories (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). Resulting gene targets were then compared among the three categories.*

Table 5.4. Target Genes from GO analysis top 10 most significant (P < 0.05) categories compared against one another. *DE-miRNAs (N = 34) identified from blood sample miRNA expression between Controls vs Severe asthmatics. Unique genes and genes identified among other GO analysis categories.*

Interaction (analysis)	Gene names
Unique to Biological Process	*CCL2, CCL5, CD44, IL13, IL3, ITGB3, TNFRSF11A, GADD45A, UBE2V2, ATP6V0D1, BTG2, C3, DUSP10, EGR3, GMFB, HERPUD1, NPTN, PLPP3, SERP1, TNFRSF10B, HBEGF, SESN1, ASPM, CD274, CXADR, HAS2, LAMC1, MCC, OSMR, PLAG1, PPM1D, PTCD3, RNF7, TES, TNFSF9, TPD52L2, USP14, CCNF, IL7R, LAMP3, MPL, NCAPG2, SETDB1, SOAT1, TGFBI, TUSC2*
Unique to Molecular Function	*CFTR, CHD1, CHD7, MYO5A, RHOBTB3, ABCC1, ABCG2, ARHGEF5, ARL4C, ARL8B, ATP13A3, ATP1A1, ATP2A2, ATP6AP1, ATP7B, BMP2K, BRD3, CDC34, CDK17, CDK19, CERK, CLCN3, COASY, CSNK1G1, DHX57, DYRK2, DYRK3, EMC7, FAM20B, GPN1, HIVEP1, HSPA12A, HSPA4, ID1, KIF1B, KIF1C, MAP4K3, MAST4, NAGK, NAV2, PANK3, RAB21, RAB27A, RAB30, RAB32, RAB40B, RAB9B, RRAGC, RRAS2, RTCB, SACS, SLFN12, SMC4, TCERG1, TOR4A, TRPS1, TUBD1, ULK2, ZNF136, ZNF281, BACE1, CPEB4, DHODH, DHRS11, ELAVL3, EWSR1, FGD6, FGF23, HACE1, KLF6, MCMBP, PPRC1, RBM12B, RBM41, RBMS2, SERTAD2, SHOC2, STEAP3, TASP1, TRAF1, TRIOBP*
Unique to Cellular Components	*ATXN2L, FANCC, HSPB7, IVNS1ABP, PBX3, SINHCAF, SMNDC1, SNRPC, SRRM2, WBP11, CENPQ, H3F3B, HIST2H4B, MEAF6, MED13L, MED22, NUP98, POLR3G, RPRD2, RRP1B, SNRPD1, AEN, HIST1H1D, SF3B2, SF3B3, SMG6, SMUG1, SNRPA1, SPTBN1, USP3, ADD3, EBNA1BP2, GRWD1, NARF, POP4, REPIN1, URB2, WRB, ZBED1, COL5A1, COLGALT1, CRTAP, CTSD, DIABLO, ERAP2, IGF2R, MRPL18, MRPS16, NUCB2, PDHX, PMPCA, AASDHPPT, AGL, ANAPC13, ANAPC16, ARAP2, ARCN1, CDCA3, CHMP3, CLINT1, CLTA, CNTRL, CRKL, DCTN3, DNAJC6, DYNC2LI1, ECHDC1, EIF1AX, ERBIN, HIST1H2BC, HIST1H2BO, HIST2H2BF, KCNH1, LSM1, LTA4H, MGST3, MYBPC1, OSTM1, PCTP, PHKA1, PHLPP1, POM121C, PPAT, PPP2R2A, QPRT, RASSF9, RPLP1, SEC24B, SH2B3, SPC25, STMN1, STXBP3, SULT1B1, SUN1, TIMM10, TIMM13, TPI1, TPM3, UGP2*
All GO analysis categories	*IGF1, JUN, SMAD3, TP53, AKT1, BCL2, CDKN1B, CDKN2D, EGFR, ILK, PTEN, TNFAIP3, VEGFA, CDKN2B, FBXW7, KAT2B, BMI1, BRCA1, CCND1, CDC25B, CTGF, EZH2, MDM2, PDGFB, RPS6KB1, SMAD4, SMAD6, SMAD7, STAT5A, APP, ATF3, CD28, CDC25A, CDK2, CDK5R1, DICER1, GATA6, HMGA2, HSP90B1, MYC, PAK1, PPARGC1A, TGFBR2, UBB, WT1, ARRB1, ATM, BCL2L1, CCNA2, CCND2, CCNE1, CCNE2, CDC27, CDK4, ECT2, FOXO1, FOXO3, HAX1, MAP2K6, MEF2C, NFKB1, NOTCH2, SENP2, SET, YAP1, CEBPB, CRK, E2F1, FXN, IRS2, KLF4, PSMC1, PSMC4, RNF111, SREBF1, TAL1, TARBP2, CDK6, EIF4A3, ESR1, NACC2, NAMPT, NFE2L2, PRKAA2, RARB, RPA1, SALL1, SMAD2, STAT1, ABCF1, CIAO1, CLOCK, E2F3, EGLN3, FASLG, FOXA1, GNL3, GUF1, IFI16, IGF2BP2, IRS1, MKNK2, MTDH, NCBP1, PTBP2, RACGAP1, ROCK1, RPS6KB2, STAT3, SYNCRIP, TAF9B, VCP, CHUK, CREG1, DDX17, E2F2, E2F5, EP300, HDAC5, MAFB, MIXL1, MYBBP1A, NR2C2, RAP1B, REL, RHOB, RNASEL, SKIL, SMARCC1, SP1, SRF, TEAD3, THRB, TK1, TRAF3*
BP and MF shared	*WNT5A, FGFR3, IL6, SOX4, TGFBR1, CAV1, IFNG, IGF1R, KRAS, LIF, NRAS, FZD5, IL10, MET, ERBB2, PTPRJ, TRIB1, TWIST1, BCL11B, GLMN, SMO, TFAP2A, ADCY9, DVL3, F3, HSF1, MAP4K2, SNX6, SOX11, SRCIN1, TWF1, JAG1, KLF13, SENP1, TIRAP, BTG1, KLF10, MXI1, PRDM1, PRDM4, RAB23, RNF187, TIA1, XRCC2, ASCC3, ASXL2, DGKZ, EN2, HSF2, MEF2D, PAK4, PCLAF, PKIA, SAFB, SOX5, USF2*
MF and CC shared	*UBE2D3, CAD, CAMK4, CASP9, CHD4, DPYSL2, GYS1, MSH3, NRBP1, PARP1, RAPGEF4, RREB1, TLE4, TOP1, TWNK, USP9X, ACTR1A, ACTR5, AK4, ARID1A, CBX6, CPSF7, DDX6, EIF2S3, MAP2K4, MYLK3, NELFCD, NOLC1, NR2F6, NSF, PCCB, PFKFB3, PIP5K1C, PKM, POLG, RAB3B, RAD51C, RHOV, SEPT7, SMARCD1, SUV39H1, TESK1, TOR1B, ACADM, AGPS, AKAP5, CA9, CASP8, CBY1, CNOT4, COL1A2, CTNND1, DCTPP1, ELP1, ERC1, GATAD2A, GDI1, HNRNPDL, HNRNPM, HOXA9, LDHA, MED8, MEIS1, NRBF2, PBXIP1, PEX5L, PHGDH, PIAS2, POU2F1, QDPR, RBM19, SNRNP70, TAF5, TAF5L, TEAD1, TERF2, TNKS1BP1, UGDH*
BP and CC shared	*CASP3, CDKN1A, PDCD10, BAG4, NF2, SPRY1, CDC14B, DUSP1, DUSP7, FKBP14, PSMD11, PSMD5, PSMD7, PSME1, RICTOR, SEC31A, TAB2, AGO1, AGO4, CARM1, CBL, DFFA, LIN28A, MCL1, NOP2, POLD4, PPT1, SKAP2, ZFP36L1, DBF4, FSCN1, INTS7, NUDC, SERPINH1, ZWINT*

Next, the miRNet results were aligned to the source miRNAs relevant to each GO analysis category and compared. Gene targets were filtered for those with more than 5 source miRNAs, and a stacked bar graph was created to show the interactions and how the gene targets were shared among the GO categories (Figure 5.23.). Of the 34 DE-miRNAs identified, thirty-one identified among all GO analysis categories with more than 5 shared gene targets or more.

Two genes *IGF1R* and *CDKN1A* had 10 source miRNAs that targeted them, with *IGF1R* identified in all GO analysis categories (BP, CC and MF) and *CDKN1A* found in both BP and MF analyses. All the categories considered (as filtered by the Venn diagram) contained one or more genes with more than 5 source miRNAs, highlighting the number of DE-miRNAs that target the same gene. Target genes found in all analysis categories generated 11 gene targets with 5 or more source miRNAs (*CCND2*, *E2F3*, *EIF4A3*, *RACGAP1*, *SP1*, *REL*, *DICER1*, *FXN*, *HMDA2*, *MDM2*, and *CCND1*).



Figure 5.23. Gene Targets and Source miRNAs from Control vs Severe GO Analysis Faceted by Analysis Interaction*. DE-miRNAs (N = 34) from blood sample analysis between Healthy Controls and Severe asthma were input into a GO analysis. The resulting genes were filtered for those with the greatest number of source miRNAs and those with more than 6 or more interactions are shown here. The highest number of miRNAs acting on one gene target were in the BP_MF interaction and target IGF1R, and combined categories targeting CCND1. The target genes identified in all GO categories combined identified the most gene targets with more than 5 source miRNAs per gene target.*

The Control vs Severe pilot, and the Control vs Severe analysis had both identified miR-32-5p and miR-1304-3p to be differentially expressed. The number of gene targets output by the Gene ontology analysis, however, varied between the studies. In the smaller pilot study, where there were 10 DE-miRNAs including miR-32-5p, miR-32-5p was associated with having six target genes, whilst in the larger cohort study, where 34 DE-miRNAs were found, miR-32-5p was found to have 36 genes as targets. Table 5.5. shows the gene names identified in both studies.

Table 5.5. Differences in Gene Targets for miR-32-5p identified between Pilot study and Broader inclusion study in Control vs Severe Cohort Biological Process Analysis*. red text highlights are gene targets identified in pilot study from the same input source DE-miRNA compared to the larger cohort study.*

Identified Gene Targets For miR-32-5p	*TRIB1*
ITGA5	*SPRY2*
ITGB3	*AXIN2*
PTEN	*ICAM1*
TWIST1	*KDR*
FBXW7	*SELE*
BCL2	*SOD2*
CASP3	*NCOA3*
CD40	*KLF4*
CDC42	*CDH1*
EDNRA	*HOXD10*
MTOR	*SERPINF2*
GSK3B	*PTX3*
KRAS	*NCOA6*
SMAD4	*TRAF3*
PAK2	*CASP7*
PDGFRB	*GPX2*
RAC1	*AP2M1*
CXCR4	*DPYSL2*
APLN	*PHLPP1*

5.3.4.2. Non-Severe vs Severe Target Analysis

All 5 DE-miRNAs identified in BaySeq analysis of Non-Severe vs Severe miRNA expression were input into miRNet database. miRNet identified 1,245 gene targets. As no targets were identified using the strict target identification criteria, for identification of potential interactions, a more relaxed threshold was permitted allowing all target identification methods. The resulting gene targets were predominantly identified using High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) and photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) methods. Gene ontology analysis followed using the results of the amended threshold. It is worth noting that these results are not ideal as they have not had the stringent experimental validation as shown in the previous section. Due to this limitation in experimental validation a selection of the 5 most significant categories were further considered for each domain. A summary figure (Figure 5.24.) combining the results across the domains is discussed below.

The Biological Process category with the highest significance was in 'regulation of cellular protein metabolic process' N = 174, $P < 0.0001$. The Molecular Function category with the highest significance was in 'enzyme binding' N = 147, $P < 0.0001$. The Cellular Component category with the highest significance was in 'nucleoplasm', N =185, $P < 0.0001$. Of the top 5 categories (all with a $P < 0.0001$), the BP category with the highest number of hits was in 'regulation of cellular protein metabolic process'. For MF (all with a $P < 0.0001$), the category with the highest number of hits was in 'enzyme binding' N = 247. For CC (all with a $P < 0.0001$), 'nucleus' had the highest number of hits, N = 576. There were a total 269 genes that made up the top 5 GO BP target categories. For MF there were a total 375 genes in the top 5 categories, and for CC there were 287

genes in the top 5. In the BP analysis there was one gene that appeared in 5 of the top categories (*VEGFA*). In the MF analysis, there was one gene that appeared in all of the top categories (*SMAD3*). In the CC analysis, 18 genes appeared in all 5 categories (*ACTB*, *AHR*, *CDK4*, *EIF4A3*, *EPAS1*, *FOXO3*, *GTF3C1*, *HIF1A*, *HMGA1*, *IVNS1ABP*, *NFATC1*, *POU5F1*, *SMAD2-6*, and *SOX2*).

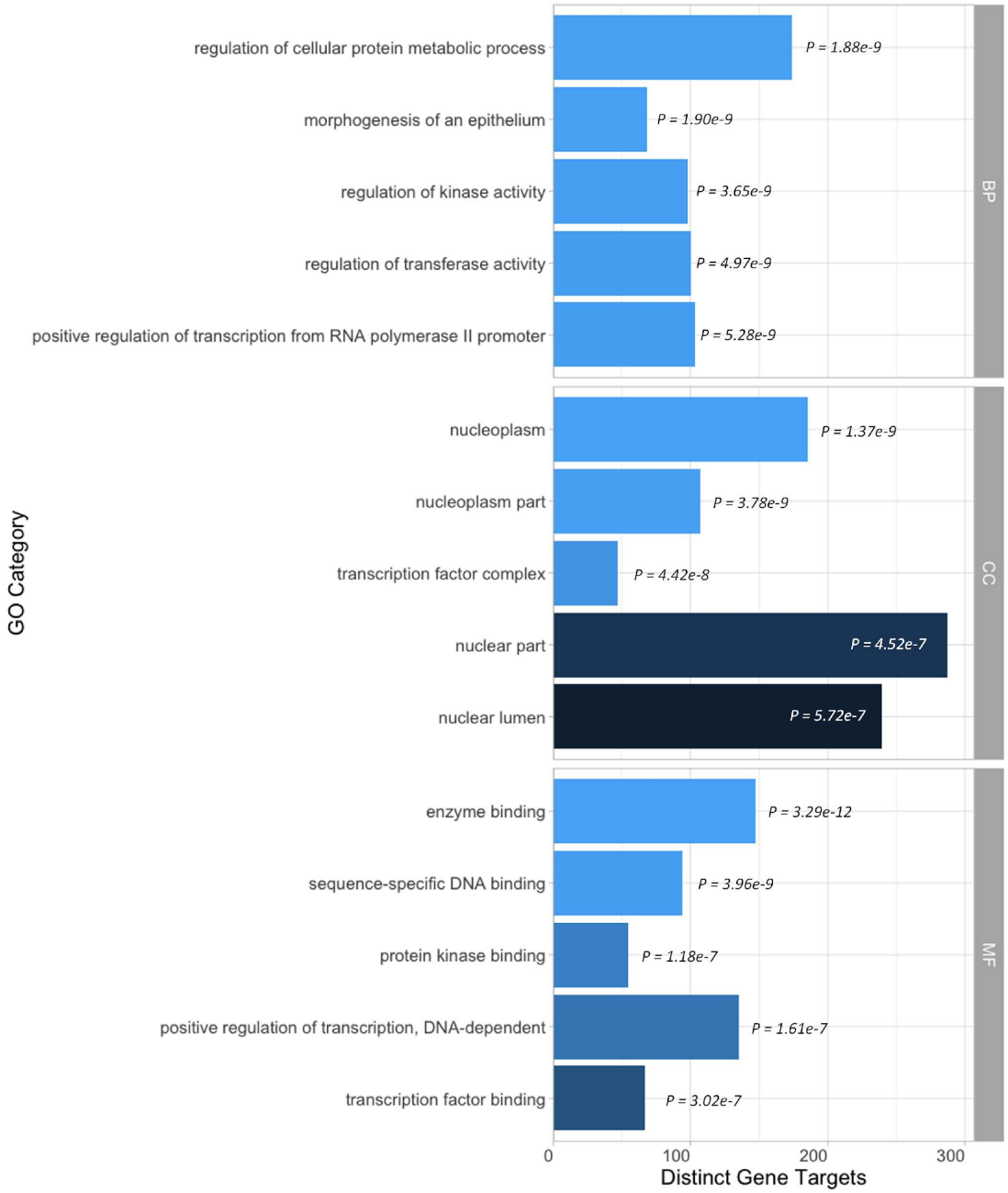


Figure 5.24. Gene Ontology Classification Summary Features and Number of Target Genes for Differentially Expressed miRNAs between Controls and Non-severe Asthmatics. *All five differentially expressed miRNAs between Non-Severe and Severe asthma were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. Only the top 5 categories for each domain are shown (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). For full details of the GO results including lists of the genes see Supplementary (Chapter 5: NSvsS). The bars represent the number of target genes in each Molecular Function category, ordered (top to bottom for each category) from most significant to least significant P-values. Interaction networks permitted computational and sequencing methods as there were insufficient experimental methods of validation for the GO analysis to be conducted.*

From the results of the GO analysis from the DE-miRNAs (N = 5), a Venn diagram was made to show crossover of gene targets among the GO-tested categories (Biological Process [BP], Molecular Function [MF], Cellular Components [CC]) top 5 most significant result categories (Figure 5.25). Just over 16% of the genes identified were shared among all tests (83 genes). Cellular Component analysis had the greatest number of unique gene targets (128 genes).

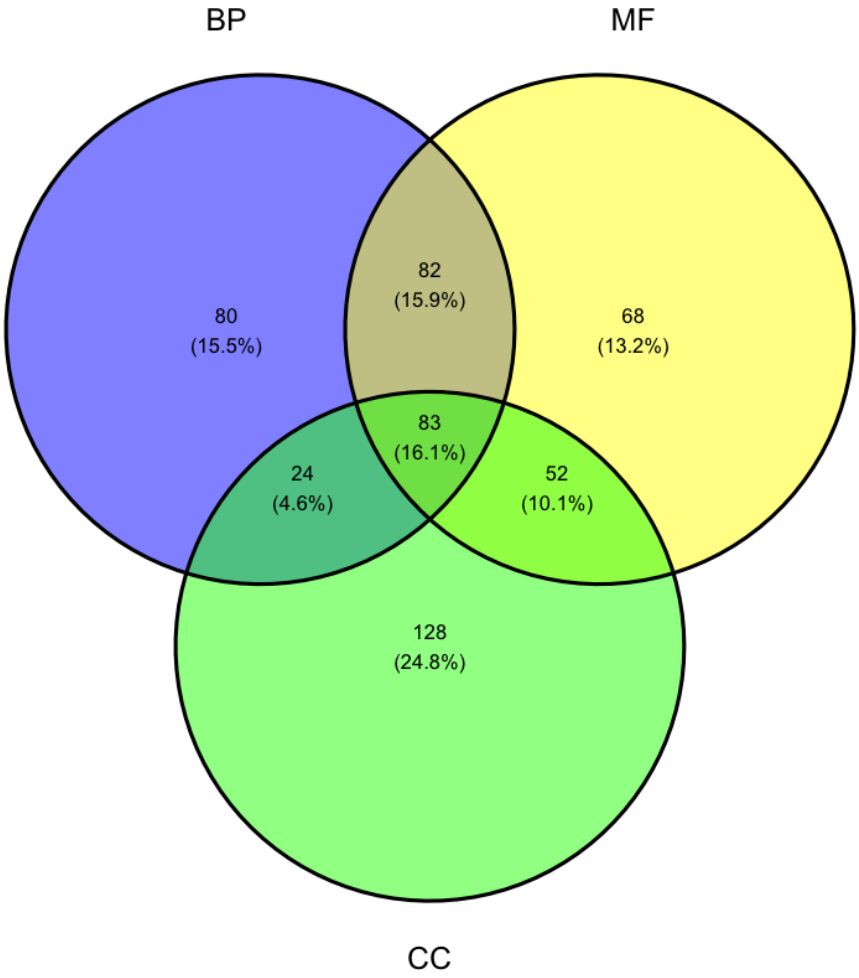


Figure 5.25. Venn Diagram of Target Genes Identified as Significant in the Top 5 Categories of each GO Analysis Category in Non-Severe vs Severe Asthma Study Blood Sample miRNA Expression*. DE-miRNAs (N = 5) were input into miRNet database and passed through GO analysis categories (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). Resulting gene targets were then compared among the three categories.*

Table 5.6. Target Genes from GO Analysis Top 10 Most Significant (P < 0.05) Category Comparison in Non-Severe vs Severe Asthma Cohort. *DE-miRNAs (N = 5) identified from blood sample miRNA expression between Non-Severe vs Severe Asthmatics. Unique genes and genes identified among other GO analysis categories.*

Interaction (analysis)	Gene names
Unique to Biological Process	*MFB, HERPUD1, IGF2, ITGB3, KRAS, MAP3K3, MBTPS1, NRAS, PREB, SNX6, SYVN1, TAB2, TGFB2, THBS1, TNFRSF10B, TRAF7, CD44, DLC1, TRIM71, VAV2, ACER2, AGO1, AGO2, AGO4, ATXN2, BUB3, COL5A1, CTGF, CTNNBIP1, DISC1, EIF4E, EIF4G2, EPHA2, FNIP1, GCN1, HBEGF, IGF2BP1, IGF2BP3, IGFBP5, IL13, KRT7, LARP4B, LUZP1, MMP12, MMP14, NTN1, NUMB, PDGFD, PLXND1, PODXL, PPP1R15B, QKI, RAB10, RICTOR, RPS4X, RPS6KB1, SEMA4C, SERINC5, SH3D19, SLC12A2, TMTC3, TNRC6A, TNRC6B, UBXN2A, USP14, WNT16, ZFP36L1*
Unique to Molecular Function	*BRSK2, CAND2, CASP9, GTF2I, MAPK6, NSD1, PDLIM5, PGAM1, PRDM1, SIK1, TIRAP, ARF6, ARIH1, ATP6V1G1, BACH1, BTG1, CARHSP1, CC2D1B, CDC42EP4, CDH2, CFTR, CRY2, CYP2C19, DLX4, DOT1L, ECM1, ELFN2, ERC1, ERG, EXPH5, FLI1, FMNL3, FZD7, GGA3, GOLGA4, HIVEP3, HOXA6, HOXC4, HOXD1, IKZF3, KCTD1, KDM6B, KLF5, MAF1, MDM4, MEIS3P1, MRPL12, MYO5A, MYRF, NDUFAF7, PDE4DIP, PLAGL2, PTGES2, RAB11FIP4, RAD18, RBFOX2, RFX6, RNF187, RTKN, SMG7, SRGAP1, STRN, STRN4, SYTL3, TCERG1, WWP2, YES1, ZNF746*
Unique to Cellular Components	*ABT1, E2F6, EYA3, GTF3C1, GTF3C5, IVNS1ABP, TAF5, TAF5L, TAF9B, TRRAP, AKAP17A, ATXN2L, CHD3, COIL, CSNK2A1, DDX39B, INTS12, INTS7, JADE1, JAZF1, MED13L, MED28, NACC1, NOLC1, PCF11, POLR2D, POLR2F, POLR3D, PRPF8, RFWD3, RPA1, SNRPC, SRSF2, TIMM50, UBAP2L, WBP11, ANKRD28, CAMK2A, CAMK2D, CDC7, CHTOP, CTDP1, DFFA, DNA2, H1F0, KPNA5, KRT8, LIN52, PRIM2, RAD21, RFC2, RRM1, RRM2, SBDS, SF3A1, SF3B4, SMC1A, STAT2, SUPT16H, TNFSF13, TTF1, ADD3, AKAP8, AURKB, BAZ2A, CD2AP, CDV3, CFL2, CIAPIN1, DDX18, DTD1, EIF3A, FIGN, G2E3, HIST1H1D, IP6K1, LMNB2, MIDN, NOM1, PGRMC1, PHF6, PNMA3, PRR13, REPIN1, RPL3, RPL4, RPL9, RPS10, RPS13, RPS24, RPS3A, SENP5, SPTBN1, UHRF2, UTP15, UTP4, UTP6, WDR3, AQR, BNIP3L, BUD13, CBX6, COPS3, COPS6, COPS8, CTDNEP1, HASPIN, HNRNPDL, IPO9, LBR, MYH9, NCKIPSD, NPC1, NRM, NUCB2, NUP155, NUP43, NUP58, PARP16, PCGF3, PTGES, RALY, SNUPN, SREK1, TUBB, TXLNG, WDR6, ZMAT5*
All GO analysis categories	*CCNC, CCNT2, EGFR, LMNA, MEN1, PAK1, SIRT1, UBA52, UBB, ZEB2, CCND1, CCND2, CDC25A, CDC25B, ERBB4, HES5, MAP2K6, MTOR, PRKCD, PTK2, SMAD3, SMAD4, SOX9, XBP1, BCL2, EIF4A3, ESR1, FOXA1, FOXO3, HAND1, HDAC2, HIF1A, MEF2C, MUC1, MYC, NCOA3, NFKB1, RPS6KA3, RXRA, SKI, SNW1, SOX2, SREBF1, STAT1, UHRF1, WDR61, AHR, BCOR, CALM3, CASP8, CREM, CRX, DDX17, E2F1, E2F5, EPAS1, ETS1, EZH2, HDAC5, HMGA2, KLF4, KMT2A, KMT2D, MDM2, MIXL1, MYO6, NANOG, NCOA1, NFATC1, NIPBL, PBX2, PIN1, POU5F1, RARB, SMAD2, SMAD5, SMAD6, SMARCC1, SP1, STAT3, SUMO1, VDR, WAC*
BP and MF shared	*IL6, PKD1, VEGFA, APP, CD40, DVL3, FZD8, HRAS, ILK, MYOCD, PDGFB, PXN, STK4, TGFBR2, CALR, CCNB2, ENG, ERBB2, F2R, GSK3A, IL10, IL6R, LYN, MAP2K7, MAP3K1, MAP3K11, MAP4K2, PIK3R1, RAF1, SFN, SOCS1, SPRED1, SRCIN1, SYK, TGFBR1, THY1, TNFAIP3, TRIB2, WNT1, ACVR1B, BCL3, CD28, DICER1, HOXB4, IFNB1, IRS1, IRS2, ISL1, JAG1, PAX2, SENP1, SOX11, ARID3A, BCL11A, CBFB, CNBP, CSRNP3, CTNND1, DNMT1, EN2, FXN, FZD6, GABPB1, GLMN, HSF1, HSF2, IL31RA, MAF, MEF2D, MLLT10, NFIC, ONECUT2, PHACTR4, PPP3CA, RYR2, SAFB, SERPINE1, SP7, SPDEF, SREBF2, USF2, VCL*
MF and CC shared	*HMGA1, ACTB, BPTF, FOXJ3, MTA1, BAZ1B, BCL2L1, CBX5, CEBPB, HOXA9, MSH3, NACC2, NR1I2, POU3F1, RAD9A, RPS6, ZFHX3, ATXN7L3, BNIP3, CAD, DHX9, E2F2, E2F3, GATAD2B, H2AFX, HDAC11, HNRNPUL1, IPO7, MAFK, MBD2, MBD4, MCM2, MEIS1, MTDH, NR1D2, NR4A1, NR6A1, NUDT21, ORC4, PARP1, PBX3, PHF12, POU3F2, RANBP2, REL, RREB1, SMARCAD1, SMARCD1, TLE4, TUT1, U2AF2, YWHAZ*
BP and CC shared	*BCCIP, CASP3, CDKN1A, DUSP1, DUSP6, DUSP7, ERN1, GPS1, MAPK11, NF2, CDK4, CFL1, EIF5AL1, FANCI, JARID2, LIN28A, MIB1, MKNK2, NCBP1, NOTCH2, OGT, PSMA2, PSMB2, SET*

The GO analysis results were aligned to the source miRNAs relevant to each GO analysis category and compared. Gene targets were filtered for those with more than 4 source miRNAs and a stacked bar graph was created to show the interactions and how the gene targets were shared among the GO categories (Figure 5.26.). All 5 DE-miRNAs were identified among all GO analysis categories with more than 5 shared gene targets or more. The *CDKN1A* gene was the target of the most source DE-miRNAs and was found to be shared between the CC and BP analysis. The combined GO analysis categories generated the highest number of gene targets (*FXN*, *HMGA2*, *KMT2D*, *MDM4*, *MYC*, *NR6A1*, and *NSD1*) with 4 source miRNAs. There were no genes with 4 or more source miRNAs exclusively found in the MF analysis or shared between the CC and BP analyses.

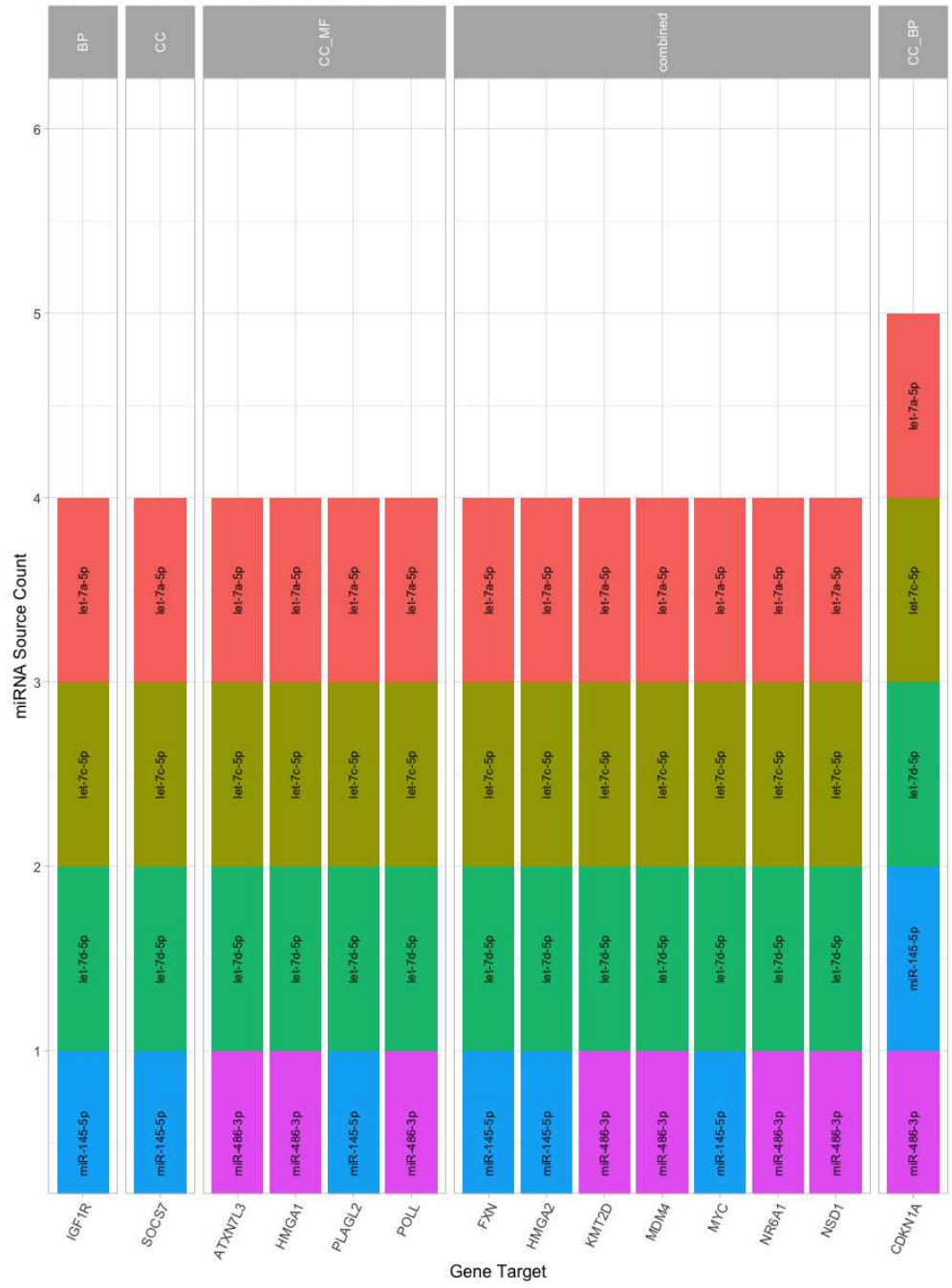


Figure 5.26. Gene Targets and Source miRNAs from Non-severe vs Severe GO Analysis Faceted by Analysis Interaction. *DE-miRNAs (N = 5) from blood sample analysis between Non-Severe and Severe Asthma were input into a GO analysis, the resulting genes were filtered for those with the greatest number of source miRNAs and those with more than 4 or more interactions are shown here. The highest number of miRNAs acting on one gene target were in the CC_BP interaction and target CDKN1A. The target genes identified in all GO categories combined identified the most gene targets with more than 4 source miRNAs per gene target.*

5.3.4.3. Control vs Non-Severe Target Analysis

As per the volcano plot (Figure 5.12.) for Control versus Non-Severe two DE-miRNAs were identified, miR-1304-3p and miR-5698, and when input into the miRNet database analysis a total of 906 target genes were identified. The interaction tables when filtered for sequencing and computational methods, however, resulted in no hits. Without any hits, no further GO analysis was possible. A relaxed threshold analysis was therefore conducted including computational methods. Gene ontology analysis followed using the results of the amended threshold. Due to this limitation, only the 5 most significant categories and genes have been considered. Figure 5.27. shows a summary of the top 5 gene categories from each domain of the GO analysis (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]).

The Biological Process category with the highest significance was in 'amino sugar metabolic process' N = 8, $P = 0.005$. For Molecular Function it was 'transition metal ion binding' N = 138, $P = 0.001$ whilst for Cellular Components 'nucleoplasm part', N =22, $P = 0.006$. Of the top 5 BP categories (all with a $P < 0.05$), the category with the highest number of hits was in 'post-translational protein modification' N = 17. MF had the highest number of hits in 'transition metal ion binding' N = 138, $P = 0.001$. For CC it was in the category 'nucleoplasm' N = 103, $P = 0.03$.

In each category of the GO (BP, MF, and CC) there were a total of (41, 247, and 113 respectively) genes making up the top 5 categories. In the BP analysis there was one gene that appeared in all 5 categories (*GNPNAT1*). There were 4 genes that appeared in all 5 categories of the MF analysis (*PRPF4*, *SF3B1*, *SRSF2*, and *WAC*). For CC one gene appeared in all 5 categories (*RAD50*).

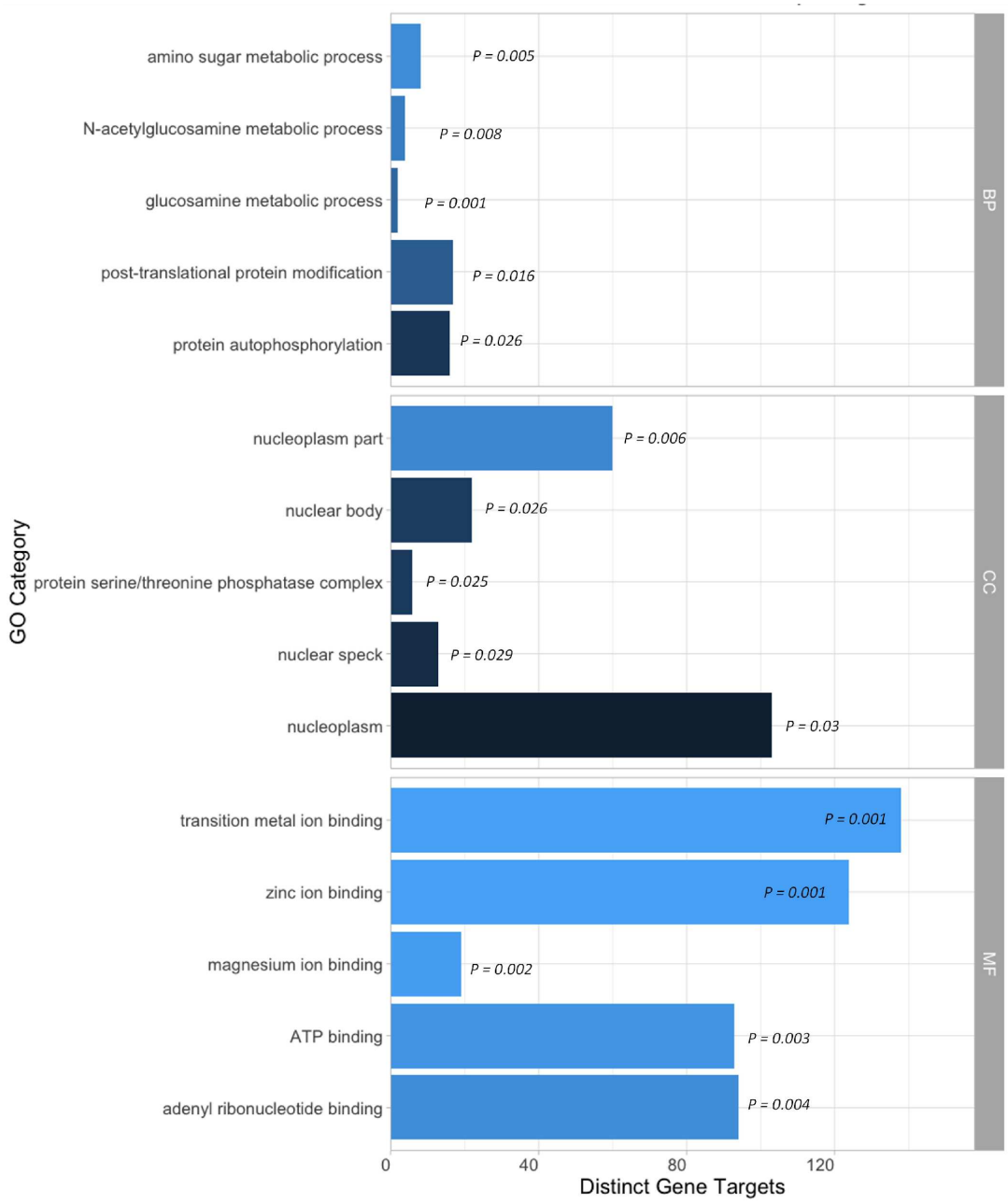


Figure 5.27. Gene Ontology Classification Summary Features and Number of Target Genes for Differentially Expressed miRNAs between Controls and Non-severe Asthmatics*. The two differentially expressed miRNAs between Controls and Non-Severe asthma were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. Only the top 5 categories for each domain are shown (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). For full details of the GO results including lists of the genes see Supplementary (Chapter 5: CvsNS). The bars represent the number of target genes in each Molecular Function category, ordered (top to bottom for each category) from most significant to least significant P-values. Interaction networks permitted computational and sequencing methods as there were insufficient experimental methods of validation for the GO analysis to be conducted.*

From the results of the GO analysis from the DE-miRNAs (N = 2), a Venn diagram was made to show crossover of gene targets among the GO-tested categories (Biological Process [BP], Molecular Function [MF], Cellular Components [CC]) top 5 most significant result categories (Figure 5.28.). Below 1% of the genes identified were shared among all tests (3 genes). MF analysis had the greatest number of unique gene targets (178 genes).

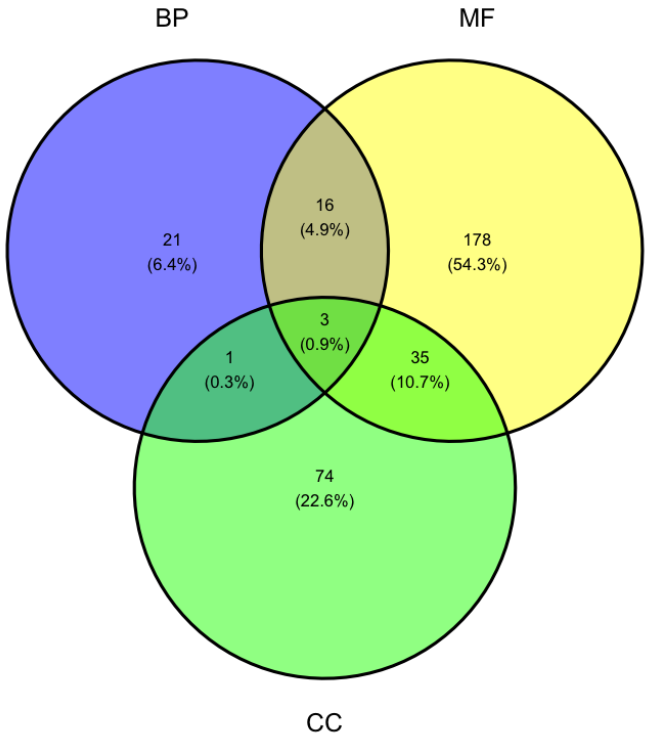


Figure 5.28. Venn Diagram of Target Genes Identified by GO Analysis in Control vs Non-severe Asthma Study. *Significant genes in the Top 5 categories of each GO analysis Category, blood sample miRNA expression. DE-miRNAs (N = 2) were input into miRNet database and passed through GO analysis categories (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). Resulting gene targets were then compared among the three categories.*

All gene targets and categories shown in the Venn Diagram are detailed in Table 5.7. The analysis results were aligned to the source miRNAs relevant to each GO analysis category and compared. Genes that were found to be targets of both DE-miRNAs are in bold and red font in Table 5.7. In the case of the BP category these amounted to 7 genes : *ALG1*, *ERGIC1*, *PIGX*, *ST3GAL1*, *ST3GAL2*, *UGGT1*, *YIPF5.* Between the BP and MF categories there were 8 genes: *CLK4*, *EEF2K*, *LYN*, *NAGK*, *SEC23B*, *SIK2*, *SMG1*, *UBE2G.* In the CC analysis there were 2 genes*: MED7*, *NACC1.* Between the CC and BP categories a single gene*: XPO5 w*hilst between CC and MF there were 2 genes: *GATA6*, *RREB1.* Among the Combined GO categories there were 2 genes: *INO80*, *MCM8.* The largest number of genes (N=18) that both DE-miRNAs targeted, belonged to the MF category: *ABL2*, *GK5*, *KCMF1*, *LIMS1*, *MSRB1*, *NDOR1*, *RAB10*, *RNF19B*, *RN222*, *SCO1*, *TMEM250*, *TUBD1*, *ZFP14*, *ZNF106*, *ZNF347*, *ZNF485*, *ZNF655*, *ZNF841.*

Table 5.7. Target Genes from GO Analysis Top 10 Most Significant (P < 0.05) Category Comparison in Control vs Non-severe Asthma Cohort. *DE-miRNAs (N = 2) identified from blood sample miRNA expression between Healthy Controls vs Non-Severe* asthmatics. Unique genes and genes identified among other GO analysis categories (from other severity comparisons). Genes that were found to be targets of both DE-miRNAs are in bold and red font.

Category	Gene Targets
Unique to Biological Process	***GNPNAT1***, *GFPT1, CHST6, GAS6, MGAT1, ST3GAL2,* ***ALG1***, *ALG14, ARSK, EDEM3, F2, FN3K, GALNT6, PIGG, PIGO,* ***PIGX***, *PTPN14, SFN,* ***ST3GAL1, UGGT1***, *XPO6*
Unique to Molecular Function	*ABL1,* ***ABL2***, *DYRK2, IDE, IRAK4, MAP2K7, PKM, RPS6KA4, ABCB5, ABCC6, ACSL6, ADAMTS4, ADCY2, ADCY9, AGAP9, AK3, ARIH1, ATAD3C, BHMT2, BMPR1A, BMPR2, BMS1, BPTF, CCS, CDADC1, CDKL1, CLPB, CNBP, CTCFL, DDX19B, DDX6, DGKE, DGKH, DNPEP, DSTYK, DTX2, ERAP2, FAM20B, FBXO40, G3BP1, GATAD1,* ***GK5***, *GRIN2B, GRK3, HHIP, HSP90AB1, HSPA4L, IKZF3, ILK,* ***KCMF1***, *KDM2B, KDM5A, KIF3B,* ***LIMS1***, *LMLN, LONP2, MAP4K2, MAT1A, MEX3A,* ***MSRB1***, *MYH14, MYLK3, MYO1F, MYO5B, NEURL1B, NFX1, OLA1, PANK2, PANK3, PARS2, PCGF3, PDLIM3, PFAS, PI4K2B, PIKFYVE, PRMT3, PRPF4B, QRSL1, RBM4B, RFFL, RNF11, RNF14, RNF157, RNF170,* ***RNF19B***, *RNF216,* ***RNF222***, *SEC23B, SLC2A4RG, SLFN12, SMYD4, SP110, SPATA5, STK35, SUV39H1, TAOK1, TARS2, TDRD1, TEP1, TESK2, TIMM10, TOR2A, TRIM44, TRIM65, TRIM66, TRIM72, TRIP13, TSHZ2, UBE2G2, UBE2N, UBE2Z, XRCC2, YME1L1, ZBTB24, ZBTB25, ZBTB7A, ZBTB8A, ZC3H15, ZDHHC22, ZDHHC8,* ***ZFP14,*** *ZFP36, ZFP82, ZKSCAN1, ZNF101,* ***ZNF106***, *ZNF135, ZNF154, ZNF207, ZNF226, ZNF281, ZNF320, ZNF333,* ***ZNF347***, *ZNF384, ZNF417, ZNF419, ZNF426, ZNF430, ZNF431, ZNF439,* ***ZNF485,*** *ZNF487, ZNF546, ZNF551, ZNF554, ZNF556, ZNF561, ZNF562, ZNF566, ZNF573, ZNF652,* ***ZNF655***, *ZNF665, ZNF682, ZNF708, ZNF730, ZNF74, ZNF786, ZNF799, ZNF8, ZNF829, ZNF841, ZNF85, AGMO, ANG, BPNT1, COQ7, CYP27C1, FXN, GTPBP10,* ***NDOR1***, *NUDT3, OGFOD1, P3H2, PRKAR2A, PSPH,* ***SCO1***
Unique to Cellular Components	*AFF2, LUC7L2, PIN1, PRPF4, RSRC1, SETD1A, SF3B1, SRSF2, SUMO1, WAC, CSTF2, GEMIN6, HAND2, LSM10,* ***NACC1***, *PHAX, SENP2, ARNT, ATF7IP, ATXN7L3, CPSF7, DFFA, E2F4, ELP2, ETV3, FANCC, FOXO6, GTF2F1, GTF3C4, GTF3C6, HMGB1, MEAF6,* ***MED7***, *MEN1, NFATC2, ONECUT3, PKNOX1, POLR2D, PPP1R9B, RBL1, RPRD2, SNW1, STRN3, TAF1B, TFDP2, BDP1, CDC14B, DDB1, DUSP7, ESF1, KHSRP, KPNA2, LIN52, MAFF, NGFR, NKD1, ORC6, POLD3, POLD4, PPP4R2, PSMB2, PSMB9, PSME3, PTPN2, REL, SF3A1, SF3B3, SNRPD1, SRF, TIMELESS, TNKS1BP1, UBTF, VHL, WDR12*
All GO analysis categories	*CAMK2A, MRE11, RAD50*
BP and MF shared	*CALR, GNE,* ***NAGK, CLK4, EEF2K,*** *FGFR1, FGFR4, IGF1R,* ***LYN***, *MELK, NLK,* ***SIK2***, ***SMG1***, *STK11, STK4, TRPM7*
MF and CC shared	*RPS6KA5, STK38, WEE1, ABCF1, ACTB, ACTR5, APTX, AR, BRPF1, CDK9, DBF4, FHL2, GATA6, GATAD2A, HINFP, INO80, KAT7, KMT2D, MAPK1, MCM4, MCM8, MTA3, PHC1, PHF12, PIAS2, PTPA, RREB1, RUNX3, SALL2, SRSF7, THRA, USP22, YY1, CYCS, FOXK1*
BP and CC shared	***XPO5***

5.3.5. Literature Comparison and Comparison with Pilot Study

After the sequencing alignment of the Control versus Severe comparison was completed, a comparison with the literature was conducted.

A literature search in the prior Chapter (Chapter 4) had identified multiple DE-miRNAs between healthy controls and severe asthmatics from biopsy samples in the study of Roffel *et al*. 2021. Table 5.8. shows the results of that study as well as the pilot study (Chapter 4) and the present study. The Roffel et al. 2021 study was used for this analysis as there were more DE-miRNAs available for comparison as the pilot would also be compared alongside.

Table 5.8. Table of DE miRNAs Identified in Literature (and the Results of Present Study)

DE-miRNAs identified	Sample size	Sample type	PMID	Reference
miR-99b-5p, miR-181b-5p, miR-141-3p, miR-374b-5p, miR-92a-3p, miR-203a-3p, miR-23b-3p, miR-100-5p, miR-150-5p, miR-199b-5p, miR-139-5p, miR-320a, miR-200c-3p, miR-92b-3p, let-7a-5p, miR-374a-5p, miR-214-3p, miR-1247-5p, let-7d-3p, miR-574-3p,miR-26a-5p, miR-190b, miR-223-3p, miR-200a-3p, let-7b-5p, miR-30d-5p, miR-30a-3p, miR-19b-3p, miR-103a-3p,miR-99a-5p, miR-200b-3p, miR-183-5p, miR-148b-3p, miR-21-5p, miR-195-5p, miR-34c-3p, miR-361-5p, miR-142-5p, let-7e-5p, miR-199b-3p, miR-182-5p,miR-34c-5p, miR-10a-5p, miR-193a-5p, let-7c-5p, miR-10b-5p, miR-652-3p, miR125a-5p, miR-429, miR-20a-5p, miR-486-5p, miR-25-3p, miR-425-5p, miR-185-5p, miR-106b-5p, miR-125b-5p, miR-15b-5p, miR-449a, miR-93-5p, miR-451a, miR-16-5p	79 asthma, 82 Healthy Controls.	Bronchial biopsy	34446467	Roffel *et al.* (2021)
miR-145-5p, miR-223-3p, miR-1304-3p, miR-98-3p, miR-26b-5p, let-7c-5p, miR-98-5p, miR-6753-5p, miR-4750-5p, miR-452-5p, miR-326, let-7a-5p, miR-5698, miR-664b-3p miR-548t-5p miR-4638-5p miR-6827-3p miR-3064-3p, miR-331-3p, let-7d-5p, miR-4655-5p, miR-4775, miR-532-3p, miR-369-3p, miR-182-3p, miR-144-5p, miR-1278, miR-425-5p, miR-374a-5p, miR-339-5p miR-15a-5p miR-32-5p miR-454-3p miR-548ap-5p|miR-548j-5p, miR-296-3p.	24 Healthy Controls, 17 Severe asthma.	Whole-blood	Controls vs Severe BaySeq DE $P < 0.05$ and FC > 0.5	Headley
miR-1304-3p, miR-5698	24 Healthy Controls, 15 Non-Severe asthma.	Whole-blood	Controls vs Non-Severe BaySeq DE $P < 0.05$ and FC > 0.5	Headley
miR-145-5p, let-7c-5p, let-7a-5p, let-7d-5p, miR-486-3p	15 Non-Severe asthma, 17 Severe asthma.	Whole-blood	Non-Severe vs Severe BaySeq DE $P < 0.05$ and FC > 0.5	Headley
miR-1304-3p, miR-4482-3p, miR-181a-5p, miR-378c, miR-374a-3p, miR-1307-5p, miR-24-2-5p, miR-32-5p, miR-224-5p, miR-191-5p , miR-17-3p, miR-378d	9 Healthy Controls, 9 Severe asthma.	Whole-blood	Controls vs Severe BaySeq DE $P < 0.05$ and FC > 0.5	Chapter 4, Section 4.3.3.

A Venn diagram was created to identify the amount of crossover between DE-miRNAs (Table 5.8.) among studies Figure 5.29. There appears to be a low similarity of findings among studies. No DE-miRNAs are shared among all studies. Between the selected published study and the Control vs Severe comparison, 5 DE-miRNAs were shared (let-7a-5p, miR-374a-5p, miR-223-3p, let-7c-5p, miR-425-5p). Between the Roffel *et al.* study and the Non-Severe vs Severe comparison, 2 DE-miRNAs were shared (let-7a-5p, let-7c-5p).

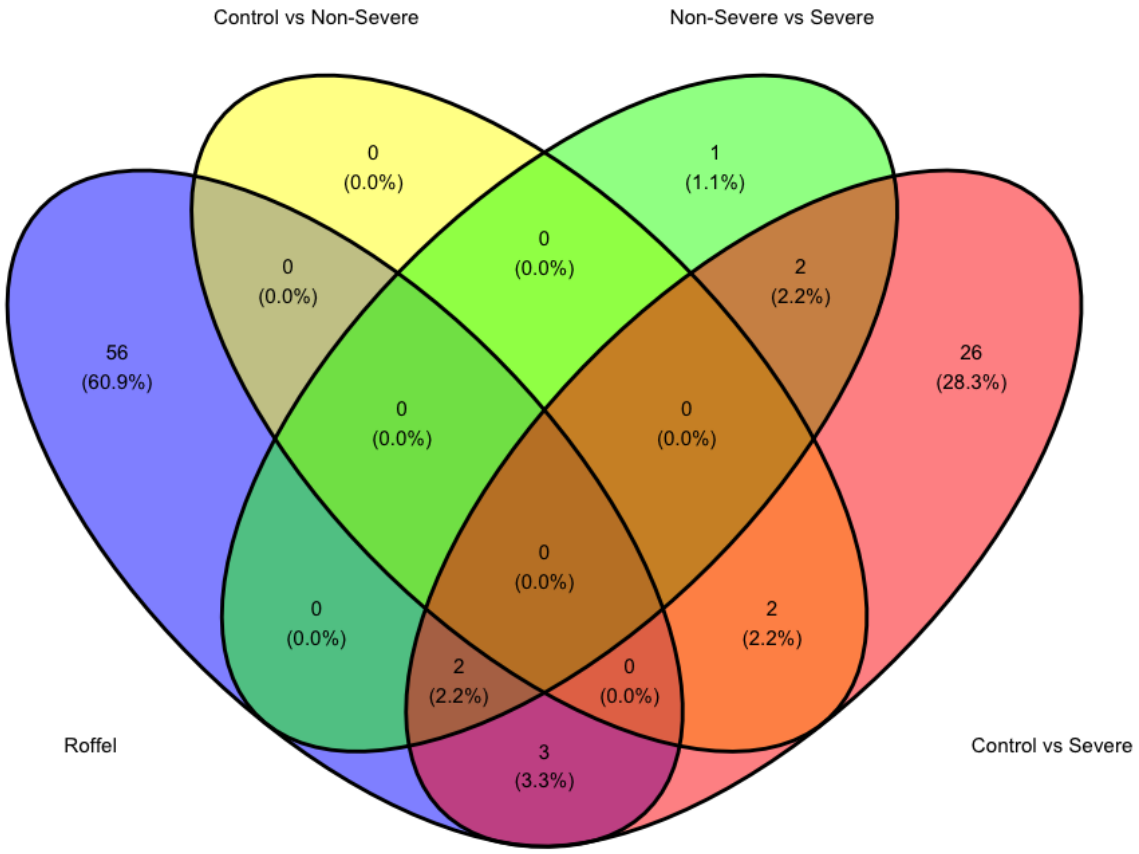


Figure 5.29. Venn Diagram of DE miRNAs Identified between Asthmatics and Healthy Controls in a Published Study and this Experimental Set (Severe Asthma, Non-severe asthma, and Healthy Control). *No miRNA crossover between all studies. Control vs Severe share 3 miRNAs with the Roffel et al. (2021) study. The Non-Severe vs Severe DE analysis shares 2 DE-miRNAs with the Roffel et al. study. And no DE-miRNAs are shared between the Roffel et al. study and the Control vs Non-Severe DE-miRNAs.*

5.3.6. mRNA/miRNA Expression Validation

To establish a protocol for miRNA expression validation through QPCR, it was crucial to identify a stably expressed miRNA across the samples to act as a reference miRNA. An analysis of average expression stability (M value) was conducted on miRNA counts data using the geNorm function in the normalisation package, ctrlGene (Chapter 2, Section 2.1.5.5.). For every sample, aligned data was used and miRNAs were filtered for counts greater than 10. A total of 131 miRNAs had a minimum of 10 reads per sample and were passed through the normalisation. The output from geNorm is a ranking score of most- (smallest M) to least-stable (largest M). The scores were then rounded to 1 decimal place. Figure 5.30. is a graph showing the most stable 3 miRNAs with their respective stability scores followed by the score of the least stable miRNA. miR-92a-3p features as the most stably expressed miRNA with a score of 0.2. As mentioned earlier in this Chapter (Section 3.2.6.3.), miR-92a-3p was the highest expressed miRNA in all samples. miR-130b-5p, and miR-25-3p follow with a stability score of 0.34 and 0.4 respectively.

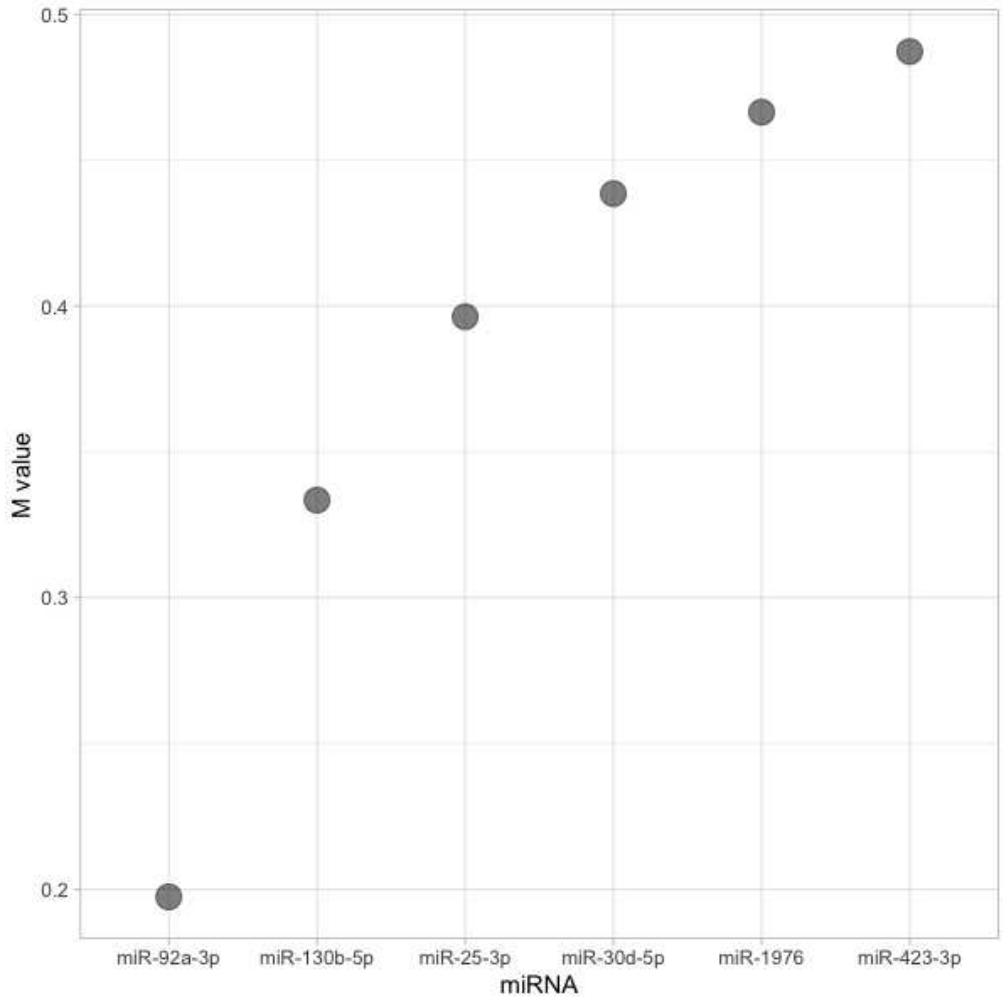


Figure 5.30. Graph of Top Stability Scores from geNorm Analysis*. miR-92a-3p appeared to have the highest stability of expression, followed by miR-130b-5p. miR-423-3p showed the least stability, though well below the recommended published range for consideration for a reference gene.*

An analysis using RefFinder (Xie, Wang and Zhang, 2023), which computes the results of multiple assessments of gene stability, suggested that miR-423-3p might be more stably expressed with a Geomean ranking value of 2, but also confirmed that the stability of miR-92a-3p and miR-25-3p were also highly stably expressed with Geomean ranking values of 2.38 and 2.59 respectively (Figure 5.31.).

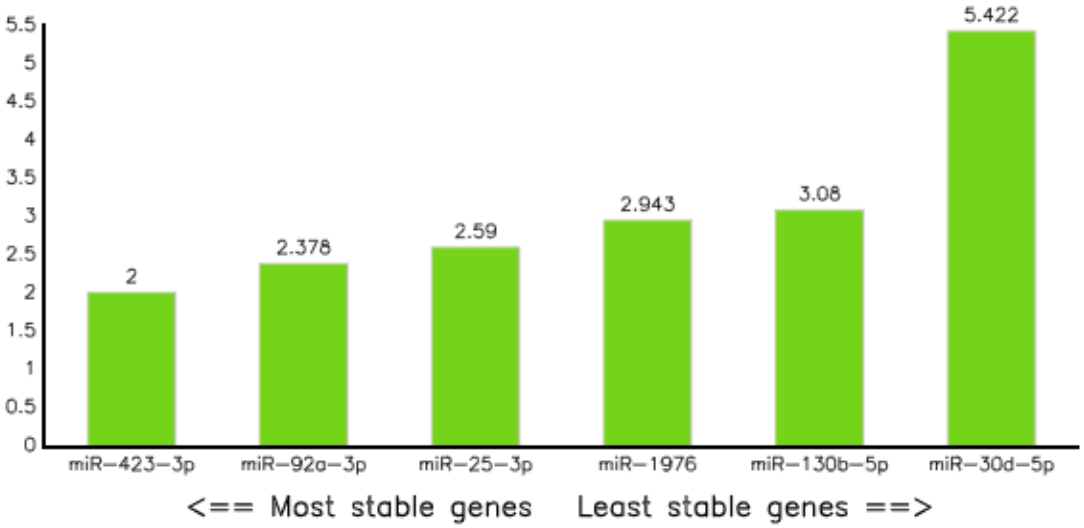


Figure 5.31. Comprehensive Gene Stability Analysis Result*. A combined analysis of gene stability using RefFinder, collates the results of gene stability tests conducted on geNorm, Normfinder, Bestkeeper, and comparative ∆-Ct methods. The y axis corresponds to the geomean of ranking values between all tested methods.*

While the same 6 miRNAs appear in both Figure 5.31. and 5.32. the confidence of the expression stability was not the same. As there is no standard method of reference miRNA identification from sequencing data and the methods above are designed for Δ-Ct results analysis a correlation analysis was used to identify the two most highly correlated miRNAs for use as reference miRNAs (Figure 5.32.). A strong correlation would indicate that the selected miRNAs would be stably expressed alongside one another to use as a pair of reference miRNAs for QPCR. miR-92a-3p and miR-25-3p had the strongest correlation of 0.94. miR-423-3p followed with a correlation of 0.93 with miR-92a-3p. Acceptable stability scores for use as reference genes have been used in the literature of M < 1.5 (Z. Zhang *et al.*, 2020). miR-92a-3p would be a strong candidate for a reference miRNA, as would either or both miR-25-3p and miR-423-3p. It would be recommended to use two reference miRNAs for normalisation of QPCR.

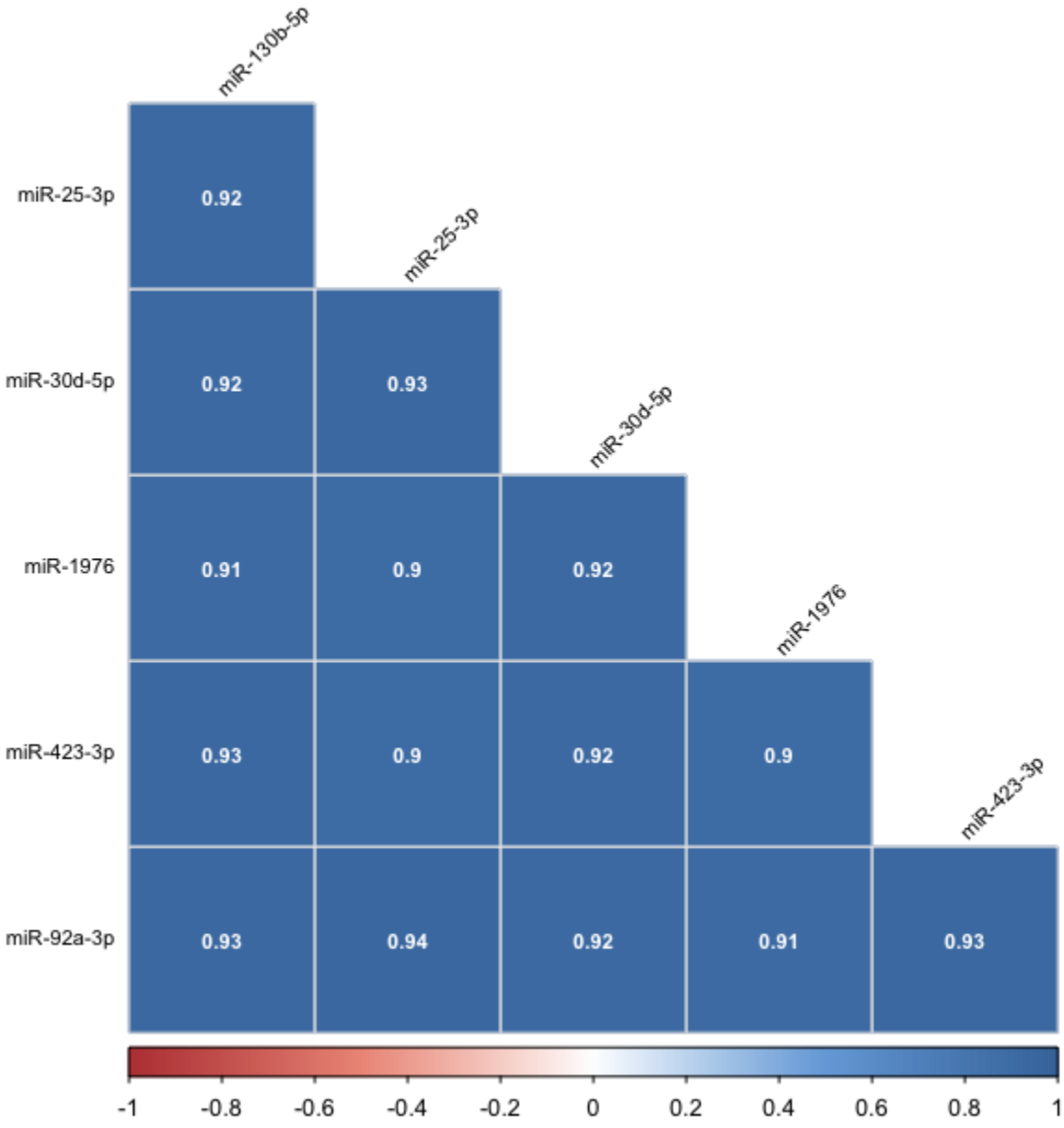


Figure 5.32. miRNA Expression Stability Correlation for the Identification of a Pair of Reference miRNAs for QPCR Validation*. miRNAs were selected based on stability M score < 0.5 and passed through a Pearson correlation. miR-92a-3p and miR-25-3p showed the strongest correlation of 0.94.*

5.4. Discussion

This Chapter has focused on the expression of miRNAs in different severities of asthma comparing mild-moderate asthmatics (Non-Severe) versus those with severe disease as well as healthy non-asthmatic controls. Comparison of several other clinical features revealed differences between test groups total Immunoglobulin E (IgE). In this study, the ratio between forced expiratory volume in one second and forced expiratory volume (FEV1/FVC ratio) was not found to be significantly different between the groups, nor was age, gender or BMI. Correlations between clinical factors were considered next and patient disease severity showed a positive moderate correlation with blood eosinophils (0.48), and weak correlations with total IgE (0.26). There were very slight negative correlations between severity and FEV1/FVC ratio (-0.15). FEV1/FVC ratio showed slight-to-moderate negative correlations with all clinical measures. Overall, this matches expectation as factors that were controlled (as best possible) for such as gender, age, and BMI.

RNA isolation and library synthesis were successfully implemented (Section 5.3.2.) although there were some technical issues regarding the Pippin Prep (key for library size selection), resulting in the replacement of the machine. Not observed on the older version of the machine, the identification of salts in library samples resulted in an amendment of the protocol to add a clean-up step prior to size-selection to remove any salt contaminants (Figure 5.5.). Salt alters the field of apparent size of the DNA, affecting the laser size identification of the Pippin Prep which in turn affects the size of DNA fragments that pass into a separate well for elution (based on the size window input into the software) (Quail *et al.*, 2012). Use of a clean-up kit was required to concentrate samples passed through size-selection on the Pippin Prep.

All libraries that passed QC thresholds underwent sequencing (two lanes) by the Sequencing Facility (Novogene) with minimum of 5 million reads generated per sample library. Immediate downstream processing (Chapter 2, Section 2.1.5.1.) confirmed the sequencing data generated to be of high quality. There were a high percentage of reads for alignment to both the genome (between 82 and 90%) and mature miRNA sense (between 69 and 90%), indicating that the samples, the libraries, and the sequencing were of good quality. Alignments to exogenous and other types of RNA (piRNA, tRNA) were low (< 1%) confirming the protocols employed for the miRNA sequencing were robust.

Elevated eosinophils have been shown to be associated with asthma but it was confirmed that this did not underpin any miRNA read variance between the three subject groups. Sample libraries were split across two lanes ensuring even distribution of the three subject sets across each lane ensuring no batch effect and was planned in the event of sequencing lane failure. The first lane was prepared with the original Pippin prep model prior to the identification of the equipment fault. The newer model of the Pippin Prep had a higher level of sensitivity and consequently required further optimisation. Despite the latter it was noted that the libraries prepared using it had a notably lower read count when sequenced by the Sequence Facility (Novogene) (Figure 5.8., Section 5.3.3.). It is not possible, however, to confirm that the change in Pippin Prep was the cause of the lower read count. Through meticulous experimental methods and QC steps throughout RNA isolation and library synthesis the libraries were expected to be the same high standard as produced in previous chapters (Chapter 3 and Chapter 4). By contrast ongoing issues highlighted at the Sequencing Facility (orphan samples Section 5.3.3) indicates the facility being the cause of lower read count.

Differential expression analysis identified several DE-miRNAs between the different subject groups. Only two DE-miRNAs, miR-5698 and miR-1304-3p, were identified in the Control versus Non-Severe comparison and these two miRNAs were two of the 35 miRNAs identified as DE for the Control versus Severe comparison. To date, there are no publications discussing any understanding of these two miRNAs directly in relation to asthma pathology. miR-5698 has been linked to lung adenocarcinoma and found to target a number of genes that are recognised as having roles in asthma pathology i.e. *PECAM1*, *CX3CR1*, *KLRD1*, and *CXCL12* (Xue *et al.*, 2023). Polymorphisms in the gene encoding *PECAM1* have been reported to be associated with asthma and potential increase the risk of asthma occurrence (Nadi *et al.*, 2012). Therefore, an impact on the expression levels via miRNA interference may also be detrimental to asthma pathology. One study identified that expression of the chemokine receptor, CX3CR1, was attenuated in monocyte subsets from asthmatic subjects suggesting that it was an indicator of severe asthma (Al-Rashoudi *et al.*, 2019), while, by contrast, the gene was linked with atopy, but not asthma, in a German population of children (Depner *et al.*, 2007). This supports the idea that children have modified gene expression compared to adults in healthy and disease states (Resztak *et al.*, 2023). CXCL12 is also associated with asthma and new findings in the CXCR4/CXCL12 signalling axis have shown that topical treatment of CXCL12 neutraligand prevented migration and accumulation of CXCR4 activated pericites from the pulmonary blood vessels into the airway in a murine model of allergic airway disease (Bignold *et al.*, 2022). As miR-5698 and miR-1304-3p were found to associate with asthma irrespective of severity they could be important diagnostically, as much as supporting evidence suggests that modifications of expression could alter inflammatory pathways.

In the Non-Severe versus Severe analysis, 5 miRNAs were found to be DE, four of which were also identified as DE in the Control versus Severe comparison. Let-7a-5p, let-7c-5p, and let-7d-5p were all upregulated between Control versus Severe, whilst miR-145-5p was downregulated but the

complete opposite was seen in the Non-Severe versus Severe comparison. This contradictory differential expression is confusing as it would be expected that if the severity of the disease might impact expression in a linear fashion, then a miRNA's expression would be found to be differentially expressed between multiple groups increasing or decreasing parallel to the disease severity. There are explanations where this could still be true from paradoxical findings in expression during disease progression as discussed by Fessel (Fessel, 2020) - admittedly in Alzheimers research - as well as variant penetrance of genes (miRNAs regulate mRNA expression after all and cannot be completely isolated from their targets own limitations) detailed in the review of (Kingdom and Wright, 2022).

Collectively considering the three comparative analyses, 36 unique miRNAs were identified as DE. A subset of these will be discussed further: five that were identified as DE between different severity comparisons, and two miRNAs that were identified in at least one of the severity analyses conducted in this chapter as well as being DE in the prior pilot study (Chapter 4 – Severe versus Healthy Controls).

Three of the let-7 family of miRNAs, let-7a-5p, let7c-5p, and let-7d-5p, were identified as DE by two of the analyses and were upregulated Control versus Severe (FC values of 1.0, 1.2, and 0.9 respectively) but in contrast were downregulated (FC values of -1.3, -1.6, and -1.1) in the Non-Severe versus Severe comparison. Let-7a-5p has been linked with asthma previously in a similar-sized study of Healthy Controls, mild asthmatics, and severe asthmatics but using QPCR for selected miRNAs (Rijavec *et al.*, 2014). Rijavec *et al*. observed a downregulation of expression in severe asthmatics, contrasting with the findings in this present study although it key to note is the sample type was different bronchial samples as opposed to peripheral blood. Nonetheless, this specific combination of let-7 family of miRNAs has been identified as targeting IL-6 and being

involved in the airway remodelling, through myofibroblast differentiation, that is seen in chronic obstructive pulmonary disease (COPD) (Di *et al.*, 2021). In COPD, a downregulation of expression of these three let-7 miRNAs was found (Di *et al.*, 2021).

Similar to the let-7 findings, miR-145-5p also exhibited contrasting expression between the two comparisons: downregulated expression in the Control versus Severe (FC = -1.3) yet upregulated in the Non-Severe versus Severe (FC = 1.6). miR-145-5p was identified as having a moderate positive correlation (0.49) with disease severity in the Non-Severe versus Severe analysis but was not identified as significant in any other correlation analysis. miR-145-5p has been well researched to date and associations of the potential impact on asthma pathology have been implicated from results in murine models. In a mouse model of airway inflammation, an antagmir of miR-145-5p was able to ameliorate airway hyperresponsiveness, eosinophilic inflammation, mucus hypersecretion, and Th2 cytokine production (Mattes *et al.*, 2009). One miRNA appeared DE between Control versus Severe and Control versus Non-Severe (miR-5689). miR-5689 has not been thoroughly considered in asthma but it has been found to correlate with survival of cancer patients with neuropathic pain in one study (H.-G. Zhang *et al.*, 2020).

Returning to the 36 unique DE-miRNAs identified in the Control vs Severe comparison, analysis for gene targets and Gene Ontology networks revealed some interesting findings (Section 5.3.4.1.). The analysis considered three aspects of gene function: Biological Processes, Molecular Function, and Cellular Components. No modifications in filtering were required. The key findings regarding the categories in which those target genes act, the shared genes between domains (which also may share some of the same gene targets), and the number of source miRNAs that share the same target will be summarised and a focussed discussion around the gene targets that were shared among the domains will follow.

Overall, around 25% of gene targets were shared across all GO categories. The top GO categories for Biological Process included regulation of cellular protein metabolic process, regulation of cell proliferation, and positive regulation of transcription from RNA polymerase II promoter. Cell proliferation and protein metabolic process are supported changes in the literature and as a self-regulatory process miNRA expression begins with transcription by the RNA polymerase II (Picado *et al.*, 1999; Puddicombe *et al.*, 2003). Figure 5.23. highlighted the number of miRNAs targeting genes with *IGF1R* having the most miRNA interactions. IGF1R has been proposed as a diagnostic marker and potential target for allergic asthma (Alfaro-Arnedo *et al.*, 2021). It was found that AHR, airway remodelling, eosinophilia and allergic inflammation, as well as altered pulmonary surfactant expression, were counteracted with an IGF1 inhibitor in a murine model of allergic airway disease.

For Non-Severe vs Severe, from the 5 DE-miRNAs that were input into the analysis, GO analysis (note thresholds were expanded Section 5.3.4.2.) revealed enrichment in the categories regulation of cellular protein metabolism, enzyme binding, and, once again nucleoplasm. Around 22% of genes were shared across GO categories, with *CDKN1A* having the most miRNA interactions. CDKN1A overexpression is known to affect cell proliferation and survival, having the potential to cause abnormal repair responses that contribute to airway inflammation and remodelling (Puddicombe *et al.*, 2003).

For Control vs Non-Severe only 2 DE-miRNAs (miR-1304-3p and miR-5698) were found. Limited GO analysis (selective evidence code filtering was modified Section 5.3.4.3.) could be conducted but the top categories included amino sugar metabolic process, nucleoplasm part and interestingly transition metal ion binding. Amino sugar metabolic process has been studied in relation to different metabolic states and linked with asthma. The underlying process is partly

elucidated, involving lactate in relation to T cell motility that could be involved in driving chronic inflammatory conditions (Haas *et al.*, 2015). Nuclear bodies comprise small membrane-less organelles within the cell nucleus including Cajal bodies, Gemini of Cajal bodies (gems), Nuclear speckles, and PML bodies. Their role within asthma has not been elucidated, however, p80-coilin (a marker protein for Cajal bodies) has been identified as having multiple roles. One key role relevant to miRNA regulation is the phosphorylation and subsequent stabilisation of DGCR8, a key protein involved in miRNA synthesis (Chapter 1, Section 1.3.1.)(Lett *et al.*, 2021).

Transition metal binding has been directly linked with asthma development and pathology, through enhancing mast cell activation, and in pollution causing changes in airway function driven by a Th17/Treg imbalance (Walczak-Drzewiecka, Wyczólkowska and Dastych, 2003; Li *et al.*, 2023; Liu *et al.*, 2023).

Among all severity comparisons, the most significant biological function categories were "regulation of cellular metabolic process" in the Control versus Severe and the Non-Severe versus Severe analysis, while the Control versus Non-Severe was for "amino sugar metabolic process". It could be inferred from these findings that the Control versus Severe and Non-Severe vs Severe are endotypically more alike even from just a few matching miRNAs. Perhaps there are more miRNAs that have similar targets between them promoting cellular metabolic process changes. Amino sugar metabolic process may be the early progression expression pattern of enabling discrimination of non-severe asthma from the healthy airway. This topic has been considered and summarised recently (Qin *et al.*, 2024). One included paper by Qin *et al.*'s publication identified that Treg and Teff cell development could be manipulated through altering glycolitic or lipid oxidative metabolism (Michalek *et al.*, 2011).

For the Cellular Component analysis, the categories focused around the “nucleoplasm”, this time with the “nucleoplasm part” specifically for Control versus Non-Severe (Section 5.3.4.). This could suggest that processes related to gene expression regulation, or nuclear functions might be affected in all tested severities of asthma. The enrichment of similar GO categories could also indicate a gradual progression or common aetiology underlying asthma severity, i.e. the extent to which the genes within the category are affected is the change with greater severity (Zhu *et al.*, 2021). Certain cellular processes or components may be initially affected, contributing to the development of asthma, and further dysregulation of these processes could lead to increased severity. This has been demonstrated previously from the interaction of macrophages in the asthmatic airway on airway remodelling and the interaction that miRNAs can contribute (Shi *et al.*, 2020).

Molecular Function analysis, on the other hand, had distinct categories for all three severity comparisons. For Control versus Severe, positive regulation of transcription (DNA dependent) was the most significant category. For Control versus Non-Severe, the most significant was transition metal binding, whilst for Non-Severe versus Severe it was enzyme binding. These are novel, as microarray analysis have identified more immune specific results for asthma GO analysis previously (Mao *et al.*, 2018). One study identified several T cell regulatory pathway specific categories, as well as immune regulation. This could suggest that sequencing may offer a broader picture of the effects of miRNAs in asthma or alternate pathways leading to inflammation outside of the better studied areas that may have informed the microarray probe selections (Li *et al.*, 2022).

There is a higher confidence in gene interactions from a particular source miRNA among multiple GO domain analyses identifying the same targets due to several factors that will now be

discussed. Firstly, reproducibility across different biological contexts suggests that the interactions are not random and are more likely to be biologically relevant. When multiple independent analyses (different GO categories) identify the same gene targets, it provides a form of cross-validation. Genes that appear in multiple GO categories are likely to be involved in several Biological Processes (Hill, 2023). This ubiquity indicates that the gene is functionally important and its regulation by miRNA is crucial across different pathways (Ju, 2011). Repeated identification of the same gene targets in different GO analyses can increase the statistical confidence in the interactions. The likelihood of observing the same genes by chance in multiple analyses is low, which strengthens the confidence in the miRNA-gene interaction (Ju, 2011). If a gene target is identified in multiple GO categories, it implies that the gene has a broad role in the cellular context and suggests functional consistency. miRNAs that regulate such genes are likely critical in maintaining cellular homeostasis and function. Put together the reproducibility of the findings across domains suggests a higher statistical confidence, forms a method of cross-validation, and helps demonstrate potential functional consistencies.

With the differences in enriched categories identified between Gene Ontology analysis test severities, it suggests that distinct functions may be associated with the varying degrees of disease severity. The enrichment of different Molecular Function categories could indicate that the underlying molecular mechanisms or pathways involved in asthma pathogenesis may differ depending on the disease severity. The differences in enriched categories may reflect a progression or escalation of molecular dysregulation as the disease severity increases. Certain molecular functions or processes may be initially affected in non-severe cases, while additional molecular functions become dysregulated as the disease progresses to a more severe state. For example, during the progression of asthma, as the airway wall hypertrophies and a

proinflammatory environment is sustained, there may be a change in the underlying expression (Pascoe *et al.*, 2017).

There could be pathophysiological mechanisms at play. For instance, "transition metal binding" enrichment between controls and non-severe cases may implicate processes related to oxidative stress or metal homeostasis, while "enzyme binding" between non-severe and severe cases could suggest a more pronounced disruption of enzymatic activities or signalling pathways. Of the gene targets identified in the Non-Severe versus Severe comparison, of those exclusively in the Molecular Function category, only two were identified as being directly related to enzyme products (*CFTR* and *CYP2C19*). *CFTR* is known to be a key gene in cystic fibrosis due to single point mutations removing the function of the chloride ion channel resulting in a thick mucus in the airways. There is heterogeneity in cystic fibrosis, and there is the potential that there is an association with the function of this ion channel in asthma as modified function of the chloride ion channel would affect mucus clearance and viscosity. This link has already been considered and while heterozygosity of a *CFTR* mutation can prevent the full expression of the recessive condition; it was identified that the result depended on multiple factors such as the type of mutation (Maurya, Awasthi and Dixit, 2012). The *CYP2C19* gene codes for a liver enzyme. CYP enzymes are often associated with drug metabolism, and CYP2C19 is associated with a slow metaboliser type, often showing a delayed breakdown of drugs into secondary components. This can result in potential adverse drug responses with subsequent doses. CYP2C19 has been identified as important in a study focused on asthmatic children's responses to an acid reflux treatment (Lansoprazole). Children in the treatment arm who were described as slow metabolisers due to reduced expression of CYP2C19, showed worsened asthma control compared to placebo. This suggests there may be a link with systemic metabolic processes and asthma inflammation control (Lang *et al.*, 2015).

Considering specific gene targets; a total of 11 gene targets were identified among all nine analyses (BP, MF, and CC of each severity comparison Control versus Severe, Control versus Non-Severe and Non-Severe versus Severe). These targets were *FXN, REL*, *ILK*, *PKM*, *DDX6*, *RREB1*, *IGF1R*, *ADCY9*, *MAP4K2*, *DUSP7*, *and DFFA*.

The question however is which are the disease-induced versus the disease-causing changes (Porcu *et al.*, 2021)? The potential benefit of conducting Gene ontology is to see if the results point in an expected direction, as well as highlight any potential avenues that have not been considered (with appropriate validation before pursuing).

Together, the findings of the DE-miRNAs between the severity test groups (Section 5.3.3.1.), the correlations with clinical outcomes (Section 5.3.3.2.), and the resulting differences between the Gene Ontology targets (Section 5.3.4.), suggest that non-severe asthmatics sit on a spectrum much closer to the expression of the healthy controls than towards the severe asthmatics. Perhaps this could provide some opportunity in prevention of disease progression if the underlying processes that push disease progression to a severe state could be understood, as there is potentially less to change. Further work to identify potential miRNA targets could be of use as they are well-understood to tweak expression, rather than apply bold changes in function, which could be ideal for the maintaining of a pharmacologically responsive asthma severity, or even undergo remission.

It is important to note, however, that this interpretation of the connections between the identified DE-miRNAs and their targets genes and functional areas from GO analysis results are putative. This is because GO annotations can be incomplete or biased, and the analysis can be influenced by factors such as the gene set size, the statistical methods used, and the specific algorithms or databases employed.

The results were compared to a published bronchial biopsy study (Roffel *et al.*, 2021) and the prior pilot study (Chapter 4). Some overlap in DE-miRNAs was observed, but also many differences, as visualised using Venn diagrams (Figure 5.29.). Differences were also seen when comparing gene targets for miR-32-5p and miR-1304-3p that were identified in the pilot study (Chapter 4) and replicated in this present study. There were differences in the number of target genes identified in the two studies. The smaller pilot, where 10 DE-miRNAs were identified, miR-32-5p had 6 target genes linked. While in the larger study where there were 33 DE-miRNAs, miR-32-5p was found to be associated with 36 target genes. This is an important consideration of some of the limitations of looking at specific DE-miRNAs in isolation as the computational analysis element has considered the expression of the DE-miRNAs as a whole. There are some reasons for this. Where there is a larger pool of DE-miRNAs, there may be more statistical power to detect miRNA-target interactions due to a larger sample size or increased biological variability (Khatri and Drăghici, 2005). Cross-talk and competition is also a consideration (Prabahar and Natarajan, 2017). When there are more DE-miRNAs there could be increased cross-talk among miRNAs for binding to the target genes, potentially resulting in changes in expression or availability of target genes, resulting in differences in the number of identified targets for a specific miRNA. In this present study of the larger cohort, there was a greater network complexity of miRNAs impacting the overall complexity of the gene targets for each miRNA (data not shown). Target prediction algorithms are not perfect and may generate false positive and false negative predictions. Variability in the occurrence of false positives and false negatives between analyses could contribute to differences in the number of identified target genes. These results need to be validated experimentally to draw firm conclusions.

miRNAs with many gene targets might indicate their involvement in regulating key pathways such as inflammation, airway remodelling, or immune response that are associated with the condition

- in this case asthma. Analysing the expression levels of these miRNAs could help identify dysregulated pathways in asthma patients compared to healthy controls. For example, miR-26b-5p had 127 identified gene targets. Although miR-26b-5p had less than a 2-fold change in expression difference, dysregulation of this miRNA could have wider reaching implications than a DE-miRNA with fewer targets and a higher fold change. An example of this has been shown in cancer research, an approximate 1.5 fold change in let-7 was linked with growth of breast cancer and colon cancer cells and some tumor cell lines (Shell *et al.*, 2007; King *et al.*, 2011; Deshmukh *et al.*, 2021). While in the research this was often induced through a change in a regulatory gene of let-7 expression, as a key regulator of angiogenesis and cell cycle progression, increasing let-7 levels (in cell lines) restored the cells ability to regulate cell growth and reduce tumor growth in mice (Deshmukh *et al.*, 2021).

Similarly, examining the total number of miRNAs that target particular genes, known to be involved in asthma pathophysiology, could provide additional insights into the condition and potential avenues of understanding or treatment. If certain genes associated with asthma are targeted by multiple miRNAs, it suggests a complex regulatory network governing their expression. Dysregulation of these miRNA-gene interactions could contribute to the development or progression of asthma. Therefore, identifying such miRNA-gene networks and understanding their dysregulation in asthma patients could uncover potential biomarkers or therapeutic targets for the disease.

Next, through computational analysis (geoNorm and RefFinder) stably expressed miRNAs for use in validation studies using the gold-standard of QPCR analyses were sought (Chapter 1, Section 1.3.2.) Three miRNAs were identified within the sample set as viable for the use as reference miRNAs (miR-25-3p, miR-423-3p, and miR-92a-3p) (Section 5.3.6.).

In conclusion, the follow-up study conducted in this Chapter has further expanded the current understanding of the differential expression of miRNAs across various severities of asthma. By including participants with severe asthma, non-severe asthma, and healthy controls, a distinct set of DE-miRNAs that exhibit significant differences in expression levels between these groups has been identified.

The findings from this analysis suggest that specific miRNA signatures may be associated with the severity of asthma. The identified DE-miRNAs could potentially serve as valuable biomarkers for distinguishing between severe and mild forms of the disease, as well as differentiating asthmatic individuals from healthy controls.

Further validation and functional characterization of these DE-miRNAs is, however, essential to elucidate their roles in the pathophysiology of asthma and their potential as diagnostic or therapeutic targets. Further validation could include QPCR including as reference the stably expressed miRNAs that have been identified in this study. Additionally, exploring the regulatory mechanisms and target genes of these miRNAs could provide insights into the molecular pathways underlying asthma severity.

-

6. Chapter 6: Circulating and Bronchial miRNA Expression in Eosinophilic Lung Disease

6.1. Introduction

Eosinophilic lung diseases (ELD) encompass a group of respiratory disorders characterized by excessive accumulation of eosinophils, a type of white blood cell, in the lungs and airways (Chapter 1, Section 1.1.2.). This eosinophilic infiltration can lead to chronic inflammation, tissue damage, and respiratory impairment. Eosinophilic lung diseases include conditions such as chronic eosinophilic pneumonia, acute eosinophilic pneumonia, and eosinophilic granulomatosis with polyangiitis (formerly known as Churg-Strauss syndrome) (Chapter 1, Section 1.1.2.1). These diseases can present with various symptoms, including cough, shortness of breath, chest pain, and fatigue, and can range in severity from mild to life-threatening (Chapter 1, Section 1.1.2.2). Despite their clinical significance, the molecular mechanisms underlying the development and progression of ELDs are not fully understood, highlighting the need for further research in this area. In particular, despite being a key diagnostic element, the underlying molecular mechanisms driving eosinophilia and lung pathology in ELD are not well understood.

MicroRNAs (miRNAs), small non-coding RNA molecules that regulate gene expression (Chapter 1, Section 1.3.), have been implicated in various respiratory diseases. Profiling of miRNA expression patterns could provide valuable insights into disease mechanisms and potentially lead to the identification of non-invasive biomarkers for improved diagnosis, monitoring, and personalized

treatment strategies. The role of miRNAs in ELD pathogenesis, however, remains unexplored (Chapter 1, Section 1.5.2).

The present study aimed to investigate the miRNA expression profiles in ELD patients compared with healthy controls, using both whole blood and lung biopsy samples. Circulating miRNAs in blood have been studied as potential biomarkers for various diseases, as they can reflect pathological changes in tissues (Zheng *et al.*, 2011). Lung biopsy samples, on the other hand, provide a direct assessment of miRNA expression in the affected tissue, although obtaining these samples via bronchoscopy is more invasive and not always feasible pending on disease severity and clinical management of patients.

As previously mentioned (Chapter 5, Section 5.4.), post the research conducted in this thesis the potential utility of miRNA expression to distinguish between asthma eosinophilic versus non-eosinophilic has been reported in one published study (Gil-Martínez *et al.*, 2022). To our knowledge, this present study, however, is the first to examine miRNA expression in ELD both in circulating blood and lung tissue biopsies to gain insights into the underlying disease mechanisms and potentially identify accessible (in the case of blood) biomarkers.

The research conducted in the prior chapters (Chapters 4 and 5) had confirmed that sufficient reads were identified using the RNA extraction methods for whole blood for an internal reference (QPCR). An effective next step in development of this process was consideration of an external control added at the library synthesis stage. Spike-in miRNAs are synthetic, non-coding RNA molecules that are added to samples at a known concentration before the library preparation and sequencing steps. These exogenous miRNAs serve as valuable references for several reasons. First, they can account for technical variations that may arise during the experimental workflow, such as differences in RNA extraction efficiency (more often used for QPCR), reverse

transcription, PCR amplification, or sequencing biases. In the case of this study, addition of spike-in was considered at the first point of normalisation i.e. library synthesis. Further details of the background rationale of use of spike-in is detailed in Chapter 1, Section 1.2.2. By normalizing the miRNA expression data to the spike-in miRNA, researchers can correct for potential technical variations and enhance the accuracy of differential expression analysis. The use of spike-in miRNAs allows for the quantification of absolute miRNA expression levels, rather than relying solely on relative abundances, which can be particularly valuable for biomarker discovery and validation studies (Sohel, 2016).

The specific objectives of this study were:

1. To compare miRNA expression profiles between whole blood and lung biopsy samples from ELD patients and Healthy Controls and how these compare with clinical parameters, such as eosinophil counts, IgE levels, and lung function measures.
2. To assess the feasibility of spike-in miRNA as a normalisation tool when conducting miRNA expression profiling by next generation sequencing with low input RNA samples.

6.2. Methods

6.2.1. Study Participants and Study Conduct

The whole blood and lung tissue samples used for this part of the thesis were obtained from participants involved in an investigation of the respiratory microbiome in eosinophilic lung diseases: the BEAM - Bronchoscopic Evaluation of the Eosinophilic Airway Microbiome - study. The BEAM study was conducted as part of the PhD thesis of Claire McBrien (former member of

the Genomic Medicine Group, NHLI - McBrien, 2020). Ethical approval for BEAM was granted by the London-Brent Research Ethics Committee (reference 17/LO/0288). All procedures were conducted in accordance with Good Clinical Practice and the recommendations for medical research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

A variety of sample types were collected as part of the BEAM study for downstream experiments and each participant also completed a questionnaire. Samples for the present study (whole blood and lung tissue) were acquired by venepuncture and bronchoscopy respectively from both subjects with eosinophilic lung diseases (ELDs) as well as healthy sex and aged matched volunteers. All subjects were non-smokers or ex-smokers for a minimum of 12 months prior to sampling with a $\leq$10 pack-years history. ELD patients required physician diagnosis of ELD by specialist respiratory assessment, plus for Chronic Eosinophilic Pneumonia (CEP) CT imaging that was concurrent with characteristic appearances. In addition, all ELD patients had to have an FEV1 $\leq$ 60% predicted. Healthy volunteers were required to be free of significant disease, with no history of asthma/atopy/rhinitis/ELD. Healthy volunteers were also required to be free of inhaled therapies/steroid or other immunotherapies and had to have a pre-bronchodilator FEV1 of $\geq$ 80% predicted (McBrien, 2020). Full inclusion and exclusion criteria are given in Appendix Figure 3. The study comprised 11 ELD subjects and 10 Healthy Controls. ELD patients were permitted to be included if currently taking inhaled steroids, omalizumab, azathioprine and methotrexate.

6.2.2. General Sample Processing

For obtaining peripheral blood RNA, whole blood samples were collected into PAXgene tubes (Chapter 2, Section 2.1.2.3.). Extraction was conducted with the PAXgene blood miRNA extraction protocol as detailed in Chapter 2, Section 2.2.2. As per Chapter 4, two elutions were performed but, on this occasion, they were combined resulting in a final volume of 12 µL. The sample was then stored at -80°C for subsequent quantification and library preparation.

For lung biopsy sampling, subjects were first given conscious sedation with midazolam and fentanyl, and topical anaesthesia with lidocaine was administered to the airway. The bronchoscope was then inserted through the subject's mouth and advanced through the trachea into the bronchial tree.

During the bronchoscopy, to obtain the biopsy specimens the bronchoscope was positioned within the right middle lobe and transbronchial biopsy forceps were passed through the working channel of the bronchoscope and small samples of tissue taken from the bronchial wall. The biopsy samples were placed immediately in RNAlater and stored at -80℃ for subsequent RNA extraction.

Lung biopsy samples were small, and their size ranged from <1 mm to 5 mm in diameter (Figure 6.1.). Lung tissue is composed of fibrous connective tissue supporting the cellular component of the lung and is known therefore of being a more challenging tissue from which to extract nucleic acids. Methods historically used for breaking down tough connective tissues include physical breakdown, e.g. pestle and mortar, and enzymatic breakdown. Therefore, prior to implementing the Zymo RNA extraction protocol (Chapter 2, Section 2.1.2.1.), a tissue disruption step was conducted.

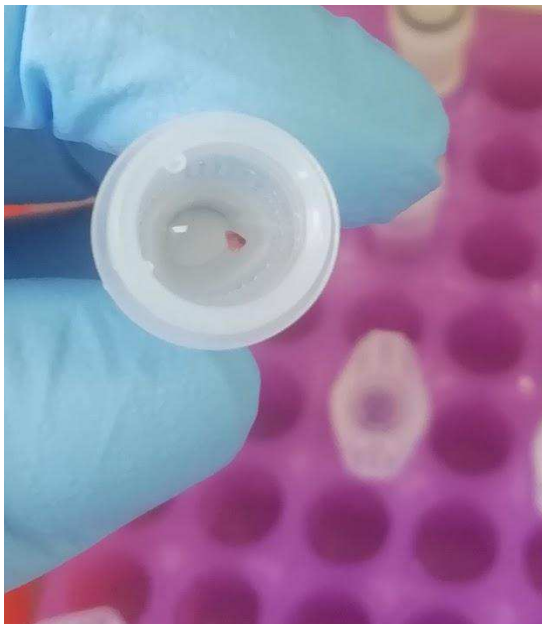


Figure 6.1. Defrosted 2 mL Cryostorage Tube Containing RNAlater and Lung Biopsy Sample (red) Obtained by Bronchoscopy. *Photo taken immediately prior to tissue disruption and subsequent RNA extraction being performed. Photo shows typical size of the BEAM lung biopsy samples.*

First, the lung biopsy sample was taken out of -80°C storage and left to defrost on the bench for 20 minutes. The sample was then transferred briefly into 500 µL of ice-cold RNAse-free water to wash off any remnants of the RNAlater. The tissue sample was then added to a 2 mL Lysing Matrix S, Bead beating tube (MP) containing six (6) 1/8" (3.175 mm) diameter stainless steel beads and 200 µL of a 50:50 mix of DNA/RNA shield and RNAse-free water was added. The tube was placed into a FastPrep-24 5G Bead beater (MP Biomedicals Ltd, California) and processed for 40 seconds at a setting of 6.0 m/s.

Next, the full contents of the tube, were transferred to a sieve filter (Costar) fitted into a 1.5 mL tube and centrifuged on a Labnet Prism microcentrifuge at 600 g for 3 minutes. This was then followed by a wash with 100 µL DNA/RNA shield and centrifugation again at 600 g for 3 minutes. As lung tissue is classed as a 'tough to lyse' sample a Proteinase K (PK) digestion step is

recommended, and reagents are provided within the Zymo RNA extraction kit to enable this. The Costar sieve filter was discarded and 30 µL PK Digestion buffer and 15 µL proteinase K (Zymo Research) were added (without premixing).

After vortexing for 5 seconds, the sample was incubated with shaking at 700 rpm at 55°C for 10 minutes. The sample was then centrifuged (Labnet Prism microcentrifuge) at 16,000 g for 2 minutes to pellet debris. The aqueous supernatant was transferred to a fresh RNAse-free tube, and an equal volume of RNA lysis buffer was added and mixed well by pipetting.

The sample then followed the same treatment as outlined in the Zymo RNA Extraction protocol (Chapter 2, Section 2.2.1.) up to and including the drying of the silica column step. After the latter, the sample was eluted by addition of a volume of 12 µL of RNase-free water directly to the column matrix followed by centrifugation (20,000 g for 1 minute on a Labnet Prism microcentrifuge). The column was reloaded with the eluent and centrifuged again and the flow-through (elution) was then stored at -80°C prior to subsequent quantification and library preparation.

Extracted total RNA samples were converted to small RNA cDNA libraries for sequencing using NEB Next small RNA library synthesis kits (Chapter 2, Section 2.4.) Quantified samples (Chapter 2, Section 2.1.3.) underwent conversion into indexed libraries using the NEBnext® Multiplex Small RNA Library Prep Set for Illumina® (Set 1) following the recommended Illumina protocol. Samples were processed in mixed batches of cases and controls. Blood and biopsy samples were processed separately due to differences in extraction methods. Details of adapter and index sequences are provided in Table 6.1. Indexed cDNA library samples were stored individually at -20°C until undergoing quality checks. Quality checking involved samples being run on the Agilent 2100 High Sensitivity DNA Bioanalyzer microfluidic chip (Agilent® Technologies, Inc.) to determine library quality and size. The expected size range of the Illumina adapter linked cDNAs

generated from the miRNA input is between 147-157 bp. An additional step was added during library synthesis. An exogenous non-human spike-in *Caenorhabditis elegans* miRNA (cel-miR-39 sourced from Norgen Biotek Corp, Canada) was added at the SR adapter ligation step. This procedure was undertaken to facilitate alternative normalization of reads. The rationale behind this was to standardize the concentration of total RNA input, thus ensuring that the subsequent cDNA synthesis amplification would be normalized during the analysis of the generated sequencing data.

The *C. elegans* miRNA (cel-miR-39) was diluted, as per manufacturer's recommendation, to a concentration of 33 fmol/ µL using nuclease free water. Aliquots of 20 µL were then stored at –20°C until required. One µL of spike-in miRNA was added to the first step of library synthesis (the SR adapter ligation step) when the experimental total RNA sample was added.

Table 6.1. Index sequences used for samples in BEAM *(Bronchoscopic Evaluation of the Eosinophilic Airway Microbiome) study (C.N. McBrien.2020).*

BEAM		
Sample number	**Index No.**	**Base Sequence**
1	1	ATCACG
2	2	CGATGT
3	3	TTAGGC
4	4	TGACCA
5	5	ACAGTG
6	6	GCCAAT
7	7	CAGATC
8	8	ACTTGA
9	9	GATCAG
10	10	TAGCTT
11	12	CTTGTA
12	13	AGTCAA
13	15	ATGTCA
14	18	GTCCGC
15	19	GTGAAA
16	20	GTGGCC
17	21	GTTTCG
18	48	TCGGCA
19	42	TAATCG
20	33	CAGGCG
21	36	CCAACA
22	39	CTATAC
23	40	CTCAGA
24	35	CATTTT
25	37	CGGAAT
26	41	GACGAC
27	25	ACTGAT
28	29	CAACTA

For each individual sample library, library size selection was conducted using the Pippin Prep size selector as described previously (Chapter 2, Section 2.5.). As with the Celtic Fire miRNA library preparation (Chapter 5), during the research conducted and detailed in this Chapter the Pippin Prep underwent replacement after a fault was identified with the original machine that could not be repaired by the company. Previously, Chapter 5, Section 5.3.2., size selection was affected by salt contamination identified in samples and an additional clean up step was conducted using the NEB Monarch ® DNA cleanup kit. Consequently, in the present study the BEAM library samples post size selection also underwent the cleanup step as detailed in Chapter 2, Section 2.6.

6.2.3. Sequencing and Analysis

As before, a pooled aliquot of sample libraries at a concentration of 500 pg/ µL as quantified using an Agilent Bioanalyzer HS DNA chip was sent to Novogene (Cambridge). The full library sent for sequencing was comprised of both blood and biopsy samples to be sequenced together on one lane. Sequencing was conducted on a Novaseq, 5 M reads per sample, 150PE (Chapter 2, Figure 2.8.)

The sequencing results were processed as previously detailed in Chapter 2, Section 2.1.5.1. In brief, the data was passed through the ExceRpt pipeline. This pipeline first involves FASTQC (to quality control the raw sequencing data); after which, steps are followed to remove adapter sequences; followed by alignment to the genome and subsequently to miRbase v21 is conducted.

As not all samples had the spike-in added (see Section 6.3.2), the samples that did contain the spike in were deemed the test set. The test samples were normalised as a percentage value based

on the expression of the total of each miRNA. The spike-in miRNA was then used to define a miRNA with a match in expression across the samples within 5% difference of each other, that could be used to normalise across the full dataset.

Differential expression (DE) analysis was conducted on TCC (Tag Count Comparison) in R. Raw alignment data was input into the programme with the corresponding metadata file indicating the group (Case or control). Factors that were considered for this study were differential expression (DE) between ELD patients (case) and Healthy Controls (control), or sample type (blood sample or biopsy). Significance was given to DE miRNAs with a $P < 0.05$.

Target genes analysis was conducted on miRNet as described in Chapter 2, Section 2.1.5.3., which was also used for the identification of Gene Ontology networks. Results were downloaded and further analysed, including visualisation, in R.

6.3. Results

6.3.1. Patient Demographics and Clinical Correlations

Whole blood and lung biopsy samples were available for 11 Eosinophilic Lung Disease patients and 10 Healthy Controls. The mean age of study participants was 51.5 years and there was no statistical difference between the study groups age (Table 6.2.). There were significant differences between all other clinical measures. There were more females in the ELD group (P = 0.02) and Healthy Controls had an overall higher BMI (P = 0.03) compared to the ELD group. For ELDs clinical measures known to be of importance including Eosinophil count and IgE (a marker of allergic ELD) and lung function (FEV1/FVC ratio) all showed as expected statistically significant differences comparing cases to controls: total IgE P <0.001, blood eosinophils P =0.01 and FEV1/FVC ratio P = 0.01.

Table 6.2. Distribution of Patients Across the Two Groups and Key Factors Considered in the Study. *Patients were selected to match as best as possible for age and gender. There were significant differences between groups regarding age. Gender, Total IgE, FEV1/FVC ratio and BMI were identified as having statistically significant differences in a Kruskal-Wallis test or Fisher's exact test. SD = standard deviation. BMI (Body mass index), FEV1 /FVC ratio (forced expiratory volume in 1 minute and forced vital capacity ratio), Eosin (eosinophil count), IgE (total serum immunoglobulin E), Sex (patient gender), Age (patient age in years).*

	Case (ELD) (N = 11)	**Control (N = 10)**
Age: *P* = 0.66		
min	34	31
max	65	62
mean (SD)	51.93 ± 9.71	51.17 ± 8.84
Gender: *P* = 0.02		
Female	8 (73)	3 (33)
Male	3 (27)	7 (67)
Total IgE kU/L: *P* < 0.001		
min	6.04	2.67
max	1006	89.1
mean (SD)	366.19 ± 402.74	37.14 ± 31.42
Blood Eosinophils: *P* = 0.01		
min	0	0
max	1.2	0.3
mean (SD)	0.45 ± 0.41	0.08 ± 0.08
FEV1/FVC ratio: *P* = 0.01		
min	55	69
max	84	86
mean (SD)	69.74 ± 10.29	79.24 ± 5.37
BMI class: *P* = 0.03		
Normal	6 (55)	4 (40)
Overweight	4 (36)	5 (50)
Obese	1 (9)	1 (10)

A correlation of clinical outcomes (Figure 6.2.) was conducted to identify differences between the groups before analysis of the sequencing data. There was no identifiable correlation with patient age and the ELD or Healthy Control groupings. Eosinophils are a key diagnostic element in ELD, and total Eosinophil counts and serum IgE were both identified as positively correlated between the groups (0.49, and 0.57 respectively). Patient gender was identified as having a moderate negative correlation with Patient Group (-0.47), as was patient BMI (-0.36).

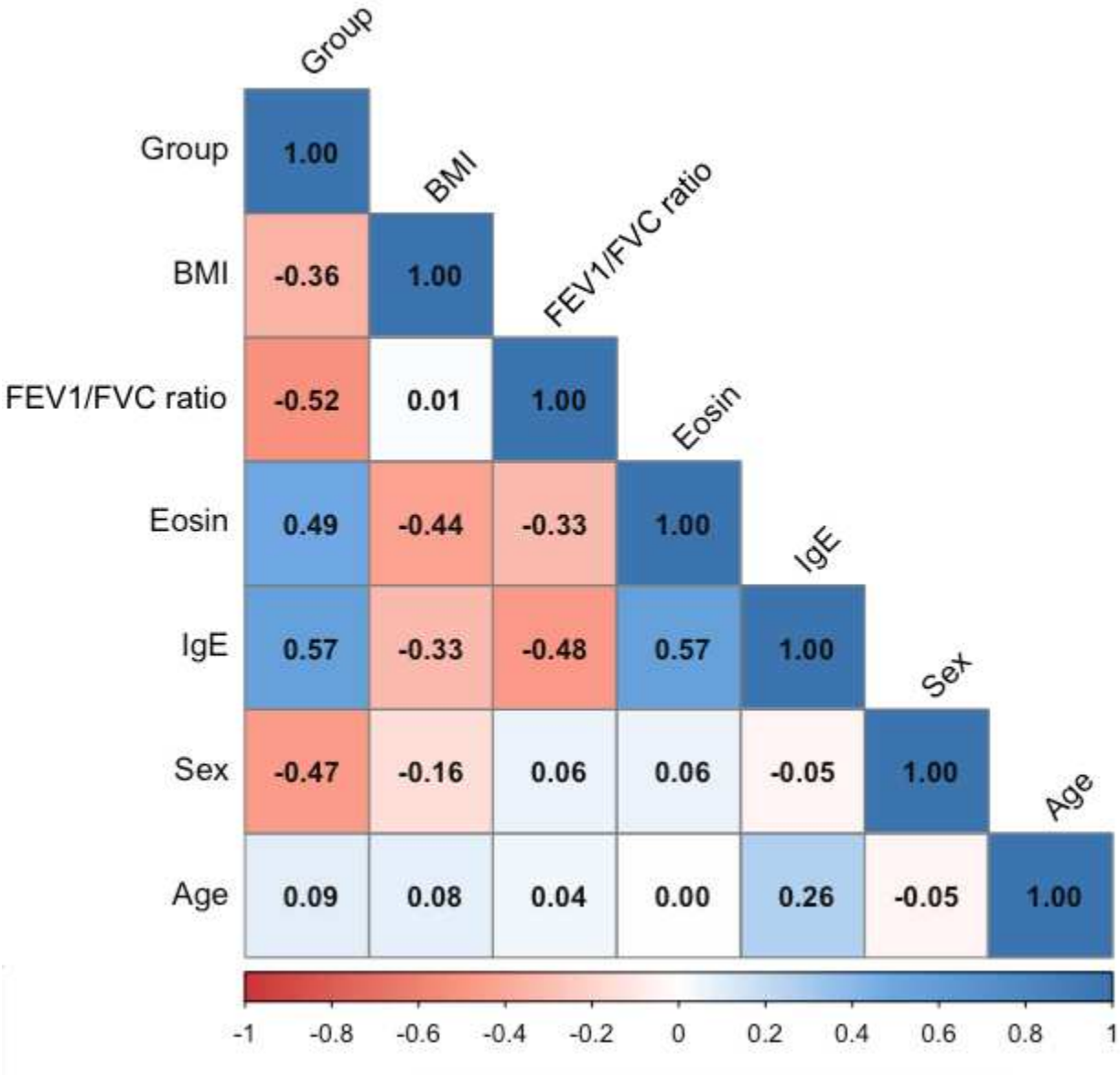


Figure 6.2. Clinical Correlation Table Showing Key Clinical Outcome Measures. *N = 21 (Ns = 11 [Case – ELD] and 10 [Controls]). Clinical measures were converted to factors where necessary and a correlation analysis was conducted. Group (Case, Control), BMI (Body mass index), FEV1FVCratio (FEV1/ FVC ratio, measure of lung function), Eosin (Total blood eosinophils), IgE (serum IgE), Sex (patient gender), Age (patient age in years). As per scale, red indicates negative correlation and blue positive correlation. Moderate positive and negative correlations are observed compared to the Group, namely Eosinophil count and total serum IgE were positively correlated with Patient Group, whilst BMI, the FEV1/FVC ratio, and patient gender showed negative correlations with Patient Group. Age showed little to no correlation with any of the clinical features examined.*

6.3.2. Sample Processing

A feature of the present study was the aim to compare miRNA expression between different tissue types (blood and biopsy) as well as identifying differences in miRNA expression comparing cases of ELD vs Healthy Controls. Each patient provided blood and biopsy samples that were then processed to extract RNA. Methods for extracting RNA from biopsies for sequencing underwent optimisation (data not shown) as lung tissue is considered a difficult tissue from which to extract RNA due to varying levels of connective tissues supporting the cellular content. Bead beating in an MP bead beater was used to disrupt connective tissue and access cellular material including RNA. The prior established RNA extraction method was then followed (Section 6.2.2.).

A total of 100 ng in up to 6 μL of total RNA was needed for input into the library synthesis protocol. With one exception, all extracted RNA samples (both from whole blood and lung biopsies) met this requirement and were put through the library synthesis protocol. A comparison of the RNA yields was conducted to see if there was a difference in total RNA acquired from different sample types

In most samples, sufficient RNA was available to meet the input range of 100 ng total RNA (in 6 μL) for small RNA library synthesis using the NEB small RNA library kit. Two samples (both blood samples, one from a Control the other from a Case [ELD]) that were within 5% of the minimum recommended input RNA concentration were included. RNA concentrations (quantified by Qubit 4, or HS Bioanalyzer) showed ranges of total RNA of 12.5-154 ng/μL.

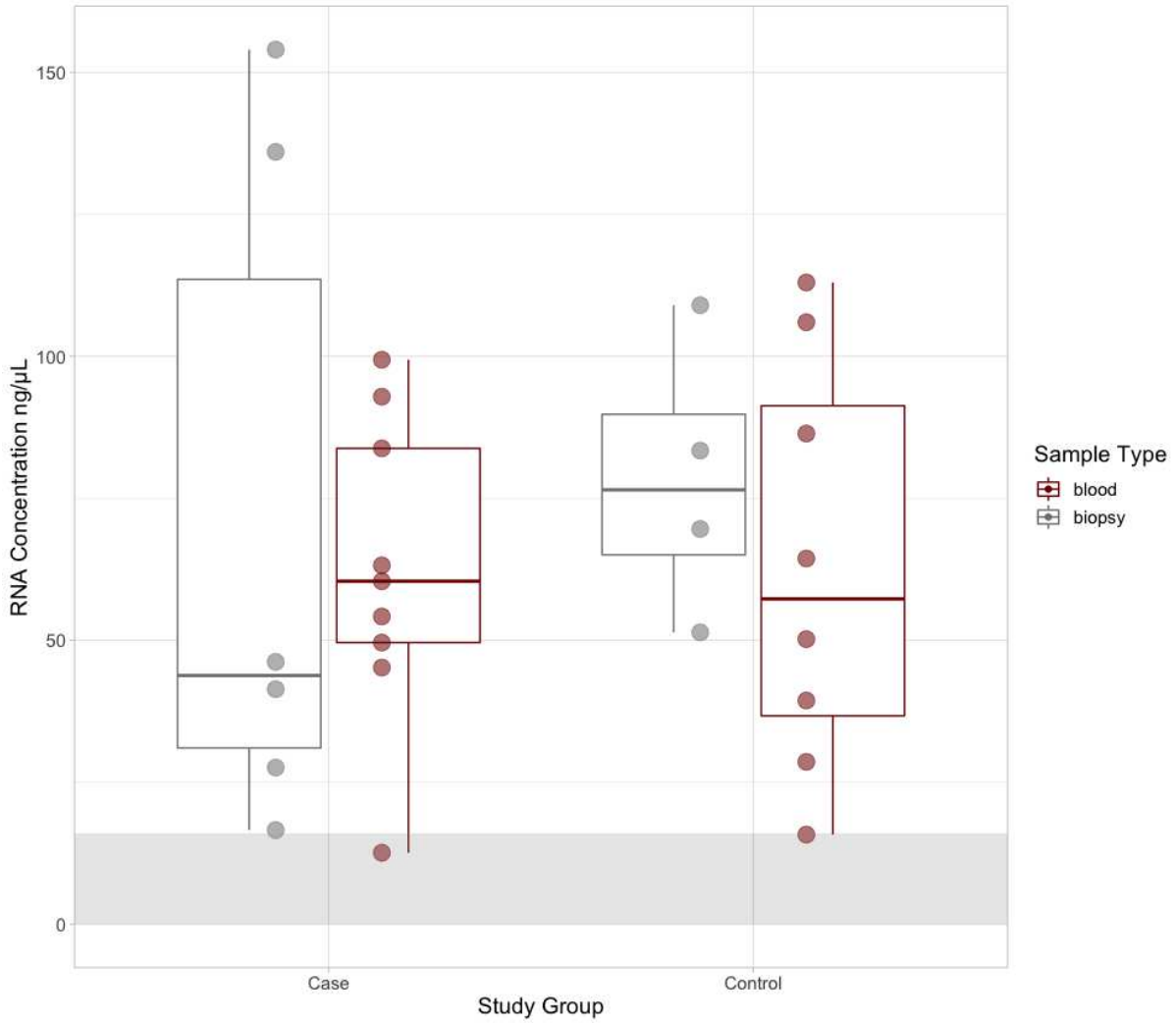


Figure 6.3. Blood and Biopsy Total RNA Concentrations from ELD and Healthy Control Subjects. *Grey bar highlights samples below minimum input concentration for cDNA library synthesis (this includes both points sitting between the white and grey demarcation. One sample (within the grey bar) was excluded from library synthesis - a blood sample from the ELD group.*

The RNA samples were converted into cDNA libraries in batches ensuring as best as possible that each batch had an even distribution of samples from cases or controls and different genders. As an additional step, a spike-in exogenous miRNA (cel-miR-39) was added at a consistent concentration to enable later normalisation (Section 6.3.2). The input concentration of spike-in had been previously established and published in the literature (Williams *et al.*, 2013).

The present study was conducted parallel to one of the experiments in Chapter 5 (preparation of libraries for Run 2), resulting in similar protocol modifications (a piece of equipment being replaced [Pippin Prep] due to a technical fault in the machine). The new machine required some optimisation and modifications in range usage. Firstly, all samples required a clean-up step to remove potential contaminants that, while not identified in the previous experiments, were being identified in samples attempted on the new Pippin Prep. Previously (Chapter 4) this step was only conducted on samples that required further concentrating to comply with the Sequencing Facility's minimum sequencing input concentration. Secondly, the range within which samples were electrophoretically isolated was modified as in applying the original range optimised on the previous machine it was found that adapter dimers remained present. The majority of samples underwent the clean-up protocol for this Chapter's experiment. Due to low volumes of input total RNA into the library synthesis protocol, samples that had been successfully converted into libraries were not repeated when the Pippin Prep was changed. Due to sample loss prior to the Pippin Prep replacement (where samples were run through the machine, or the machine failed to isolate a lane) it was decided to abstain from adding the spike-in miRNA to any further attempts at the same sample library. Further details of the samples (including total number) with spike-in are detailed during the sequencing results.

Library synthesis was attempted for a total of 42 samples. Twenty-seven samples met QC criteria after synthesis and size selection and Figure 6.4. shows these, the sample type (biopsy or blood) and status (case or control). The final pool of 27 sample libraries consisted of 6 Case (ELD) blood – biopsy paired samples, an additional 3 Case (ELD) blood samples whilst for the Control group, libraries were successfully synthesised for 4 blood-biopsy pairs in addition to 4 blood samples.

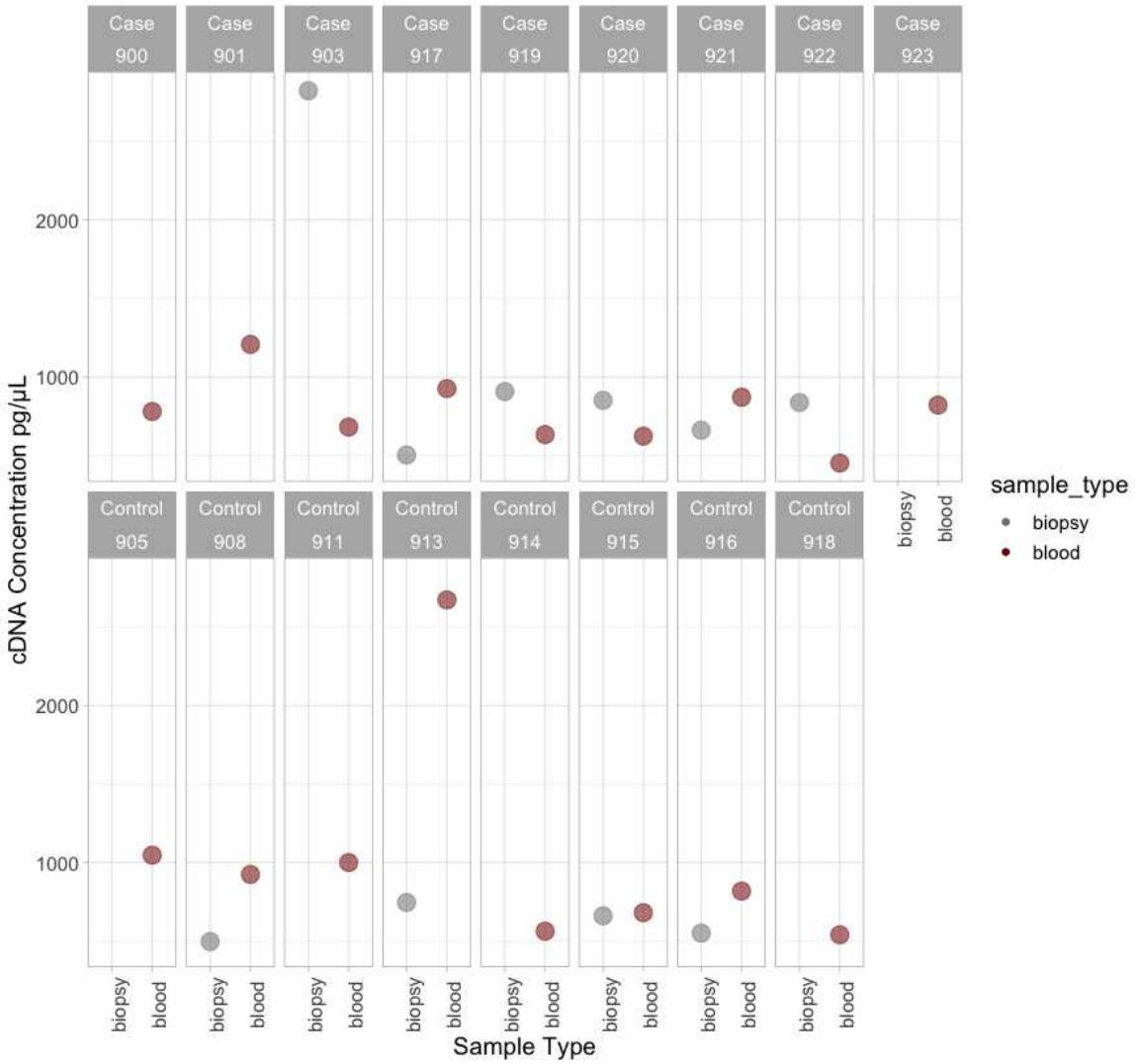


Figure 6.4. Overview of Individual Subject cDNA Library Information. *Results of individual biopsy and blood cDNA libraries that met the Sequencing Facility (Novogene) input requirements on Bioanalyzer traces. Each subject had a cDNA library prepared for sequencing from one biopsy sample and one blood sample. Those that met the sequencing criteria and were pooled together are shown above (N = 27).*

6.3.3. Sequencing Analysis

The final library (pool of 27 sample libraries) was approved by Novogene (the Sequencing Facility) as it passed the QC requirements off-site as well as meeting the in-house QC (Chapter 2, Section 2.1.4.2.. Post the library sequencing and data transfer, correspondence with Novogene highlighted that there was an unexpected lack of sequencing output from the ELD BEAM study libraries. Previous studies conducted with the same protocols, equipment and reagents had resulted in greater read counts per library.

In 9 samples, fewer than 20 total reads were identified. Twelve samples had reads between 80 and 4.5 M while 5 samples had reads greater than the 5M requested for each sample. One sequenced sample from a healthy control blood sample had just over 10M reads. Figure 6.5. provides an overview of the total reads identified across the groups and sample types.

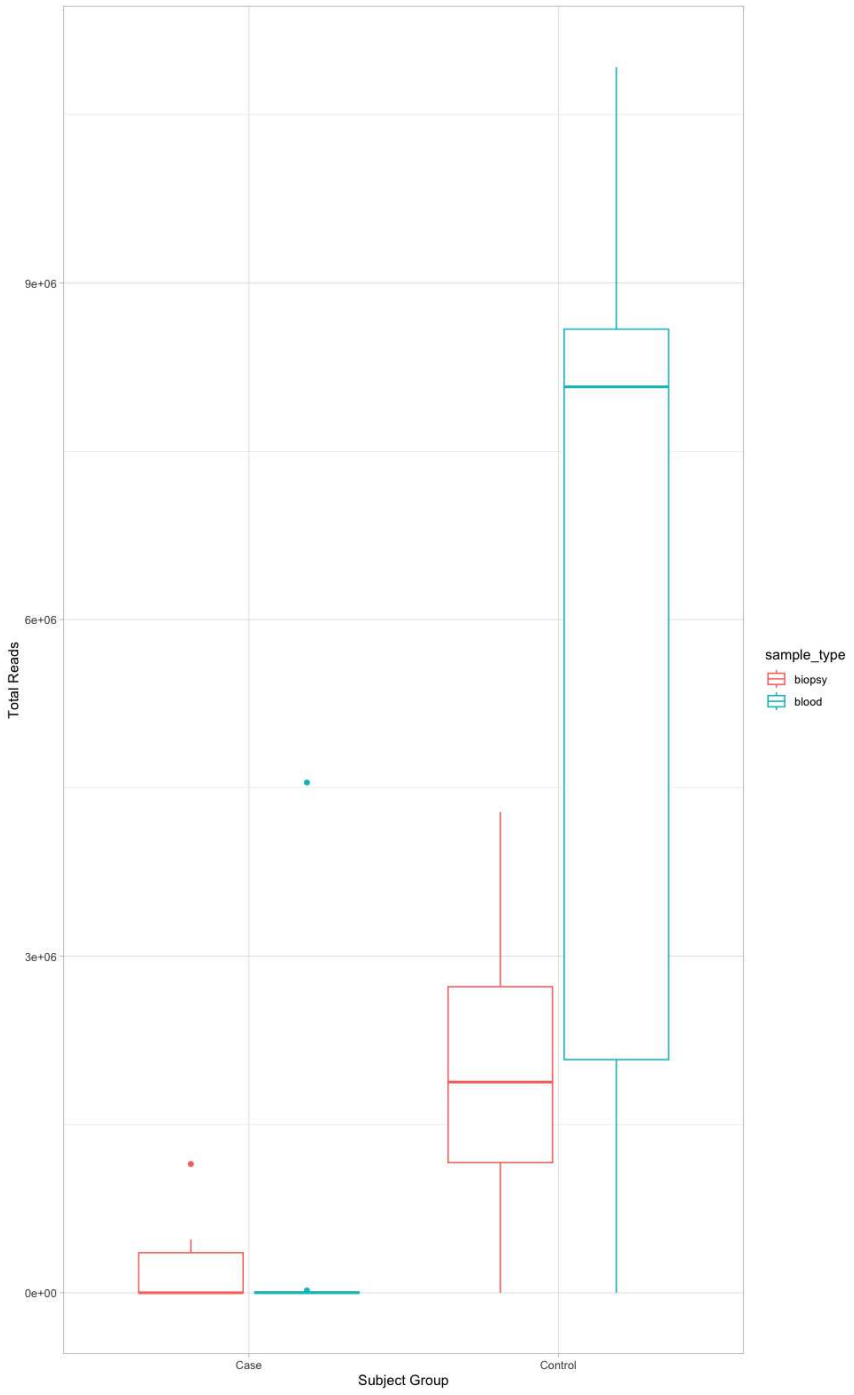


Figure 6.5. Boxplot of Raw Reads Sequencing Results from ELD vs Healthy Control Whole Blood and Biopsy Samples. *A total of 27 samples were sequenced (4 biopsy Controls, 8 blood Controls, 6 biopsy Cases [ELD], 9 blood Cases [ELD]). Read counts ranged from 0 to > 10,000,000.*

To identify areas that may have impacted the success of sample sequencing outcomes, an analysis of sample processing and QC was conducted. Adapters are key elements added at library synthesis to enable binding to the flow cell and to enable the strand to be identified and sequenced. Identification of adapter sequences in sequencing results is one method of ascertaining which samples were not successfully sequenced. These are different from adapter-dimers, which are dimers of themselves and can interfere with sequencing.

Figure 6.6. shows a dot plot of RNA and cDNA concentrations coloured by whether adapters were identified to be present or absent. There appeared to be no pattern of adaptor identification in relation to either the RNA or cDNA library concentrations. As sample input concentration into the cDNA library was normalised it should not be expected to see any linear relationship with RNA concentration and final cDNA concentration. QC steps measuring concentration would ensure sufficient RNA was being input into the library synthesis protocol (to ensure adapter binding).

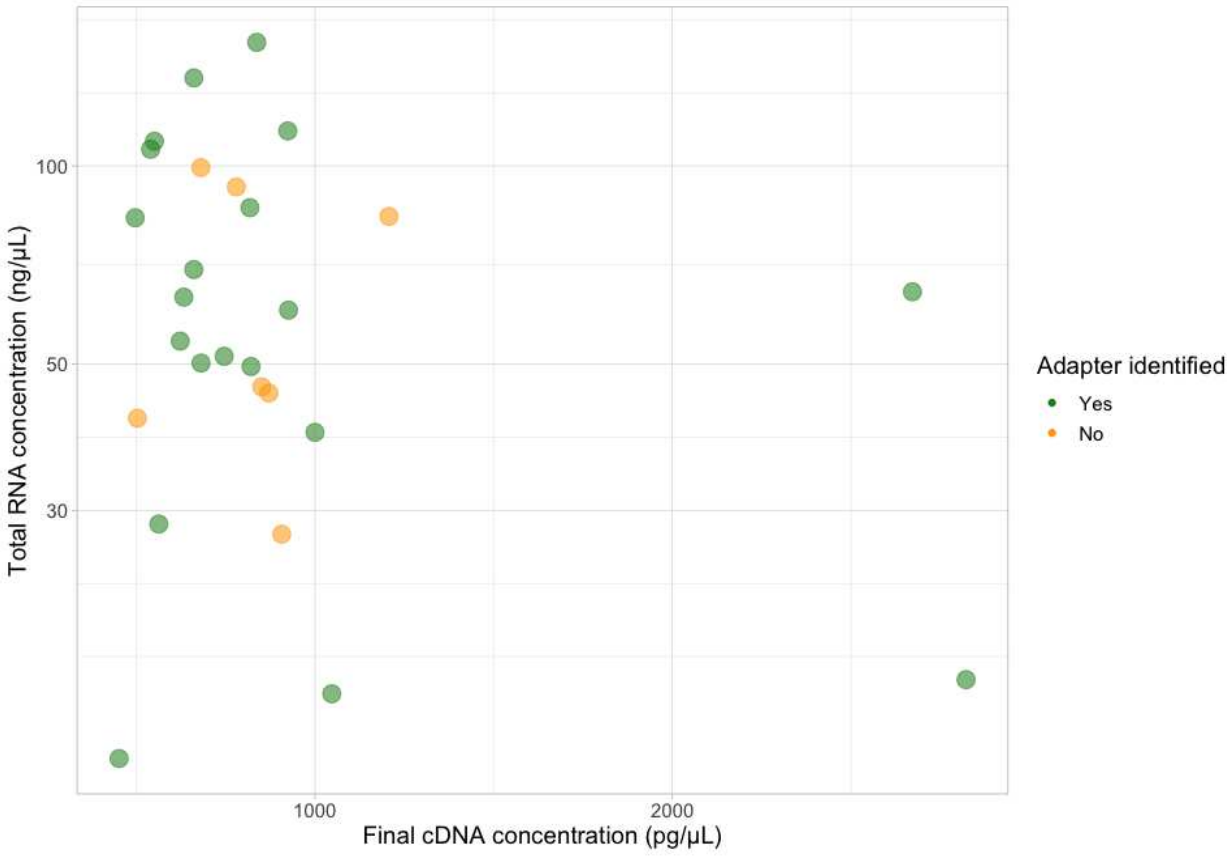


Figure 6.6. Individual Sample Total RNA and Final cDNA Library Concentrations. *Coloured by adapter identification at initial raw read alignment step. Seven samples did not have adapter identified and were not able to be further analysed. All remaining samples were further analysed, including the two samples that had a predicted low chance of success as they did not meet the initial library synthesis criteria. As sample input concentration into the cDNA library was normalised it should not be expected to see any linear relationship with RNA concentration and final cDNA concentration.*

The samples that did not have adapter sequence identified from the sequencing reads output were analysed for day-of processing (to see if there was a particular processing day where adapters were "missed" during library synthesis). Figure 6.7. shows example traces of Agilent Bioanalyzer High Sensitivity DNA chips, both with similar clean peaks and concentrations, and highlighted in the accompanying dot plot the successful identification of one library's adapters while the other did not. This emphasises that as best possible the quality of the samples sent for sequencing was consistent and met the sequencing facility input quality requirements.

It is possible that settings on the sequencing machine at Novogene may have been altered for another customer run and had not been fully adjusted for this run. This may also have been the scenario of Run 2 of the larger Celtic Fire Study of Chapter 5 (Section 5.3.3.) that had also have a paucity of sequence reads. This alternative suggestion was identified during a conversation with the manager of the Imperial College Biomedical Research Centre Genomics Facility, Dr Nik Matthews, as Novogene were not willing to engage or discuss further potential errors that may have occurred at their facility.

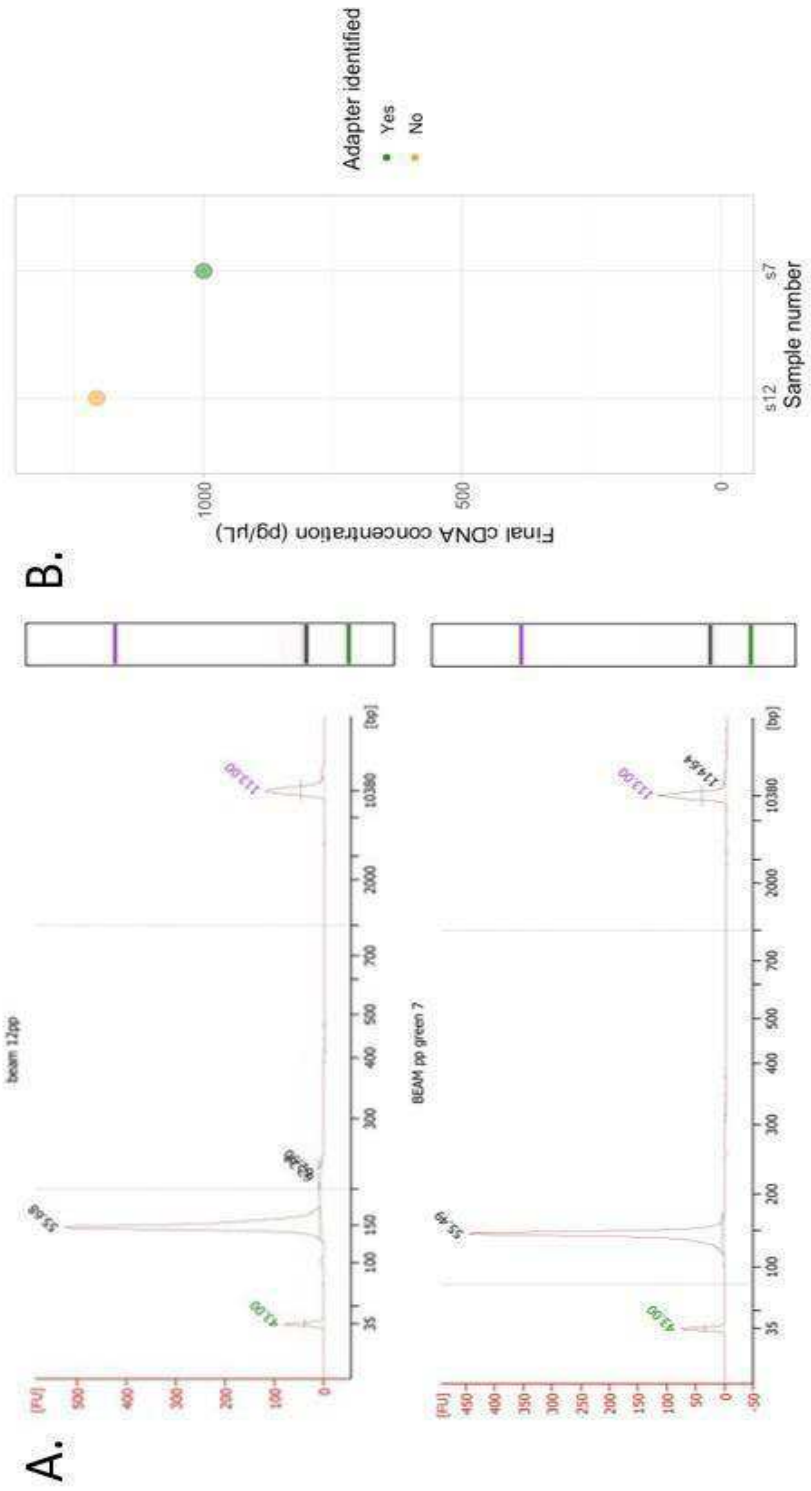


Figure 6.7. Alignment of Sequence Data From Two Libraries to Establish Presence of Adapter. *A. Example of two samples cDNA library (post-size selection) Bioanalyzer trace sample 12 above, and sample 7 below, and B. Outcome of sequencing adapter was not identified in sample 12.*

Of the samples that were positive for adapter identification and could therefore be aligned to the genome, an analysis of read depth was conducted. Due to being sequenced at a similar time as the samples in the previous Chapter (where an issue with low read counts was also identified for Run 2 – Chapter 5, Section 5.3.3.) there was an expectation of possible low read counts in this study.

Paired samples (blood and biopsy) remained only for three subjects in the Control group. Disappointingly, no pairs for within subject comparison in the ELD group (Case) made it through sequencing and alignment. The analysis conducted, therefore, focused on comparison of miRNA expression of Case (ELD) vs Control for the sample types with reads above 1,000 (Table 6.3.). A very significant difference ($P < 0.0001$ Wilcox-signed rank test) for the total read distribution (per miRNA identified) between the runs (Figure 6.8.).

Table 6.3. Total Aligned Reads for Each Library by Sample Type and by Group*. Study_IDs are unique sample code reference for identification of subject to identify sample type pairs from shared donor. All data was anonymised for most clinical outcome measure, only sample type and group were identifiable to enable a fair spread of samples between experiment steps. Aligned reads shows reads post Genboree alignment.*

Study ID	Group	Sample type	Aligned Reads
900	ELD	Blood	0
901	ELD	Blood	0
903	ELD	Blood	0
903	ELD	Biopsy	475,472
905	Control	Blood	10,924,676
908	Control	Blood	21
908	Control	Biopsy	1,546,773
911	Control	Blood	8,601,888
913	Control	Blood	8,584,285
913	Control	Biopsy	0
914	Control	Blood	7,927,633
915	Control	Blood	556
915	Control	Biopsy	4,286,221
916	Control	Blood	8,221,169
916	Control	Biopsy	2,208,342
917	ELD	Blood	4,548,051
917	ELD	Biopsy	0
918	Control	Blood	2,769,809
919	ELD	Blood	85
919	ELD	Biopsy	0
920	ELD	Blood	6,666
920	ELD	Biopsy	0
921	ELD	Blood	0
921	ELD	Biopsy	1,147,526
922	ELD	Blood	20,331
922	ELD	Biopsy	2
923	ELD	Blood	463

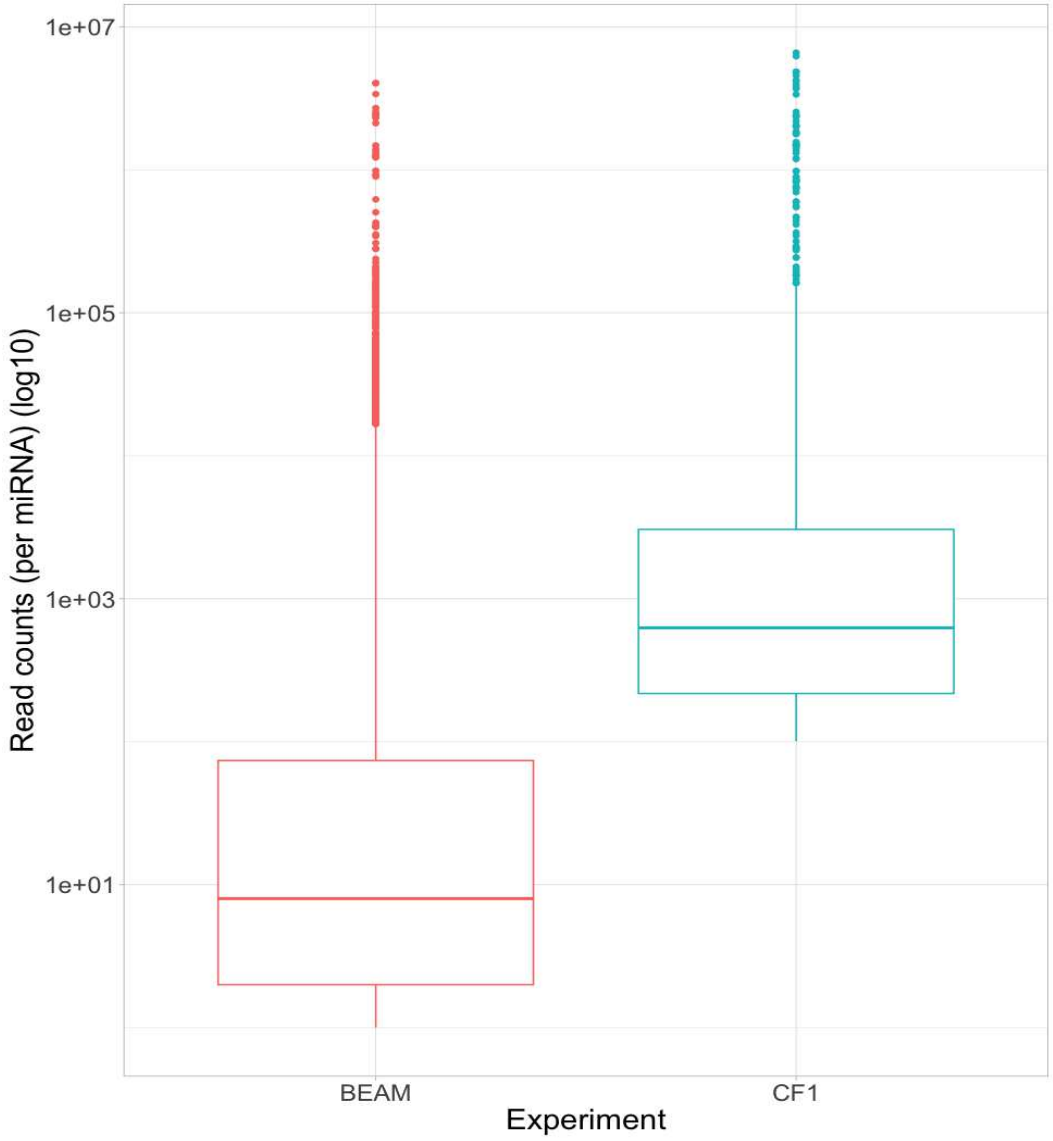


Figure 6.8. Total Aligned miRNA Counts: Comparison of ELD vs Healthy controls (BEAM) Compared to Severe Asthma vs Healthy Controls (CF1). *Comparison of miRNAs successfuly obtained from this thesis (BEAM) with the exemplar run from Chapter 4. Box centre line marks the median reads, with the outer top and bottom demarcating the interquartile range. Outliers are marked with dots. Wilcox Signed rank test identified a significant difference between the groups (P < 0.0001).*

As well as the total number, the quality of the reads was also significantly impacted. Fewer reads passed quality controls, and fewer were aligned (Table 6.3) to the miRNA databases. Figure 6.9 highlights the differences with regard to successful alignment of individual samples in the BEAM study versus those of Chapter 4 – the Celtic Fire (CF1) Study Severe asthma vs Healthy Controls (Chapter 4, Section 4.3.1.4). There were no notable alignments failing the quality filter in the pilot Severe Asthma vs Healthy Control study. This is in contrast to the BEAM samples, of this present study, where there were a notable number of sequencing reads that failed the quality filter. As a result, fewer samples remained for alignment to miRNA and the human genome. In addition, samples where the adapter was not identified and therefore not clipped had to be excluded from further analysis. Five samples (S01, S05, S33, S39, and S40) in the BEAM study were not mapped to the genome (contained < 5% of reads). The remaining samples had between 7-93% of reads aligning to the genome, this is notably lower than the 92-97% of aligned reads to the genome seen in the Celtic Fire pilot study in Chapter 4. Six samples (S15 in addition to those mentioned above) aligned <5% of reads to other defined libraries (miRNAs, for example). All BEAM samples contained exogenous rRNA (between 2% and 58% of all reads). The extent of this was considerable compared to that for the CF1 study.

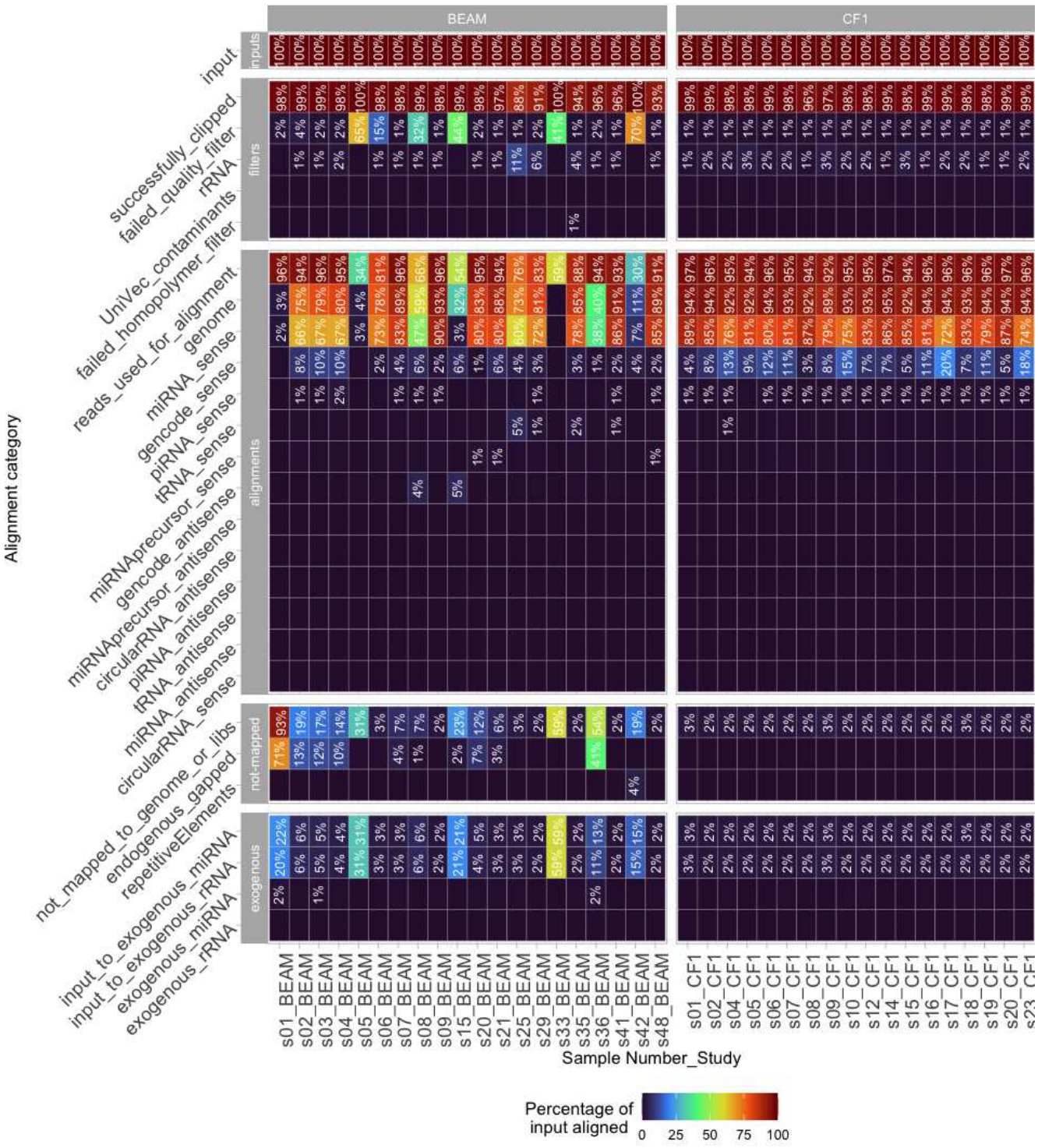


Figure 6.9. Comparison of Read Count from BEAM and CF1*. Total Aligned reads for each experiment set. CF: Celtic Fire, Severe asthma and Healthy Controls (N = 18 - Chapter 4), and for this experiment (BEAM: Eosinophilic Lung Disease vs Healthy Controls). Data is normalised to input reads for the highest reads. Only samples with successful adapter identification (clipping) included (BEAM, N = 20).*

It is well understood that bacterial species are identifiable within the airways and it therefore might be expected to see some alignment to exogenous genomes (Moffatt and Cookson, 2017) since lung biopsy samples were included in the present study. To understand whether the tissue type contributed potentially to this exogenous genome contamination or was the source of the poor-quality reads was, therefore, explored. Only samples where the adapter was identified (clipped/trimmed for alignment) were included (N = 20). Of those samples between 88-100% of reads were clipped. Of the total input between 1-70% of reads failed the quality filters. Between 30-96% of sample reads were put through into the alignment, with <1-91% of reads aligning to the genome. MiRNA alignment also had a broad range of alignments with <1-90% alignments across samples.

Figure 6.10., shows the alignment details for the BEAM scaled alignment with details of Group status, sample type, and spike-in. T-tests were conducted to compare blood and biopsy samples at different alignment and filtration steps to see if the apparent differences were significantly different between the two sample types (Table 6.4.). There were statistically significant differences between contaminants and exogenous alignments, as well as differences between genome alignments and miRNA alignments. There were statistically higher contaminant hits and genome misalignment in the biopsy samples compared to the blood samples. Due to the spike-in of *C. elegans* miRNA, as expected some samples contained exogenous miRNA. Highlighting that while there was a significant difference in the quality of the sequencing compared to the previously reported runs in Chapter 4 and 5, there was also a significant disparity between the tissue types.

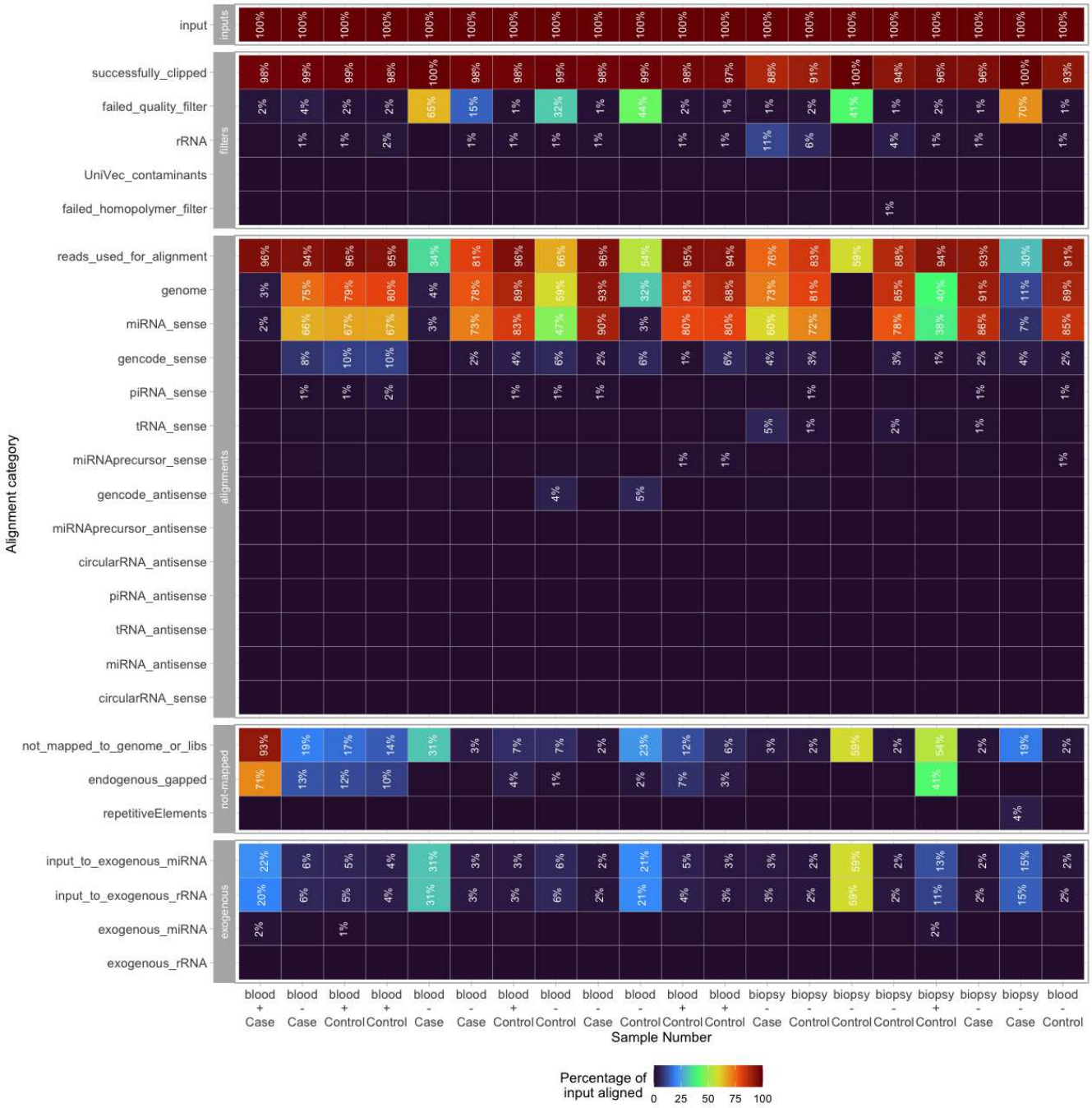


Figure 6.10. Scaled Alignment of Reads. *Comparison of read count from BEAM blood and biopsy samples are highlighted. Total Aligned reads for each blood or biopsy samples prior to selection for differential expression analysis to understand if tissue type had an impact on the outcome of the sequencing methods. Only samples where adapter was identified for clipping are included. Reads are scaled to input. Samples treated with C. elegans miRNA spike-in denoted with a +, or those without a -. Percentages shown are between 1-100%, below 1% are left blank. A broad range of percentage of reads (6-91%) successfully passed through the filters and aligned to the genome (>= 89%) and to sense miRNA (<1-91%). Between 1-91% of reads aligned to contaminants or failed the quality filters.*

Table 6.4. P Values of T Tests Comparing Sample Type (Blood or Biopsy) for the BEAM Study. *T tests were conducted for each alignment step from the raw sequencing reads. P values < 0.05 viewed significant and are highlighted in red. Alignment steps are in alphabetical order.*

Variable	***P* value**
circularRNA_antisense	0.024
circularRNA_sense	0.352
endogenous_gapped	0.044
exogenous_genomes	0.006
exogenous_miRNA	0.03
exogenous_rRNA	0.058
failed_homopolymer_filter	0.123
failed_quality_filter	0.004
gencode_antisense	0.002
gencode_sense	0.035
genome	0.004
input	0.003
input_to_exogenous_genomes	0.009
input_to_exogenous_miRNA	0.01
input_to_exogenous_rRNA	0.009
miRNA_antisense	0.136
miRNA_sense	0.004
miRNAprecursor_antisense	0.002
miRNAprecursor_sense	0.05
not_mapped_to_genome_or_libs	0.028
piRNA_antisense	0.011
piRNA_sense	0.08
rRNA	0.006
reads_used_for_alignment	0.003
repetitiveElements	0.004
successfully_clipped	0.003
tRNA_antisense	0.087
tRNA_sense	0.063
UniVec_contaminants	0.023

6.3.4. Differential Expression Analysis

One goal of the research conducted in this Chapter was to test the application of a spike-in exogenous miRNA as a method of normalisation (Vigneron *et al.*, 2016). The initial aim was to treat all samples with the spike-in (during library synthesis) but due to the problems encountered in successful library generation only a small sub-set (N = 7) of the libraries ultimately contained the spike-in. Despite this, exploration of application of the spike-in was continued.

The spike-in was added at a consistent concentration across the samples at the first stage of library synthesis and would therefore be able to account for any differences that arose because of subsequent variations (e.g. the impact of PCR amplification). Several attempts to use the raw read data for the samples (N = 7) with successfully identified spike-in were made. One sample was excluded from the analysis due to low sequencing read counts (N = 9,208) (data not shown). No other miRNA was stably expressed in relation to the expression of the spike-in miRNA across any of the samples treated with the exogenous spike-in within a 5% variation. There were no identified miRNAs with similar expression until a threshold of 65% variation was permitted, at which point two miRNAs were identified as potential matches (miR-139-3p, and miR-199a-5p). Therefore, there were no miRNAs with a sufficiently similar level of expression between groups to be used as a normaliser.

As a result of these findings, as in the previous chapters, the normalisation of reads was conducted using quantile normalisation and then miRNA expression of Case (ELD) vs Control for the sample types with reads above 1,000 (Table 6.3.) was conducted. The blood sample groups

were composed of 6 Controls and 3 ELD Case samples, while the biopsy groups were composed of 3 Controls and 2 ELD Case samples. Differential expression was then conducted (Section 6.2.3.)

Considering differences between the ELD and Control miRNA expression in the blood no DE-miRNAs were identified as shown by the volcano plot, Figure 6.11. There were 570 miRNAs that did show a fold change of greater than 0.5 between the Healthy Controls and ELD group but none proved to be significant (lowest FDR value 0.44). To narrow the search to only the most likely miRNAs the P value < 0.05 was used. Sixty-one miRNAs met the $P < 0.05$ and were considered for further analysis: miR-654-5p, miR-6882-3p, miR-6733-5p, miR-4753-3p, miR-1178-3p, miR-4429, miR-1306-3p, miR-6780a-3p, miR-7854-3p, miR-6820-3p, miR-6789-3p, miR-4677-5p, miR-1255a, miR-3690, miR-190a-5p, miR-6877-5p, miR-548q, miR-202-5p, miR-6750-5p, miR-1236-3p, miR-337-3p, miR-6743-3p, miR-491-5p, miR-548ac, miR-6737-3p, miR-4435, miR-6843-3p, miR-628-5p, miR-582-3p, miR-4668-5p, miR-6514-5p, miR-6767-5p, miR-191-5p, miR-3157-5p, miR-4753-5p, miR-381-3p, miR-6735-3p, miR-4661-5p, miR-4504, miR-4738-3p, miR-943, miR-548at-3p, miR-2277-5p, miR-1288-3p, miR-6849-3p, miR-1538, miR-3177-3p, miR-660-3p, miR-548ay-3p, miR-581, miR-4677-3p, miR-26a-2-3p, miR-4511, miR-25-5p, miR-495-3p, miR-548o-3p, miR-212-5p, miR-125b-5p, miR-3190-3p, miR-758-3p, miR-6842-5p.

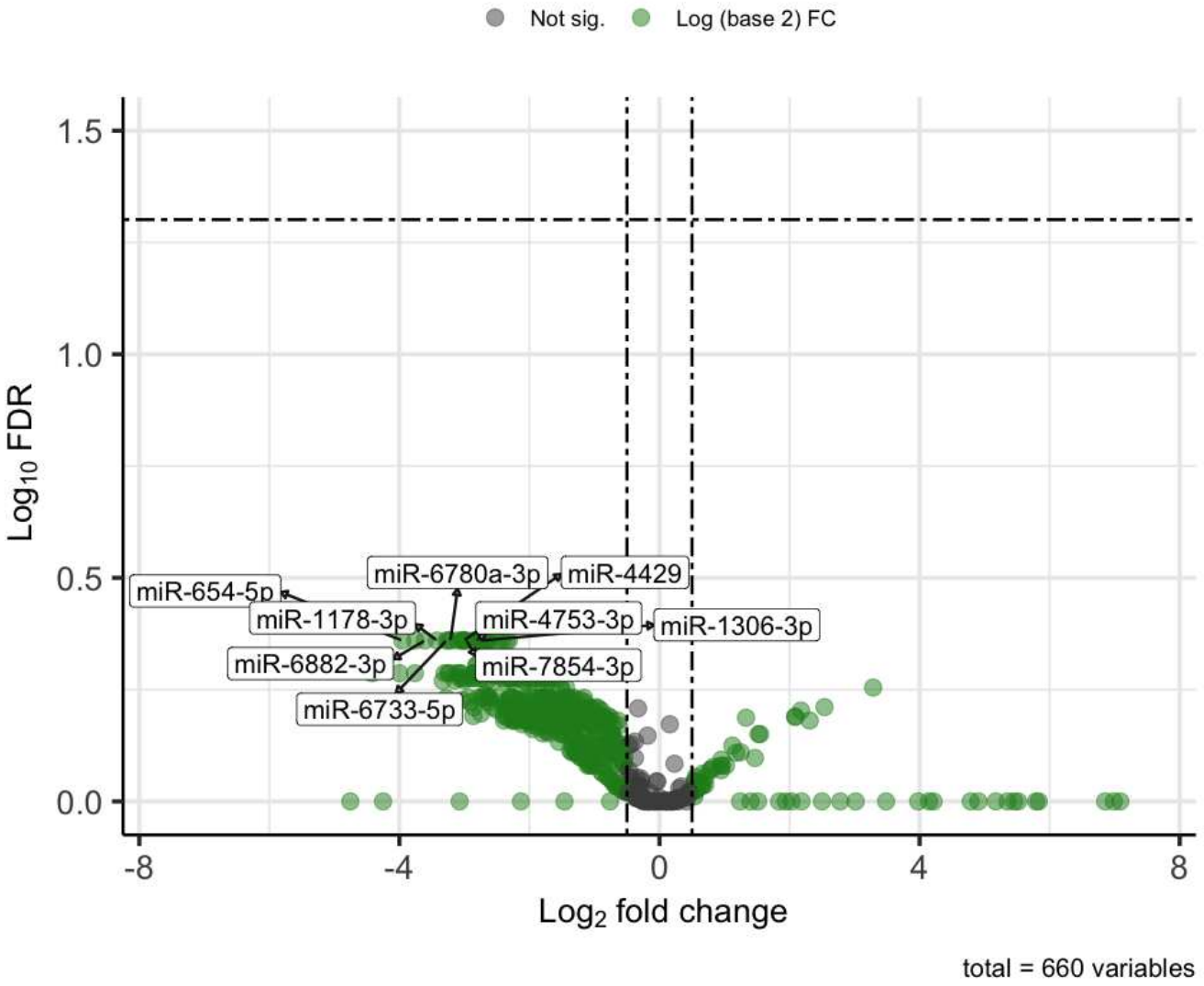


Figure 6.11. Volcano Plot of DE Analysis Results of Blood Sample miRNA Expression between Controls and ELDs. *FDR and Fold Changes in BaySeq DE analysis. No miRNAs were identified with an FDR (Q value) < 0.05, while 570 miRNAs were identified to have a fold change greater than 0.5. Blood sample analysis made up of 6 Controls and 3 ELD Case samples. Top 9 trending miRNAs labelled.*

Similarly, for the 5 biopsy samples, post normalisation, there were no identified DE-miRNAs (volcano plot Figure 6.12.). Again, the lowest FDR identified between miRNA expression between groups was 0.09. While there were no significantly identified miRNAs using FDR values, there were 366 miRNAs that had a fold change > 0.5. To narrow the search to only the most likely miRNAs, the *P* value <0.05 was used. Thirty-one miRNAs met the $P < 0.05$ and were considered for further analysis: miR-223-3p, miR-486-3p, miR-584-5p, miR-486-5p, miR-203a-3p, miR-31-5p, miR-652-3p, miR-202-5p, miR-503-5p, miR-4732-3p, miR-3679-5p, miR-615-3p, miR-187-3p, miR-15b-5p, miR-203b-3p, miR-3940-3p, miR-196b-5p, miR-224-5p, miR-4446-3p, miR-5196-3p, miR-143-5p, miR-223-5p, miR-455-3p, miR-642a-3p, miR-577, miR-485-5p, miR-133a-3p, let-7b-5p, miR-3605-5p, miR-185-3p, miR-184.

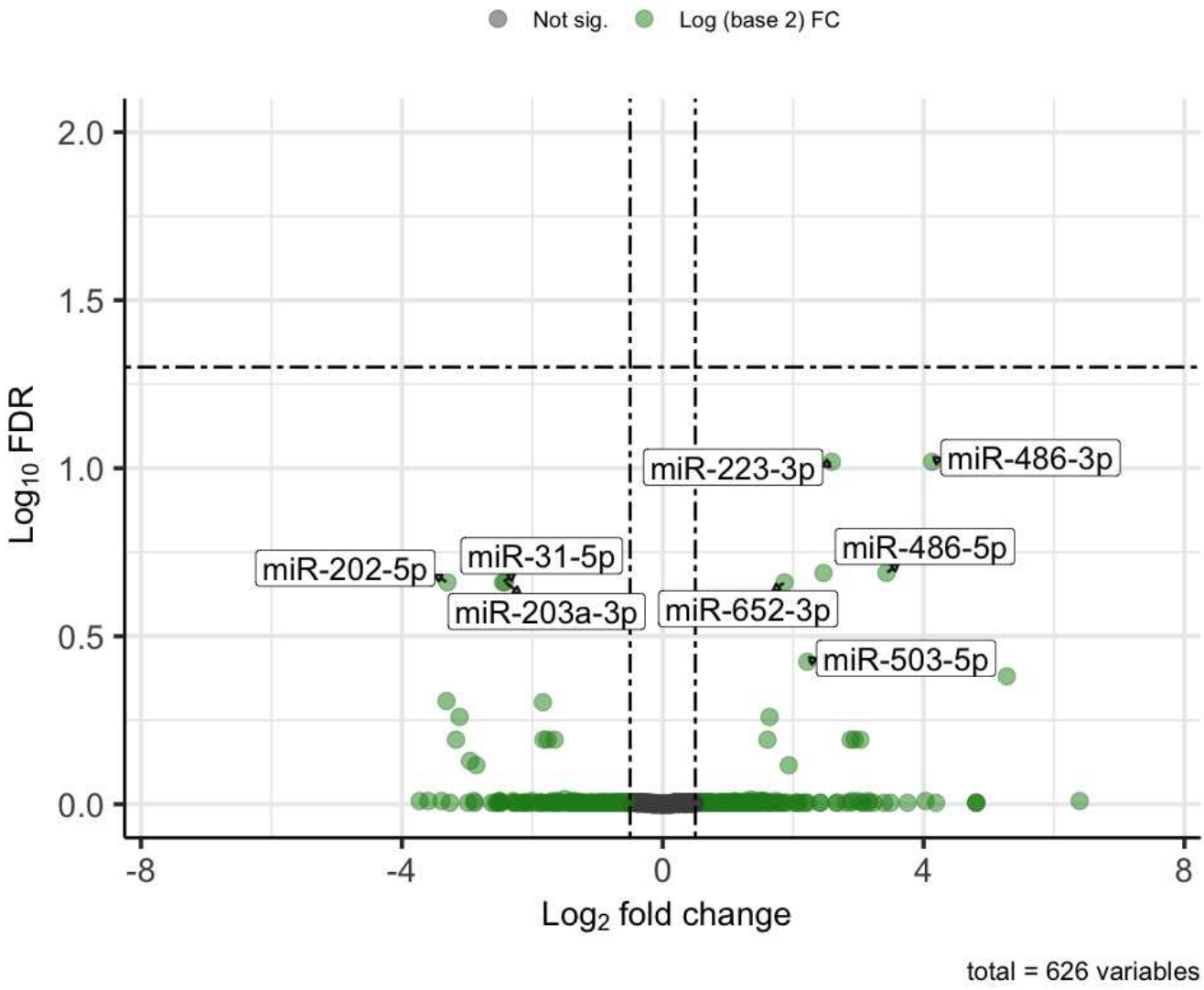


Figure 6.12. Volcano Plot of DE Analysis Results of Biopsy Sample miRNA Expression between Controls and ELDs. *FDR and Fold Changes in BaySeq DE analysis. No miRNAs were identified with an FDR (Q value) < 0.05. Three hundred and 66 miRNAs were identified to have a fold change greater than 0.5. Biopsy sample analysis made up of 3 Controls and 2 ELD Case samples. Top 8 trending miRNAs are labelled.*

One of the goals of this set of experiments was to look at the expression of miRNAs between tissues. This might give a better understanding of not only the processes within tissues but also whether more distal or associated but more accessible sample types can be used effectively as surrogates for tissues difficult to access. In this example if blood samples could be a surrogate for biopsy samples. Due to issues with the sequencing run, sequencing data of a sufficient quality was only available for biopsy and blood samples from a single individual – a male healthy control. Looking at the paired blood and biopsy sample (and as would be done analysing a larger dataset and using a more standard DE analysis), the reads were filtered for greater than 10 reads per miRNA identified. This reduced the total number of miRNAs from 1,583 to 284. A Wilcox test (P = 0.034) identified a difference in blood and biopsy miRNA expression. Figure 6.13. shows the top 10 miRNAs with the highest fold change in the tissue (blood versus biopsy) comparison from the subject. In all cases shown in the figure the biopsy expression was found to be higher. Of the 282 miRNAs identified 147 had FC > 0.5.

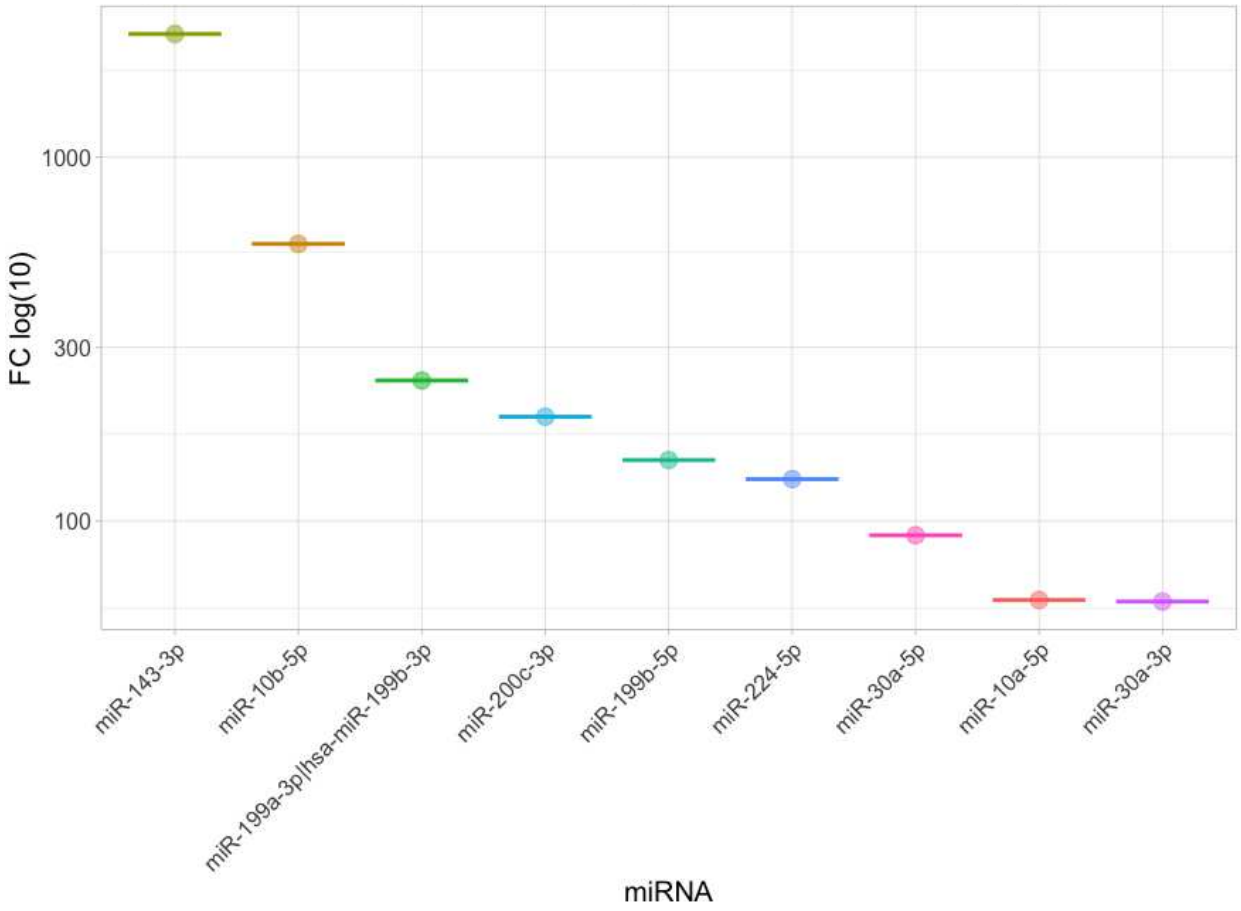


Figure 6.13. Top 10 miRNAs with Highest Fold Change (FC) Between Biopsy and Blood sample. *N = 1. A single pair of sequenced samples isolated from blood and biopsy samples from the same healthy male control with sufficient reads and quality were used for analysing fold change. A significant difference was identified between the tissue types, and 147 miRNAs had a FC greater than 0.5. The top 10 miRNAs with a fold change of between 60 and 2,183 are shown.*

6.3.5. miRNA Correlation between Healthy Controls and ELDs

While no miRNAs met both FDR and FC significance thresholds, a modified threshold was considered using $P < 0.05$.miRNAs were identified with $P < 0.05$, for both the ELD vs Healthy Control blood analysis (N = 61), and for the biopsy analysis (N = 31). The miRNAs with $P < 0.05$ were further considered in relation to clinical outcomes associated with disease. Heatmaps were generated to better understand the correlations of the top 9 miRNAs to enable visualisation of correlation values of the most significant interactions, for the blood samples analysis; miR-654-5p, miR-6882-3p, miR-6733-5p, miR-4753-3p, miR-1178-3p, miR-4429, miR-1306-3p, miR-6780a-3p, miR-7854-3p; and for the biopsy sample analysis; miR-223-3p, miR-486-3p, miR-584-5p, miR-486-5p, miR-203a-3p, miR-31-5p, miR-652-3p, miR-202-5p, miR-503-5p, and the clinical outcomes. Data was factorised and Spearman correlations were created from raw-counts data (Figure 6.14.). Correlation strength ranges applied are given in Chapter 2, Section 2.1.5.2.

A strong (>0.6) positive correlation was observed between the sample groups and all the identified ranking miRNAs. There were no correlations with BMI, FEV1/FVC ratio, or Eosinophil counts or total IgE. There were moderate-to-strong (0.53) positive correlations with the identified miRNAs and patient gender, and, by contrast, weak-to-strong (-0.41 - -0.42) negative correlations regarding age. All nine considered miRNAs showed strong correlations among themselves (0.89-0.99).

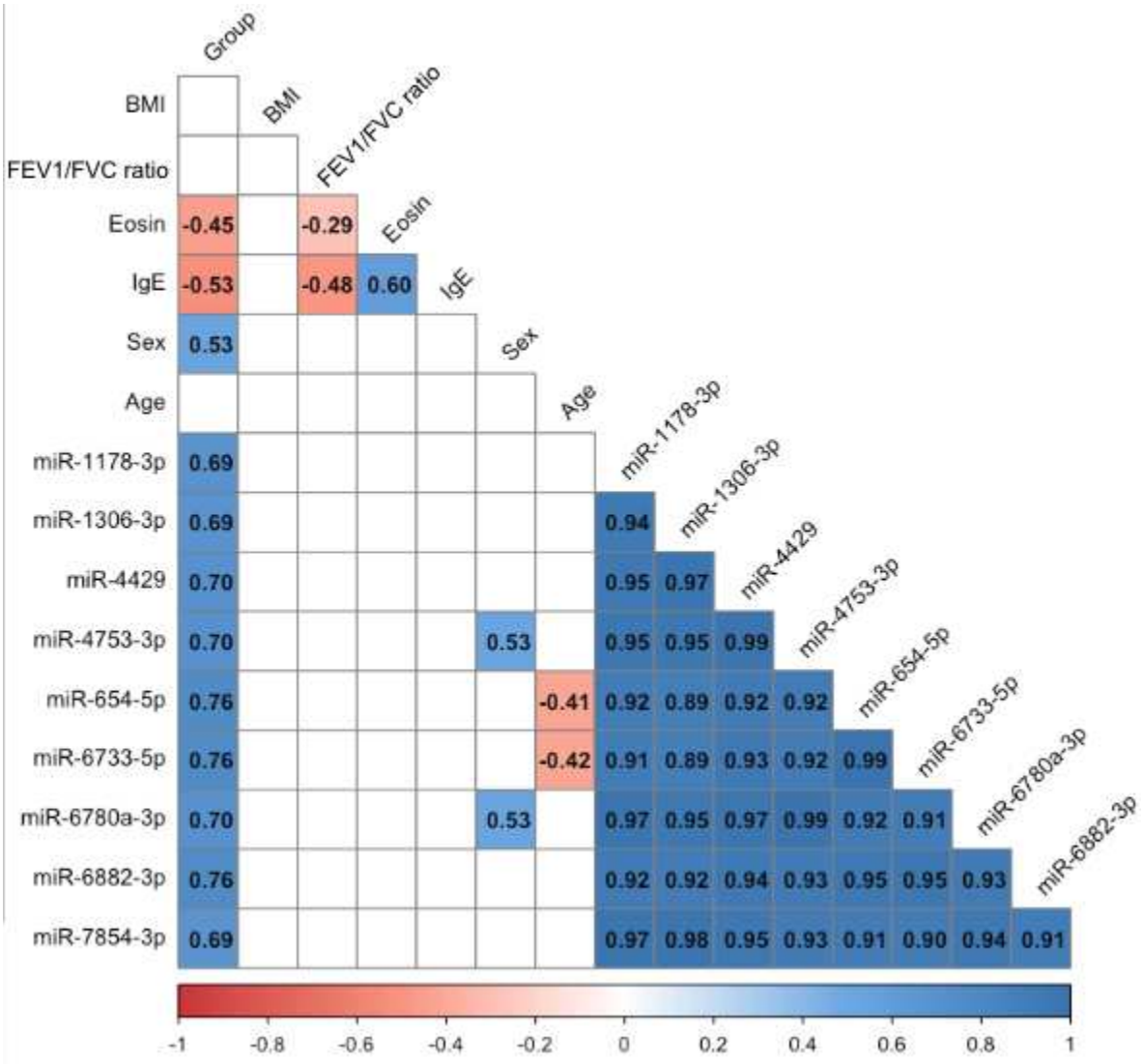


Figure 6.14. Correlation of Clinical Measures vs Highest Ranking miRNAs Expressed in Blood Samples of ELD vs Healthy Controls. *(N = 9 [6 Controls, 3 Case ELD samples]). Clinical measures were converted to factors where necessary, and a correlation analysis conducted. Group (Case, Control), BMI (Body mass index), FEV1 /FVC ratio (FEV1/FVC ratio, measure of lung function), Eosin (Total blood eosinophils), IgE (serum IgE), Sex (patient gender), Age (patient age in years). As per scale, red indicates negative correlation and blue positive correlation. Correlations shown P < 0.05, input miRNAs were not significantly differently expressed between groups, were the highest-ranking miRNAs for differential expression only. Moderate positive correlation was seen between identified miRNAs and patient Group, and gender. Mild to moderate negative correlation with patient age. No significant correlations identified for BMI. Negative mild to moderate lung function (FEV1/FVC ratio), or inflammatory markers (Eosinophil count, IgE). There were also strong correlations identified between the 9 considered miRNAs.*

A correlation analysis was also conducted for the top 9 biopsy sample miRNAs that were ranked most highly but were not significantly DE (Figure 6.15.). Patient Group was identified as having a significant correlation with miR-202-5p (0.87). Lung function measure (FEV1/FVC ratio) did not show a significant correlation with any of the ranked miRNAs. Inflammatory measures associated with ELD (Eosinophil count and total IgE) showed a strong negative or weak positive correlation, respectively, with different miRNAs. miR-202-5p showed strong positive correlations with Group (0.68). miR-223-3p showed a weak negative correlation with Age (-0.28), and a strong correlation with gender (0.77). miR-503-5p showed a strong positive correlation with gender (0.69). There were also strong positive correlations identified between some of the 9 considered miRNAs (for example a perfect correlation of (1) miR-203a-3p with miR-31-5p) (-0.87 to -0.97).

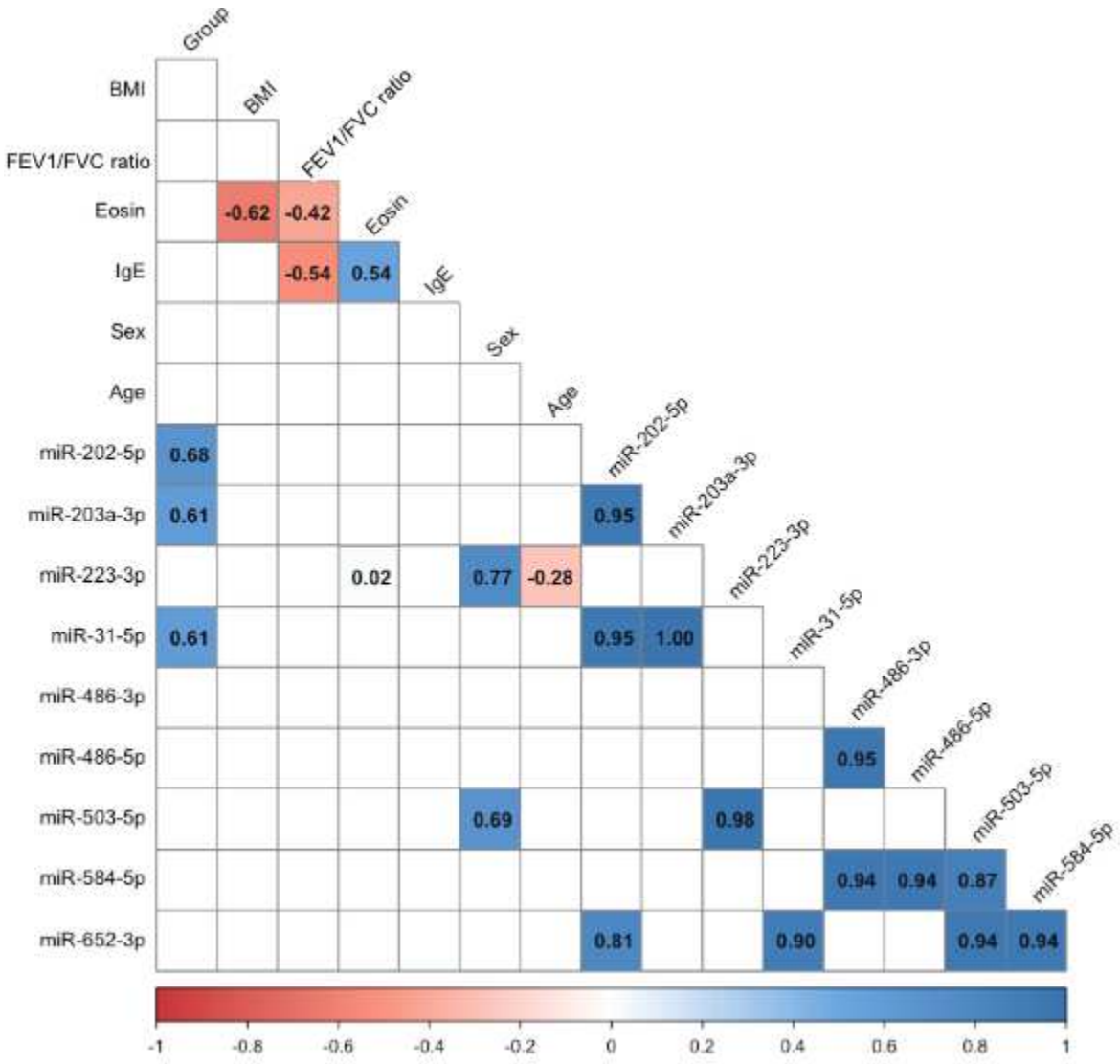


Figure 6.15. Correlation of Clinical Measures vs Highest Ranking miRNAs Expressed in Biopsy Samples of ELD vs Healthy Controls. *(N = 5 (3 Controls, 2 Case samples)). Clinical measures were converted to factors where necessary, and a correlation analysis conducted. Group (Case, Control), BMI (Body mass index), FEV1/FVC ratio (FEV1/FVC ratio, measure of lung function), Eosin (Total blood eosinophils), IgE (serum IgE), Sex (patient gender), Age (patient age in years). As per scale, red indicates negative correlation and blue positive correlation Correlations shown P < 0.05, input miRNAs were not significantly differently expressed between groups, were the highest-ranking miRNAs for differential expression only. MiR-202-5p showed strong positive correlations with Group (0.68). miR-223-3p showed a weak negative correlation with Age (-0.28), and a strong correlation with gender (0.77). miR-503-5p showed a strong positive correlation with gender (0.69). There were also strong positive correlations identified between some of the 9 considered miRNAs (for example a perfect correlation of (1) miR-203a-3p with miR-31-5p).*

6.3.6. Target Identification

All miRNAs that were trending towards DE with a P <0.05 (but not an FDR < 0.05): 61 miRNAs ELD versus Healthy Controls blood samples, 31 miRNAs for the biopsy sample comparison, were analysed for targets using miRNet (Chang et al. 2020).

As per prior Chapters (4 and 5) only experimentally validated mRNA targets of the input miRNAs were used for the initial target identification and all sequencing related target identification was excluded from the results. The target and category results are a useful stepping stone in understanding the weight of evidence supporting specific gene targets or pathways (how unique the targets are) and therefore help identify miRNAs that are more selectively affected in ELD for further research purposes.

6.3.6.1. Blood Gene Ontology Analysis

From the input 61 miRNAs (from the blood samples DE analysis) there were a total of 4,737 gene targets identified. Interaction tables were filtered for experiment type, where all sequencing and computational methods were filtered out leaving only experimental methods. Filtering the network by degree for all network nodes resulted in 1,867 targets (see Chapter 2 Section 2.1.5.3.).

For the following results the top five categories were considered, all of which had an adjusted *P* value < 0.05 for all target categories (as calculated by the GO algorithm used by miRNet). All categories identified in the analysis are available within the Supplementary materials (Supplementary: Chapter 6: (BEAM) Supplementary: [BEAM_GO_sig_tables]).

The Biological Process (BP) analysis identified 62 target genes with a *P* value of < 0.05. The most significant target gene category was for 'negative regulation of apoptotic process' N = 22 genes (*P* < 0.0001). The category with the highest number of target genes was in 'apoptotic process' and "programmed cell death", N = 33, (*P* < 0.0001). There were 45 target genes that made up the top five most significant Biological Process target categories. There were 17 genes identified across 3 of the top 5 categories (*AKT1*, *BCL2L1*, *CDKN1B*, *CEBPB*, *DICER1*, *EGFR*, *ERBB3*, *HMGA2*, *IGF1*, *IGF1R*, *IL6R*, *KDR*, *MDM4*, *MYC*, *RAF1*, *TGFBR1*, and *TP53*). The top section of the graph in Figure 6.16. shows the distribution of the 45 target genes across the categories.

The Molecular Function (MF) target analysis identified 54 target genes with a *P* value of < 0.05. The most significant target gene category was for 'negative regulation of transcription, DNA dependent' N = 19 genes (*P* < 0.0001). The category with the highest number of target genes was in 'transcription from RNA polymerase II promoter', N = 20, (*P* < 0.0001). There were 35 target

genes that made up the top five most significant Molecular Function target categories. Of the categories, 3 were related to transcription (Figure 6.16. middle section). Two genes (*HMGA2*, *SMAD4*) appeared in four of the five categories.

The Cellular Component (CC) target analysis identified 59 target genes with a *P* value of < 0.05. The most significant target gene category was for 'cytosol' N = 27 (*P* < 0.0001). The category with the highest number of target genes was in 'macromolecular complex' N = 36, (*P* < 0.0001). There were 49 target genes that made up the top five most significant Cellular Component target categories (Figure 6.16.). Nine genes appeared in all 5 of the top ten most significant gene target categories (i.e. genes can occur in more than one category) (*AKT1*, *CCNE1*, *CSNK2A1*, *HMGA1*, *MAPK14*, *RPS6KA1*, *SET*, *SMAD4*, and *TP53*).

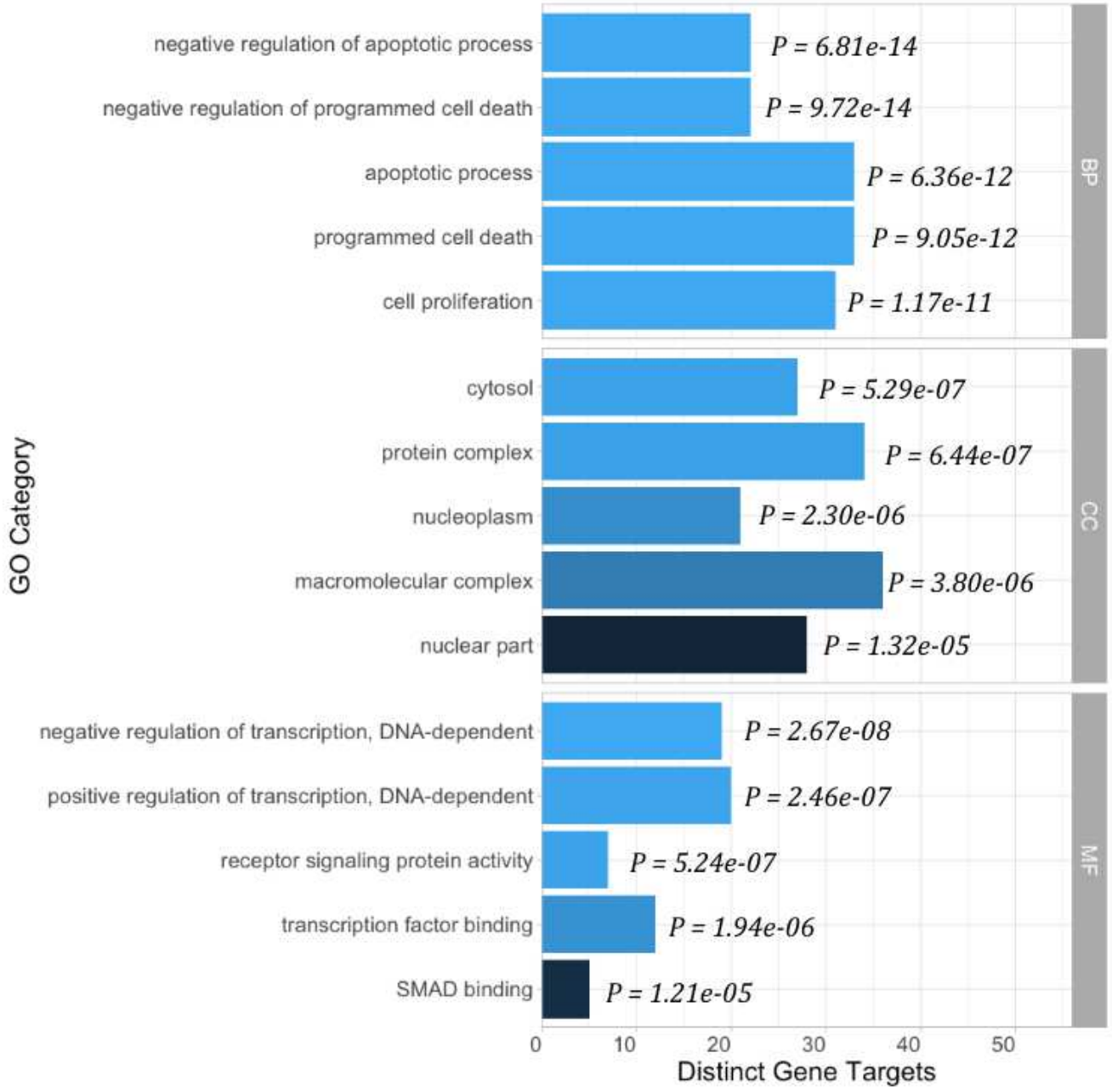


Figure 6.16. ELD vs Healthy Control Blood sample miRNAs Gene Ontology Classification Summary Results (Biological Process, Molecular Function and Cellular Components). *features and number of target genes for differentially expressed miRNAs between Blood sample miRNA expression in ELD patients and Healthy controls. The top 61 expressed miRNAs trending towards DE between ELD and Healthy Controls were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. The top 5 most significant categories were selected for display. The bars represent the number of target genes in each category, ordered top to bottom from most significant to least P-value. (Biological Process = BP, Molecular Function = MF, and Cellular Components = CC).*

From the results of the GO analysis from the blood sample trending-towards-DE-miRNAs, a Venn diagram was made to show crossover of gene targets among the GO-tested categories (BP, MF and CC – for each top 5 most significant result categories) (Figure 6.19.). Thirty nine percent of the genes identified were shared among all tests (23 genes). Each GO-tested category had unique gene targets. These as well as the genes shared by all three categories and those shared between two categories are listed in Table 6.5.

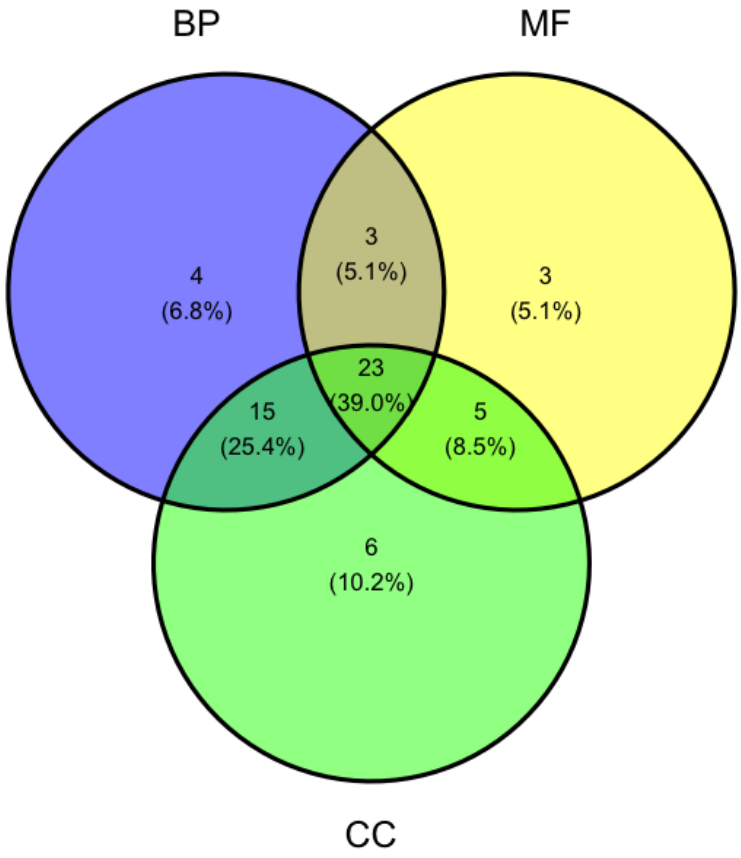


Figure 6.17. Venn Diagram of Target Genes Identified as Significant in the Top 5 Categories of each GO Analysis Category for Blood Samples. *More than a third (39%) of the genes identified were present in all three categories. MiRNAs with (P < 0.05) (N = 61) were input into miRNet database and passed through GO analysis categories (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). Resulting gene targets were then compared among the three categories.*

Table 6.5. ELD Target Genes from GO Analysis Categories (Blood). *Target Genes from GO analysis top 10 most significant (P < 0.05) categories compared against one another. MiRNAs P < 0.05 (N = 61) identified from blood sample miRNA expression between Healthy Controls (N = 6) vs ELDs (N = 3). Unique genes and genes identified among other GO analysis categories.*

Interaction (analysis)	*Gene names*
Unique to Biological Process	*IGF1R, CD44, DRAM2, IGF2*
Unique to Molecular Function	*MAF, ARID3A, DKK3*
Unique to Cellular Components	*IGF2BP1, PIK3CD, RAP1A, TNPO2, PIK3CB, VPS4B*
All GO analysis categories	*AKT1, CDKN1B, CEBPB, DICER1, EGFR, HMGA2, IGF1, MYC, RAF1, TGFBR1, TP53, BCL3, RPS6KA1, SET, TGFBR2, AXIN2, E2F2, E2F3, HMGA1, MAPK14, SMAD4, STAT3, VDR*
BP and MF shared	*KDR, MDM4, TWIST1*
MF and CC shared	*PHF8, BASP1, CCNE1, CHD4, NCOR2*
BP and CC shared	*BCL2L1, ERBB3, IL6R, MCL1, BBC3, CDK6, CSNK2A1, DUSP6, EIF5A2, GAB2, GRIN2A, MTA3, PHLPP1, TJP1, TRIM71*

To identify the interaction between the miRNAs input and the gene targets identified as significant within specific categories of each GO analysis, the trending miRNAs were aligned to their corresponding associated target genes selecting only those relevant to each GO analysis. By doing this, there was also the potential of identifying whether there were any specific processes or pathways preferentially targeted.

Gene targets were filtered for those with more than 5 source miRNAs and a stacked bar graph was created to show the interactions and how the gene targets were shared among the GO categories (Figure 6.18.). Of the 61 miRNAs, 19 miRNAs were identified among all GO analysis categories with ≥5 shared gene targets. Of the 19 miRNAs identified that made up the 5 source miRNAs for that gene, each miRNA was linked to between 1 and 5 genes (Figure 6.18.). Shared among all three categoryies was the highest frequency of gene target interactions, followed by the BP and MF shared gene targets, and CC and BP shared gene targets taking up some of the interactions. There were no more than 5 source miRNAs for gene targets within the remaining categories (exclusive to either CC, BP, or MF, or CC_MF shared). The gene target with the highest number of source miRNAs was *IL6R* (with 6 miRNA sources). The remaining gene targets were linked with 5 source miRNAs each (*E2F2*, *CDK6*, *MCL1*, *IGF1R*).

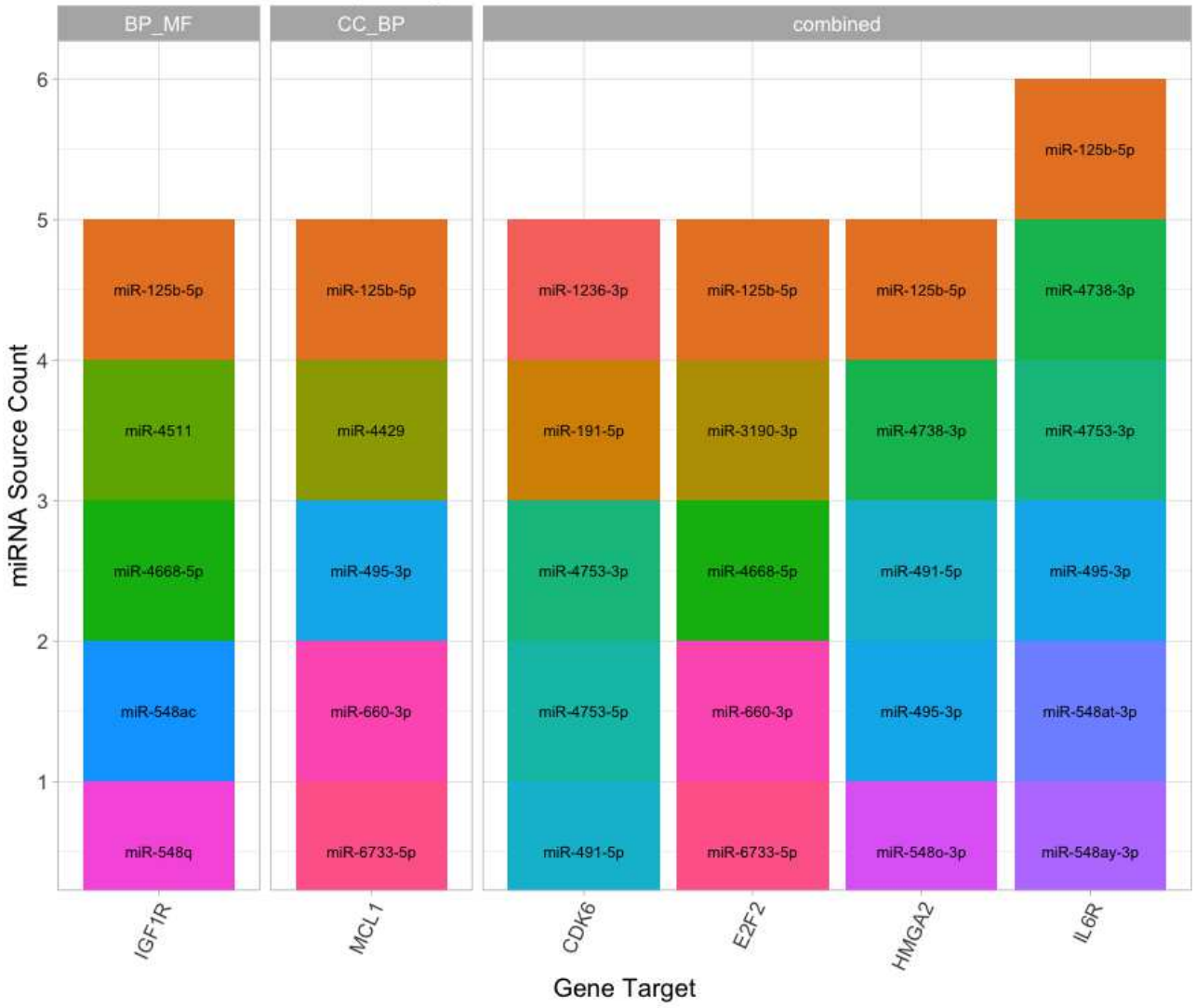


Figure 6.18. Gene Targets and Source miRNAs from GO analysis Faceted by Analysis Interaction. *Trending towards DE-miRNAs (N =61) from blood sample analysis between ELD and Healthy Controls were input into a GO analysis, the resulting genes were filtered for those with the greatest number of source miRNAs and those with more than 5 or more interactions are shown here. The bars represent the number of source miRNAs that target the gene denoted in the x axis, ordered from most significant (top) to least (bottom) P-value within each category. The highest number of miRNAs acting on one gene target were in the combined interaction (all categories of GO tested) and target IL6R.*

6.3.6.2. Biopsy Gene Ontology Analysis

Similar to the blood DE-miRNA analysis, for the biopsy samples ELD versus Healthy Contrls none were found to attain significance but 31 were found to trend towards significance. These 31 miRNAs were therefore input into the mirNet database. Five thousand and 185 gene targets were identified and these were passed through a degree filter for all network nodes (Chapter 2, Section 2.1.5.3.) resulting in 1,738 gene targets remaining.

There were very few interaction table results that remained after filtering out computational and sequencing target identification. A more relaxed threshold was used allowing inclusion of "Proteomics" research methods as well as "Cross-linking and immunoprecipitation" (CLIP), and "Cross-linking ligation and sequencing of hybrids" (CLASH). The results of the miRNet analysis then underwent Gene Ontology analysis to identify the genes associated with Molecular Function (MF), Biological Processes (BP) and Cellular Components (CC).

As for the blood analysis, for the following results the top ten categories were considered, all of which had an adjusted *P* value < 0.05 for all target categories (as calculated by the GO algorithm used by miRNet). All categories identified in the analysis are available within the Supplementary file (Chapter 6: (BEAM) Supplementary: [BEAM_GO_sig_tables]).

For the GO analysis BP, the category with the most significant genes was 'viral reproductive process' N = 105 $P < 0.001$, and, of the top 5 most significant categories (Figure 6.19.), the gene category with the largest number of interacting genes was in 'regulation of cellular protein metabolic process' N = 223 genes. The top 5 categories were made up of 386 unique gene targets. There was one gene that appeared in all top 5 categories (*TNF*).

For the GO analysis MF, the description with the largest number of interacting genes was 'nucleotide binding' N = 340, and, of the top 5 most significant categories (Figure 6.19.), this category also had the most targets. Of the MF categories making up the 5 most significant, 3 categories were made up of transcription. There were 551 genes that made up the top 5 most significant categories. Of these, 4 genes were identified across all 5 of the categories (*EEF1A1*, *HSP90B1*, *TPR*, *TUT1*).

For the GO analysis CC, the description with the largest number of interacting genes was 'membrane enclosed lumen' N = 465 $P < 0.0001$, and, of the top 5 most significant categories (Figure 6.19.), the gene category with the most significant genes was in 'nucleoplasm' N = 262 $P < 0.001$. There was a total of 663 genes that made up the top 5 categories. There were 71 genes that appeared in all 5 categories.

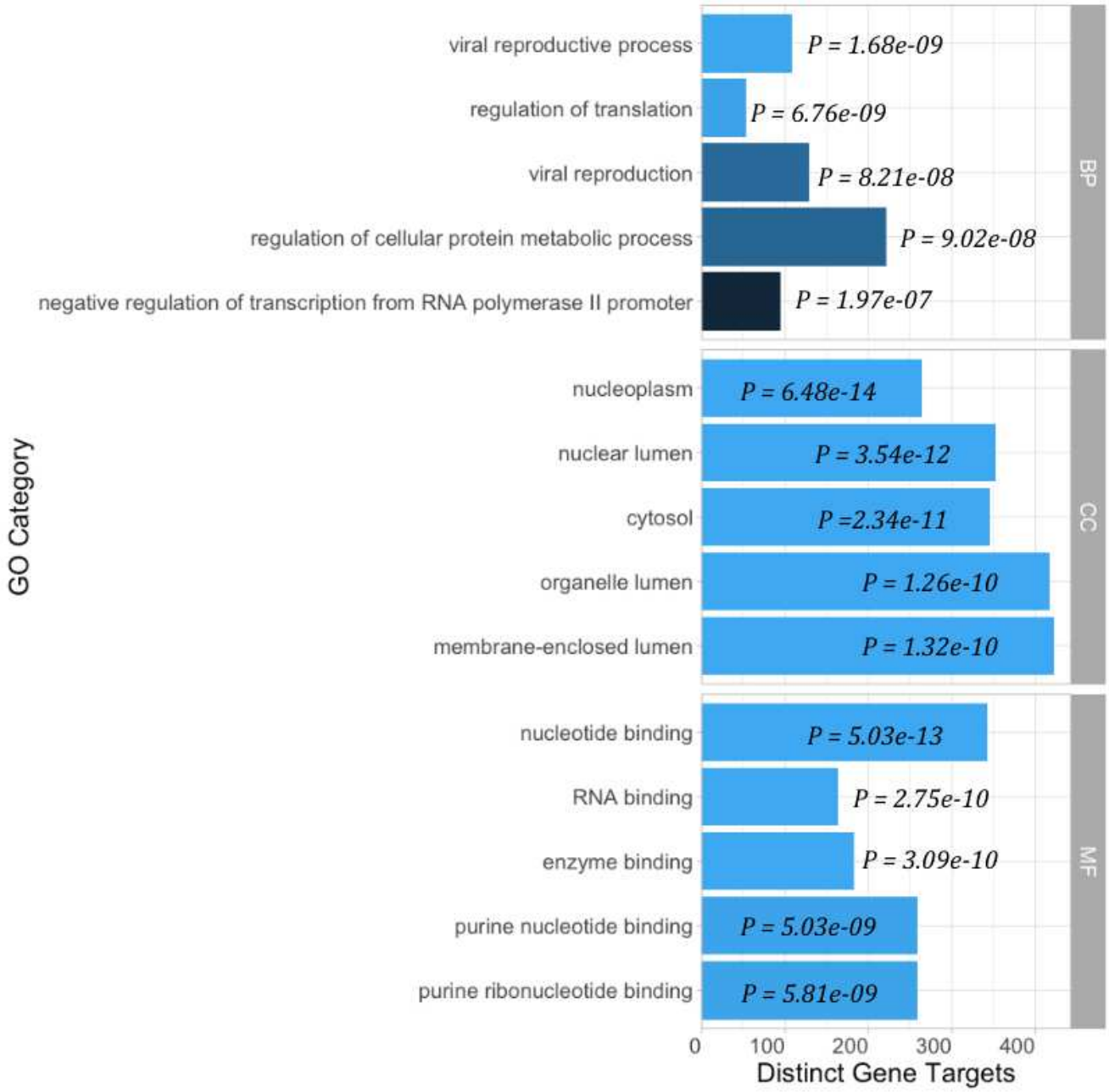


Figure 6.19. ELD vs Healthy Control Biopsy sample miRNAs Gene Ontology Classification Summary Results (Biological Process, Molecular Function and Cellular Components). *Features and number of target genes for differentially expressed miRNAs between Biopsy sample miRNA expression in ELD patients and Healthy Controls. The top 31 expressed miRNAs trending towards DE between ELD and Healthy Controls were assessed for targets in miRTarBase 8.0, and the resulting targets were analysed through GO classification. The top 5 most significant categories were selected for display. The bars represent the number of target genes in each category, ordered top to bottom from most significant to least P-value. (Biological Process = BP, Molecular Function = MF, and Cellular Components = CC)*

As for the blood miRNAs, further comparisons of the GO analyses for the biopsy miRNAs were conducted with construction of a Venn diagram (Figure 6.20.) to show crossover of gene targets among the three GO-tested categories (BP, MF and CC) top 5 most significant result categories. Sixteen percent of the genes identified were shared among all tests (154 genes). In BP, 82 unique targets were identified, 59 targets shared with MF, 91 targets shared with CC. In MF 159 unique targets and 179 shared with CC were identified. In CC 239 unique targets were identified. The full gene targets and categories identified are available in Supplementary materials (Supplementary: Chapter 6: (BEAM) Supplementary: [Biopsy_Gene_Ontology_GeneList]).

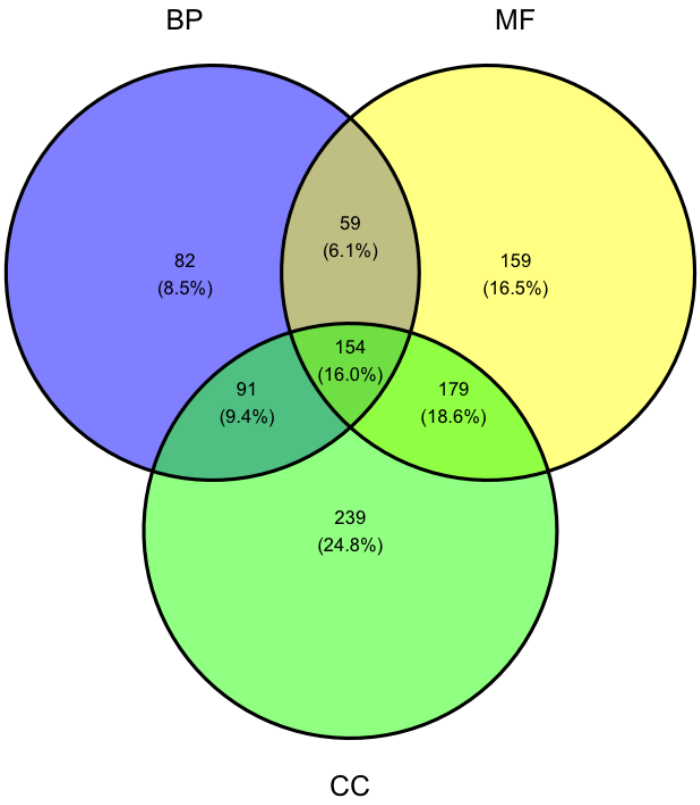


Figure 6.20. Venn Diagram of Target Genes identified as Significant in the Top 10 Categories of each GO Analysis Category for Biopsy Analysis. *Genes identified as Significant in the Top 5 categories of each GO analysis Category in Healthy Controls vs ELD study biopsy sample miRNA expression. Sixteen percent of the identified genes were found among all three categories. trending towards DE-miRNAs (N = 31) were input into miRNet database and passed through GO analysis categories (Biological Process [BP], Molecular Function [MF], and Cellular Components [CC]). Resulting gene targets were then compared among the three categories. BP identified 82 unique targets, 59 targets shared with MF, 91 targets shared with CC. MF identified 159 unique targets and 179 shared with CC. CC identified 239 unique targets. One hundred and fifty-four shared targets among the categories.*

Next, the biopsy miRNAs were aligned to their corresponding associated target genes selecting only those relevant to each GO analysis. Gene targets were filtered for those with more than 5 source miRNAs and a stacked bar graph was created to show the interactions and how the gene targets were shared among the GO categories (Figure 6.25.). Of the 31 miRNAs identified as trending towards differential expression between ELDs and Healthy Controls in biopsy samples, 23 miRNAs were identified among all GO analysis categories with ≥5 shared gene targets. The *IGF1R* gene had the most miRNAs that share it as a target, followed by *CDKN1A*, with both found shared between the CC and BP analysis. CC analysis exclusive interactions were the only GO analysis category with gene targets with more than 5 source miRNAs shared for the targets *AGO1,* and *PPP2R2A*. The combined GO analysis categories generated the highest number of gene targets with 5 or more source miRNAs (*E2F3*, *SMAD2*, *BCL2L1,* and *BCL2*). The other combined category of MF and BP identified a further 2 gene targets that shared 5 or more source miRNAs (*ZEB1*, *SPRED1*). There were no gene targets with 5 or more source miRNAs in the BP exclusive categories, or the CC and MF shared category.

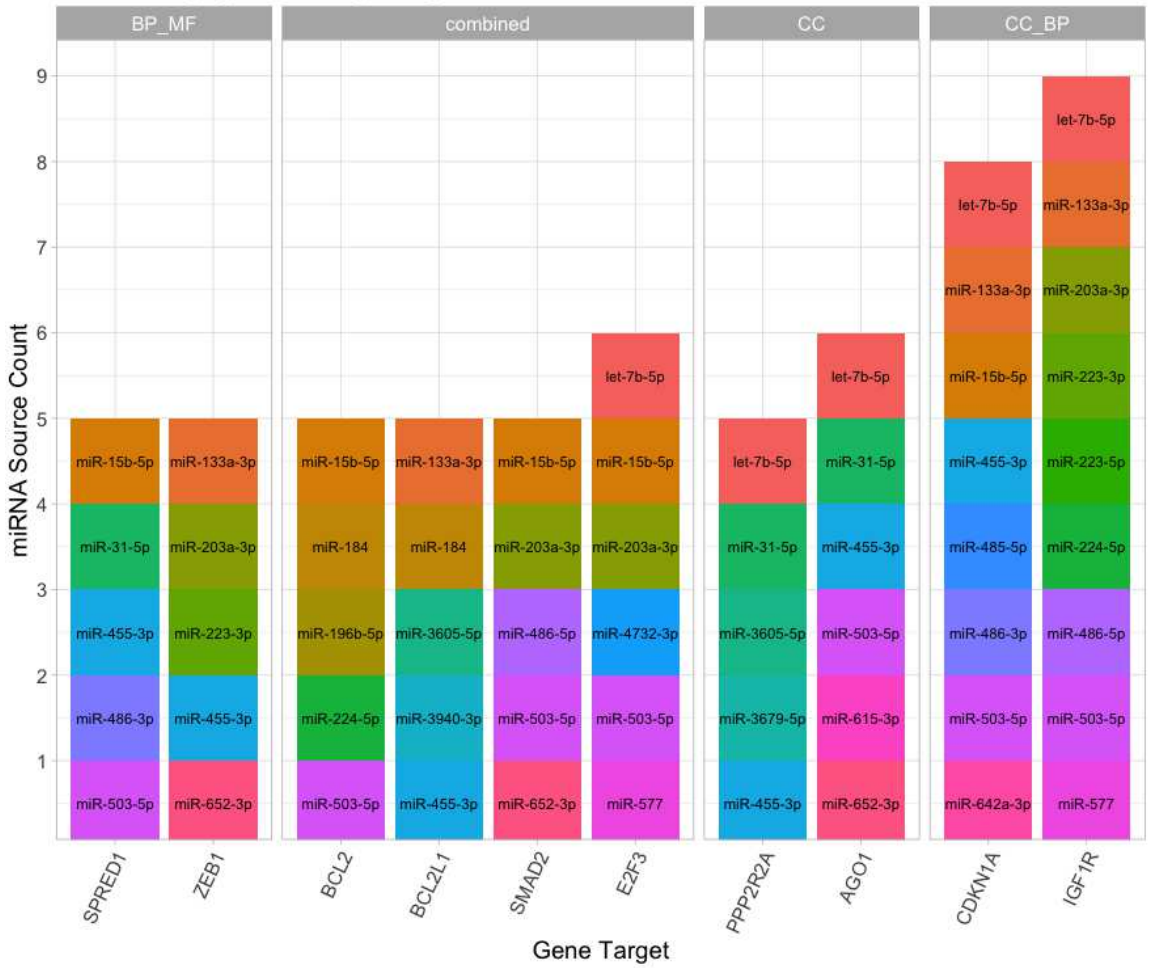


Figure 6.21. Gene Targets and Source miRNAs from GO Analysis Faceted by Analysis Interaction*. Trending towards DE-miRNAs (N = 31) from biopsy sample analysis between ELD and Healthy controls were input into a GO analysis, the resulting genes were filtered for those with the greatest number of source miRNAs and those with more than 5 or more interactions are shown here. Twenty-three unique miRNAs were identified in the interactions with the greatest number of miRNAs per gene target. The highest number of miRNAs acting on one gene target were in the CC_BP interaction and target IGF1R. CC was the only unique GO category with more than 5 miRNAs to one gene target.*

6.3.7. Reference Gene Analysis

An analysis of potential internal validation miRNAs was considered as a back-up where spike-in might not be viable. Additionally, a goal of sequencing analysis is to validate using QPCR. As discussed in Chapter 1 (Section 1.3.2.), miRNAs do not show consistently stable expression making selection of reference miRNAs with which to normalise expression more complicated. An analysis was conducted to identify reference miRNAs from all ELD vs Healthy Control samples (biopsy and blood) to see if any miRNAs were stably expressed between tissues/sample types. Ideally, a second analysis of sample type would have been conducted but due to the limited sample size per sample type group this was not feasible.

There were 14 samples (5 biopsy [3 Healthy Controls and 2 ELDs], 9 blood (6 Healthy Controls and 3 ELDs) with reads greater than 0, in at least 1 miRNA identified, that could be input into the analysis. A total of 9 miRNAs met these requirements and were passed through the geNorm package (Chapter 2, Section 2.1.5.5). geNorm was designed to identify reference mRNAs, however, this package has also been used for identification of reference miRNAs (Veryaskina. 2022). Five of the miRNAs had a stability value of expression of less than 1.5 (which is considered acceptable for use as a reference miRNA) (Vandesomepele *et al*., 2002). The top (most stably expressed) miRNAs (let-7a-5p, let-7i-5p, miR-30c-5p, miR-191-5p, miR-25-3p) and their stability values are shown below (Figure 6.22).

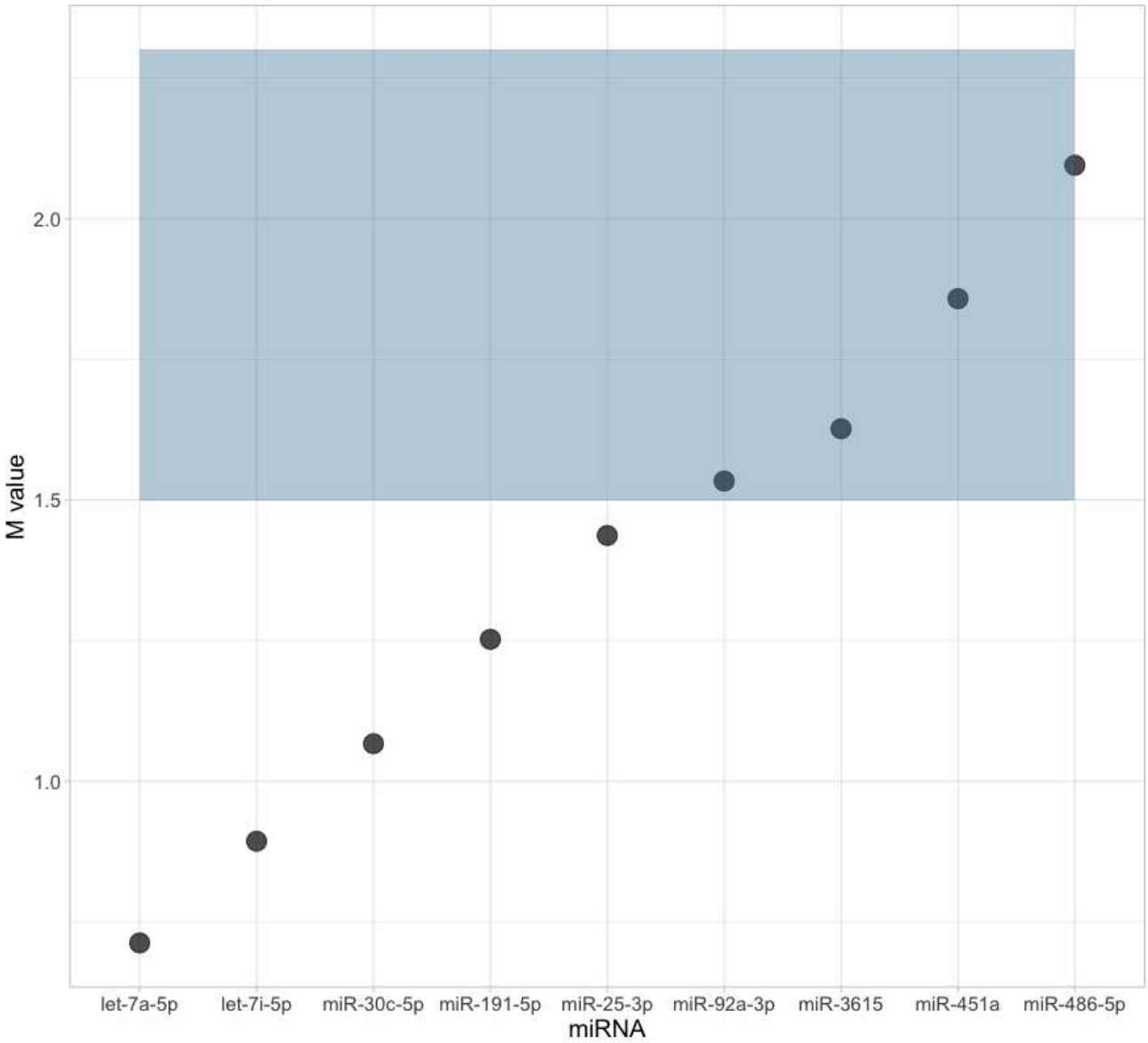


Figure. 6.22. Reference Gene Analysis for ELD vs Healthy Controls Blood and Biopsy Samples*. Read count data was analysed for potential reference genes. Samples and reads were filtered for reads greater than 0, and miRNAs were filtered for rows (across samples) with reads greater than 0. A total of 9 miRNAs remained across 17 samples. geNorm analysis identified M values between 0.7 and 2.1. Literature suggests an M value (stability score) of up to 1.5 for reference miRNAs, blue box indicates cut-off for genes that are not stably expressed for use as reference. Let-7a-5p shows the most stable expression as a reference miRNA candidate*

6.4. Discussion

This is the first time that miRNA expression of blood and lung biopsy samples has been considered in understanding the underlying processes of Eosinophilic Lung Disease compared to Healthy Controls.

Comparing ELD versus Healthy Controls, there was as anticipated a significant difference in certain clinical variables such as total IgE, eosinophil counts, and lung function measures (FEV1/FCV ratio) (Table 6.2.). There were, however, also statistically significant differences between patient gender, with more female ELD patients, and ELD subjects having an overall lower BMI than the Healthy Controls. The age of the groups was well balanced and there was no difference identified between the groups. These findings were echoed in the correlation analysis with eosinophil counts and IgE correlating positively with the ELD group (Figure 6.2.). There is not a significant amount currently published on the subject of gender due to the overlap with other eosinophilic airway conditions, however, one study also identified a higher ratio of females to males being affected, specifically in Chronic Eosinophilic Pneumonia (CEP) - a subset of ELD (Xu *et al.*, 2022). By contrast, there appears to be conflicting data in the literature regarding BMI, as increased BMI has been reported to be associated with positive correlations with blood eosinophilia (Sunadome *et al.*, 2020).

As the key aim of the present study was to conduct a comparison of miRNA expression between tissue types in ELD, both whole blood and biopsy samples were sourced, and methods to extract total RNA from these samples for further analysis were optimised and employed. Lung biopsy

samples are known difficult samples from which to extract total RNA as they are primarily made up of connective tissue that is resistant to disruption and can impede the separation of the RNA available for isolation (Brown *et al.*, 2018) hence a bead beating step was included at the start of the extraction protocol (Section 6.2.2.).

There was, overall, some difficulty in processing samples for sequencing. There was a notable loss of samples as they did not meet minimum input criteria at key steps. Complications in this experiment regarding the equipment used for size selection and errors that arose that were initially put down to human error but turned out to be a faulty electrophoresis size selection device [Pippin Prep]. This caused significant delay and loss of samples. Factors considered while testing issues initially perceived to be human error, were sample preparation methods, reagents and reagent temperatures, calibration and other Pippin Prep run set up steps. Until all the steps (as recommended by the manufacturer) were conducted and evidenced, there was no suggestion that the device itself was to blame. In some cases, a size selection run would pass all calibration steps only to fail to elute, or the Pippin Prep would fail to calibrate altogether. As a result, it was not always known when a test sample should be used as opposed to the samples that were key to this study. Sample loss was avoided as best as possible by setting up the Pippin Prep ensuring the reagents were within the correct temperatures and only removing samples from storage when all calibrations were validated. Several reruns of failed samples were conducted to try and attempt to submit all initial samples and sample types for sequencing.

Ultimately 27 out of 42 samples were submitted to Novogene for sequencing and in addition to the in-house rigorous quality controls steps the samples passed Novogene's QC.

The quality of the sequencing reads generated, however, by Novogene was very poor. Several samples failed to have their adapters identified resulting in no reads at all. Of those that did have reads, a significant number of reads did not pass quality filters during alignment. Insufficient information could be obtained from Novogene as to what happened when they conducted the sequencing of the submitted libraries. There were some concerns as one of the runs of the broader Celtic Fire study (Chapter 5, Run 2) had similarly had a reduction in reads in contrast to the other run in that study (Run 1) and the sequencing run of the Celtic Fire pilot study (Chapter 4, run CF1). The Celtic Fire study (Chapter 5), less successful run, and the ELD study were part of the same order submitted to the Sequencing Facility. It was suggested in a communication with the Imperial Biomedical Research Centre Genomic Facility that there could have been at Novogene a setting error that may not have been changed for both runs that could explain the issue, or even a computer error on the sequencing machine that could cause this.

The addition of the spike-in miRNA (cel-miR-29) was implemented for some samples, the libraries of which were included in the final pooled library. Initially, as libraries were prepared in small batches, the spike-in miRNA was added to all samples but then as issues with the processing and the remaining amount of input sample became scarce it was decided to make libraries without the addition of the spike-in. Hence although the inclusion of the spike-in was explored (Section 6.3.4.) ultimately as per prior Chapters quantile normalisation was implemented. The goal of the spike-in had been to use it as a control for miRNA expression hence its addition at the point before reverse transcription and the subsequent PCR amplification. Williams *et al*. discuss the benefits of adding a spike-in to act as a normaliser of differences in numbers of miRNA reads identified between samples that are a result of library preparation rather than being differences

between the biological samples (Williams *et al.*, 2013). Ultimately, this was too small a study to draw conclusions whether using the spike in in future studies is recommended.

One of the reasons that despite strong FC differences being identified there were no DE-miRNAs identified could be down to the lack of statistical strength because of low counts as many of the samples failed to achieve the target 5M reads during sequencing. Future studies may be more fortunate in experimental conditions in generating sufficient reads for analysis, and the addition of a spike-in miRNA is an important consideration (Williams *et al.*, 2013). Overall, regarding the use of spike in miRNA as a normaliser may also necessitate achieving greater than 5M reads per samples to account for the altered percentage of reads that will be spike-in relative to those belonging to the sample.

One of the aims of the study was a comparison of miRNA expression between blood and biopsy to identify differences in expression between the tissues, and any similarities. If the latter was found, and as blood is more accessible than lung tissue, there would be blood biomarker potential for ELD. Disappointingly, due to the problems encountered and highlighted above, this proved not to be possible as no blood-biopsy pairs from an ELD patient were successfully processed and sequenced.. There are, to date, no publications comparing miRNA expression identified with a sequencing approach between lung tissue and peripheral blood for any respiratory condition or including in health. In this present study looking at the lung biopsy-blood pair that was available for a single healthy control (Figure 6.13.)a total of 284 DE-miRNAs were identified with more than 10 reads with 10 of the most significant shown in the Figure. miR-143-3p had the most significant difference in expression between tissues of > 1500 FC (log10). Insufficient data was present (lack of repeats) for a fair comparison of similarities in expression.

Correlation analyses (one for the blood samples and one for the biopsy samples) were conducted to consider the interaction between the identified miRNAs that trended towards differential expression and clinical measures of ELD (Section 5.3.5., Figures 6.14. [blood] and 6.15. [biopsy]).

For the blood samples, all nine miRNAs examined showed a positive correlation with patient group (between 0.6 and 0.76). None showed correlations with clinical factors that are key to diagnosis such as eosinophil counts or total blood IgE. Two of the nine had moderate correlations with patient gender (0.53) and two showed negative correlations with patient age (-0.41 to -0.42). All these miRNAs showed strong correlations to one another with a minimum correlation of 0.89 to the highest correlation of 0.99 (Figure 6.14.). It could be inferred that gender and age were the clinical factors that might have generated most of the correlation. This is partially supported by findings in other respiratory research where DE-miRNAs were found to be correlated with age in asthma, but contrastingly also correlated with systemic inflammation measures (FeNO) and TNFα levels (Wardzyńska *et al.*, 2021). It may be that despite eosinophils and IgE being tenets in ELD diagnosis there may be other markers and pathways to consider.

For the biopsy samples, for the nine miRNAs selected miR-202-5p showed a positive correlation with patient group (0.68) (Figure 6.15.). There were also strong correlations between these trending miRNAs, with miR-31-5p having correlations between 0.95 and 1.0 with miR-202-5p and miR-203a-3p. The precursor of miR-31-5p has been identified as influencing Dicer processing (S. Ma *et al.*, 2023). Dicer is a key enzyme in the regulation of miRNAs themselves through cleavage of double stranded RNA, such as hairpin loops (latter mentioned in Chapter 1, Section 1.2.1.), into the active form of miRNA. With more miR-31-5p being produced, there would also

have been an effect on pre-miR-31-5p potentially leading to an overall impact on the regulatory processes of many other miRNAs.

A Gene Ontology analysis was conducted to identify, for the trending towards DE-miRNAs, the potential gene targets and target categories. This was carried out separately for both blood and biopsy samples. For the blood sample analysis, the gene categories with the highest significance were in 'negative regulation of apoptotic process' (for Biological Process), whilst for Molecular Function (MF) the most significant category was 'negative regulation of transcription, DNA dependent', and for Cellular Components (CC) 'cytosol' was the most significant category. There were 38 gene targets in common across all three analyses (Figure 6.16. and Table 6.5.).

For the GO analysis for the biopsy samples, the strict threshold was limiting for the analysis and consequently a more relaxed threshold had to be applied but notably this generated numerous gene targets. Further experimental validation with QPCR would need to be conducted to strengthen confidence in findings. The gene target areas that were significant were in 'viral reproductive process' for BP, 'nucleotide binding' for MF, and 'nucleoplasm' for CC (Figure 6.19., and (Supplementary: Chapter 6: (BEAM) Supplementary: [Biopsy_Gene_Ontology_GeneList]). Of note, during the correlation discussion earlier in this Chapter, miR-31-5p (that was correlated with two of the other miRNAs) was identified as having an interaction with DICER in the literature. This is further supported by the GO analysis findings where DICER1 was identified as a target in all GO categories (Supplementary: Chapter 6: (BEAM) Supplementary: [Biopsy_Gene_Ontology_GeneList]). This could suggest that the regulation of miRNAs themselves and subsequent mRNA translation are important to the pathology of ELD.

An analysis of potential internal validation miRNAs was considered as a back-up where spike-in might not be viable. miR-1-3p has previously been identified as the most (lung) tissue specific stably expressed miRNA (Ludwig *et al.*, 2016). By comparison, the same study identified miR-3960 as having the most ubiquitous expression across all tissues tested. In this present study neither miR-1-3p or miR-3690 were included in the nine miRNAs that reached the input criteria for the geoNorm analysis. The latter revealed Let-7a-5p to have the most stable expression as a reference miRNA candidate for future studies implementing internal validation.

The present study is the first of its kind in identifying circulating and biopsy miRNA expression between Healthy Controls and Eosinophilic Lung Disease. Despite a number of unforeseen challenges that impacted the final sample numbers (particularly with regard to paired samples) there nonetheless have been some interesting findings. Some of the latter would be worthy of further exploration in further cohorts of ELD patients with age and gender matched healthy controls.

7. Chapter 7: Discussion

Work detailed in this thesis commenced in September 2018 but was impacted by the COVID-19 pandemic and therefore ultimately has ended up being comprised of two research projects.

The first aimed to optimize and assess the feasibility of a novel nasal sampling method for miRNA sequencing, with the goal of performing a subsequent study making a comparison of nasal miRNA profiles of Asthmatics versus Healthy Controls using the developed methodology. Whilst the methodology was successfully developed (Chapter 3 and Section 7.1. below), the COVID-19 pandemic prevented recruitment of asthmatic and healthy controls for the second element hence the study applying the developed methodology could not take place.

A change in research direction therefore had to be adopted with instead a study of circulating miRNAs in asthmatics, healthy controls and patients with eosinophilic lung disease (ELD) being conducted using previous collected and banked samples.

The findings of these two research projects will now be discussed.

7.1. Nasosorption Sampling for small RNA Sequencing Analysis

Previous studies have shown that miRNA expression is dysregulated in asthma (Maes *et al.*, 2016). Various methods of sampling and measures for asthma research have been established,

such as sputum induction, exhaled breath condensate collection, and blood sampling, in addition to more invasive methods such as bronchoscopy (Bakakos *et al.*, 2011; Ramshankar and Krishnamurthy, 2013; Mitchell *et al.*, 2024). Nasosorption, a minimally invasive method of nasal airway sampling, has been found to be effective at detecting variations in inflammatory mediators and viral load enabling identification of differences between tested groups (Thwaites *et al.*, 2017) e.g. asthmatics versus healthy controls. This suggested that nasosorption could be a sensitive method for the sampling of nasal miRNAs the latter having the potential to act as a surrogate for the lower airways.

As yet few studies have been conducted using sequencing to identify differentially expressed miRNAs in the airways (S. Zhang *et al.*, 2018). The overall aim was to use an unbiased sequencing approach (as opposed to targeted arrays discussed in Chapter 1, Section 1.3.2.) to obtain a broad picture of the miRNA expression profiles in nasal samples of subjects. This involved the development and optimization of a nasosorption sequencing methodology with planned subsequent evaluation of said methodology in samples to be taken from a cohort of severe asthmatics and healthy controls.

7.1.1. Relevance of the Findings of Nasosorption sampling for Sequencing Analysis

Optimization of processing of Nasosorption samples was conducted for key aspects from sample storage conditions through RNA isolation and cDNA library synthesis, to preparation of cDNA library for sequencing (Chapter 3, Section 3.3.).

Optimization of the initial sampling and immediate processing and subsequent RNA extraction (through modifications of protectant fluids, tubes and reagent usage) resulted in a significant increase in RNA yield from 5-10 ng/ μL to > 150 ng/ μL of total RNA (Chapter 3, Figures 3.2. and 3.8.). This was key as concentrations of RNA less than 100 ng/ μL are insufficient input for library synthesis protocols suitable for small RNA sequencing. Further optimization was conducted for library synthesis and size selection, shown in Bioanalyzer traces of size selected libraries (Chapter 3, Figure 3.10.), to ensure the library for sequencing met the Sequencing Facility requirement of 500 ng/μL.

Samples collected using the nasosorption device have not until this project been used for miRNA sequencing. The developed method also appears to be superior to nasal tissue sampling where yields of total RNA are lower than that achieved in this thesis (Gata *et al.*, 2023). Gata *et al.*, whose study involved nasal tissue sampling, were able to achieve a 50 ng/ μL of total RNA yield which is insufficient for current next generation sequencing library protocols.

From the single healthy Nasosorption sample processed and sequenced in Chapter 3, the results revealed 49 miRNAs with reads exceeding 1,000 (Chapter 3, Figure 3.17.) sufficient for differential expression analysis. Eighteen of the 49 miRNAs matched healthy lung miRNA expression profiles reported in the literature (Chapter 3, Figure 3.18.). An unexpected finding, however, was that 70% of the reads were tRNAs (Chapter 3, Figure 3.16.) highlighting a significant technical challenge and indicating a need for refined methodologies to ensure the accuracy and relevance of the miRNA data collected from nasal samples. There are presently no publications stating the relative expression of miRNAs compared to tRNAs in nasal swab samples from sequencing analysis. There are publications highlighting that tRNAs and miRNAs are identifiable alongside one another through small RNA-seq methodologies due to their relative

size (Skeparnias *et al.*, 2020). Further work needs to be conducted to identify if this is a repeatable result. As both miRNAs and tRNAs contribute towards mRNA regulation this could be used in future to identify further interactions as well as identify pathways targeted through both small RNA species.

After successful optimization of a nasosorption sampling method for miRNA sequencing from nasal mucosal samples the research project had to be terminated, despite ethics being obtained. Recruitment of subjects for the case-control cohort was not possible due to the COVID-19 pandemic.

Therefore studies of circulatory miRNAs in asthmatics, healthy controls and eosinophilic lung diseases were performed and first the findings in relation to asthma will be discussed.

7.2. Asthma

While direct sampling of the lung is presumed to give the most accurate detailed overview of cellular and pathological processes, bronchoscopies are not best suited for repeated sampling due to their invasive nature and associated costs. This has led to a drive towards the use of liquid biopsies, as has been done with cancers, to identify differences between asthma sufferers and healthy controls (Ono *et al.*, 2015).

To address the limited scope of miRNA selection in QPCR or microarray studies, this thesis aimed to apply an unbiased sequencing approach (Belair *et al.*, 2019) to profile miRNA expression in peripheral blood samples. For the asthma study, subjects were obtained via the Celtic Fire Study (Chapter 4 Section 4.2.1.). The pilot study focused on patients with Severe asthma versus Healthy

Controls to identify if miRNA expression profiles were differentially expressed between the extreme phenotypes (Chapter 4). Following on from the pilot study, a larger study considering asthmatics of varying severities and again including healthy controls (Chapter 5) was conducted with miRNA expression profiles compared between the different groups.

It was hypothesized that there would be sufficient miRNAs in number and quantity to enable differential expression analysis. Furthermore, it was hypothesized that differentially expressed miRNAs would be observed between healthy controls and asthmatics, and that there would be differences in miRNA signatures between different severities of asthma. The goal being for Chapter 4 to be a pilot study comparing severe asthmatics with healthy controls, and Chapter 5 being a follow up larger study to identify if non-severe asthma miRNA expression was identifiable compared to healthy controls and severe asthmatics.

7.2.1. Relevance of Findings of miRNA Expression in Asthma Compared to in Health

For the pilot study circulating miRNA expression between nine severe asthmatics and nine healthy controls from the Celtic Fire Study were compared (Chapter 4). The study identified 10 differentially expressed (DE) miRNAs: miR-1304, miR-4482-3p, miR-32-5p, miR-374a-3p, miR-224-5p, miR-378c, miR-24-2-5p, miR-1307-5p, miR-17-3p, miR-378d (Chapter 4, Figure 4.8.). Gene ontology analyses were performed to explore potential target genes and biological pathways (Chapter 4, Figures 4.10., 4.11. and 4.12. and discussed later in Section 7.7. below). Of the DE-miRNAs, miR-1304 showed the highest FC (8.2) between the severe asthmatics and

healthy controls and the pre-miRNA of miR-1304 has been associated with allergic asthma in relation to methacholine challenge but with a fold change of 0.105 (Bélanger *et al.*, 2020). The same miRNA (miR-1304) has also been proposed as one of a panel of miRNAs involved in pulmonary arterial hypertension (Chen *et al.*, 2023).

The broader study in Chapter 5 identified 35 unique DE-miRNAs (shown in the Volcano plot Figure 5.11.) between the Healthy Controls vs Severe asthmatics (miR-145-5p, miR-223-3p, miR-1304-3p, miR-98-3p, miR-26b-5p, let-7c-5p, miR-98-5p, miR-6753-5p, miR-4750-5p, miR-452-5p, miR-326, let-7a-5p, miR-5698, miR-664b-3p miR-548t-5p miR-4638-5p miR-6827-3p miR-3064-3p, miR-331-3p, let-7d-5p, miR-4655-5p, miR-4775, miR-532-3p, miR-369-3p, miR-182-3p, miR-144-5p, miR-1278, miR-425-5p, miR-374a-5p, miR-339-5p miR-15a-5p miR-32-5p miR-454-3p miR-548ap-5p|miR-548j-5p, miR-296-3p), whilst five DE-miRNAs (miR-145-5p, let-7c-5p, let-7a-5p, let-7d-5p, miR-486-3p) were identified between the non-severe asthmatics and severe asthmatics, and two DE-miRNAs (miR-1304-3p, miR-5698) between the healthy controls and non-severe asthmatics.

The largest fold change (FC 8.2) miRNA of the pilot study, miR-1304-3p (Chapter 4), was replicated in the Severe versus Healthy Controls comparison in the larger study (Chapter 5) of differing severities and had a FC of 6.6. The only other miRNA seen in both the Severe versus Healthy Controls comparisons of Chapters 4 and 5 was miR-32-5p with a FC in the pilot (Chapter 4) of 2.1 and a FC of 2.7 in the larger study (Chapter 5). Considering other data obtained for the study subjects in relation to these two DE-miRNAs, in the work in the pilot study (Chapter 4) there was a significant positive correlation was identified between expression of miR-32-5p and lung function (FEV1/FVC (0.66) (Chapter 4, Figure 4.10.), but no correlation was observed with total eosinophil counts (Chapter 4, Figure 4.9.). In the follow up study (Chapter 5), however, there

was no significant correlation between miR-32-5p and any of the clinical observations (Chapter 5, Figure 5.15.). This suggests that there is still a gap in understanding of what correlations are functionally related to the alterations in miRNA expression.

Post completion of experimental work in this thesis Gil-Martinez *et al.,* (2022) published contrasting findings from serum samples of eosinophilic asthmatics. While the researchers did identify DE-miRNAs between asthmatics and healthy controls, they found no significant difference in expression of miR-32-5p between their study cohorts.

Findings from both studies were compared with relevant published literature (Chapter 4, Section 4.3.5. and Chapter 5, Section 5.3.5.) highlighting similarities and differences in miRNA expression patterns and potential targets. While miR-32-5p and miR-1304-3p were shared between the pilot study and the follow up Severe asthma vs Healthy Controls study. There is little presently in the literature presenting any association of these miRNAs with asthma.

Between the larger cohort study in Chapter 5 and the published literature (Roffel *et al.*, 2021) five DE-miRNAs were identified to be in common: let-7a-5p, miR-374a-5p, miR-223-3p, let-7c-5p, miR-425-5p. miR-223 is worth noting as it was the most highly expressed miRNA from the tested sample (healthy control) in the Nasosorption study (Chapter 3) and was identified as being expressed in the pilot Severe asthma study but not significantly different between groups (FC= 0.49, FDR = 0.97). miR-223-3p has also been identified in the studies of Maes *et al*. and Panganiban *et al*. (both discussed in Chapter 1, Section 1.5.1.1.) as a DE-miRNA negatively correlated with lung function and found to be localised to neutrophils (Maes *et al.*, 2016; Panganiban *et al.*, 2019). miR-223-3p appears to be consistent in DE-miRNA status between sample types analysed as the Panganiban study considered circulating miRNAs (as did this thesis) whilst the Maes *et al.* study considered sputum samples.

7.3. Eosinophilic Lung Disease

Eosinophilic Lung Diseases (ELD) are less common than asthma, although severe eosinophilic asthma is often clinically considered to be an ELD feature (Cottin, 2016). Consequently, the presentation, symptoms, and resulting treatments invariably similarly focus around those used for asthma (Chapter 1, Section 1.1.2.2.). Understanding the drivers of this group of eosinophilic inflammatory conditions could shed light on underlying processes towards improved outcomes through modification of treatments or developments of novel interventions. No miRNA expression sequencing analysis studies specifically in ELD have been published to date. One ELD study using QPCR on RNA from blood samples identified miR-132-3p as a potential diagnostic marker of allergic bronchopulmonary aspergillosis (ABPA) a condition classified as an ELD (Esawy *et al.*, 2021).

The goals of this thesis (Chapter 6) were to identify if there was a difference in miRNA expression between Healthy Controls and ELDs (sourced from the BEAM Study, Chapter 6, Section 6.2.1.) in whole blood as well as from biopsies from the same individuals in a paired design. In addition, the study sought to assess the feasibility of a spike-in miRNA for normalization. It was hypothesized that significantly different miRNA expression between Healthy Controls and ELDs would be observed. The study also investigated correlations between miRNA expression and clinical parameters, such as eosinophil counts, IgE levels, and lung function measures.

7.3.1. Relevance of Findings of miRNA Expression in ELD Compared to in Health

While a larger sample size was planned, due to issues with the Sequencing Facility ultimately sufficient sequencing data was only generated from blood samples for 9 subjects (6 healthy controls and 3 ELD) , and biopsy samples for 5 subjects (3 healthy controls and 2 ELDs). Five hundred and 70 miRNAs were identified with a FC greater than 0.5 when circulating miRNA expression was compared between ELDs (N = 3) and Healthy Controls (N = 6). No miRNAs, however, met the requirement of FDR to qualify as DE-miRNAs (Chapter 6, Figure 6.12.). A total of 61 miRNAs were identified that had a $P < 0.05$ and were further analysed for Gene Ontology interactions (discussed below in Section 7.7). Similarly, although 366 miRNAs were identified as having a FC of greater than 0.5 in the biopsy miRNA expression comparison between (ELDs versus Healthy Controls, 2 vs 3), none met the FDR requirement to qualify as DE-miRNAs (Chapter 6, Figure 6.12.). A total of 31 miRNAs, however, had a $P < 0.05$ and were therefore further analysed for Gene Ontology interactions (discussed below in Section 7.7).

Conservative results are welcome where limited sample size is present but clearly further work is required with a larger cohort to provide the statistical power to generate a clearer picture of miRNA expression in ELD, and differences between miRNA expression in the lung and the blood.

Chapter 6 also considered the feasibility of using a spike-in miRNA (cel-miR-29) as a normalization tool for low-input samples in the context of the ELD study. While the spike-in miRNA was identified in some samples, the overall low sequencing depth limited its utility for normalization purposes. Other researchers have had success with the application of a spike-in

with promising results suggesting the proposed benefits of their use could better assist with the development of diagnostic or prognostic markers for disease (Khamina *et al.*, 2022).

7.4. miRNA Expression between Tissues

While it was not possible to complete the comparison of miRNA expression between biopsies and whole blood in a number of healthy controls and patients with ELD (Chapter 6, Section 6.3.4.), there was the opportunity to compare the differences in miRNA expression in a single healthy control. In the single within subject (healthy control) biopsy and blood sample comparison there were some significant differences in miRNA expression (FC) with miR-143-3p having a log 10 FC of 2,183. Expression was higher in biopsy samples across all shown miRNAs (Chapter 6, Figure 6.14.) which could suggest a stronger signal from the tissue of interest. While miRNA expression did show some differentially expressed miRNAs between tissue types (biopsy versus whole blood), there were also several miRNAs that were highly expressed and comparable between tissues (Chapter 6, Section 6.3.4., Figures 6.13.) This could suggest that for expression of some miRNAs, distal or whole blood sampling could be further developed for diagnostic or monitoring purposes.

To the author's knowledge currently, within the published literature, there are no studies of miRNA sequencing comparison of paired lung biopsy and blood sample from either healthy individuals or individuals with ELD. To date, most miRNA expression analysis experiments conducted have used selected panels of miRNAs that have been assessed either by QPCR or microarrays. One such study conducted on samples from cadavers (a patient that died from

natural causes and one from cancer) using microarray showed there were identifiable differences between tissues, more so than between individuals (Ludwig *et al.*, 2016). miR-143-3p was not mentioned in the paper but may not have been included in the microarray used. Only 143 miRNAs were identified in each tissue tested in the Ludwig study, compared to the present study which by using an unbiased approach of sequencing identified 1,583. This supports further comparison of blood and biopsy samples in disease conditions to understand the underlying differences in expression and for the potential extrapolation of use as diagnostic biomarkers.

7.5. miRNA Expression between Inflammatory Respiratory Conditions

miRNAs have been shown play a critical role in regulating gene expression associated eosinophillic inflammation (Huo *et al.*, 2016). For asthma, specific miRNA signatures, including the let-7 family, have been identified to differentiate between severe and mild forms of the disease (Rijavec *et al.*, 2014). Similarly in ELD, distinct miR-132 expression patterns were observed in QPCR and microarray studies when comparing patients to healthy controls suggesting a potential role in disease mechanisms (Esawy *et al.*, 2021).

In the asthma studies, miRNAs, including let-7a-5p, let-7c-5p, and let-7d-5p, were differentially expressed but not correlated with severity (Chapter 5, Figures 5.10. and 5.14.). These miRNAs were upregulated in severe asthma compared to healthy controls but downregulated when comparing severe to non-severe asthma (Chapter 5, Figure 5.11.). The expression of the let-7 family has previously been found to be dysregulated in asthma (Rijavec *et al.*, 2014). As this particular family of miRNAs are well studied, they have also been found to be dysregulated in

other conditions including various cancers and COPD (Shell *et al.*, 2007; Deshmukh *et al.*, 2021; Di *et al.*, 2021) which may limit their application as a diagnostic or prognostic measure where comorbidities are present.

miR-145-5p, by contrast, showed an inverse expression pattern being downregulated in severe asthma compared to healthy controls and upregulated when comparing non-severe to severe asthma (Chapter 5, Figures 5.10. and 5.11.). This suggests that expression is dynamic based on the disease severity and that there may be further interactions regulating expression. While this has not been shown in miRNA expression in asthma previously, findings in epilepsy research have identified changes during developmental phases, some of which reflect dysregulated states that lead to epilepsy (Bencurova *et al.*, 2021). There is support for this theory in COPD research of the developing lung with downregulation of miR-145-5p expression linked to reduced growth and early decline of lung function (Tiwari *et al.*, 2021).

In the ELD study, differences (trending towards significance) in miRNA expression were observed between ELD patients and healthy controls, with a noted impact on clinical variables such as IgE levels, eosinophil counts, and lung function measures (FEV1/FVC ratio). From the biopsy ELD analysis, nine miRNAs trending towards differential expression showed strong correlations with patient gender, BMI and age but did not correlate significantly with eosinophil counts or total IgE (Chapter 6, Figure 6.14.).

Several DE or trending miRNAs were in common between the asthma and ELD cohorts. For example, miR-223-3p (already discussed above in Section 7.2.1.) was the miRNA in blood samples from ELD patients that trended towards the most significant difference in expression.

Thematically, there are nascent associations of the role of miRNAs involved in inflammatory pathways and immune regulation in both asthma and ELD (discussed in Chapter 1, Section 1.5. and further below). In relation to the asthma studies, the let-7 family miRNAs are known to be involved in airway remodelling and immune response modulation (Levänen *et al.*, 2013). These miRNAs target IL-6 and play a role in myofibroblast differentiation, contributing to the chronic nature of asthma (Lee *et al.*, 2015). The differential expression patterns suggest a complex regulatory mechanism, of the let-7 family of miRNAs in this present study, that may not follow a linear progression with asthma disease severity. For the ELD patients, a strong influence of gender and age on miRNA expression was observed which might suggest a differential response to inflammatory stimuli in different demographic groups. Additionally, miRNAs in ELD, such as miR-202-5p, showed correlations with clinical measures like eosinophil counts and BMI, highlighting their potential role in disease pathogenesis and progression (Chapter 6, Section 6.3.5., Figure 6.15.). miR-202-5p has been previously reported to be elevated in peripheral blood mononuclear cells in patients with allergic rhinitis and linked to inflammation through regulatory T cell functions (Wang *et al.*, 2019). This could suggest that miR-202-5p contributes to the regulation of T cells and may contribute to the steroid resistance that is associated in some cases of the ELD variant, Chronic Eosinophilic Pneumonia (CEP).

7.6. The Relationship between miRNA Expression and other Clinical Variables

The clinical variables sex, age and BMI were controlled for during subject selection (Chapters 4 and 5) as they are known risk factors for asthma and may therefore contribute to miRNA

expression profiles in a deleterious manner (Peters, Dixon and Forno, 2018; Sinyor and Perez, 2023). For the pilot study (Severe asthma versus Healthy Controls, Chapter 4), a moderate correlation was seen with eosinophils and miR-378d and miR-378c (0.63, 0.61 respectively), and FEV1/FVC ratio (0.66) (Chapter 4, Section 4.10.). There were no significant correlations with any DE-miRNAs in the broader cohort study for the Control vs Severe comparison (Chapter 5). For the Control vs Non-Severe and Non-Severe vs Severe only a single correlation between a miRNA and BMI was observed (Non-Severe: miR-1304-3p [-0.34] and Severe: miR-145-5p [0.38] respectively). This is not a surprise as these factors were for the most part controlled for during sample selection to avoid mistaking interactions of risk factors with miRNA expression in particular due to whole blood being circulatory potentially having contributions of miRNAs from tissues other than the lung (Iacomino and Siani, 2017; Llera-Oyola *et al.*, 2024). This also tells us that the DE-miRNAs identified were not strongly associated with core markers of asthma (eosinophils) for the most part. This raises questions around correlation and causation. While key diagnostic markers are not correlated, it could be a pathway not considered in the diagnostic pathway that is dysregulated.

7.7. The Relationship between miRNA Expression and Gene Ontology Results

All three miRNA expression studies, the pilot study (Chapter 4), the larger cohort study (Chapter 5) and the ELD study (Chapter 6), employed Gene Ontology (GO) analyses. The purpose of these analyses being to investigate the potential target genes and associated biological pathways of the

identified DE-miRNAs and thereby provide insights into the molecular mechanisms underlying asthma and ELD pathogenesis and severity.

For the pilot asthma study (Chapter 4), in the Cellular Components (CC) category of the Gene Ontology analysis membrane and cell surface markers made up the majority of the gene targets. This suggests that the cellular interaction with the environment is a strong component underpinning the changes seen in miRNA expression between healthy controls and severe asthmatics. A comparison between the two asthma studies (Chapters 4 and 5) identified gene targets *TRAF3* and *PTEN* were shared between the pilot study and Control vs Severe follow up study. Interestingly *TRAF3* has been experimentally linked with asthma in relation to TGF-β1 driven airway inflammation and remodelling. It has been shown that *TRAF3* is deubiquinated by *OTUB1* to activate the *NLRP3* inflammasome, resulting in inflammation and remodelling of TGF-β1-induced cells (Shang *et al.*, 2023). This could link the association with the GO CC result mentioned above (cell surface marker gene targets) if cells need to be responsive to external inflammatory signals. PTEN has also been identified as downregulated in asthma and was linked with airway hyperresponsiveness as a consequence of decreased synthesis of the protein (Kwak *et al.*, 2003). This supports the expected change that would occur due to increased miRNA degradation of PTEN synthesis. That there is support from published literature of the results of this thesis is encouraging and shows that the integrated and non-biased method of miRNA analysis applied in this thesis could raise awareness of key targets or diagnostic methods that might not have previously been of significant interest.

For the asthma studies key Gene ontology Biological Process (BP) categories were identified as being stress response categories i.e. Regulation of phosphorylation, regulation of metabolic process, regulation of Molecular Function. Cellular stress can activate pathways that both inhibit

apoptosis and regulate protein metabolism. Such pathways allow the cell to adapt and survive under adverse conditions. One of the targets found in the key BP categories was *ICAM1* (Chapter 4, Section 4.3.4.). The protein ICAM1 is recognised to play a key role with regard to inflammation through regulation of the cellular response (Zhang *et al.*, 2019; Bui, Wiesolek and Sumagin, 2020) but links with ORMDL3, the major genetic risk factor for childhood asthma (Chapter 1, Section 1.1.1.). Knockdown of ORMDL3 expression has been shown to result in significant reduced expression of ICAM1 ((Zhang *et al.*, 2019; Bui, Wiesolek and Sumagin, 2020)*et al., 2019*). A study using the microarray approach for identifying associated miRNAs for severe asthma identified more specific inflammatory categories in the GO analysis conducted such as T cell activation and regulation and leukocyte cell-cell adhesion (Li *et al.*, 2022). While the HumanRef-8 v3 microarray contains probes for 22,000 transcripts this includes lncRNA, miRNA, mRNA and others it is not a complete representation of the human genome and notably the smaller RNA species. This likely underpins the difference seen between the present unbiased sequencing study and that of Li *et al.*

In the blood ELD comparison (Chapter 6) 61.3% of gene targets were the same across all three GO domains tested (Cellular Components [CC], Biological Process [BP], and Molecular Function [MF]) from the ELD blood trending towards significant miRNAs (N = 61) (Chapter 6, Figure 6.16.) . Gene targets with the highest number of shared source miRNAs included *IL-6R*, *HMGA2*, *E2F2*, *CDK6*, *MCL1*, and *IGF1R*. In the case of the biopsy ELD cohort (Chapter 6) approximately 16% of genes identified in common between the GO domains (BP, MF, CC) (Figure 6.20.) from the input trending biopsy miRNAs (N = 31). Several gene targets were identified that shared several source miRNAs (*IGF1*, *CDKN1A*, *AGO1*, *PPP2R2A*, *E2F2*, *SMAD2*, *BCL2L1*, *BCL2*, *ZEB1*, *SPRED1*). IGF1 has been proposed as a potential target in allergic asthma previously (Jiang *et al.*, 2023). The researchers identified that circulating IGF1 may be causally related to a lower risk of asthma. This supports the findings of this present study as if multiple miRNAs are targeting and therefore

downregulating IGF1 expression then there will be increased symptoms. Another gene of note that shared several source miRNAs was *AGO1*. The gene encodes Argonaute 1 a key protein in miRNA function (previously discussed Chapter 1, Section 1.3.1.). It has been proposed that Argonaute 1 could act as a negative regulatory pathway to reduce miRNA regulation of gene expression and it has been suggested that this could be a method of intercellular communication to have a tissue or organism wide effect of self-regulation (Turchinovich and Burwinkel, 2012). It could be therefore speculated that being a target of several source miRNAs would deplete its expression thereby impacting the expression of other miRNAs whose target genes contribute to sustaining pathways that play important roles in the ELD phenotype.

There were no similar GO domain categories between the blood and biopsy cohorts (Chapter 6, Figures 6.16. and 6.19.). In the blood ELD study key categories from each domain were:- BP: negative regulation of apoptotic process (22 gene targets), MF: Negative regulation of transcription, DNA dependent (19 gene targets), CC: cytosol (27 gene targets). Whilst in the biopsy ELD study key domain categories were:- BP: Viral reproductive process (105 gene targets), MF: nucleotide binding (340 gene targets), CC: nucleoplasm (260 gene targets). Gene Ontology results should be interpreted with caution in light of lowering the threshold with regard to the identification of the DE-miRNAs.

7.8. Strengths and Limitations of the Studies

The miRNA expression studies conducted in this thesis demonstrate several significant strengths. The use of an unbiased sequencing approach allowed for a comprehensive analysis of miRNA expression profiles, providing a broad and detailed picture of the miRNA landscape in these conditions. This approach facilitated the identification of novel miRNAs that may play critical

roles in disease pathogenesis. The optimization of RNA extraction, library synthesis, and size selection techniques ensured high-quality data, which enhances the reliability of the findings. Moreover, the studies' focus on both asthma and ELD allowed for comparative analysis, offering insights into the common and distinct molecular mechanisms underlying these inflammatory lung conditions. These strengths collectively contribute to the advancement of miRNA research in respiratory diseases paving the way for future diagnostic and therapeutic developments.

The presented work in this thesis did identify some limitations that impacted the findings and their generalizability. One significant limitation was the sample size, particularly for the ELD study. A larger sample size could give more power to detect significant DE-miRNAs. As ELD incidence is low, a multicentre as opposed to single centre (that the BEAM study was) study involving recruitment of more patients would be required. With successful optimisation of the SAM processing RNA extraction and sequencing pipeline this readily affords the opportunity to increment sample sizes of controls, asthmatics and those with ELD by nature of it being a less invasive procedure. SAMs have also been successfully used for sampling of the lower airways through bronchoscopy and are more tolerated than brushing or biopsy of the airways. This means that there are opportunities to conduct further larger scale studies sampling and comparing nasal, lower airway and blood miRNA profiles in an unbiased manner through next generation sequencing.

One limitation of this study is its reliance on whole blood counts, which provide an overall picture of immune cell populations but lack the granularity to pinpoint specific cell-type changes (Devonshire *et al.*, 2021). However, the use of whole blood may allow for detection of other relevant pathways that may have been missed by more focussed sequencing. Consequently, it's challenging to determine whether the observed shifts in miRNA expression are attributable to

disease progression or fluctuations in the proportions of various cell types within the blood. To address this, future studies could employ cellular deconvolution techniques. These computational methods can estimate the relative abundance of different cell types within a mixed population, like blood, based on gene expression or other molecular profiles (White *et al.*, 2024). This approach may help identify whether specific miRNAs are associated with changes in particular cell types, offering more precise insights into the underlying biological mechanisms. Furthermore, interrogation of publicly available blood and lung cell atlases could provide valuable context, potentially revealing cell-type-specific miRNA expression patterns and their relevance to disease pathogenesis (Sikkema *et al.*, 2023).

There were some challenges in the sample preparation for sequencing. One of the significant technical challenges encountered was the difficulty in lysing samples, especially lung tissue, for RNA extraction. This step is crucial for obtaining high-quality RNA and improvements in this area could enhance the reliability of results. The use of a bead beating device did facilitate good RNA yields (Chapter 6, Section 6.3.), demonstrating its effectiveness in overcoming some of the extraction challenges identified in biopsy RNA extraction. The synthesis of the RNA library and the selection of spike-in controls were carefully managed, although optimization is required to adapt these steps to specific laboratory conditions. Size selection using the Pippin Prep system was very precise, and the removal of salts is recommended to ensure the best results. Further optimization of pre-size selection methods could improve the efficiency and reliability of this step (Chapter 6, Section 6.4., and Section 7.9). Specifically, QPCR of raw libraries, if possible with small volumes, would ensure the presence of miRNAs and the relative concentration of spike-in to ensure that spike in is not the sole miRNA present.

Read depth is also an important consideration. If too deep cost efficacy wanes, while too shallow it limits analysis and the potential identification of all present miRNAs. One study showed that an average sequencing depth of ~5 million (M) reads per sample generated DE-miRNAs with RPM (reads per million) of greater than 1 million being reproducible (R = 0.88) when the same samples were sequenced to greater depths (Campbell *et al.*, 2015). For the BEAM study (Chapter 6), reads were overall below 5M in the samples that were successfully identified (N = 13) with only 5 samples achieving greater than 5M reads.

Due to challenges with equipment and Sequencing Facility issues, there was a significant impact on the yield of reads generated for the ELD versus Healthy Controls study (Chapter 6). A more relaxed threshold was applied to the ELD study samples to accommodate this limitation and enable further data analysis. Although no significant differences as far as the strict FC of > 0.5 and FDR > 0.05 were not met together, there were numerous miRNAs with a significant FC between ELDs and Healthy Controls (blood N = 660, biopsy N = 626) (number potentially inflated due to the relaxed threshold). As illustrated by Figure 6.10. (Chapter 6), there were also lower total numbers of miRNAs in the affected run. This suggests that ensuring a minimum read count is crucial for both detecting and obtaining a true representation of miRNA expression within a sample (Conesa et al., 2016).

In the miRNA target analysis, there was often no option for empirical sampling available on other platforms to enable comparison with the results in the literature (Fan et al., 2016). A lack of transferable or comparable results within the literature affects the level by which studies can be compared and could result in non-replicability or inaccurate conclusions that could hinder progress in identification of key molecules involved in the condition of interest (particularly as we are now able to consider to a single cell level). This limitation highlights the need for standardized

methodologies and platforms to facilitate cross-study comparisons and validate findings across different research contexts.

Addressing these limitations in future studies by increasing sample sizes, optimizing technical procedures, and standardizing methodologies will enhance the reliability and applicability of miRNA research in asthma and ELD. This will ultimately contribute to a more comprehensive understanding of these diseases and their underlying molecular mechanisms.

7.9. Future Work

As highlighted above, the results of the Nasosorption study conducted in Chapter 3 affords several opportunities for conducting comparison studies of miRNAs between asthmatics and healthy controls as well as explorations of other respiratory diseases. The methodology is suitable for repeated application without discomfort. As the SAMs have been successfully used for cytokine analysis there is the potential for assessment of both miRNAs and key cytokines pending respiratory disease of interest including allergen challenge studies.

For the circulating miRNA studies conducted in Chapter 4 and 5, future research should aim to include larger and more diverse populations. Instead of excluding smokers entirely, studies could include smokers and in the downstream differential expression analysis incorporate smoking status as a covariate. This approach would control for smoking effects while increasing the sample size and the applicability of the findings to real-world patient populations.

In addition, conducting longitudinal studies would provide a clearer understanding of whether miRNA expression differences between healthy controls and individuals with asthma or ELD are consistent over time or related in a linear fashion as the disease progresses. Such studies would

help in determining associations or drivers of miRNA changes in relation to disease development and progression. This could be particularly helpful in understanding any changes in miRNA expression for example during allergen challenge in asthma, or to understand how interventions and common treatments alter expression, as discussed in Chapter 1 (Section 1.1.6.).

A more integrated view of the molecular mechanisms involved could be garnered by conducting within the same sample/subject comprehensive global transcriptomic (mRNA) sequencing in combination with miRNA sequencing thereby providing a more complete picture of the gene expression profiles and the potential subsequent biological impact of any observed differences in miRNA levels.

Regarding technical processes, not only were protectant fluids an invaluable addition to protocols despite literature suggesting the stability of miRNAs (Chapter 3), but further work to establish the addition of spike-ins for normalisation (to control for variance in how samples respond to the library synthesis protocols) is needed. Size selection of libraries was critical for success of sequencing but this could potentially be circumvented in the future as methods of library preparation are evolving rapidly including the development of circular libraries (Lama *et al.*, 2019).

Following up with validation experiments is crucial, particularly for findings with Gene Ontology results that lack sufficient experimental support. Future studies should also consider additional factors and clinical outcomes, such as fractional exhaled nitric oxide (FeNO), and interactions with bacterial and fungal infections, which are known to impact miRNA expression (Muhammad *et al.*, 2015; Mukherjee *et al.*, 2020; Wardzyńska *et al.*, 2021). Incorporating these elements will provide a more comprehensive understanding of the disease mechanisms and their practical implications in how the body reacts wholistically to the disease.

There is also the issue of validating DE-miRNA findings from both microarray and sequencing approaches. Although QPCR is the gold standard for measuring miRNA expression levels, other molecular genetic methods that can assess the absolute number of miRNA molecules do exist. One multiplex system used for measuring gene expression is NanoString nCounter (Veryaskina *et al.*, 2022). This technique uses digital barcoding, requires neither mRNA or miRNA-to-cDNA conversion by reverse transcription nor cDNA amplification by PCR (Tay *et al.*, 2017).

As advancements in computational methods and artificial intelligence develop and are adopted, these will no doubt enable the inclusion of broader and more diverse populations that are more representative of real-world conditions. These methods can help manage complex datasets and control for multiple variables, enhancing the relevance and applicability of the research findings.

Put together, expanding future research to include larger and more diverse populations, incorporating longitudinal study designs, and integrating comprehensive genomic analyses will significantly advance the understanding the role of miRNAs in health and respiratory diseases such as asthma and ELD.

7.10. Summary

Through a series of studies involving next-generation sequencing and bioinformatic analyses, differentially expressed (DE) miRNAs were identified in several comparisons conducted between Healthy Controls, Non-Severe asthmatics, Severe asthmatics, and individuals with ELDs.

The initial study focused and successfully optimized a nasosorption sampling method for miRNA sequencing from nasal mucosal samples. This non-invasive approach demonstrated the feasibility

of obtaining sufficient RNA for miRNA profiling and revealed shared miRNAs with a published lung biopsy study, supporting the approach's potential utility for respiratory disease research.

Subsequent studies analysed circulating miRNA expression in whole blood samples, first in a pilot study comparing severe asthmatics and healthy controls, and then in a larger cohort including non-severe asthmatics. These studies identified distinct miRNA signatures associated with asthma severity, such as miR-1304-3p, miR-32-5p, and members of the let-7 family. Gene ontology analyses of the DE-miRNAs implicated their involvement in processes like protein metabolic process, cellular metabolism, and enzyme binding, providing insights into potential underlying molecular mechanisms of disease.

The final study explored miRNA expression in both whole blood and lung biopsy samples from individuals with ELD and Healthy Controls. Although no DE-miRNAs were identified, likely due to limitations in sample size and sequencing depth, correlation analyses and Gene Ontology analyses revealed not significant but potential associations between trending miRNAs and clinical features, as well as miRNA involvement in processes like apoptosis regulation, cell proliferation, and transcriptional regulation. The identification of stably expressed miRNAs as potential reference genes for normalization purposes was also addressed, highlighting the need for methodological refinement in miRNA expression studies.

While further validation and functional characterization are necessary, these findings contribute to the understanding of miRNA dysregulation in respiratory diseases and their potential as biomarkers or therapeutic targets. The studies also underscore the complexity of miRNA regulatory networks and the need for integrative approaches to decipher their roles in disease pathogenesis.

The use of next-generation sequencing (NGS) and the resulting quality of reads and alignments further validate the methodology, as reflected in the high alignment percentages to both the genome and mature miRNA sense sequences. The depth and breadth of data that is provided by sequencing and analysis could be key for understanding the underlying mechanisms, advancing the field and developing new diagnostic tools and treatments for asthma.

In conclusion, this thesis provides a comprehensive analysis of miRNA expression in asthma, revealing important regulatory roles and potential biomarkers for the disease. The study's robust methodology and alignment with existing literature underscore its contributions to the field. While the goals of the thesis were to identify miRNA expression between various respiratory conditions, the resulting workflow of sample processing and analysis has also been developed to address the unique needs of miRNAseq research. Future research building on these findings will likely lead to more effective and personalized approaches to asthma diagnosis and treatment.

Moving forward, the immediate next steps in relation to the work conducted in this thesis would be:

1. Validate findings with QPCR, using at least two reference miRNAs identified in the sample set as being stably expressed as detailed above (Section 7.10.).
2. Seek replication of findings in larger cohorts to confirm the robustness and generalizability of the identified miRNA signatures across diverse populations.

With long term aims being to:

1. Conduct functional studies to investigate the biological roles and mechanisms of action of the identified DE-miRNAs in the context of asthma pathogenesis.

2. Integrate the miRNA data with other omics data (e.g., transcriptomics, proteomics) to gain a comprehensive understanding of the molecular networks and pathways involved in asthma severity.
3. Investigate the potential therapeutic applications of miRNA modulation, either through miRNA mimics or inhibitors, as a novel approach for targeting specific miRNAs implicated in asthma severity.

By furthering an understanding of the DE-miRNAs in asthma, there is the potential for development of valuable tools for improving the diagnosis, prognosis, and potential treatment strategies for individuals with varying severities of asthma.

8. References

Aaron, S.D., Vandemheen, K.L., Whitmore, G.A., Bergeron, C., Boulet, L.-P., Côté, A., McIvor, R.A., Penz, E., Field, S.K., Lemière, C., Mayers, I., Bhutani, M., Azher, T., Lougheed, M.D., Gupta, S., Ezer, N., Licskai, C.J., Hernandez, P., Ainslie, M., Alvarez, G.G., Mulpuru, S. and UCAP Investigators (2024) 'Early Diagnosis and Treatment of COPD and Asthma - A Randomized, Controlled Trial', *The New England journal of medicine*, 390(22), pp. 2061–2073.

Agache, I., Beltran, J., Akdis, C., Akdis, M., Canelo-Aybar, C., Canonica, G.W., Casale, T., Chivato, T., Corren, J., Del Giacco, S., Eiwegger, T., Firinu, D., Gern, J.E., Hamelmann, E., Hanania, N., Mäkelä, M., Hernández-Martín, I., Nair, P., O'Mahony, L., Papadopoulos, N.G., Papi, A., Park, H.-S., Pérez de Llano, L., Posso, M., Rocha, C., Quirce, S., Sastre, J., Shamji, M., Song, Y., Steiner, C., Schwarze, J., Alonso-Coello, P., Palomares, O. and Jutel, M. (2020) 'Efficacy and safety of treatment with biologicals (benralizumab, dupilumab, mepolizumab, omalizumab and reslizumab) for severe eosinophilic asthma. A systematic review for the EAACI Guidelines - recommendations on the use of biologicals in severe asthma', *Allergy*, 75(5), pp. 1023–1042.

Alfaro-Arnedo, E., López, I.P., Piñeiro-Hermida, S., Ucero, Á.C., González-Barcala, F.J., Salgado, F.J. and Pichel, J.G. (2021) 'IGF1R as a Potential Pharmacological Target in Allergic Asthma', *Biomedicines*, 9(8). Available at: https://doi.org/10.3390/biomedicines9080912.

Alipoor, S.D., Adcock, I.M., Garssen, J., Mortaz, E., Varahram, M., Mirsaeidi, M. and Velayati, A. (2016) 'The roles of miRNAs as potential biomarkers in lung diseases', *European journal of pharmacology*, 791, pp. 395–404.

Al-Rashoudi, R., Moir, G., Al-Hajjaj, M.S., Al-Alwan, M.M., Wilson, H.M. and Crane, I.J. (2019) 'Differential expression of CCR2 and CX3CR1 on CD16+ monocyte subsets is associated with asthma severity', *Allergy, asthma, and clinical immunology: official journal of the Canadian Society of Allergy and Clinical Immunology*, 15, p. 64.

Anauate, A.C., Leal, M.F., Wisnieski, F., Santos, L.C., Gigek, C.O., Chen, E.S., Geraldis, J.C., Calcagno, D.Q., Assumpção, P.P., Demachki, S., Arasaki, C.H., Lourenço, L.G., Artigiani, R., Burbano, R.R. and Smith, M.A.C. (2017) 'Identification of suitable reference genes for miRNA expression normalization in gastric cancer', *Gene*, 621, pp. 59–68.

Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. (2004) '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 research*, 64(15), pp. 5245–5250.

Andrews, S. and Others (2010) 'FastQC: a quality control tool for high throughput sequence data'. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

An, T., Fan, T., Zhang, X.Q., Liu, Y.-F., Huang, J., Liang, C., Lv, B.-H., Wang, Y.-Q., Zhao, X.-G., Liu, J.-X., Fu, Y.-H. and Jiang, G.-J. (2020) 'Publisher Correction: Comparison of Alterations in miRNA Expression in Matched Tissue and Blood Samples during Spinal Cord Glioma Progression', *Scientific reports*, 10(1), p. 6871.

Argyropoulos, C., Etheridge, A., Sakhanenko, N. and Galas, D. (2017) 'Modeling bias and variation in the stochastic processes of small RNA sequencing', *Nucleic acids research*, 45(11), p. e104.

Bagley, D., Fox, K., Redd, P.F., Joseph, M., Ortiz-Zapater, E., Parsons, M. and Rosenblatt, J. (2019) 'Blocking Cell Extrusion Prevents Bronchoconstriction-Induced Airway Epithelial Damage and Inflammation'. Available at: https://doi.org/10.2139/ssrn.3493955.

Bakakos, P., Schleich, F., Alchanatis, M. and Louis, R. (2011) 'Induced sputum in asthma: from bench to bedside', *Current medicinal chemistry*, 18(10), pp. 1415–1422.

Bao, Z., Lim, S., Liao, W., Lin, Y., Thiemermann, C., Leung, B.P. and Wong, W.S.F. (2007) 'Glycogen synthase kinase-3beta inhibition attenuates asthma in mice', *American journal of respiratory and critical care medicine*, 176(5), pp. 431–438.

Barbato, C., Arisi, I., Frizzo, M.E., Brandi, R., Da Sacco, L. and Masotti, A. (2009) 'Computational challenges in miRNA target predictions: to be or not to be a true target?', *Journal of biomedicine & biotechnology*, 2009, p. 803069.

Barnes, P.J. (2006) 'Corticosteroid effects on cell signalling', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 27(2), pp. 413–426.

Barnes, P.J. (2006) 'How corticosteroids control inflammation: Quintiles Prize Lecture 2005', *British journal of pharmacology*, 148(3), pp. 245–254.

Barnes, P.J. (2016) 'Asthma mechanisms', *Medicine*, 44(5), pp. 265–270.

Barnes, P.J. (2018) 'Targeting cytokines to treat asthma and chronic obstructive pulmonary disease', *Nature reviews. Immunology*, 18(7), pp. 454–466.

Barnes, P.J. and Pedersen, S. (1993) 'Efficacy and Safety of Inhaled Corticosteroids in Asthma', *The American review of respiratory disease*, 148(4_pt_2), pp. S1–S26.

Beasley, R., Semprini, A. and Mitchell, E.A. (2015) 'Risk factors for asthma: is prevention possible?', *The Lancet*, 386(9998), pp. 1075–1085.

Belair, C.D., Hu, T., Chu, B., Freimer, J.W., Cooperberg, M.R. and Blelloch, R.H. (2019) 'High-throughput, Efficient, and Unbiased Capture of Small RNAs from Low-input Samples for Sequencing', *Scientific reports*, 9(1), p. 2262.

Bélanger, É., Madore, A.-M., Boucher-Lafleur, A.-M., Simon, M.-M., Kwan, T., Pastinen, T. and Laprise, C. (2020) 'Eosinophil microRNAs Play a Regulatory Role in Allergic Diseases Included in the Atopic March', *International journal of molecular sciences*, 21(23). Available at: https://doi.org/10.3390/ijms21239011.

Bencurova, P., Baloun, J., Hynst, J., Oppelt, J., Kubova, H., Pospisilova, S. and Brazdil, M. (2021) 'Dynamic miRNA changes during the process of epileptogenesis in an infantile and adult-onset model', *Scientific reports*, 11(1), p. 9649.

Bhankhur, D., Singla, N., Aggarwal, D. and Chander, J. (2019) 'Prevalence of allergic bronchopulmonary aspergillosis among patients with severe bronchial asthma in a tertiary care hospital in Northern India', *Indian journal of pathology & microbiology*, 62(1), pp. 111–113.

Bignold, R., Shammout, B., Rowley, J.E., Repici, M., Simms, J. and Johnson, J.R. (2022) 'Chemokine CXCL12 drives pericyte accumulation and airway remodeling in allergic airway disease', *Respiratory research*, 23(1), p. 183.

Booton, R. and Lindsay, M.A. (2014) 'Emerging role of MicroRNAs and long noncoding RNAs in respiratory disease', *Chest*, 146(1), pp. 193–204.

Bossley, C.J., Saglani, S., Kavanagh, C., Payne, D.N.R., Wilson, N., Tsartsali, L., Rosenthal, M., Balfour-Lynn, I.M., Nicholson, A.G. and Bush, A. (2009) 'Corticosteroid responsiveness and clinical characteristics in childhood difficult asthma', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 34(5), pp. 1052–1059.

Bousquet, J., Chanez, P., Vignola, A.M., Lacoste, J.Y. and Michel, F.B. (1994) 'Eosinophil inflammation in asthma', *American journal of respiratory and critical care medicine*, 150(5 Pt 2), pp. S33–8.

Brommage, R., Powell, D.R. and Vogel, P. (2019) 'Predicting human disease mutations and identifying drug targets from mouse gene knockout phenotyping campaigns', *Disease models & mechanisms*, 12(5). Available at: https://doi.org/10.1242/dmm.038224.

Brown, R.A.M., Epis, M.R., Horsham, J.L., Kabir, T.D., Richardson, K.L. and Leedman, P.J. (2018) 'Total RNA extraction from tissues for microRNA and target gene expression analysis: not all kits are created equal', *BMC biotechnology*, 18(1), p. 16.

Bui, T.M., Wiesolek, H.L. and Sumagin, R. (2020) 'ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis', *Journal of leukocyte biology*, 108(3), pp. 787–799.

Buschmann, D., Haberberger, A., Kirchner, B., Spornraft, M., Riedmaier, I., Schelling, G. and Pfaffl, M.W. (2016) 'Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow', *Nucleic acids research*, 44(13), pp. 5995–6018.

Busse, W., Corren, J., Lanier, B.Q., McAlary, M., Fowler-Taylor, A., Cioppa, G.D., van As, A. and Gupta, N. (2001) 'Omalizumab, anti-IgE recombinant humanized monoclonal antibody, for the treatment of severe allergic asthma', *The journal of allergy and clinical immunology*, 108(2), pp. 184–190.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F. and Croce, C.M. (2002) 'Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), pp. 15524–15529.

Calvén, J., Ax, E. and Rådinger, M. (2020) 'The Airway Epithelium-A Central Player in Asthma Pathogenesis', *International journal of molecular sciences*, 21(23). Available at: https://doi.org/10.3390/ijms21238907.

Campbell, J.D., Liu, G., Luo, L., Xiao, J., Gerrein, J., Juan-Guardela, B., Tedrow, J., Alekseyev, Y.O., Yang, I.V., Correll, M., Geraci, M., Quackenbush, J., Sciurba, F., Schwartz, D.A., Kaminski, N., Johnson, W.E., Monti, S., Spira, A., Beane, J. and Lenburg, M.E. (2015) 'Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data', *RNA* , 21(2), pp. 164–171.

Carrington, C.B., Addington, W.W., Goff, A.M., Madoff, I.M., Marks, A., Schwaber, J.R. and Gaensler, E.A. (1969) 'Chronic eosinophilic pneumonia', *The New England journal of medicine*, 280(15), pp. 787–798.

Carr, T.F., Moore, W.C., Kraft, M., Brusselle, G., Castro, M., Chupp, G.L., Wechsler, M.E., Hunter, G., Lindsley, A.W., Llanos, J.-P., Burke, L.K., Chandarana, S. and Ambrose, C.S. (2024) 'Efficacy of tezepelumab in patients with severe, uncontrolled asthma across multiple clinically relevant subgroups in the NAVIGATOR study', *Advances in therapy*, 41(7), pp. 2978–2990.

Chang, L., Zhou, G., Soufan, O. and Xia, J. (2020) 'miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology', *Nucleic acids research*, 48(W1), pp. W244–W251.

Chaplin, S. (2018) 'Diagnosis, monitoring and management of chronic asthma', *The Prescriber*, 29. Available at: https://doi.org/10.1002/PSB.1665.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W. and Hannon, G.J. (2010) 'A dicer-independent miRNA biogenesis pathway that requires Ago catalysis', *Nature*, 465(7298), pp. 584–589.

Chen, S., Zhong, J., Hu, B., Shao, N. and Deng, C. (2023) 'Whole-genome microRNA sequencing analysis in patients with pulmonary hypertension', *Frontiers in genetics*, 14. Available at: https://doi.org/10.3389/fgene.2023.1250629.

Chowdhury, N.U., Guntur, V.P., Newcomb, D.C. and Wechsler, M.E. (2021) 'Sex and gender in asthma', *European respiratory review: an official journal of the European Respiratory Society*, 30(162). Available at: https://doi.org/10.1183/16000617.0067-2021.

Christophe Dessimoz, N.Š. (ed.) (2017) *The Gene Ontology Handbook*. Springer New York.

Chua, F. and Lai, D. (2007) 'Acute severe asthma: Triage, treatment and thereafter', *Current anaesthesia and critical care*, 18(2), pp. 61–68.

Churg, J. and Strauss, L. (1951) 'Allergic granulomatosis, allergic angiitis, and periarteritis nodosa', *The American journal of pathology*, 27(2), pp. 277–301.

Coenye, T. (2021) 'Do results obtained with RNA-sequencing require independent verification?', *Biofilms*, 3, p. 100043.

Collison, A., Mattes, J., Plank, M. and Foster, P.S. (2011) 'Inhibition of house dust mite–induced allergic airways disease by antagonism of microRNA-145 is comparable to glucocorticoid treatment', *The Journal of allergy and clinical immunology*, 128(1), pp. 160–167.e4.

Cordero, F., Beccuti, M., Arigoni, M., Donatelli, S. and Calogero, R.A. (2012) 'Optimizing a massive parallel sequencing workflow for quantitative miRNA expression analysis', *PloS one*, 7(2), p. e31630.

Cottin, V. (2016) 'Eosinophilic Lung Diseases', *Clinics in chest medicine*, 37(3), pp. 535–556.

Crowe, M., Robinson, D., Sagar, M., Chen, L. and Ghamande, S. (2019) 'Chronic eosinophilic pneumonia: clinical perspectives', *Therapeutics and clinical risk management*, 15, pp. 397–403.

Darbà, J., Ramírez, G., Sicras, A., García-Bujalance, L., Torvinen, S. and Sánchez-de la Rosa, R. (2016) 'Identification of factors involved in medication compliance: incorrect inhaler technique of asthma treatment leads to poor compliance', *Patient preference and adherence*, 10, pp. 135–145.

Davis, B.E., Blais, C.M. and Cockcroft, D.W. (2018) 'Methacholine challenge testing: comparative pharmacology', *Journal of asthma and allergy*, 11, pp. 89–99.

Denayer, T., Stöhr, T. and Van Roy, M. (2014) 'Animal models in translational medicine: Validation and prediction', *New Horizons in Translational Medicine*, 2(1), pp. 5–11.

Deng, H., Dokshin, G.A., Lei, J., Goldsmith, A.M., Bitar, K.N., Fingar, D.C., Hershenson, M.B. and Bentley, J.K. (2008) 'Inhibition of glycogen synthase kinase-3beta is sufficient for airway smooth muscle hypertrophy', *The Journal of biological chemistry*, 283(15), pp. 10198–10207.

Depner, M., Kormann, M.S.D., Klopp, N., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Combadière, C. and Kabesch, M. (2007) 'CX3CR1 polymorphisms are associated with atopy but not asthma in German children', *International archives of allergy and immunology*, 144(1), pp. 91–94.

Deshmukh, S.K., Srivastava, S.K., Zubair, H., Khan, M.A., Singh, A.P. and Singh, S. (2021) 'Resistin induces LIN28A-mediated Let-7a repression in breast cancer cells leading to IL-6 and STAT3 upregulation', *Cancers*,

13(18), p. 4498.

De Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L. and Fink, C. (2015) 'Reference gene validation for RT-qPCR, a note on different available software packages', *PloS one*, 10(3), p. e0122515.

Devonshire, A.L., Fan, H., Pujato, M., Paranjpe, A., Gursel, D., Schipma, M., Dunn, J.M., Andorf, S., Pongracic, J.A., Kottyan, L.C. and Kumar, R. (2021) 'Whole blood transcriptomics identifies gene expression associated with peanut allergy in infants at high risk', *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology*, 51(10), pp. 1396–1400.

Di, T., Yang, Y., Fu, C., Zhang, Z., Qin, C., Sai, X., Liu, J., Hu, C., Zheng, M., Wu, Y. and Bian, T. (2021) 'Let-7 mediated airway remodelling in chronic obstructive pulmonary disease via the regulation of IL-6', *European journal of clinical investigation*, 51(4), p. e13425.

Di, Y., Schafer, D.W., Cumbie, J.S. and Chang, J.H. (2011) 'The NBP negative binomial model for assessing differential gene expression from RNA-Seq', *Statistical applications in genetics and molecular biology*, 10(1). Available at: https://doi.org/10.2202/1544-6115.1637.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), pp. 15–21.

Drake, M.G., Cook, M., Fryer, A.D., Jacoby, D.B. and Scott, G.D. (2021) 'Airway Sensory Nerve Plasticity in Asthma and Chronic Cough', *Frontiers in physiology*, 12, p. 720538.

D'silva, L., Cook, R.J., Allen, C.J., Hargreave, F.E. and Parameswaran, K. (2007) 'Changing pattern of sputum cell counts during successive exacerbations of airway disease', *Respiratory medicine*, 101(10), pp. 2217–2220.

Edwards, M.R., Saglani, S., Schwarze, J., Skevaki, C., Smith, J.A., Ainsworth, B., Almond, M., Andreakos, E., Belvisi, M.G., Chung, K.F., Cookson, W., Cullinan, P., Hawrylowicz, C., Lommatzsch, M., Jackson, D., Lutter, R., Marsland, B., Moffatt, M., Thomas, M., Virchow, J.C., Xanthou, G., Edwards, J., Walker, S., Johnston, S.L. and members of the EARIP WP2 working group (2017) 'Addressing unmet needs in understanding asthma mechanisms: From the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators: From the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 49(5), p. 1602448.

ElKashef, S.M.M.A.E., Ahmad, S.E.-A., Soliman, Y.M.A. and Mostafa, M.S. (2021) 'Role of microRNA-21 and microRNA-155 as biomarkers for bronchial asthma', *Innate immunity*, 27(1), pp. 61–69.

Esawy, M.M., Baioumy, S.A., Ismail, N.A. and Shabana, M.A. (2021) 'Role of circulating microRNA-132 in allergic bronchopulmonary aspergillosis: A case-control study', *Immunobiology*, 226(3), p. 152074.

Fadiel, A. and Naftolin, F. (2003) 'Microarray applications and challenges: a vast array of possibilities', *Int Arch Biosci* [Preprint]. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2a1cba82d25ef50f3557ac1dcd99bb81a06d9709.

Fahy, J.V. (2015) 'Type 2 inflammation in asthma--present in most, absent in many', *Nature reviews. Immunology*, 15(1), pp. 57–65.

Farne, H.A., Wilson, A., Milan, S., Banchoff, E., Yang, F. and Powell, C.V. (2022) 'Anti-IL-5 therapies for asthma', *Cochrane database of systematic reviews*, 7(7), p. CD010834.

Farrell, R.E., Jr (2009) *RNA Methodologies: Laboratory Guide for Isolation and Characterization*. Academic Press.

Farr, R.J., Rootes, C.L., Stenos, J., Foo, C.H., Cowled, C. and Stewart, C.R. (2022) 'Detection of SARS-CoV-2 infection by microRNA profiling of the upper respiratory tract', *PloS one*, 17(4), p. e0265670.

Fessel, J. (2020) 'The paradox of opposite directions of gene expressions in MCI and AD suggests possible therapy to prevent progression of MCI to AD', *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 6(1), p. e12003.

Fish, J.E., Santoro, M.M., Morton, S.U., Yu, S., Yeh, R.-F., Wythe, J.D., Ivey, K.N., Bruneau, B.G., Stainier, D.Y.R. and Srivastava, D. (2008) 'miR-126 regulates angiogenic signaling and vascular integrity', *Developmental cell*, 15(2), pp. 272–284.

Friedman, Y., Balaga, O. and Linial, M. (2013) 'Working together: combinatorial regulation by microRNAs', *Advances in experimental medicine and biology*, 774, pp. 317–337.

Gao, X., Wang, Y., Zhao, H., Wei, F., Zhang, X., Su, Y., Wang, C., Li, H. and Ren, X. (2016) 'Plasma miR-324-3p and miR-1285 as diagnostic and prognostic biomarkers for early stage lung squamous cell carcinoma', *Oncotarget*, 7(37), pp. 59664–59675.

Garmire, L.X. and Subramaniam, S. (2012) 'Evaluation of normalization methods in mammalian microRNA-Seq data', *RNA* , 18(6), pp. 1279–1288.

Garrett, J.K., Jameson, S.C., Thomson, B., Collins, M.H., Wagoner, L.E., Freese, D.K., Beck, L.A., Boyce, J.A., Filipovich, A.H., Villanueva, J.M., Sutton, S.A., Assa'ad, A.H. and Rothenberg, M.E. (2004) 'Anti-interleukin-5 (mepolizumab) therapy for hypereosinophilic syndromes', *The Journal of allergy and clinical immunology*, 113(1), pp. 115–119.

Gata, A., Neagoe, I.B., Leucuta, D.-C., Budisan, L., Raduly, L., Trombitas, V.E. and Albu, S. (2023) 'MicroRNAs: Potential Biomarkers of Disease Severity in Chronic Rhinosinusitis with Nasal Polyps', *Medicina* , 59(3). Available at: https://doi.org/10.3390/medicina59030550.

Gautam, A., Kumar, R., Dimitrov, G., Hoke, A., Hammamieh, R. and Jett, M. (2016) 'Identification of extracellular miRNA in archived serum samples by next-generation sequencing from RNA extracted using multiple methods', *Molecular biology reports*, 43(10), pp. 1165–1178.

Geller, D.E., Kaplowitz, H., Light, M.J. and Colin, A.A. (1999) 'Allergic bronchopulmonary aspergillosis in cystic fibrosis: reported prevalence, regional distribution, and patient characteristics. Scientific Advisory Group, Investigators, and Coordinators of the Epidemiologic Study of Cystic Fibrosis', *Chest*, 116(3), pp. 639–646.

Gemicioglu, B., Musellim, B., Dogan, I. and Guven, K. (2014) 'Fractional exhaled nitric oxide (FeNo) in different asthma phenotypes', *Allergy & rhinology* , 5(3), pp. 157–161.

Gershon, A.S., Victor, J.C., Guan, J., Aaron, S.D. and To, T. (2010) 'Use Of Pulmonary Function Tests In The Diagnosis Of Asthma', in *B21. HETEROGENEITY AND COMORBIDITY IN ASTHMA*. *American Thoracic Society 2010 International Conference, May 14-19, 2010 • New Orleans*, American Thoracic Society, pp. A2545–A2545.

Ghassemian, A., Park, J.J., Tsoulis, M.W. and Kim, H. (2021) 'Targeting the IL-5 pathway in eosinophilic asthma: a comparison of mepolizumab to benralizumab in the reduction of peripheral eosinophil counts', *Allergy, asthma, and clinical immunology: official journal of the Canadian Society of Allergy and Clinical Immunology*, 17(1), p. 3.

Gil-Martínez, M., Rodrigo-Muñoz, J.M., Sastre, B., Cañas, J.A., García-Latorre, R., Redondo, N., de la Fuente, L., Mínguez, P., Mahíllo-Fernández, I., Sastre, J., Quirce, S., Caballero, M.L., Olaguibel, J.M. and Del Pozo, V. (2022) 'Serum microRNAs Catalog Asthma Patients by Phenotype', *Journal of investigational allergology & clinical immunology: official organ of the International Association of Asthmology and Sociedad Latinoamericana de Alergia e Inmunologia*, 32(6), pp. 471–478.

Glinge, C., Clauss, S., Boddum, K., Jabbari, R., Jabbari, J., Risgaard, B., Tomsits, P., Hildebrand, B., Kääb, S., Wakili, R., Jespersen, T. and Tfelt-Hansen, J. (2017) 'Stability of Circulating Blood-Based MicroRNAs - Pre-Analytic Methodological Considerations', *PloS one*, 12(2), p. e0167969.

Global Strategy for Asthma Management and Prevention (2021) *GINA Full Report 2021*, *Global Initiative for Asthma*. Available at: www.ginasthma.org (Accessed: 14 October 2021).

Greenberger, P.A. (2002) 'Allergic bronchopulmonary aspergillosis', *The Journal of allergy and clinical immunology*, 110(5), pp. 685–692.

Greenberger, P.A. (2013) 'When to suspect and work up allergic bronchopulmonary aspergillosis', *Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology*, 111(1), pp. 1–4.

de Groot, J.C., Ten Brinke, A. and Bel, E.H.D. (2015) 'Management of the patient with eosinophilic asthma: a new era begins', *ERJ open research*, 1(1). Available at: https://doi.org/10.1183/23120541.00024-2015.

Guida, G., Bagnasco, D., Carriero, V., Bertolini, F., Ricciardolo, F.L.M., Nicola, S., Brussino, L., Nappi, E., Paoletti, G., Canonica, G.W. and Heffler, E. (2022) 'Critical evaluation of asthma biomarkers in clinical practice', *Frontiers in medicine*, 9, p. 969243.

Guideline, N. (2017) 'Asthma: diagnosis, monitoring andAsthma: diagnosis, monitoring and chronic asthma managementchronic asthma management'. Available at: https://www.nice.org.uk/guidance/ng80/resources/asthma-diagnosis-monitoring-and-chronic-asthma-management-pdf-1837687975621.

Gulyaeva, L.F. and Kushlinskiy, N.E. (2016) 'Regulatory mechanisms of microRNA expression', *Journal of translational medicine*, 14(1), p. 143.

Guo, F.H., De Raeve, H.R., Rice, T.W., Stuehr, D.J., Thunnissen, F.B. and Erzurum, S.C. (1995) 'Continuous nitric oxide synthesis by inducible nitric oxide synthase in normal human airway epithelium in vivo', *Proceedings of the National Academy of Sciences of the United States of America*, 92(17), pp. 7809–7813.

Haas, R., Smith, J., Rocher-Ros, V., Nadkarni, S., Montero-Melendez, T., D'Acquisto, F., Bland, E.J., Bombardieri, M., Pitzalis, C., Perretti, M., Marelli-Berg, F.M. and Mauro, C. (2015) 'Lactate regulates metabolic and pro-inflammatory circuits in control of T cell migration and effector functions', *PLoS biology*, 13(7), p. e1002202.

Ha, M. and Kim, V.N. (2014) 'Regulation of microRNA biogenesis', *Nature reviews. Molecular cell biology*, 15(8), pp. 509–524.

Hanania, N.A., Korenblat, P., Chapman, K.R., Bateman, E.D., Kopecky, P., Paggiaro, P., Yokoyama, A., Olsson, J., Gray, S., Holweg, C.T.J., Eisner, M., Asare, C., Fischer, S.K., Peng, K., Putnam, W.S. and Matthews, J.G. (2016) 'Efficacy and safety of lebrikizumab in patients with uncontrolled asthma (LAVOLTA I and LAVOLTA II): replicate, phase 3, randomised, double-blind, placebo-controlled trials', *The Lancet. Respiratory medicine*, 4(10), pp. 781–796.

Hansen, J. and Desjardins, P. (2004) 'Novel spectrophotometer enables rapid quantitation of one-microliter

samples', *PharmaGenomics*, 4, p. S23+.

Harris, T.A., Yamakuchi, M., Ferlito, M., Mendell, J.T. and Lowenstein, C.J. (2008) 'MicroRNA-126 regulates endothelial expression of vascular cell adhesion molecule 1', *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp. 1516–1521.

Hashimoto, Y., Akiyama, Y. and Yuasa, Y. (2013) 'Multiple-to-multiple relationships between microRNAs and target genes in gastric cancer', *PloS one*, 8(5), p. e62589.

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A.Y.M., T Flåm, S., Leithaug, M., Dalland, M., Farmer, A., Henderson, J.M., Hussong, M.A., Moll, P., Nguyen, L., McNulty, A., Shaffer, J.M., Shore, S., Yip, H.K., Vitkovska, J., Rayner, S., Lie, B.A. and Gilfillan, G.D. (2020) 'Systematic assessment of commercially available low-input miRNA library preparation kits', *RNA biology*, 17(1), pp. 75–86.

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A.Y.M., T Flåm, S., Leithaug, M., Dalland, M., Rayner, S., Lie, B.A. and Gilfillan, G.D. (2020) 'An extension to: Systematic assessment of commercially available low-input miRNA library preparation kits', *RNA biology*, 17(9), pp. 1284–1292.

Herrera-Luis, E., Martin-Almeida, M. and Pino-Yanes, M. (2024) 'Asthma-genomic advances toward risk prediction', *Clinics in chest medicine*, 45(3), pp. 599–610.

Hill, P. (2023) *Gene Ontology: A Resource for Analysis and Interpretation of Biological Database*. Syrawood Publishing House.

Hinson, K.F.W., Moon, A.J. and Plummer, N.S. (1952) 'Broncho-pulmonary aspergillosis; a review and a report of eight new cases', *Thorax*, 7(4), pp. 317–333.

Hoefig, K.P. and Heissmeyer, V. (2008) 'MicroRNAs grow up in the immune system', *Current opinion in immunology*, 20(3), pp. 281–287.

Holmes, A.M., Solari, R. and Holgate, S.T. (2011) 'Animal models of asthma: value, limitations and opportunities for alternative approaches', *Drug discovery today*, 16(15-16), pp. 659–670.

Hoshino, M. and Ohtawa, J. (2012) 'Effects of adding omalizumab, an anti-immunoglobulin E antibody, on airway wall thickening in asthma', *Respiration; international review of thoracic diseases*, 83(6), pp. 520–528.

Huang, H.-Y., Lin, Y.-C.-D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., Wang, W., Li, J., Ni, J., Ruan, Y., Li, L., Chen, Y., Xie, Y., Zhu, Z., Cai, X., Chen, X., Yao, L., Chen, Y., Luo, Y., LuXu, S., Luo, M., Chiu, C.-M., Ma, K., Zhu, L., Cheng, G.-J., Bai, C., Chiang, Y.-C., Wang, L., Wei, F., Lee, T.-Y. and Huang, H.-D. (2022) 'miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions', *Nucleic acids research*, 50(D1), pp. D222–D230.

Hücker, S.M., Fehlmann, T., Werno, C., Weidele, K., Lüke, F., Schlenska-Lange, A., Klein, C.A., Keller, A. and Kirsch, S. (2021) 'Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells', *Nature communications*, 12(1), p. 4316.

Hughes, T.R. (2009) '"Validation" in genome-scale research', *Journal of biology*, 8(1), p. 3.

Humbert, M., Taillé, C., Mala, L., Le Gros, V., Just, J., Molimard, M. and STELLAIR investigators (2018) 'Omalizumab effectiveness in patients with severe allergic asthma according to blood eosinophil count: the STELLAIR study', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 51(5). Available at: https://doi.org/10.1183/13993003.02523-2017.

Huntley, R.P., Sitnikov, D., Orlic-Milacic, M., Balakrishnan, R., D'Eustachio, P., Gillespie, M.E., Howe, D., Kalea,

A.Z., Maegdefessel, L., Osumi-Sutherland, D., Petri, V., Smith, J.R., Van Auken, K., Wood, V., Zampetaki, A., Mayr, M. and Lovering, R.C. (2016) 'Guidelines for the functional annotation of microRNAs using the Gene Ontology', *RNA* , 22(5), pp. 667–676.

Huo, X., Zhang, K., Yi, L., Mo, Y., Liang, Y., Zhao, J., Zhang, Z., Xu, Y. and Zhen, G. (2016) 'Decreased epithelial and plasma miR-181b-5p expression associates with airway eosinophilic inflammation in asthma', *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology*, 46(10), pp. 1281–1290.

Hur, J., Rhee, C.K., Lee, S.Y., Kim, Y.K. and Kang, J.Y. (2021) 'MicroRNA-21 inhibition attenuates airway inflammation and remodelling by modulating the transforming growth factor β-Smad7 pathway', *The Korean journal of internal medicine*, 36(3), pp. 706–720.

Iacomino, G. and Siani, A. (2017) 'Role of microRNAs in obesity and obesity-related diseases', *Genes & nutrition*, 12(1), p. 23.

Ilmarinen, P. and Kankaanranta, H. (2014) 'Eosinophil apoptosis as a therapeutic target in allergic asthma', *Basic & clinical pharmacology & toxicology*, 114(1), pp. 109–117.

James, A.L., Elliot, J.G., Jones, R.L., Carroll, M.L., Mauad, T., Bai, T.R., Abramson, M.J., McKay, K.O. and Green, F.H. (2012) 'Airway smooth muscle hypertrophy and hyperplasia in asthma', *American journal of respiratory and critical care medicine*, 185(10), pp. 1058–1064.

Jansson, M.D. and Lund, A.H. (2012) 'MicroRNA and cancer', *Molecular oncology*, 6(6), pp. 590–610.

Jat, K.R., Vaidya, P.C., Mathew, J.L., Jondhale, S. and Singh, M. (2018) 'Childhood allergic bronchopulmonary aspergillosis', *Lung India: official organ of Indian Chest Society*, 35(6), pp. 499–507.

Jia, G., Erickson, R.W., Choy, D.F., Mosesova, S., Wu, L.C., Solberg, O.D., Shikotra, A., Carter, R., Audusseau, S., Hamid, Q., Bradding, P., Fahy, J.V., Woodruff, P.G., Harris, J.M., Arron, J.R. and Bronchoscopic Exploratory Research Study of Biomarkers in Corticosteroid-refractory Asthma (BOBCAT) Study Group (2012) 'Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients', *The journal of allergy and clinical immunology*, 130(3), pp. 647–654.e10.

Jiang, S., Xu, Z., Shi, Y., Liang, S., Jiang, X., Xiao, M., Wang, K. and Ding, L. (2023) 'Circulating insulin-like growth factor-1 and risk of lung diseases: A Mendelian randomization analysis', *Frontiers in endocrinology*, 14, p. 1126397.

Ju, C. (2011) 'Determining overrepresentation of gene ontology terms using the hypergeometric distribution'. Available at: https://pages.pomona.edu/~jsh04747/Student%20Theses/ChristineJu_Scripps11.pdf.

Kaya, H., Gümüş, S., Uçar, E., Aydoğan, M., Muşabak, U., Tozkoparan, E. and Bilgiç, H. (2012) 'Omalizumab as a steroid-sparing agent in chronic eosinophilic pneumonia', *Chest*, 142(2), pp. 513–516.

Keown, A. (2018) *AstraZeneca Drops Phase III Asthma Drug as Company Posts Disappointing 2017 Financials*, *BioSpace*. Available at: https://www.biospace.com/unique-astrazeneca-drops-phase-iii-asthma-drug-as-company-posts-disappointing-2017-financials (Accessed: 30 August 2024).

Kermani, N.Z., Li, C.-X., Versi, A., Badi, Y., Sun, K., Abdel-Aziz, M.I., Bonatti, M., Maitland-van der Zee, A.-H., Djukanovic, R., Wheelock, Å., Dahlen, S.-E., Howarth, P., Guo, Y., Chung, K.F., Adcock, I.M. and U-BIOPRED Project Team (2024) 'Endotypes of severe neutrophilic and eosinophilic asthma from multi-omics integration of U-BIOPRED sputum samples', *Clinical and translational medicine*, 14(7), p. e1771.

Keuls, R.A. and Parchem, R. (2020) 'Preparation of Small RNA Libraries for Sequencing from Early Mouse Embryos', *Journal of visualized experiments: JoVE* [Preprint], (164). Available at: https://doi.org/10.3791/61086.

Khamina, K., Diendorfer, A.B., Skalicky, S., Weigl, M., Pultar, M., Krammer, T.L., Fournier, C.A., Schofield, A.L., Otto, C., Smith, A.T., Buchtele, N., Schoergenhofer, C., Jilma, B., Frank, B.J.H., Hofstaetter, J.G., Grillari, R., Grillari, J., Ruprecht, K., Goldring, C.E., Rehrauer, H., Glaab, W.E. and Hackl, M. (2022) 'A MicroRNA Next-generation-Sequencing Discovery Assay (miND) for genome-scale analysis and absolute quantitation of circulating MicroRNA biomarkers', *International journal of molecular sciences*, 23(3), p. 1226.

Khatri, P. and Drăghici, S. (2005) 'Ontological analysis of gene expression data: current tools, limitations, and open problems', *Bioinformatics* , 21(18), pp. 3587–3595.

Kim, R.Y., Horvat, J.C., Pinkerton, J.W., Starkey, M.R., Essilfie, A.T., Mayall, J.R., Nair, P.M., Hansbro, N.G., Jones, B., Haw, T.J., Sunkara, K.P., Nguyen, T.H., Jarnicki, A.G., Keely, S., Mattes, J., Adcock, I.M., Foster, P.S. and Hansbro, P.M. (2017) 'MicroRNA-21 drives severe, steroid-insensitive experimental asthma by amplifying phosphoinositide 3-kinase-mediated suppression of histone deacetylase 2', *The Journal of allergy and clinical immunology*, 139(2), pp. 519–532.

Kim, Y.-K. (2015) 'Extracellular microRNAs as Biomarkers in Human Disease', *Chonnam medical journal*, 51(2), pp. 51–57.

King, C.E., Cuatrecasas, M., Castells, A., Sepulveda, A.R., Lee, J.-S. and Rustgi, A.K. (2011) 'LIN28B promotes colon cancer progression and metastasis', *Cancer research*, 71(12), pp. 4260–4268.

Kingdom, R. and Wright, C.F. (2022) 'Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts', *Frontiers in genetics*, 13, p. 920390.

Kips, J.C., Anderson, G.P., Fredberg, J.J., Herz, U., Inman, M.D., Jordana, M., Kemeny, D.M., Lötvall, J., Pauwels, R.A., Plopper, C.G., Schmidt, D., Sterk, P.J., Van Oosterhout, A.J.M., Vargaftig, B.B. and Chung, K.F. (2003) 'Murine models of asthma', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 22(2), pp. 374–382.

Klain, A., Dinardo, G., Salvatori, A., Indolfi, C., Contieri, M., Brindisi, G., Decimo, F., Zicari, A.M. and Miraglia Del Giudice, M. (2022) 'An overview on the primary factors that contribute to non-allergic asthma in children', *Journal of clinical medicine*, 11(21), p. 6567.

Kozera, B. and Rapacz, M. (2013) 'Reference genes in real-time PCR', *Journal of applied genetics*, 54(4), pp. 391–406.

Kozomara, A. and Griffiths-Jones, S. (2014) 'miRBase: annotating high confidence microRNAs using deep sequencing data', *Nucleic acids research*, 42(Database issue), pp. D68–73.

Kwak, Y.-G., Song, C.H., Yi, H.K., Hwang, P.H., Kim, J.-S., Lee, K.S. and Lee, Y.C. (2003) 'Involvement of PTEN in airway hyperresponsiveness and inflammation in bronchial asthma', *The journal of clinical investigation*, 111(7), pp. 1083–1092.

Ladjemi, M.Z., Gras, D., Dupasquier, S., Detry, B., Lecocq, M., Garulli, C., Fregimilicka, C., Bouzin, C., Gohy, S., Chanez, P. and Pilette, C. (2018) 'Bronchial Epithelial IgA Secretion Is Impaired in Asthma. Role of IL-4/IL-13', *American journal of respiratory and critical care medicine*, 197(11), pp. 1396–1409.

Lama, L., Cobo, J., Buenaventura, D. and Ryan, K. (2019) 'Small RNA-seq: The RNA 5'-end adapter ligation problem and how to circumvent it', *Journal of biological methods*, 6(1). Available at: https://doi.org/10.14440/jbm.2019.269.

Lamba, V., Ghodke-Puranik, Y., Guan, W. and Lamba, J.K. (2014) 'Identification of suitable reference genes for hepatic microRNA quantitation', *BMC research notes*, 7(1), p. 129.

Lambert, K.A., Roff, A.N., Panganiban, R.P., Douglas, S. and Ishmael, F.T. (2018) 'MicroRNA-146a is induced by inflammatory stimuli in airway epithelial cells and augments the anti-inflammatory effects of glucocorticoids', *PloS one*, 13(10), p. e0205434.

Lambrecht, B.N., Hammad, H. and Fahy, J.V. (2019) 'The Cytokines of Asthma', *Immunity*, 50(4), pp. 975–991.

Lang, J.E., Holbrook, J.T., Mougey, E.B., Wei, C.Y., Wise, R.A., Teague, W.G., Lima, J.J. and American Lung Association-Airways Clinical Research Centers (2015) 'Lansoprazole Is Associated with Worsening Asthma Control in Children with the CYP2C19 Poor Metabolizer Phenotype', *Annals of the American Thoracic Society*, 12(6), pp. 878–885.

Lee, C.-C., Wang, C.-N., Lee, Y.-L., Tsai, Y.-R. and Liu, J.-J. (2015) 'High mobility group box 1 induced human lung myofibroblasts differentiation and enhanced migration by activation of MMP-9', *PloS one*, 10(2), p. e0116393.

Lee, J.-H., Wang, L.-C., Lin, Y.-T., Yang, Y.-H., Yu, H.-H., Hu, Y.-C. and Chiang, B.-L. (2023) 'Differentially expressed microRNAs in peripheral blood cell are associated with downregulated expression of IgE in nonallergic childhood asthma', *Scientific reports*, 13(1), p. 6381.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75(5), pp. 843–854.

Leite-de-Moraes, M., Hammad, H. and Dy, M. (2012) 'Crosstalk between innate and adaptive cells on allergic process', *Journal of allergy*, 2012, p. 720568.

Leiten, E.O., Martinsen, E.M.H., Bakke, P.S., Eagan, T.M.L. and Grønseth, R. (2016) 'Complications and discomfort of bronchoscopy: a systematic review', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 3, p. 33324.

Lett, K.E., Logan, M.K., McLaurin, D.M. and Hebert, M.D. (2021) 'Coilin enhances phosphorylation and stability of DGCR8 and promotes miRNA biogenesis', *Molecular biology of the cell*, 32(20), p. br4.

Levänen, B., Bhakta, N.R., Torregrosa Paredes, P., Barbeau, R., Hiltbrunner, S., Pollack, J.L., Sköld, C.M., Svartengren, M., Grunewald, J., Gabrielsson, S., Eklund, A., Larsson, B.-M., Woodruff, P.G., Erle, D.J. and Wheelock, Å.M. (2013) 'Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients', *The Journal of allergy and clinical immunology*, 131(3), pp. 894–903.

Li, J., Panganiban, R., Kho, A.T., McGeachie, M.J., Farnam, L., Chase, R.P., Weiss, S.T., Lu, Q. and Tantisira, K.G. (2020) 'Circulating MicroRNAs and Treatment Response in Childhood Asthma', *American journal of respiratory and critical care medicine*, 202(1), pp. 65–72.

Lim, H.F. and Nair, P. (2015) 'Efficacy and safety of reslizumab in patients with moderate to severe eosinophilic asthma', *Expert review of respiratory medicine*, 9(2), pp. 135–142.

Li, M., Li, M., Hou, Y., He, H., Jiang, R., Wang, C. and Sun, S. (2023) 'Ferroptosis triggers airway inflammation in asthma', *Therapeutic advances in respiratory disease*, 17, p. 17534666231208628.

Li, M., Zhu, W., Saeed, U., Sun, S., Fang, Y., Wang, C. and Luo, Z. (2022) 'Identification of the molecular subgroups in asthma by gene expression profiles: airway inflammation implications', *BMC pulmonary medicine*, 22(1), p. 29.

Lin, C., Li, X., Zhang, Y., Guo, Y., Zhou, J., Gao, K., Dai, J., Hu, G., Lv, L., Du, J. and Zhang, Y. (2015) 'The microRNA feedback regulation of p63 in cancer progression', *Oncotarget*, 6(11), pp. 8434–8453.

Lindsay, M.A. (2008) 'microRNAs and the immune response', *Trends in immunology*, 29(7), pp. 343–351.

Liu, C., Chu, D., Kalantar-Zadeh, K., George, J., Young, H.A. and Liu, G. (2021) 'Cytokines: From Clinical Significance to Quantification', *Advancement of science*, 8(15), p. e2004433.

Liu, N., Bezprozvannaya, S., Williams, A.H., Qi, X., Richardson, J.A., Bassel-Duby, R. and Olson, E.N. (2008) 'microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart', *Genes & development*, 22(23), pp. 3242–3254.

Liu, N., Wang, D., Tian, J., Wang, X., Shi, H., Wang, C., Jiang, Y., Pang, M., Fan, X., Zhao, J., Liu, L., Wu, H., Guan, L., Zheng, H., Shi, D. and Zhang, Z. (2023) 'PM2.5-bound metals and blood metals are associated with pulmonary function and Th17/Treg imbalance: A panel study of asthmatic adults', *Chemosphere*, 340(139869), p. 139869.

Livak, K.J. and Schmittgen, T.D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods (San Diego, Calif.)*, 25(4), pp. 402–408.

Li, X., Nair, A., Wang, S. and Wang, L. (2015) 'Quality control of RNA-seq experiments', *Methods in molecular biology* , 1269, pp. 137–146.

Li, Y. and Kowdley, K.V. (2012) 'Method for microRNA isolation from clinical serum samples', *Analytical biochemistry*, 431(1), pp. 69–75.

Li, Z., Chen, D., Wang, Q., Tian, H., Tan, M., Peng, D., Tan, Y., Zhu, J., Liang, W. and Zhang, L. (2021) 'mRNA and microRNA stability validation of blood samples under different environmental conditions', *Forensic science international. Genetics*, 55, p. 102567.

Llera-Oyola, J., Carceller, H., Andreu, Z., Hidalgo, M.R., Soler-Sáez, I., Gordillo, F., Gómez-Cabañes, B., Roson, B., de la Iglesia-Vayá, M., Mancuso, R., Guerini, F.R., Mizokami, A. and García-García, F. (2024) 'The role of microRNAs in understanding sex-based differences in Alzheimer's disease', *Biology of sex differences*, 15(1), p. 13.

Ludwig, N., Becker, M., Schumann, T., Speer, T., Fehlmann, T., Keller, A. and Meese, E. (2017) 'Bias in recent miRBase annotations potentially associated with RNA quality issues', *Scientific reports*, 7(1), p. 5162.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E. and Keller, A. (2016) 'Distribution of miRNA expression across human tissues', *Nucleic acids research*, 44(8), pp. 3865–3877.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R. and Golub, T.R. (2005) 'MicroRNA expression profiles classify human cancers', *Nature*, 435(7043), pp. 834–838.

Lu, T.X., Hartner, J., Lim, E.-J., Fabry, V., Mingler, M.K., Cole, E.T., Orkin, S.H., Aronow, B.J. and Rothenberg, M.E. (2011) 'MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity', *Journal of immunology* , 187(6), pp. 3362–3373.

Lynch, H.E., Sanchez, A.M., D'Souza, M.P., Rountree, W., Denny, T.N., Kalos, M. and Sempowski, G.D. (2014) 'Development and implementation of a proficiency testing program for Luminex bead-based cytokine assays', *Journal of immunological methods*, 409, pp. 62–71.

MacPherson, J.C., Comhair, S.A.A., Erzurum, S.C., Klein, D.F., Lipscomb, M.F., Kavuru, M.S., Samoszuk, M.K. and Hazen, S.L. (2001) 'Eosinophils are a major source of nitric oxide-derived oxidants in severe asthma: characterization of pathways available to eosinophils for generating reactive nitrogen species', *The Journal of Immunology*, 166(9), pp. 5763–5772.

Maes, T., Cobos, F.A., Schleich, F., Sorbello, V., Henket, M., De Preter, K., Bracke, K.R., Conickx, G., Mesnil, C., Vandesompele, J., Lahousse, L., Bureau, F., Mestdagh, P., Joos, G.F., Ricciardolo, F.L.M., Brusselle, G.G. and Louis, R. (2016) 'Asthma inflammatory phenotypes show differential microRNA expression in sputum', *The Journal of allergy and clinical immunology*, 137(5), pp. 1433–1446.

Mahmoud, A.A.F., Austen, K.F. and Simon, A.S. (1980) *The Eosinophil in Health and Disease*. Grune & Stratton.

Mao, Z., Shi, Y., Cao, Q., Chen, Y., Sun, Y., Liu, Z., Zhang, Q. and Huang, M. (2018) 'Transcriptional regulation on the gene expression signature in combined allergic rhinitis and asthma syndrome', *Epigenomics*, 10(2), pp. 119–131.

Maqueda, J.J., Giovanazzi, A., Rocha, A.M., Rocha, S., Silva, I., Saraiva, N., Bonito, N., Carvalho, J., Maia, L., Wauben, M.H.M. and Oliveira, C. (2023) 'Adapter dimer contamination in sRNA-sequencing datasets predicts sequencing failure and batch effects and hampers extracellular vesicle-sRNA analysis', *Journal of extracellular biology*, 2(6). Available at: https://doi.org/10.1002/jex2.91.

Marchand, E. and Cordier, J.-F. (2006) 'Idiopathic chronic eosinophilic pneumonia', *Orphanet journal of rare diseases*, 1, p. 11.

Martinez-Nunez, R.T., Rupani, H., Platé, M., Niranjan, M., Chambers, R.C., Howarth, P.H. and Sanchez-Elsner, T. (2018) 'Genome-Wide Posttranscriptional Dysregulation by MicroRNAs in Human Asthma as Revealed by Frac-seq', *Journal of immunology* , 201(1), pp. 251–263.

Ma, S., Kotar, A., Grote, S., Rouskin, S. and Keane, S.C. (2023) 'Structure of pre-miR-31 reveals an active role in Dicer processing', *bioRxiv : the preprint server for biology* [Preprint]. Available at: https://doi.org/10.1101/2023.01.03.519659.

Mattes, J., Collison, A., Plank, M., Phipps, S. and Foster, P.S. (2009) 'Antagonism of microRNA-126 suppresses the effector function of TH2 cells and the development of allergic airways disease', *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), pp. 18704–18709.

Mattiuzzi, C. and Lippi, G. (2020) 'Worldwide asthma epidemiology: insights from the Global Health Data Exchange database', *International forum of allergy & rhinology*, 10(1), pp. 75–80.

Maurya, N., Awasthi, S. and Dixit, P. (2012) 'Association of CFTR gene mutation with bronchial asthma', *The Indian journal of medical research*, 135(4), pp. 469–478.

Ma, X., Lou, C., Pan, J., Zhou, C., Zhao, X., Li, N., Tian, H. and Meng, X. (2023) 'The diagnostic potential of a circRNA-miRNA network in non-small cell lung cancer', *Journal of molecular medicine* , 101(6), pp. 671–684.

Mc Brien, C.N. (2020) *The Bacterial and Fungal Communities in the Airways of Adults with Asthma and Eosinophilic Lung Diseases*. Edited by M.M.W.C. Menzies-Gow. PhD. Imperial College London.

Meijer, H.A., Smith, E.M. and Bushell, M. (2014) 'Regulation of miRNA strand selection: follow the leader?', *Biochemical Society transactions*, 42(4), pp. 1135–1140.

Michalek, R.D., Gerriets, V.A., Jacobs, S.R., Macintyre, A.N., MacIver, N.J., Mason, E.F., Sullivan, S.A., Nichols,

A.G. and Rathmell, J.C. (2011) 'Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets', *The journal of immunology*, 186(6), pp. 3299–3303.

Mitchell, M.I., Ben-Dov, I.Z., Ye, K., Liu, C., Shi, M., Sadoughi, A., Shah, C., Siddiqui, T., Okorozo, A., Gutierrez, M., Unawane, R., Biamonte, L., Parikh, K., Spivack, S. and Loudig, O. (2024) 'Exhaled breath condensate contains extracellular vesicles (EVs) that carry miRNA cargos of lung tissue origin that can be selectively purified and analyzed', *Journal of extracellular vesicles*, 13(4), p. e12440.

Moffatt, M.F. and Cookson, W.O. (2017) 'The lung microbiome in health and disease', *Clinical medicine*, 17(6), pp. 525–529.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., Heinzmann, A., Simma, B., Frischer, T., Willis-Owen, S.A.G., Wong, K.C.C., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Abecasis, G.R., Farrall, M., Gut, I.G., Lathrop, G.M. and Cookson, W.O.C. (2007) 'Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma', *Nature*, 448(7152), pp. 470–473.

Moncada, S., Palmer, R.M. and Higgs, E.A. (1991) 'Nitric oxide: physiology, pathophysiology, and pharmacology', *Pharmacological reviews*, 43(2), pp. 109–142.

Moretti, F., D'Antona, P., Finardi, E., Barbetta, M., Dominioni, L., Poli, A., Gini, E., Noonan, D.M., Imperatori, A., Rotolo, N., Cattoni, M. and Campomenosi, P. (2017) 'Systematic review and critique of circulating miRNAs as biomarkers of stage I-II non-small cell lung cancer', *Oncotarget*, 8(55), pp. 94980–94996.

Muhammad, S.A., Fatima, N., Syed, N.-I.-H., Wu, X., Yang, X.F. and Chen, J.Y. (2015) 'MicroRNA Expression Profiling of Human Respiratory Epithelium Affected by Invasive Candida Infection', *PloS one*, 10(8), p. e0136454.

Mukherjee, K., Amsel, D., Kalsy, M., Billion, A., Dobrindt, U. and Vilcinskas, A. (2020) 'MicroRNAs regulate innate immunity against uropathogenic and commensal-like Escherichia coli infections in the surrogate insect model Galleria mellonella', *Scientific reports*, 10(1), p. 2570.

Mukherjee, M., Sehmi, R. and Nair, P. (2014) 'Anti-IL5 therapy for asthma and beyond', *The World Allergy Organization journal*, 7(1), p. 32.

Mullard, A. (2022) 'FDA approves first-in-class TSLP-targeted antibody for severe asthma', *Nature reviews. Drug discovery*, 21(2), p. 89.

Munafo, D., Hoda, S., Ettwiller, L. and Langhorst, B. (2013) *Automated Size Selection of NEBNext Small RNA Libraries with the Sage Pippin Prep*. Available at: http://www.sagescience.com/wp-content/uploads/2014/04/Sage_AppNote_NEBNextRNALib_Final_12_20_13_3.pdf.

Nadi, E., Hajilooi, M., Babakhani, D. and Rafiei, A. (2012) 'Platelet endothelial cell adhesion molecule-1 polymorphism in patients with bronchial asthma', *Iranian journal of allergy, asthma, and immunology*, 11(4), pp. 276–281.

Nair, P. and O'Byrne, P.M. (2019) 'The interleukin-13 paradox in asthma: effective biology, ineffective biologicals', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology*, 53(2), p. 1802250.

Newton, A.C. and Trotman, L.C. (2014) 'Turning off AKT: PHLPP as a drug target', *Annual review of pharmacology and toxicology*, 54, pp. 537–558.

Nilsen, T.W. (2013) 'The fundamentals of RNA purification', *Cold Spring Harbor protocols*, 2013(7), pp. 618–624.

Novosad, J., Krčmová, I., Bartoš, V., Drahošová, M., Vaník, P., Růžičková-Kirchnerová, O., Teřl, M. and Krejsek, J. (2020) 'Serum periostin levels in asthma patients in relation to omalizumab therapy and presence of chronic rhinosinusitis with nasal polyps', *Postepy dermatologii i alergologii*, 37(2), pp. 240–249.

O'Brien, J., Hayder, H., Zayed, Y. and Peng, C. (2018) 'Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation', *Frontiers in endocrinology*, 9, p. 402.

O'Connell, R.M., Rao, D.S., Chaudhuri, A.A. and Baltimore, D. (2010) 'Physiological and pathological roles for microRNAs in the immune system', *Nature reviews. Immunology*, 10(2), pp. 111–122.

Odgers, T. and Gow, A. (2014) 'P166 - Inducible nitric oxide synthase is responsible for macrophage recruitment and activation following bleomycin mediated lung injury', *Nitric oxide: biology and chemistry / official journal of the Nitric Oxide Society*, 42, pp. 135–136.

Oglesby, I.K., Bray, I.M., Chotirmall, S.H., Stallings, R.L., O'Neill, S.J., McElvaney, N.G. and Greene, C.M. (2010) 'miR-126 is downregulated in cystic fibrosis airway epithelial cells and regulates TOM1 expression', *Journal of immunology* , 184(4), pp. 1702–1709.

Oliphant, C.J., Barlow, J.L. and McKenzie, A.N.J. (2011) 'Insights into the initiation of type 2 immune responses: Initiation of type 2 responses', *Immunology*, 134(4), pp. 378–385.

Ono, S., Lam, S., Nagahara, M. and Hoon, D.S.B. (2015) 'Circulating microRNA Biomarkers as Liquid Biopsy for Cancer Patients: Pros and Cons of Current Assays', *Journal of clinical medicine research*, 4(10), pp. 1890–1907.

Pandey, G., Pandey, O.P., Rogers, A.J., Ahsen, M.E., Hoffman, G.E., Raby, B.A., Weiss, S.T., Schadt, E.E. and Bunyavanich, S. (2018) 'A Nasal Brush-based Classifier of Asthma Identified by Machine Learning Analysis of Nasal RNA Sequence Data', *Scientific reports*, 8(1), p. 8826.

Panek, M.G., Karbownik, M.S., Górski, K.M., Koćwin, M., Kardas, G., Marynowski, M. and Kuna, P. (2022) 'New insights into the regulation of TGF-β/Smad and MPK signaling pathway gene expressions by nasal allergen and methacholine challenge test in asthma', *Clinical and translational allergy*, 12(7), p. e12172.

Panganiban, R.P., Lambert, K.A., Hsu, M.-H., Laryea, Z. and Ishmael, F.T. (2019) 'Isolation and profiling of plasma microRNAs: Biomarkers for asthma and allergic rhinitis', *Methods* , 152, pp. 48–54.

Panganiban, R.P., Wang, Y., Howrylak, J., Chinchilli, V.M., Craig, T.J., August, A. and Ishmael, F.T. (2016) 'Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma', *The Journal of allergy and clinical immunology*, 137(5), pp. 1423–1432.

Papiris, S., Kotanidou, A., Malagari, K. and Roussos, C. (2002) 'Clinical review: severe asthma', *Critical care / the Society of Critical Care Medicine*, 6(1), pp. 30–44.

Park, J.H., Chang, H.S., Park, C.-S., Jang, A.-S., Park, B.L., Rhim, T.Y., Uh, S.-T., Kim, Y.H., Chung, I.Y. and Shin, H.D. (2007) 'Association analysis of CD40 polymorphisms with asthma and the level of serum total IgE', *American journal of respiratory and critical care medicine*, 175(8), pp. 775–782.

Pascoe, C.D., Obeidat, M. 'en, Arsenault, B.A., Nie, Y., Warner, S., Stefanowicz, D., Wadsworth, S.J., Hirota, J.A., Jasemine Yang, S., Dorscheid, D.R., Carlsten, C., Hackett, T.L., Seow, C.Y. and Paré, P.D. (2017) 'Gene expression analysis in asthma using a targeted multiplex array', *BMC pulmonary medicine*, 17(1), p. 189.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. and Ruvkun, G. (2000) 'Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA', *Nature*, 408(6808), pp. 86–89.

Patterson, R., Greenberger, P.A., Lee, T.M., Liotta, J.L., O'Neill, E.A., Roberts, M. and Sommers, H. (1987) 'Prolonged evaluation of patients with corticosteroid-dependent asthma stage of allergic bronchopulmonary aspergillosis', *The Journal of allergy and clinical immunology*, 80(5), pp. 663–668.

Pavlidis, P., Jensen, J.D., Stephan, W. and Stamatakis, A. (2012) 'A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans', *Molecular biology and evolution*, 29(10), pp. 3237–3248.

Pavord, I.D., Hoyte, F.C.L., Lindsley, A.W., Ambrose, C.S., Spahn, J.D., Roseti, S.L., Cook, B., Griffiths, J.M., Hellqvist, Å., Martin, N., Llanos, J.-P., Martin, N., Colice, G. and Corren, J. (2023) 'Tezepelumab reduces exacerbations across all seasons in patients with severe, uncontrolled asthma (NAVIGATOR)', *Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology*, 131(5), pp. 587–597.e3.

Peake, H.L., Currie, A.J., Stewart, G.A. and McWilliam, A.S. (2003) 'Nitric oxide production by alveolar macrophages in response to house dust mite fecal pellets and the mite allergens, Der p 1 and Der p 2', *The Journal of allergy and clinical immunology*, 112(3), pp. 531–537.

Pelaia, C., Paoletti, G., Puggioni, F., Racca, F., Pelaia, G., Canonica, G.W. and Heffler, E. (2019) 'Interleukin-5 in the Pathophysiology of Severe Asthma', *Frontiers in Physiology*, 10. Available at: https://doi.org/10.3389/fphys.2019.01514.

Peng, J., Wang, H., Lu, J., Hui, W., Wang, Y. and Shang, X. (2017) 'Identifying term relations cross different gene ontology categories', *BMC bioinformatics*, 18(Suppl 16), p. 573.

Peters, U., Dixon, A.E. and Forno, E. (2018) 'Obesity and asthma', *The Journal of allergy and clinical immunology*, 141(4), pp. 1169–1179.

Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P. (2004) 'Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations', *Biotechnology letters*, 26(6), pp. 509–515.

Phipps, S., Lam, C.E., Kaiko, G.E., Foo, S.Y., Collison, A., Mattes, J., Barry, J., Davidson, S., Oreo, K., Smith, L. and Others (2009) 'Toll-interleukin-1 signalling is critical for house dust mite-specific Th2 and Th17 responses', *American journal of respiratory and critical care medicine*, 179, pp. 883–893.

Piatek, M.J. and Werner, A. (2014) 'Endogenous siRNAs: regulators of internal affairs', *Biochemical Society transactions*, 42(4), pp. 1174–1179.

Picado, C. (1992) 'Early and late-phase asthmatic reactions: a hypothesis', *Allergy*, 47(4 Pt 1), pp. 331–333.

Picado, C., Deulofeu, R., Lleonart, R., Agustí, M., Casals, E., Quintó, L. and Mullol, J. (1999) 'Lipid and protein metabolism in asthma. Effects of diet and corticosteroid therapy', *Allergy*, 54(6), pp. 569–575.

Porcu, E., Sadler, M.C., Lepik, K., Auwerx, C., Wood, A.R., Weihs, A., Sleiman, M.S.B., Ribeiro, D.M., Bandinelli, S., Tanaka, T., Nauck, M., Völker, U., Delaneau, O., Metspalu, A., Teumer, A., Frayling, T., Santoni, F.A., Reymond, A. and Kutalik, Z. (2021) 'Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome', *Nature communications*, 12(1), p. 5647.

Prabahar, A. and Natarajan, J. (2017) 'MicroRNA mediated network motifs in autoimmune diseases and its crosstalk between genes, functions and pathways', *Journal of immunological methods*, 440, pp. 19–26.

Pre-Analytix (2020) *PAXgene® Blood RNA Kit Instructions for Use (Handbook)*. Qiagen. Available at: https://www.qiagen.com/us/resources/download.aspx?id=9a4cdd13-8b72-4b33-877c-ba60a83b87be&lang=en.

Prechl, J., Papp, K. and Erdei, A. (2010) 'Antigen microarrays: descriptive chemistry or functional immunomics?', *Trends in immunology*, 31(4), pp. 133–137.

Prinz, F., Jonas, K., Balihodzic, A., Klec, C., Reicher, A., Barth, D.A., Riedl, J., Gerger, A., Kiesslich, T., Mayr, C., Rinner, B., Kargl, J. and Pichler, M. (2022) 'MicroRNA mimics can distort physiological microRNA effects on immune checkpoints by triggering an antiviral interferon response', *RNA biology*, 19(1), pp. 1305–1315.

Pua, H.H. and Ansel, K.M. (2015) 'MicroRNA regulation of allergic inflammation and asthma', *Current opinion in immunology*, 36, pp. 101–108.

Puddicombe, S.M., Torres-Lozano, C., Richter, A., Bucchieri, F., Lordan, J.L., Howarth, P.H., Vrugt, B., Albers, R., Djukanovic, R., Holgate, S.T., Wilson, S.J. and Davies, D.E. (2003) 'Increased expression of p21(waf) cyclin-dependent kinase inhibitor in asthmatic bronchial epithelium', *American journal of respiratory cell and molecular biology*, 28(1), pp. 61–68.

Qin, Z., Chen, Y., Wang, Y., Xu, Y., Liu, T., Mu, Q. and Huang, C. (2024) 'Immunometabolism in the pathogenesis of asthma', *Immunology*, 171(1), pp. 1–17.

Qi, P., Cheng, S.-Q., Wang, H., Li, N., Chen, Y.-F. and Gao, C.-F. (2011) 'Serum microRNAs as biomarkers for hepatocellular carcinoma in Chinese patients with chronic hepatitis B virus infection', *PloS one*, 6(12), p. e28486.

qPCR Reference Gene Selection Protocol (no date). Available at: https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/genomics/qpcr/qpcr-reference-gene-selection (Accessed: 12 September 2024).

Quail, M.A., Gu, Y., Swerdlow, H. and Mayho, M. (2012) 'Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation', *Electrophoresis*, 33(23), pp. 3521–3528.

Ramshankar, V. and Krishnamurthy, A. (2013) 'Lung cancer detection by screening - presenting circulating miRNAs as a promising next generation biomarker breakthrough', *Asian Pacific journal of cancer prevention: APJCP*, 14(4), pp. 2167–2172.

Raponi, M., Dossey, L., Jatkoe, T., Wu, X., Chen, G., Fan, H. and Beer, D.G. (2009) 'MicroRNA classifiers for predicting prognosis of squamous cell lung cancer', *Cancer research*, 69(14), pp. 5776–5783.

Razzak, R., Bédard, E.L.R., Kim, J.O., Gazala, S., Guo, L., Ghosh, S., Joy, A., Nijjar, T., Wong, E. and Roa, W.H. (2016) 'MicroRNA expression profiling of sputum for the detection of early and locally advanced non-small-cell lung cancer: a prospective case-control study', *Current oncology* , 23(2), pp. e86–94.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D. and Bader, G.D. (2019) 'Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap', *Nature protocols*, 14(2), pp. 482–517.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) 'The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans', *Nature*,

403(6772), pp. 901–906.

Resztak, J.A., Choe, J., Nirmalan, S., Wei, J., Bruinsma, J., Houpt, R., Alazizi, A., Mair-Meijers, H.E., Wen, X., Slatcher, R.B., Zilioli, S., Pique-Regi, R. and Luca, F. (2023) 'Analysis of transcriptional changes in the immune system associated with pubertal development in a longitudinal cohort of children with asthma', *Nature communications*, 14(1), p. 230.

Rick, E.-M., Woolnough, K.F., Seear, P.J., Fairs, A., Satchwell, J., Richardson, M., Monteiro, W.R., Craner, M., Bourne, M., Wardlaw, A.J. and Pashley, C.H. (2020) 'The airway fungal microbiome in asthma', *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology*, 50(12), pp. 1325–1341.

Rijavec, M., Korošec, P., Žavbi, M., Kern, I. and Malovrh, M.M. (2014) 'Let-7a is differentially expressed in bronchial biopsies of patients with severe asthma', *Scientific reports*, 4, p. 6103.

Robinson, M.D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', *Genome biology*, 11(3), p. R25.

Roche, N., Aggarwal, B., Boucot, I., Mittal, L., Martin, A. and Chrystyn, H. (2022) 'The impact of inhaler technique on clinical outcomes in adolescents and adults with asthma: A systematic review', *Respiratory medicine*, 202, p. 106949.

Roffel, M.P., Boudewijn, I.M., van Nijnatten, J.L.L., Faiz, A., Vermeulen, C.J., van Oosterhout, A.J., Affleck, K., Timens, W., Bracke, K.R., Maes, T., Heijink, I.H., Brandsma, C.-A. and van den Berge, M. (2021) 'Identification of asthma associated microRNAs in bronchial biopsies', *The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology* [Preprint]. Available at: https://doi.org/10.1183/13993003.01294-2021.

Ronchetti, D., Manzoni, M., Todoerti, K., Neri, A. and Agnelli, L. (2016) 'In Silico Characterization of miRNA and Long Non-Coding RNA Interplay in Multiple Myeloma', *Genes*, 7(12). Available at: https://doi.org/10.3390/genes7120107.

Rosenberg, M., Patterson, R., Mintzer, R., Cooper, B.J., Roberts, M. and Harris, K.E. (1977) 'Clinical and immunologic criteria for the diagnosis of allergic bronchopulmonary aspergillosis', *Annals of internal medicine*, 86(4), pp. 405–414.

Rothenberg, M.E. and Hogan, S.P. (2006) 'The eosinophil', *Annual review of immunology*, 24, pp. 147–174.

Rotival, M., Siddle, K.J., Silvert, M., Pothlichet, J., Quach, H. and Quintana-Murci, L. (2020) 'Population variation in miRNAs and isomiRs and their impact on human immunity to infection', *Genome biology*, 21(1), p. 187.

Rozowsky, J., Kitchen, R.R., Park, J.J., Galeev, T.R., Diao, J., Warrell, J., Thistlethwaite, W., Subramanian, S.L., Milosavljevic, A. and Gerstein, M. (2019) 'exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling', *Cell systems*, 8(4), pp. 352–357.e3.

Ruby, J.G., Jan, C.H. and Bartel, D.P. (2007) 'Intronic microRNA precursors that bypass Drosha processing', *Nature*, 448(7149), pp. 83–86.

Sada, K.-E., Amano, K., Uehara, R., Yamamura, M., Arimura, Y., Nakamura, Y., Makino, H. and Research Committee on Intractable Vasculitides, the Ministry of Health, Labour, Welfare of Japan (2014) 'A nationwide survey on the epidemiology and clinical features of eosinophilic granulomatosis with polyangiitis (Churg-Strauss) in Japan', *Modern rheumatology / the Japan Rheumatism Association*, 24(4), pp. 640–644.

Saeedavi, M., Goudarzi, M., Mehrzadi, S., Basir, Z., Hasanvand, A. and Hosseinzadeh, A. (2022) 'Sinapic acid ameliorates airway inflammation in murine ovalbumin-induced allergic asthma by reducing Th2 cytokine production', *Life sciences*, 307, p. 120858.

Samant, M., Ziemniak, J. and Paolini, J.F. (2023) 'First-in-Human Phase 1 Randomized Trial with the Anti-CD40 Monoclonal Antibody KPL-404: Safety, Tolerability, Receptor Occupancy, and Suppression of T-Cell-Dependent Antibody Response', *The Journal of pharmacology and experimental therapeutics*, 387(3), pp. 306–314.

Sarkis, E., Patel, S., Burns, K., Batarseh, H. and Mador, M.J. (2020) 'Anti-interleukin (IL)-5 as a steroid-sparing agent in chronic eosinophilic pneumonia', *The Journal of asthma: official journal of the Association for the Care of Asthma*, 57(1), pp. 82–86.

Sastre, I., Blanco, J., Mata, H. and García, F. (2012) 'A case of allergic bronchopulmonary aspergillosis treated with omalizumab', *Journal of investigational allergology & clinical immunology: official organ of the International Association of Asthmology and Sociedad Latinoamericana de Alergia e Inmunologia*, 22(2), pp. 145–147.

Sevignani, C., Calin, G.A., Siracusa, L.D. and Croce, C.M. (2006) 'Mammalian microRNAs: a small world for fine-tuning gene expression', *Mammalian genome: official journal of the International Mammalian Genome Society*, 17(3), pp. 189–202.

Shah, A. (2007) 'Allergic bronchopulmonary aspergillosis: an Indian perspective', *Current opinion in pulmonary medicine*, 13(1), pp. 72–80.

Shah, A., Bhagat, R. and Panchal, N. (1993) 'Allergic bronchopulmonary aspergillosis with clubbing and cavitation', *Indian pediatrics*, 30(2), pp. 248–251.

Shang, L., Du, Y., Zhao, Y., Zhang, Y. and Liu, C. (2023) 'The interaction of OTUB1 and TRAF3 mediates NLRP3 inflammasome activity to regulate TGF-β1-induced BEAS-2B cell injury', *Applied biochemistry and biotechnology*, 195(11), pp. 7060–7074.

Shell, S., Park, S.-M., Radjabi, A.R., Schickel, R., Kistner, E.O., Jewell, D.A., Feig, C., Lengyel, E. and Peter, M.E. (2007) 'Let-7 expression defines two differentiation stages of cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 104(27), pp. 11400–11405.

Shi, J., Chen, M., Ouyang, L., Wang, Q., Guo, Y., Huang, L. and Jiang, S. (2020) 'miR-142-5p and miR-130a-3p regulate pulmonary macrophage polarization and asthma airway remodeling', *Immunology and cell biology*, 98(9), pp. 715–725.

Shirshova, A.N., Shamovskaya, D.A., Boyarskikh, U.A., Kushlinskii, N.E. and Filipenko, M.L. (2018) 'One-phase phenol-free method for microRNA isolation from blood plasma', *MethodsX*, 5, pp. 737–743.

Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L.-E., Ji, Y., Ansari, M., Arguel, M.-J., Apperloo, L., Banchero, M., Bécavin, C., Berg, M., Chichelnitskiy, E., Chung, M.-I., Collin, A., Gay, A.C.A., Gote-Schniering, J., Hooshiar Kashani, B., Inecik, K., Jain, M., Kapellos, T.S., Kole, T.M., Leroy, S., Mayr, C.H., Oliver, A.J., von Papen, M., Peter, L., Taylor, C.J., Walzthoeni, T., Xu, C., Bui, L.T., De Donno, C., Dony, L., Faiz, A., Guo, M., Gutierrez, A.J., Heumos, L., Huang, N., Ibarra, I.L., Jackson, N.D., Kadur Lakshminarasimha Murthy, P., Lotfollahi, M., Tabib, T., Talavera-López, C., Travaglini, K.J., Wilbrey-Clark, A., Worlock, K.B., Yoshida, M., Lung Biological Network Consortium, van den Berge, M., Bossé, Y., Desai, T.J., Eickelberg, O., Kaminski, N., Krasnow, M.A., Lafyatis, R., Nikolic, M.Z., Powell, J.E., Rajagopal, J., Rojas, M., Rozenblatt-Rosen, O., Seibold, M.A., Sheppard, D., Shepherd, D.P., Sin, D.D., Timens, W., Tsankov, A.M., Whitsett, J., Xu, Y., Banovich, N.E., Barbry, P., Duong, T.E., Falk, C.S., Meyer, K.B., Kropski, J.A., Pe'er, D., Schiller,

H.B., Tata, P.R., Schultze, J.L., Teichmann, S.A., Misharin, A.V., Nawijn, M.C., Luecken, M.D. and Theis, F.J. (2023) 'An integrated cell atlas of the lung in health and disease', *Nature medicine*, 29(6), pp. 1563–1577.

Silver, N., Best, S., Jiang, J. and Thein, S.L. (2006) 'Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR', *BMC molecular biology*, 7(1), p. 33.

Simpson, J.L., Scott, R.J., Boyle, M.J. and Gibson, P.G. (2005) 'Differential proteolytic enzyme activity in eosinophilic and neutrophilic asthma', *American journal of respiratory and critical care medicine*, 172(5), pp. 559–565.

Sinyor, B. and Perez, L.C. (2023) *Pathophysiology Of Asthma*. StatPearls Publishing.

Skeparnias, I., Anastasakis, D., Grafanaki, K., Kyriakopoulos, G., Alexopoulos, P., Dougenis, D., Scorilas, A., Kontos, C.K. and Stathopoulos, C. (2020) 'Contribution of miRNAs, tRNAs and tRFs to Aberrant Signaling and Translation Deregulation in Lung Cancer', *Cancers*, 12(10). Available at: https://doi.org/10.3390/cancers12103056.

Smulders, T., Van Der Schee, M.P., Maitland-Van Der Zee, A.H., Dikkers, F.G. and Van Drunen, C.M. (2024) 'Influence of the gut and airway microbiome on asthma development and disease', *Pediatric allergy and immunology: official publication of the European Society of Pediatric Allergy and Immunology*, 35(3), p. e14095.

Sohel, M.H. (2016) 'Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges', *Achievements in the Life Sciences*, 10(2), pp. 175–186.

Sokol, K.C., Sharma, G., Lin, Y.-L. and Goldblum, R.M. (2015) 'Choosing wisely: adherence by physicians to recommended use of spirometry in the diagnosis and management of adult asthma', *The American journal of medicine*, 128(5), pp. 502–508.

Solberg, O.D., Ostrin, E.J., Love, M.I., Peng, J.C., Bhakta, N.R., Hou, L., Nguyen, C., Solon, M., Nguyen, C., Barczak, A.J., Zlock, L.T., Blagev, D.P., Finkbeiner, W.E., Ansel, K.M., Arron, J.R., Erle, D.J. and Woodruff, P.G. (2012) 'Airway epithelial miRNA expression is altered in asthma', *American journal of respiratory and critical care medicine*, 186(10), pp. 965–974.

Soliman, M.M. (2018) *COMPARISON OF RNA PRESERVATION TECHNIQUES*. Edited by D.T. Harris. University of Arizona, Tucson, USA.

Specjalski, K. and Niedoszytko, M. (2020) 'MicroRNAs: future biomarkers and targets of therapy in asthma?', *Current opinion in pulmonary medicine*, 26(3), pp. 285–292.

Steele, A.M., Meuret, A.E., Millard, M.W. and Ritz, T. (2012) 'Discrepancies between lung function and asthma control: asthma perception and association with demographics and anxiety', *Allergy and asthma proceedings: the official journal of regional and state allergy societies*, 33(6), pp. 500–507.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M.A., Pourmand, N. and Sanford, J.R. (2013) 'Frac-seq reveals isoform-specific recruitment to polyribosomes', *Genome research*, 23(10), pp. 1615–1623.

Sunadome, H., Matsumoto, H., Izuhara, Y., Nagasaki, T., Kanemitsu, Y., Ishiyama, Y., Morimoto, C., Oguma, T., Ito, I., Murase, K., Muro, S., Kawaguchi, T., Tabara, Y., Chin, K., Matsuda, F. and Hirai, T. (2020) 'Correlation between eosinophil count, its genetic background and body mass index: The Nagahama Study', *Allergology international: official journal of the Japanese Society of Allergology*, 69(1), pp. 46–52.

Sun, J., Nishiyama, T., Shimizu, K. and Kadota, K. (2013) 'TCC: an R package for comparing tag count data

with robust normalization strategies', *BMC bioinformatics*, 14, p. 219.

Suyundikov, A., Stevens, J.R., Corcoran, C., Herrick, J., Wolff, R.K. and Slattery, M.L. (2015) 'Incorporation of subject-level covariates in quantile normalization of miRNA data', *BMC genomics*, 16, p. 1045.

Suzuki, Y. and Suda, T. (2018) 'Long-term management and persistent impairment of pulmonary function in chronic eosinophilic pneumonia: A review of the previous literature', *Allergology international: official journal of the Japanese Society of Allergology*, 67(3), pp. 334–340.

Takayama, G., Arima, K., Kanaji, T., Toda, S., Tanaka, H., Shoji, S., McKenzie, A.N.J., Nagai, H., Hotokebuchi, T. and Izuhara, K. (2006) 'Periostin: a novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals', *The journal of allergy and clinical immunology*, 118(1), pp. 98–104.

Tam, S., Tsao, M.-S. and McPherson, J.D. (2015) 'Optimization of miRNA-seq data preprocessing', *Briefings in bioinformatics*, 16(6), pp. 950–963.

Tang, M., Sun, J., Shimizu, K. and Kadota, K. (2015) 'Evaluation of methods for differential expression analysis on multi-group RNA-seq count data', *BMC bioinformatics*, 16, p. 361.

Tay, J.W., James, I., Hughes, Q.W., Tiao, J.Y. and Baker, R.I. (2017) 'Identification of reference miRNAs in plasma useful for the study of oestrogen-responsive miRNAs associated with acquired Protein S deficiency in pregnancy', *BMC research notes*, 10(1), p. 312.

Thomeer, M.J., Costabe, U., Rizzato, G., Poletti, V. and Demedts, M. (2001) 'Comparison of registries of interstitial lung diseases in three European countries', *The European respiratory journal. Supplement*, 32, p. 114s–118s.

Thwaites, R.S., Ito, K., Chingono, J.M.S., Coates, M., Jarvis, H.C., Tunstall, T., Anderson-Dring, L., Cass, L., Rapeport, G., Openshaw, P.J., Nadel, S. and Hansel, T.T. (2017) 'Nasosorption as a Minimally Invasive Sampling Procedure: Mucosal Viral Load and Inflammation in Primary RSV Bronchiolitis', *The Journal of infectious diseases*, 215(8), pp. 1240–1244.

Thwaites, R.S., Jarvis, H.C., Singh, N., Jha, A., Pritchard, A., Fan, H., Tunstall, T., Nanan, J., Nadel, S., Kon, O.M., Openshaw, P.J. and Hansel, T.T. (2018) 'Absorption of Nasal and Bronchial Fluids: Precision Sampling of the Human Respiratory Mucosa and Laboratory Processing of Samples', *Journal of visualized experiments: JoVE* [Preprint], (131). Available at: https://doi.org/10.3791/56413.

Tian, M., Ji, Y., Wang, T., Zhang, W., Zhou, Y. and Cui, Y. (2018) 'Changes in circulating microRNA-126 levels are associated with immune imbalance in children with acute asthma', *International journal of immunopathology and pharmacology*, 32, p. 2058738418779243.

Tillie-Leblond, I., Germaud, P., Leroyer, C., Tétu, L., Girard, F., Devouassoux, G., Grignet, J.-P., Prudhomme, A., Dusser, D. and Wallaert, B. (2011) 'Allergic bronchopulmonary aspergillosis and omalizumab', *Allergy*, 66(9), pp. 1254–1256.

Tiwari, A., Li, J., Kho, A.T., Sun, M., Lu, Q., Weiss, S.T., Tantisira, K.G. and McGeachie, M.J. (2021) 'COPD-associated miR-145-5p is downregulated in early-decline FEV1 trajectories in childhood asthma', *The journal of allergy and clinical immunology*, 147(6), pp. 2181–2190.

Tonga, K.O., King, G.G., Farah, C.S., Thamrin, C., Tang, F.S., Santos, J., Sharma, P., Chapman, D.G. and Oliver, B.G. (2018) 'Steroid insensitive fixed airflow obstruction is not related to airway inflammation in older non-smokers with asthma', *Respiratory research*, 19(1), p. 176.

Turchinovich, A. and Burwinkel, B. (2012) 'Distinct AGO1 and AGO2 associated miRNA profiles in human

cells and blood plasma', *RNA biology*, 9(8), pp. 1066–1075.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. and Johnson, R. (2018) 'Towards a complete map of the human long non-coding RNA transcriptome', *Nature reviews. Genetics*, 19(9), pp. 535–548.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', *Genome biology*, 3(7), p. RESEARCH0034.

Van Goethem, A., Yigit, N., Everaert, C., Moreno-Smith, M., Mus, L.M., Barbieri, E., Speleman, F., Mestdagh, P., Shohet, J., Van Maerken, T. and Vandesompele, J. (2016) 'Depletion of tRNA-halves enables effective small RNA sequencing of low-input murine serum samples', *Scientific reports*, 6, p. 37876.

Vernon, M.K., Wiklund, I., Bell, J.A., Dale, P. and Chapman, K.R. (2012) 'What do we know about asthma triggers? a review of the literature', *The Journal of asthma: official journal of the Association for the Care of Asthma*, 49(10), pp. 991–998.

Veryaskina, Y.A., Titov, S.E., Ivanov, M.K., Ruzankin, P.S., Tarasenko, A.S., Shevchenko, S.P., Kovynev, I.B., Stupak, E.V., Pospelova, T.I. and Zhimulev, I.F. (2022) 'Selection of reference genes for quantitative analysis of microRNA expression in three different types of cancer', *PloS one*, 17(2), p. e0254304.

Vigneron, N., Meryet-Figuière, M., Guttin, A., Issartel, J.-P., Lambert, B., Briand, M., Louis, M.-H., Vernon, M., Lebailly, P., Lecluse, Y., Joly, F., Krieger, S., Lheureux, S., Clarisse, B., Leconte, A., Gauduchon, P., Poulain, L. and Denoyelle, C. (2016) 'Towards a new standardized method for circulating miRNAs profiling in clinical studies: Interest of the exogenous normalization to improve miRNA signature accuracy', *Molecular oncology*, 10(7), pp. 981–992.

Walczak-Drzewiecka, A., Wyczólkowska, J. and Dastych, J. (2003) 'Environmentally relevant metal and transition metal ions enhance Fc epsilon RI-mediated mast cell activation', *Environmental health perspectives*, 111(5), pp. 708–713.

Wang, J.-S., Lin, H.-Y., Cheng, M.-L. and Wong, M.-K. (2007) 'Chronic intermittent hypoxia modulates eosinophil- and neutrophil-platelet aggregation and inflammatory cytokine secretion caused by strenuous exercise in men', *Journal of applied physiology*, 103(1), pp. 305–314.

Wang, K., Yuan, Y., Cho, J.-H., McClarty, S., Baxter, D. and Galas, D.J. (2012) 'Comparing the MicroRNA spectrum between serum and plasma', *PloS one*, 7(7), p. e41561.

Wang, L., Yang, X., Li, W., Song, X. and Kang, S. (2019) 'MiR-202-5p/MATN2 are associated with regulatory T-cells differentiation and function in allergic rhinitis', *Human cell*, 32(4), pp. 411–417.

Wang, Z. (2011) 'The guideline of the design and validation of MiRNA mimics', *Methods in molecular biology (Clifton, N.J.)*, 676, pp. 211–223.

Wang, Z., Pianosi, P.T., Keogh, K.A., Zaiem, F., Alsawas, M., Alahdab, F., Almasri, J., Mohammed, K., Larrea-Mantilla, L., Farah, W., Daraz, L., Barrionuevo, P., Morrow, A.S., Prokop, L.J. and Murad, M.H. (2018) 'The Diagnostic Accuracy of Fractional Exhaled Nitric Oxide Testing in Asthma: A Systematic Review and Meta-analyses', *Mayo Clinic proceedings. Mayo Clinic*, 93(2), pp. 191–198.

Ward, C., Pais, M., Bish, R., Reid, D., Feltis, B., Johns, D. and Walters, E.H. (2002) 'Airway inflammation, basement membrane thickening and bronchial hyperresponsiveness in asthma', *Thorax*, 57(4), pp. 309–316.

Wardzyńska, A., Pawełczyk, M., Rywaniak, J., Makowska, J., Jamroz-Brzeska, J. and Kowalski, M.L. (2021)

'Circulating miRNA expression in asthmatics is age-related and associated with clinical asthma parameters, respiratory function and systemic inflammation', *Respiratory research*, 22(1), p. 177.

Wenzel, S.E. (2012) 'Asthma phenotypes: the evolution from clinical to molecular approaches', *Nature medicine*, 18(5), pp. 716–725.

Wenzel, S., Ford, L., Pearlman, D., Spector, S., Sher, L., Skobieranda, F., Wang, L., Kirkesseli, S., Rocklin, R., Bock, B., Hamilton, J., Ming, J.E., Radin, A., Stahl, N., Yancopoulos, G.D., Graham, N. and Pirozzi, G. (2013) 'Dupilumab in persistent asthma with elevated eosinophil levels', *The New England journal of medicine*, 368(26), pp. 2455–2466.

White, B.S., de Reyniès, A., Newman, A.M., Waterfall, J.J., Lamb, A., Petitprez, F., Lin, Y., Yu, R., Guerrero-Gimenez, M.E., Domanskyi, S., Monaco, G., Chung, V., Banerjee, J., Derrick, D., Valdeolivas, A., Li, H., Xiao, X., Wang, S., Zheng, F., Yang, W., Catania, C.A., Lang, B.J., Bertus, T.J., Piermarocchi, C., Caruso, F.P., Ceccarelli, M., Yu, T., Guo, X., Bletz, J., Coller, J., Maecker, H., Duault, C., Shokoohi, V., Patel, S., Liliental, J.E., Simon, S., Tumor Deconvolution DREAM Challenge consortium, Saez-Rodriguez, J., Heiser, L.M., Guinney, J. and Gentles, A.J. (2024) 'Community assessment of methods to deconvolve cellular composition from bulk gene expression', *Nature communications*, 15(1), p. 7362.

Williams, A.E., Larner-Svensson, H., Perry, M.M., Campbell, G.A., Herrick, S.E., Adcock, I.M., Erjefalt, J.S., Chung, K.F. and Lindsay, M.A. (2009) 'MicroRNA expression profiling in mild asthmatic human airways and effect of corticosteroid therapy', *PloS one*, 4(6), p. e5889.

Williams, Z., Ben-Dov, I.Z., Elias, R., Mihailovic, A., Brown, M., Rosenwaks, Z. and Tuschl, T. (2013) 'Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations', *Proceedings of the National Academy of Sciences of the United States of America*, 110(11), pp. 4255–4260.

Willis-Owen, S.A.G., Cookson, W.O.C. and Moffatt, M.F. (2018) 'The genetics and genomics of asthma', *Annual review of genomics and human genetics*, 19(1), pp. 223–246.

Wong LLK, Lee K, Russell I, Chen C. (2010) *Endogenous Controls for Real-Time Quantitation of miRNA UsingTaqMan MicroRNA Assays. Applied Biosystems Application Note*. Available at: https://assets.thermofisher.com/TFS-Assets/LSG/Application-Notes/cms_044972.pdf.

Wong, M., Forno, E. and Celedón, J.C. (2022) 'Asthma interactions between obesity and other risk factors', *Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology*, 129(3), pp. 301–306.

Wu, W.-S., Brown, J.S., Chen, T.-T., Chu, Y.-H., Huang, W.-C., Tu, S. and Lee, H.-C. (2019) 'piRTarBase: a database of piRNA targeting sites and their roles in gene regulation', *Nucleic acids research*, 47(D1), pp. D181–D187.

Wynn, T.A. (2015) 'Type 2 cytokines: mechanisms and therapeutic strategies', *Nature reviews. Immunology*, 15(5), pp. 271–282.

Xie, F., Wang, J. and Zhang, B. (2023) 'RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes', *Functional & integrative genomics*, 23(2), p. 125.

Xue, F., You, W., Liu, Y. and Zhong, L. (2023) 'Identification of the key miRNA-mRNA regulatory network in lung adenocarcinoma', *Journal of thoracic disease*, 15(4), pp. 2037–2050.

Xu, Y., Wang, Z., Li, W., Zhu, Q. and Liang, Z. (2022) 'Clinical characteristics of patients with chronic eosinophilic pneumonia in a Chinese tertiary-care hospital: A 6-year retrospective study', *The clinical*

respiratory journal, 16(1), pp. 35–42.

Yaghoubi, M., Adibi, A., Safari, A., FitzGerald, J.M. and Sadatsafavi, M. (2019) 'The Projected Economic and Health Burden of Uncontrolled Asthma in the United States', *American journal of respiratory and critical care medicine*, 200(9), pp. 1102–1112.

Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D. and Lai, E.C. (2010) 'Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), pp. 15163–15168.

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q. and Tomari, Y. (2010) 'ATP-dependent human RISC assembly pathways', *Nature structural & molecular biology*, 17(1), pp. 17–23.

Yssel, Abbal, Pène and Bousquet (1998) 'The role of IgE in asthma', *Clinical & Experimental Allergy*, 28, pp. 104–109.

Zhang, H.-G., Liu, L., Song, Z.-P. and Zhang, D.-Y. (2020) 'Bioinformatics Analysis of the Regulatory lncRNA-miRNA-mRNA Network in Patients with Neuropathic Pain', *Research Square*. Research Square. Available at: https://doi.org/10.21203/rs.3.rs-54799/v1.

Zhang, K., Zhang, X., Cai, Z., Zhou, J., Cao, R., Zhao, Y., Chen, Z., Wang, D., Ruan, W., Zhao, Q., Liu, G., Xue, Y., Qin, Y., Zhou, B., Wu, L., Nilsen, T., Zhou, Y. and Fu, X.-D. (2018) 'A novel class of microRNA-recognition elements that function only within open reading frames', *Nature structural & molecular biology* [Preprint]. Available at: https://doi.org/10.1038/s41594-018-0136-3.

Zhang, S., Laryea, Z., Panganiban, R.P.L., Lambert, K.A. and Ishmael, F. (2018) 'Plasma microRNA profiles identify distinct clinical phenotypes in human asthmatics'. Available at: https://doi.org/10.20517/jtgg.2018.22.

Zhang, X.-Y., Simpson, J.L., Powell, H., Yang, I.A., Upham, J.W., Reynolds, P.N., Hodge, S., James, A.L., Jenkins, C., Peters, M.J., Lin, J.-T. and Gibson, P.G. (2014) 'Full blood count parameters for the detection of asthma inflammatory phenotypes', *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology*, 44(9), pp. 1137–1145.

Zhang, Y., Jing, Y., Qiao, J., Luan, B., Wang, X., Wang, L. and Song, Z. (2017) 'Activation of the mTOR signaling pathway is required for asthma onset', *Scientific reports*, 7(1), p. 4532.

Zhang, Y., Willis-Owen, S.A.G., Spiegel, S., Lloyd, C.M., Moffatt, M.F. and Cookson, W.O.C.M. (2019) 'The ORMDL3 Asthma Gene Regulates ICAM1 and Has Multiple Effects on Cellular Inflammation', *American journal of respiratory and critical care medicine*, 199(4), pp. 478–488.

Zhang, Z., Li, C., Zhang, J., Chen, F., Gong, Y., Li, Y., Su, Y., Wei, Y. and Zhao, Y. (2020) 'Selection of the Reference Gene for Expression Normalization in Papaver somniferum L. under Abiotic Stress and Hormone Treatment', *Genes*, 11(2). Available at: https://doi.org/10.3390/genes11020124.

Zhao, S., Gordon, W., Du, S., Zhang, C., He, W., Xi, L., Mathur, S., Agostino, M., Paradis, T., von Schack, D., Vincent, M. and Zhang, B. (2017) 'QuickMIRSeq: a pipeline for quick and accurate quantification of both known miRNAs and isomiRs by jointly processing multiple samples from microRNA sequencing', *BMC bioinformatics*, 18(1), p. 180.

Zheng, D., Haddadin, S., Wang, Y., Gu, L.-Q., Perry, M.C., Freter, C.E. and Wang, M.X. (2011) 'Plasma microRNAs as novel biomarkers for early detection of lung cancer', *International journal of clinical and experimental pathology*, 4(6), pp. 575–586.

Zhu, W., Zhou, K., Zha, Y., Chen, D., He, J., Ma, H., Liu, X., Le, H. and Zhang, Y. (2016) 'Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer', *PloS one*, 11(4), p. e0153046.

Zhu, Z., Hasegawa, K., Camargo, C.A., Jr and Liang, L. (2021) 'Investigating asthma heterogeneity through shared and distinct genetics: Insights from genome-wide cross-trait analysis', *The Journal of allergy and clinical immunology*, 147(3), pp. 796–807.

Zoorob, R.J. and Cender, D. (1998) 'A different look at corticosteroids', *American family physician*, 58(2), pp. 443–450.

Aaron, S.D., Vandemheen, K.L., Whitmore, G.A., Bergeron, C., Boulet, L.-P., Côté, A., McIvor, R.A., Penz, E., Field, S.K., Lemière, C., Mayers, I., Bhutani, M., Azher, T., Lougheed, M.D., Gupta, S., Ezer, N., Licskai, C.J., Hernandez, P., Ainslie, M., Alvarez, G.G., Mulpuru, S. and UCAP Investigators (2024) 'Early Diagnosis and Treatment of COPD and Asthma - A Randomized, Controlled Trial', The New England journal of medicine, 390(22), pp. 2061–2073.

Agache, I., Beltran, J., Akdis, C., Akdis, M., Canelo-Aybar, C., Canonica, G.W., Casale, T., Chivato, T., Corren, J., Del Giacco, S., Eiwegger, T., Firinu, D., Gern, J.E., Hamelmann, E., Hanania, N., Mäkelä, M., Hernández-Martín, I., Nair, P., O'Mahony, L., Papadopoulos, N.G., Papi, A., Park, H.-S., Pérez de Llano, L., Posso, M., Rocha, C., Quirce, S., Sastre, J., Shamji, M., Song, Y., Steiner, C., Schwarze, J., Alonso-Coello, P., Palomares, O. and Jutel, M. (2020) 'Efficacy and safety of treatment with biologicals (benralizumab, dupilumab, mepolizumab, omalizumab and reslizumab) for severe eosinophilic asthma. A systematic review for the EAACI Guidelines - recommendations on the use of biologicals in severe asthma', Allergy, 75(5), pp. 1023–1042.

Alfaro-Arnedo, E., López, I.P., Piñeiro-Hermida, S., Ucero, Á.C., González-Barcala, F.J., Salgado, F.J. and Pichel, J.G. (2021) 'IGF1R as a Potential Pharmacological Target in Allergic Asthma', Biomedicines, 9(8). Available at: https://doi.org/10.3390/biomedicines9080912.

Alipoor, S.D., Adcock, I.M., Garssen, J., Mortaz, E., Varahram, M., Mirsaeidi, M. and Velayati, A. (2016) 'The roles of miRNAs as potential biomarkers in lung diseases', European journal of pharmacology, 791, pp. 395–404.

Al-Rashoudi, R., Moir, G., Al-Hajjaj, M.S., Al-Alwan, M.M., Wilson, H.M. and Crane, I.J. (2019) 'Differential expression of CCR2 and CX3CR1 on CD16+ monocyte subsets is associated with asthma severity', Allergy, asthma, and clinical immunology: official journal of the Canadian Society of Allergy and Clinical Immunology, 15, p. 64.

Anauate, A.C., Leal, M.F., Wisnieski, F., Santos, L.C., Gigek, C.O., Chen, E.S., Geraldis, J.C., Calcagno, D.Q., Assumpção, P.P., Demachki, S., Arasaki, C.H., Lourenço, L.G., Artigiani, R., Burbano, R.R. and Smith, M.A.C. (2017) 'Identification of suitable reference genes for miRNA expression normalization in gastric cancer', Gene, 621, pp. 59–68.

Andersen, C.L., Jensen, J.L. and Ørntoft, T.F. (2004) '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 research, 64(15), pp. 5245–5250.

Andrews, S. and Others (2010) 'FastQC: a quality control tool for high throughput sequence data'. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

An, T., Fan, T., Zhang, X.Q., Liu, Y.-F., Huang, J., Liang, C., Lv, B.-H., Wang, Y.-Q., Zhao, X.-G., Liu, J.-X., Fu, Y.-H. and Jiang, G.-J. (2020) 'Publisher Correction: Comparison of Alterations in miRNA Expression in Matched Tissue and Blood Samples during Spinal Cord Glioma Progression', Scientific reports, 10(1), p. 6871.

Argyropoulos, C., Etheridge, A., Sakhanenko, N. and Galas, D. (2017) 'Modeling bias and variation in the stochastic processes of small RNA sequencing', Nucleic acids research, 45(11), p. e104.

Bagley, D., Fox, K., Redd, P.F., Joseph, M., Ortiz-Zapater, E., Parsons, M. and Rosenblatt, J. (2019) 'Blocking Cell Extrusion Prevents Bronchoconstriction-Induced Airway Epithelial Damage and Inflammation'. Available at: https://doi.org/10.2139/ssrn.3493955.

Bakakos, P., Schleich, F., Alchanatis, M. and Louis, R. (2011) 'Induced sputum in asthma: from bench to bedside', Current medicinal chemistry, 18(10), pp. 1415–1422.

Bao, Z., Lim, S., Liao, W., Lin, Y., Thiemermann, C., Leung, B.P. and Wong, W.S.F. (2007) 'Glycogen synthase kinase-3beta inhibition attenuates asthma in mice', American journal of respiratory and critical care medicine, 176(5), pp. 431–438.

Barbato, C., Arisi, I., Frizzo, M.E., Brandi, R., Da Sacco, L. and Masotti, A. (2009) 'Computational challenges in miRNA target predictions: to be or not to be a true target?', Journal of biomedicine & biotechnology, 2009, p. 803069.

Barnes, P.J. (2006) 'Corticosteroid effects on cell signalling', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 27(2), pp. 413–426.

Barnes, P.J. (2006) 'How corticosteroids control inflammation: Quintiles Prize Lecture 2005', British journal of pharmacology, 148(3), pp. 245–254.

Barnes, P.J. (2016) 'Asthma mechanisms', Medicine, 44(5), pp. 265–270.

Barnes, P.J. (2018) 'Targeting cytokines to treat asthma and chronic obstructive pulmonary disease', Nature reviews. Immunology, 18(7), pp. 454–466.

Barnes, P.J. and Pedersen, S. (1993) 'Efficacy and Safety of Inhaled Corticosteroids in Asthma', The American review of respiratory disease, 148(4_pt_2), pp. S1–S26.

Beasley, R., Semprini, A. and Mitchell, E.A. (2015) 'Risk factors for asthma: is prevention possible?', The Lancet, 386(9998), pp. 1075–1085.

Belair, C.D., Hu, T., Chu, B., Freimer, J.W., Cooperberg, M.R. and Blelloch, R.H. (2019) 'High-throughput, Efficient, and Unbiased Capture of Small RNAs from Low-input Samples for Sequencing', Scientific reports, 9(1), p. 2262.

Bélanger, É., Madore, A.-M., Boucher-Lafleur, A.-M., Simon, M.-M., Kwan, T., Pastinen, T. and Laprise, C. (2020) 'Eosinophil microRNAs Play a Regulatory Role in Allergic Diseases Included in the Atopic March', International journal of molecular sciences, 21(23). Available at: https://doi.org/10.3390/ijms21239011.

Bencurova, P., Baloun, J., Hynst, J., Oppelt, J., Kubova, H., Pospisilova, S. and Brazdil, M. (2021) 'Dynamic miRNA changes during the process of epileptogenesis in an infantile and adult-onset model', Scientific reports, 11(1), p. 9649.

Bhankhur, D., Singla, N., Aggarwal, D. and Chander, J. (2019) 'Prevalence of allergic bronchopulmonary aspergillosis among patients with severe bronchial asthma in a tertiary care hospital in Northern India', Indian journal of pathology & microbiology, 62(1), pp. 111–113.

Bignold, R., Shammout, B., Rowley, J.E., Repici, M., Simms, J. and Johnson, J.R. (2022) 'Chemokine CXCL12 drives pericyte accumulation and airway remodeling in allergic airway disease', Respiratory research, 23(1), p. 183.

Booton, R. and Lindsay, M.A. (2014) 'Emerging role of MicroRNAs and long noncoding RNAs in respiratory disease', Chest, 146(1), pp. 193–204.

Bossley, C.J., Saglani, S., Kavanagh, C., Payne, D.N.R., Wilson, N., Tsartsali, L., Rosenthal, M., Balfour-Lynn, I.M., Nicholson, A.G. and Bush, A. (2009) 'Corticosteroid responsiveness and clinical characteristics in childhood difficult asthma', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 34(5), pp. 1052–1059.

Bousquet, J., Chanez, P., Vignola, A.M., Lacoste, J.Y. and Michel, F.B. (1994) 'Eosinophil inflammation in asthma', American journal of respiratory and critical care medicine, 150(5 Pt 2), pp. S33–8.

Brommage, R., Powell, D.R. and Vogel, P. (2019) 'Predicting human disease mutations and identifying drug targets from mouse gene knockout phenotyping campaigns', Disease models & mechanisms, 12(5). Available at: https://doi.org/10.1242/dmm.038224.

Brown, R.A.M., Epis, M.R., Horsham, J.L., Kabir, T.D., Richardson, K.L. and Leedman, P.J. (2018) 'Total RNA extraction from tissues for microRNA and target gene expression analysis: not all kits are created equal', BMC biotechnology, 18(1), p. 16.

Bui, T.M., Wiesolek, H.L. and Sumagin, R. (2020) 'ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis', Journal of leukocyte biology, 108(3), pp. 787–799.

Buschmann, D., Haberberger, A., Kirchner, B., Spornraft, M., Riedmaier, I., Schelling, G. and Pfaffl, M.W. (2016) 'Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow', Nucleic acids research, 44(13), pp. 5995–6018.

Busse, W., Corren, J., Lanier, B.Q., McAlary, M., Fowler-Taylor, A., Cioppa, G.D., van As, A. and Gupta, N. (2001) 'Omalizumab, anti-IgE recombinant humanized monoclonal antibody, for the treatment of severe allergic asthma', The journal of allergy and clinical immunology, 108(2), pp. 184–190.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F. and Croce, C.M. (2002) 'Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', Proceedings of the National Academy of Sciences of the United States of America, 99(24), pp. 15524–15529.

Calvén, J., Ax, E. and Rådinger, M. (2020) 'The Airway Epithelium-A Central Player in Asthma Pathogenesis', International journal of molecular sciences, 21(23). Available at: https://doi.org/10.3390/ijms21238907.

Campbell, J.D., Liu, G., Luo, L., Xiao, J., Gerrein, J., Juan-Guardela, B., Tedrow, J., Alekseyev, Y.O., Yang, I.V., Correll, M., Geraci, M., Quackenbush, J., Sciurba, F., Schwartz, D.A., Kaminski, N., Johnson, W.E., Monti, S., Spira, A., Beane, J. and Lenburg, M.E. (2015) 'Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data', RNA , 21(2), pp. 164–171.

Carrington, C.B., Addington, W.W., Goff, A.M., Madoff, I.M., Marks, A., Schwaber, J.R. and Gaensler, E.A. (1969) 'Chronic eosinophilic pneumonia', The New England journal of medicine, 280(15), pp. 787–798.

Carr, T.F., Moore, W.C., Kraft, M., Brusselle, G., Castro, M., Chupp, G.L., Wechsler, M.E., Hunter, G., Lindsley, A.W., Llanos, J.-P., Burke, L.K., Chandarana, S. and Ambrose, C.S. (2024) 'Efficacy of tezepelumab in patients with severe, uncontrolled asthma across multiple clinically relevant subgroups in the NAVIGATOR study', Advances in therapy, 41(7), pp. 2978–2990.

Chang, L., Zhou, G., Soufan, O. and Xia, J. (2020) 'miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology', Nucleic acids research, 48(W1), pp. W244–W251.

Chaplin, S. (2018) 'Diagnosis, monitoring and management of chronic asthma', The Prescriber, 29. Available at: https://doi.org/10.1002/PSB.1665.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W. and Hannon, G.J. (2010) 'A dicer-independent miRNA biogenesis pathway that requires Ago catalysis', Nature, 465(7298), pp. 584–589.

Chen, S., Zhong, J., Hu, B., Shao, N. and Deng, C. (2023) 'Whole-genome microRNA sequencing analysis in patients with pulmonary hypertension', Frontiers in genetics, 14. Available at: https://doi.org/10.3389/fgene.2023.1250629.

Chowdhury, N.U., Guntur, V.P., Newcomb, D.C. and Wechsler, M.E. (2021) 'Sex and gender in asthma', European respiratory review: an official journal of the European Respiratory Society, 30(162). Available at: https://doi.org/10.1183/16000617.0067-2021.

Christophe Dessimoz, N.Š. (ed.) (2017) The Gene Ontology Handbook. Springer New York.

Chua, F. and Lai, D. (2007) 'Acute severe asthma: Triage, treatment and thereafter', Current anaesthesia and critical care, 18(2), pp. 61–68.

Churg, J. and Strauss, L. (1951) 'Allergic granulomatosis, allergic angiitis, and periarteritis nodosa', The American journal of pathology, 27(2), pp. 277–301.

Coenye, T. (2021) 'Do results obtained with RNA-sequencing require independent verification?', Biofilms, 3, p. 100043.

Collison, A., Mattes, J., Plank, M. and Foster, P.S. (2011) 'Inhibition of house dust mite–induced allergic airways disease by antagonism of microRNA-145 is comparable to glucocorticoid treatment', The Journal of allergy and clinical immunology, 128(1), pp. 160–167.e4.

Cordero, F., Beccuti, M., Arigoni, M., Donatelli, S. and Calogero, R.A. (2012) 'Optimizing a massive parallel sequencing workflow for quantitative miRNA expression analysis', PloS one, 7(2), p. e31630.

Cottin, V. (2016) 'Eosinophilic Lung Diseases', Clinics in chest medicine, 37(3), pp. 535–556.

Crowe, M., Robinson, D., Sagar, M., Chen, L. and Ghamande, S. (2019) 'Chronic eosinophilic pneumonia: clinical perspectives', Therapeutics and clinical risk management, 15, pp. 397–403.

Darbà, J., Ramírez, G., Sicras, A., García-Bujalance, L., Torvinen, S. and Sánchez-de la Rosa, R. (2016) 'Identification of factors involved in medication compliance: incorrect inhaler technique of asthma treatment leads to poor compliance', Patient preference and adherence, 10, pp. 135–145.

Davis, B.E., Blais, C.M. and Cockcroft, D.W. (2018) 'Methacholine challenge testing: comparative pharmacology', Journal of asthma and allergy, 11, pp. 89–99.

Denayer, T., Stöhr, T. and Van Roy, M. (2014) 'Animal models in translational medicine: Validation and prediction', New Horizons in Translational Medicine, 2(1), pp. 5–11.

Deng, H., Dokshin, G.A., Lei, J., Goldsmith, A.M., Bitar, K.N., Fingar, D.C., Hershenson, M.B. and Bentley, J.K. (2008) 'Inhibition of glycogen synthase kinase-3beta is sufficient for airway smooth muscle hypertrophy', The Journal of biological chemistry, 283(15), pp. 10198–10207.

Depner, M., Kormann, M.S.D., Klopp, N., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Combadière, C. and Kabesch, M. (2007) 'CX3CR1 polymorphisms are associated with atopy but not asthma in German children', International archives of allergy and immunology, 144(1), pp. 91–94.

Deshmukh, S.K., Srivastava, S.K., Zubair, H., Khan, M.A., Singh, A.P. and Singh, S. (2021) 'Resistin induces LIN28A-mediated Let-7a repression in breast cancer cells leading to IL-6 and STAT3 upregulation', Cancers, 13(18), p. 4498.

De Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L. and Fink, C. (2015) 'Reference gene validation for RT-qPCR, a note on different available software packages', PloS one, 10(3), p. e0122515.

Di, T., Yang, Y., Fu, C., Zhang, Z., Qin, C., Sai, X., Liu, J., Hu, C., Zheng, M., Wu, Y. and Bian, T. (2021) 'Let-7 mediated airway remodelling in chronic obstructive pulmonary disease via the regulation of IL-6', European journal of clinical investigation, 51(4), p. e13425.

Di, Y., Schafer, D.W., Cumbie, J.S. and Chang, J.H. (2011) 'The NBP negative binomial model for assessing differential gene expression from RNA-Seq', Statistical applications in genetics and molecular biology, 10(1). Available at: https://doi.org/10.2202/1544-6115.1637.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) 'STAR: ultrafast universal RNA-seq aligner', Bioinformatics (Oxford, England), 29(1), pp. 15–21.

Drake, M.G., Cook, M., Fryer, A.D., Jacoby, D.B. and Scott, G.D. (2021) 'Airway Sensory Nerve Plasticity in Asthma and Chronic Cough', Frontiers in physiology, 12, p. 720538.

D'silva, L., Cook, R.J., Allen, C.J., Hargreave, F.E. and Parameswaran, K. (2007) 'Changing pattern of sputum cell counts during successive exacerbations of airway disease', Respiratory medicine, 101(10), pp. 2217–2220.

Edwards, M.R., Saglani, S., Schwarze, J., Skevaki, C., Smith, J.A., Ainsworth, B., Almond, M., Andreakos, E., Belvisi, M.G., Chung, K.F., Cookson, W., Cullinan, P., Hawrylowicz, C., Lommatzsch, M., Jackson, D., Lutter, R., Marsland, B., Moffatt, M., Thomas, M., Virchow, J.C., Xanthou, G., Edwards, J., Walker, S., Johnston, S.L. and members of the EARIP WP2 working group (2017) 'Addressing unmet needs in understanding asthma mechanisms: From the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators: From the European Asthma Research and Innovation Partnership (EARIP) Work Package (WP)2 collaborators', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 49(5), p. 1602448.

ElKashef, S.M.M.A.E., Ahmad, S.E.-A., Soliman, Y.M.A. and Mostafa, M.S. (2021) 'Role of microRNA-21 and microRNA-155 as biomarkers for bronchial asthma', Innate immunity, 27(1), pp. 61–69.

Esawy, M.M., Baioumy, S.A., Ismail, N.A. and Shabana, M.A. (2021) 'Role of circulating microRNA-132 in allergic bronchopulmonary aspergillosis: A case-control study', Immunobiology, 226(3), p. 152074.

Fadiel, A. and Naftolin, F. (2003) 'Microarray applications and challenges: a vast array of possibilities', Int Arch Biosci [Preprint]. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2a1cba82d25ef50f3557ac1dcd99bb81a06d9709.

Fahy, J.V. (2015) 'Type 2 inflammation in asthma--present in most, absent in many', Nature reviews. Immunology, 15(1), pp. 57–65.

Farne, H.A., Wilson, A., Milan, S., Banchoff, E., Yang, F. and Powell, C.V. (2022) 'Anti-IL-5 therapies for asthma', Cochrane database of systematic reviews, 7(7), p. CD010834.

Farrell, R.E., Jr (2009) RNA Methodologies: Laboratory Guide for Isolation and Characterization. Academic Press.

Farr, R.J., Rootes, C.L., Stenos, J., Foo, C.H., Cowled, C. and Stewart, C.R. (2022) 'Detection of SARS-CoV-2 infection by microRNA profiling of the upper respiratory tract', PloS one, 17(4), p. e0265670.

Fessel, J. (2020) 'The paradox of opposite directions of gene expressions in MCI and AD suggests possible therapy to prevent progression of MCI to AD', Alzheimer's & dementia: the journal of the Alzheimer's Association, 6(1), p. e12003.

Fish, J.E., Santoro, M.M., Morton, S.U., Yu, S., Yeh, R.-F., Wythe, J.D., Ivey, K.N., Bruneau, B.G., Stainier, D.Y.R. and Srivastava, D. (2008) 'miR-126 regulates angiogenic signaling and vascular integrity', Developmental cell, 15(2), pp. 272–284.

Friedman, Y., Balaga, O. and Linial, M. (2013) 'Working together: combinatorial regulation by microRNAs', Advances in experimental medicine and biology, 774, pp. 317–337.

Gao, X., Wang, Y., Zhao, H., Wei, F., Zhang, X., Su, Y., Wang, C., Li, H. and Ren, X. (2016) 'Plasma miR-324-3p and miR-1285 as diagnostic and prognostic biomarkers for early stage lung squamous cell carcinoma', Oncotarget, 7(37), pp. 59664–59675.

Garmire, L.X. and Subramaniam, S. (2012) 'Evaluation of normalization methods in mammalian microRNA-Seq data', RNA , 18(6), pp. 1279–1288.

Garrett, J.K., Jameson, S.C., Thomson, B., Collins, M.H., Wagoner, L.E., Freese, D.K., Beck, L.A., Boyce, J.A., Filipovich, A.H., Villanueva, J.M., Sutton, S.A., Assa'ad, A.H. and Rothenberg, M.E. (2004) 'Anti-interleukin-5 (mepolizumab) therapy for hypereosinophilic syndromes', The Journal of allergy and clinical immunology, 113(1), pp. 115–119.

Gata, A., Neagoe, I.B., Leucuta, D.-C., Budisan, L., Raduly, L., Trombitas, V.E. and Albu, S. (2023) 'MicroRNAs: Potential Biomarkers of Disease Severity in Chronic Rhinosinusitis with Nasal Polyps', Medicina , 59(3). Available at: https://doi.org/10.3390/medicina59030550.

Gautam, A., Kumar, R., Dimitrov, G., Hoke, A., Hammamieh, R. and Jett, M. (2016) 'Identification of extracellular miRNA in archived serum samples by next-generation sequencing from RNA extracted using multiple methods', Molecular biology reports, 43(10), pp. 1165–1178.

Geller, D.E., Kaplowitz, H., Light, M.J. and Colin, A.A. (1999) 'Allergic bronchopulmonary aspergillosis in cystic fibrosis: reported prevalence, regional distribution, and patient characteristics. Scientific Advisory Group, Investigators, and Coordinators of the Epidemiologic Study of Cystic Fibrosis', Chest, 116(3), pp. 639–646.

Gemicioglu, B., Musellim, B., Dogan, I. and Guven, K. (2014) 'Fractional exhaled nitric oxide (FeNo) in different asthma phenotypes', Allergy & rhinology , 5(3), pp. 157–161.

Gershon, A.S., Victor, J.C., Guan, J., Aaron, S.D. and To, T. (2010) 'Use Of Pulmonary Function Tests In The Diagnosis Of Asthma', in B21. HETEROGENEITY AND COMORBIDITY IN ASTHMA. American Thoracic Society 2010 International Conference, May 14-19, 2010 • New Orleans, American Thoracic Society, pp. A2545–A2545.

Ghassemian, A., Park, J.J., Tsoulis, M.W. and Kim, H. (2021) 'Targeting the IL-5 pathway in eosinophilic asthma: a comparison of mepolizumab to benralizumab in the reduction of peripheral eosinophil counts', Allergy, asthma, and clinical immunology: official journal of the Canadian Society of Allergy and Clinical Immunology, 17(1), p. 3.

Gil-Martínez, M., Rodrigo-Muñoz, J.M., Sastre, B., Cañas, J.A., García-Latorre, R., Redondo, N., de la Fuente, L., Mínguez, P., Mahíllo-Fernández, I., Sastre, J., Quirce, S., Caballero, M.L., Olaguibel, J.M. and Del Pozo, V.

(2022) ‘Serum microRNAs Catalog Asthma Patients by Phenotype’, Journal of investigational allergology & clinical immunology: official organ of the International Association of Asthmology and Sociedad Latinoamericana de Alergia e Inmunologia, 32(6), pp. 471–478.

Glinge, C., Clauss, S., Boddum, K., Jabbari, R., Jabbari, J., Risgaard, B., Tomsits, P., Hildebrand, B., Kääb, S., Wakili, R., Jespersen, T. and Tfelt-Hansen, J. (2017) ‘Stability of Circulating Blood-Based MicroRNAs - Pre-Analytic Methodological Considerations’, PloS one, 12(2), p. e0167969.

Global Strategy for Asthma Management and Prevention (2021) GINA Full Report 2021, Global Initiative for Asthma. Available at: www.ginasthma.org (Accessed: 14 October 2021).

Greenberger, P.A. (2002) ‘Allergic bronchopulmonary aspergillosis’, The Journal of allergy and clinical immunology, 110(5), pp. 685–692.

Greenberger, P.A. (2013) ‘When to suspect and work up allergic bronchopulmonary aspergillosis’, Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology, 111(1), pp. 1–4.

de Groot, J.C., Ten Brinke, A. and Bel, E.H.D. (2015) ‘Management of the patient with eosinophilic asthma: a new era begins’, ERJ open research, 1(1). Available at: https://doi.org/10.1183/23120541.00024-2015.

Guida, G., Bagnasco, D., Carriero, V., Bertolini, F., Ricciardolo, F.L.M., Nicola, S., Brussino, L., Nappi, E., Paoletti, G., Canonica, G.W. and Heffler, E. (2022) ‘Critical evaluation of asthma biomarkers in clinical practice’, Frontiers in medicine, 9, p. 969243.

Guideline, N. (2017) ‘Asthma: diagnosis, monitoring andAsthma: diagnosis, monitoring and chronic asthma managementchronic asthma management’. Available at: https://www.nice.org.uk/guidance/ng80/resources/asthma-diagnosis-monitoring-and-chronic-asthma-management-pdf-1837687975621.

Gulyaeva, L.F. and Kushlinskiy, N.E. (2016) ‘Regulatory mechanisms of microRNA expression’, Journal of translational medicine, 14(1), p. 143.

Guo, F.H., De Raeve, H.R., Rice, T.W., Stuehr, D.J., Thunnissen, F.B. and Erzurum, S.C. (1995) ‘Continuous nitric oxide synthesis by inducible nitric oxide synthase in normal human airway epithelium in vivo’, Proceedings of the National Academy of Sciences of the United States of America, 92(17), pp. 7809–7813.

Haas, R., Smith, J., Rocher-Ros, V., Nadkarni, S., Montero-Melendez, T., D’Acquisto, F., Bland, E.J., Bombardieri, M., Pitzalis, C., Perretti, M., Marelli-Berg, F.M. and Mauro, C. (2015) ‘Lactate regulates metabolic and pro-inflammatory circuits in control of T cell migration and effector functions’, PLoS biology, 13(7), p. e1002202.

Ha, M. and Kim, V.N. (2014) ‘Regulation of microRNA biogenesis’, Nature reviews. Molecular cell biology, 15(8), pp. 509–524.

Hanania, N.A., Korenblat, P., Chapman, K.R., Bateman, E.D., Kopecky, P., Paggiaro, P., Yokoyama, A., Olsson, J., Gray, S., Holweg, C.T.J., Eisner, M., Asare, C., Fischer, S.K., Peng, K., Putnam, W.S. and Matthews, J.G. (2016) ‘Efficacy and safety of lebrikizumab in patients with uncontrolled asthma (LAVOLTA I and LAVOLTA II): replicate, phase 3, randomised, double-blind, placebo-controlled trials’, The Lancet. Respiratory medicine, 4(10), pp. 781–796.

Hansen, J. and Desjardins, P. (2004) ‘Novel spectrophotometer enables rapid quantitation of one-microliter samples’, PharmaGenomics, 4, p. S23+.

Harris, T.A., Yamakuchi, M., Ferlito, M., Mendell, J.T. and Lowenstein, C.J. (2008) 'MicroRNA-126 regulates endothelial expression of vascular cell adhesion molecule 1', Proceedings of the National Academy of Sciences of the United States of America, 105(5), pp. 1516–1521.

Hashimoto, Y., Akiyama, Y. and Yuasa, Y. (2013) 'Multiple-to-multiple relationships between microRNAs and target genes in gastric cancer', PloS one, 8(5), p. e62589.

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A.Y.M., T Flåm, S., Leithaug, M., Dalland, M., Farmer, A., Henderson, J.M., Hussong, M.A., Moll, P., Nguyen, L., McNulty, A., Shaffer, J.M., Shore, S., Yip, H.K., Vitkovska, J., Rayner, S., Lie, B.A. and Gilfillan, G.D. (2020) 'Systematic assessment of commercially available low-input miRNA library preparation kits', RNA biology, 17(1), pp. 75–86.

Heinicke, F., Zhong, X., Zucknick, M., Breidenbach, J., Sundaram, A.Y.M., T Flåm, S., Leithaug, M., Dalland, M., Rayner, S., Lie, B.A. and Gilfillan, G.D. (2020) 'An extension to: Systematic assessment of commercially available low-input miRNA library preparation kits', RNA biology, 17(9), pp. 1284–1292.

Herrera-Luis, E., Martin-Almeida, M. and Pino-Yanes, M. (2024) 'Asthma-genomic advances toward risk prediction', Clinics in chest medicine, 45(3), pp. 599–610.

Hill, P. (2023) Gene Ontology: A Resource for Analysis and Interpretation of Biological Database. Syrawood Publishing House.

Hinson, K.F.W., Moon, A.J. and Plummer, N.S. (1952) 'Broncho-pulmonary aspergillosis; a review and a report of eight new cases', Thorax, 7(4), pp. 317–333.

Hoefig, K.P. and Heissmeyer, V. (2008) 'MicroRNAs grow up in the immune system', Current opinion in immunology, 20(3), pp. 281–287.

Holmes, A.M., Solari, R. and Holgate, S.T. (2011) 'Animal models of asthma: value, limitations and opportunities for alternative approaches', Drug discovery today, 16(15-16), pp. 659–670.

Hoshino, M. and Ohtawa, J. (2012) 'Effects of adding omalizumab, an anti-immunoglobulin E antibody, on airway wall thickening in asthma', Respiration; international review of thoracic diseases, 83(6), pp. 520–528.

Huang, H.-Y., Lin, Y.-C.-D., Cui, S., Huang, Y., Tang, Y., Xu, J., Bao, J., Li, Y., Wen, J., Zuo, H., Wang, W., Li, J., Ni, J., Ruan, Y., Li, L., Chen, Y., Xie, Y., Zhu, Z., Cai, X., Chen, X., Yao, L., Chen, Y., Luo, Y., LuXu, S., Luo, M., Chiu, C.-M., Ma, K., Zhu, L., Cheng, G.-J., Bai, C., Chiang, Y.-C., Wang, L., Wei, F., Lee, T.-Y. and Huang, H.-D. (2022) 'miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions', Nucleic acids research, 50(D1), pp. D222–D230.

Hücker, S.M., Fehlmann, T., Werno, C., Weidele, K., Lüke, F., Schlenska-Lange, A., Klein, C.A., Keller, A. and Kirsch, S. (2021) 'Single-cell microRNA sequencing method comparison and application to cell lines and circulating lung tumor cells', Nature communications, 12(1), p. 4316.

Hughes, T.R. (2009) '"Validation" in genome-scale research', Journal of biology, 8(1), p. 3.

Humbert, M., Taillé, C., Mala, L., Le Gros, V., Just, J., Molimard, M. and STELLAIR investigators (2018) 'Omalizumab effectiveness in patients with severe allergic asthma according to blood eosinophil count: the STELLAIR study', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 51(5). Available at: https://doi.org/10.1183/13993003.02523-2017.

Huntley, R.P., Sitnikov, D., Orlic-Milacic, M., Balakrishnan, R., D'Eustachio, P., Gillespie, M.E., Howe, D., Kalea, A.Z., Maegdefessel, L., Osumi-Sutherland, D., Petri, V., Smith, J.R., Van Auken, K., Wood, V., Zampetaki, A.,

Mayr, M. and Lovering, R.C. (2016) 'Guidelines for the functional annotation of microRNAs using the Gene Ontology', RNA , 22(5), pp. 667–676.

Huo, X., Zhang, K., Yi, L., Mo, Y., Liang, Y., Zhao, J., Zhang, Z., Xu, Y. and Zhen, G. (2016) 'Decreased epithelial and plasma miR-181b-5p expression associates with airway eosinophilic inflammation in asthma', Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology, 46(10), pp. 1281–1290.

Hur, J., Rhee, C.K., Lee, S.Y., Kim, Y.K. and Kang, J.Y. (2021) 'MicroRNA-21 inhibition attenuates airway inflammation and remodelling by modulating the transforming growth factor β-Smad7 pathway', The Korean journal of internal medicine, 36(3), pp. 706–720.

Iacomino, G. and Siani, A. (2017) 'Role of microRNAs in obesity and obesity-related diseases', Genes & nutrition, 12(1), p. 23.

Ilmarinen, P. and Kankaanranta, H. (2014) 'Eosinophil apoptosis as a therapeutic target in allergic asthma', Basic & clinical pharmacology & toxicology, 114(1), pp. 109–117.

James, A.L., Elliot, J.G., Jones, R.L., Carroll, M.L., Mauad, T., Bai, T.R., Abramson, M.J., McKay, K.O. and Green, F.H. (2012) 'Airway smooth muscle hypertrophy and hyperplasia in asthma', American journal of respiratory and critical care medicine, 185(10), pp. 1058–1064.

Jansson, M.D. and Lund, A.H. (2012) 'MicroRNA and cancer', Molecular oncology, 6(6), pp. 590–610.

Jat, K.R., Vaidya, P.C., Mathew, J.L., Jondhale, S. and Singh, M. (2018) 'Childhood allergic bronchopulmonary aspergillosis', Lung India: official organ of Indian Chest Society, 35(6), pp. 499–507.

Jia, G., Erickson, R.W., Choy, D.F., Mosesova, S., Wu, L.C., Solberg, O.D., Shikotra, A., Carter, R., Audusseau, S., Hamid, Q., Bradding, P., Fahy, J.V., Woodruff, P.G., Harris, J.M., Arron, J.R. and Bronchoscopic Exploratory Research Study of Biomarkers in Corticosteroid-refractory Asthma (BOBCAT) Study Group (2012) 'Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients', The journal of allergy and clinical immunology, 130(3), pp. 647–654.e10.

Jiang, S., Xu, Z., Shi, Y., Liang, S., Jiang, X., Xiao, M., Wang, K. and Ding, L. (2023) 'Circulating insulin-like growth factor-1 and risk of lung diseases: A Mendelian randomization analysis', Frontiers in endocrinology, 14, p. 1126397.

Ju, C. (2011) 'Determining overrepresentation of gene ontology terms using the hypergeometric distribution'. Available at: https://pages.pomona.edu/~jsh04747/Student%20Theses/ChristineJu_Scripps11.pdf.

Kaya, H., Gümüş, S., Uçar, E., Aydoğan, M., Muşabak, U., Tozkoparan, E. and Bilgiç, H. (2012) 'Omalizumab as a steroid-sparing agent in chronic eosinophilic pneumonia', Chest, 142(2), pp. 513–516.

Keown, A. (2018) AstraZeneca Drops Phase III Asthma Drug as Company Posts Disappointing 2017 Financials, BioSpace. Available at: https://www.biospace.com/unique-astrazeneca-drops-phase-iii-asthma-drug-as-company-posts-disappointing-2017-financials (Accessed: 30 August 2024).

Kermani, N.Z., Li, C.-X., Versi, A., Badi, Y., Sun, K., Abdel-Aziz, M.I., Bonatti, M., Maitland-van der Zee, A.-H., Djukanovic, R., Wheelock, Å., Dahlen, S.-E., Howarth, P., Guo, Y., Chung, K.F., Adcock, I.M. and U-BIOPRED Project Team (2024) 'Endotypes of severe neutrophilic and eosinophilic asthma from multi-omics integration of U-BIOPRED sputum samples', Clinical and translational medicine, 14(7), p. e1771.

Keuls, R.A. and Parchem, R. (2020) 'Preparation of Small RNA Libraries for Sequencing from Early Mouse Embryos', Journal of visualized experiments: JoVE [Preprint], (164). Available at: https://doi.org/10.3791/61086.

Khamina, K., Diendorfer, A.B., Skalicky, S., Weigl, M., Pultar, M., Krammer, T.L., Fournier, C.A., Schofield, A.L., Otto, C., Smith, A.T., Buchtele, N., Schoergenhofer, C., Jilma, B., Frank, B.J.H., Hofstaetter, J.G., Grillari, R., Grillari, J., Ruprecht, K., Goldring, C.E., Rehrauer, H., Glaab, W.E. and Hackl, M. (2022) 'A MicroRNA Next-generation-Sequencing Discovery Assay (miND) for genome-scale analysis and absolute quantitation of circulating MicroRNA biomarkers', International journal of molecular sciences, 23(3), p. 1226.

Khatri, P. and Drăghici, S. (2005) 'Ontological analysis of gene expression data: current tools, limitations, and open problems', Bioinformatics , 21(18), pp. 3587–3595.

Kim, R.Y., Horvat, J.C., Pinkerton, J.W., Starkey, M.R., Essilfie, A.T., Mayall, J.R., Nair, P.M., Hansbro, N.G., Jones, B., Haw, T.J., Sunkara, K.P., Nguyen, T.H., Jarnicki, A.G., Keely, S., Mattes, J., Adcock, I.M., Foster, P.S. and Hansbro, P.M. (2017) 'MicroRNA-21 drives severe, steroid-insensitive experimental asthma by amplifying phosphoinositide 3-kinase-mediated suppression of histone deacetylase 2', The Journal of allergy and clinical immunology, 139(2), pp. 519–532.

Kim, Y.-K. (2015) 'Extracellular microRNAs as Biomarkers in Human Disease', Chonnam medical journal, 51(2), pp. 51–57.

King, C.E., Cuatrecasas, M., Castells, A., Sepulveda, A.R., Lee, J.-S. and Rustgi, A.K. (2011) 'LIN28B promotes colon cancer progression and metastasis', Cancer research, 71(12), pp. 4260–4268.

Kingdom, R. and Wright, C.F. (2022) 'Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts', Frontiers in genetics, 13, p. 920390.

Kips, J.C., Anderson, G.P., Fredberg, J.J., Herz, U., Inman, M.D., Jordana, M., Kemeny, D.M., Lötvall, J., Pauwels, R.A., Plopper, C.G., Schmidt, D., Sterk, P.J., Van Oosterhout, A.J.M., Vargaftig, B.B. and Chung, K.F. (2003) 'Murine models of asthma', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 22(2), pp. 374–382.

Klain, A., Dinardo, G., Salvatori, A., Indolfi, C., Contieri, M., Brindisi, G., Decimo, F., Zicari, A.M. and Miraglia Del Giudice, M. (2022) 'An overview on the primary factors that contribute to non-allergic asthma in children', Journal of clinical medicine, 11(21), p. 6567.

Kozera, B. and Rapacz, M. (2013) 'Reference genes in real-time PCR', Journal of applied genetics, 54(4), pp. 391–406.

Kozomara, A. and Griffiths-Jones, S. (2014) 'miRBase: annotating high confidence microRNAs using deep sequencing data', Nucleic acids research, 42(Database issue), pp. D68–73.

Kwak, Y.-G., Song, C.H., Yi, H.K., Hwang, P.H., Kim, J.-S., Lee, K.S. and Lee, Y.C. (2003) 'Involvement of PTEN in airway hyperresponsiveness and inflammation in bronchial asthma', The journal of clinical investigation, 111(7), pp. 1083–1092.

Ladjemi, M.Z., Gras, D., Dupasquier, S., Detry, B., Lecocq, M., Garulli, C., Fregimilicka, C., Bouzin, C., Gohy, S., Chanez, P. and Pilette, C. (2018) 'Bronchial Epithelial IgA Secretion Is Impaired in Asthma. Role of IL-4/IL-13', American journal of respiratory and critical care medicine, 197(11), pp. 1396–1409.

Lama, L., Cobo, J., Buenaventura, D. and Ryan, K. (2019) 'Small RNA-seq: The RNA 5'-end adapter ligation problem and how to circumvent it', Journal of biological methods, 6(1). Available at: https://doi.org/10.14440/jbm.2019.269.

Lamba, V., Ghodke-Puranik, Y., Guan, W. and Lamba, J.K. (2014) 'Identification of suitable reference genes for hepatic microRNA quantitation', BMC research notes, 7(1), p. 129.

Lambert, K.A., Roff, A.N., Panganiban, R.P., Douglas, S. and Ishmael, F.T. (2018) 'MicroRNA-146a is induced by inflammatory stimuli in airway epithelial cells and augments the anti-inflammatory effects of glucocorticoids', PloS one, 13(10), p. e0205434.

Lambrecht, B.N., Hammad, H. and Fahy, J.V. (2019) 'The Cytokines of Asthma', Immunity, 50(4), pp. 975–991.

Lang, J.E., Holbrook, J.T., Mougey, E.B., Wei, C.Y., Wise, R.A., Teague, W.G., Lima, J.J. and American Lung Association-Airways Clinical Research Centers (2015) 'Lansoprazole Is Associated with Worsening Asthma Control in Children with the CYP2C19 Poor Metabolizer Phenotype', Annals of the American Thoracic Society, 12(6), pp. 878–885.

Lee, C.-C., Wang, C.-N., Lee, Y.-L., Tsai, Y.-R. and Liu, J.-J. (2015) 'High mobility group box 1 induced human lung myofibroblasts differentiation and enhanced migration by activation of MMP-9', PloS one, 10(2), p. e0116393.

Lee, J.-H., Wang, L.-C., Lin, Y.-T., Yang, Y.-H., Yu, H.-H., Hu, Y.-C. and Chiang, B.-L. (2023) 'Differentially expressed microRNAs in peripheral blood cell are associated with downregulated expression of IgE in nonallergic childhood asthma', Scientific reports, 13(1), p. 6381.

Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', Cell, 75(5), pp. 843–854.

Leite-de-Moraes, M., Hammad, H. and Dy, M. (2012) 'Crosstalk between innate and adaptive cells on allergic process', Journal of allergy, 2012, p. 720568.

Leiten, E.O., Martinsen, E.M.H., Bakke, P.S., Eagan, T.M.L. and Grønseth, R. (2016) 'Complications and discomfort of bronchoscopy: a systematic review', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 3, p. 33324.

Lett, K.E., Logan, M.K., McLaurin, D.M. and Hebert, M.D. (2021) 'Coilin enhances phosphorylation and stability of DGCR8 and promotes miRNA biogenesis', Molecular biology of the cell, 32(20), p. br4.

Levänen, B., Bhakta, N.R., Torregrosa Paredes, P., Barbeau, R., Hiltbrunner, S., Pollack, J.L., Sköld, C.M., Svartengren, M., Grunewald, J., Gabrielsson, S., Eklund, A., Larsson, B.-M., Woodruff, P.G., Erle, D.J. and Wheelock, Å.M. (2013) 'Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients', The Journal of allergy and clinical immunology, 131(3), pp. 894–903.

Li, J., Panganiban, R., Kho, A.T., McGeachie, M.J., Farnam, L., Chase, R.P., Weiss, S.T., Lu, Q. and Tantisira, K.G. (2020) 'Circulating MicroRNAs and Treatment Response in Childhood Asthma', American journal of respiratory and critical care medicine, 202(1), pp. 65–72.

Lim, H.F. and Nair, P. (2015) 'Efficacy and safety of reslizumab in patients with moderate to severe eosinophilic asthma', Expert review of respiratory medicine, 9(2), pp. 135–142.

Li, M., Li, M., Hou, Y., He, H., Jiang, R., Wang, C. and Sun, S. (2023) 'Ferroptosis triggers airway inflammation in asthma', Therapeutic advances in respiratory disease, 17, p. 17534666231208628.

Li, M., Zhu, W., Saeed, U., Sun, S., Fang, Y., Wang, C. and Luo, Z. (2022) 'Identification of the molecular subgroups in asthma by gene expression profiles: airway inflammation implications', BMC pulmonary medicine, 22(1), p. 29.

Lin, C., Li, X., Zhang, Y., Guo, Y., Zhou, J., Gao, K., Dai, J., Hu, G., Lv, L., Du, J. and Zhang, Y. (2015) 'The microRNA feedback regulation of p63 in cancer progression', Oncotarget, 6(11), pp. 8434–8453.

Lindsay, M.A. (2008) 'microRNAs and the immune response', Trends in immunology, 29(7), pp. 343–351.

Liu, C., Chu, D., Kalantar-Zadeh, K., George, J., Young, H.A. and Liu, G. (2021) 'Cytokines: From Clinical Significance to Quantification', Advancement of science, 8(15), p. e2004433.

Liu, N., Bezprozvannaya, S., Williams, A.H., Qi, X., Richardson, J.A., Bassel-Duby, R. and Olson, E.N. (2008) 'microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart', Genes & development, 22(23), pp. 3242–3254.

Liu, N., Wang, D., Tian, J., Wang, X., Shi, H., Wang, C., Jiang, Y., Pang, M., Fan, X., Zhao, J., Liu, L., Wu, H., Guan, L., Zheng, H., Shi, D. and Zhang, Z. (2023) 'PM2.5-bound metals and blood metals are associated with pulmonary function and Th17/Treg imbalance: A panel study of asthmatic adults', Chemosphere, 340(139869), p. 139869.

Livak, K.J. and Schmittgen, T.D. (2001) 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', Methods (San Diego, Calif.), 25(4), pp. 402–408.

Li, X., Nair, A., Wang, S. and Wang, L. (2015) 'Quality control of RNA-seq experiments', Methods in molecular biology , 1269, pp. 137–146.

Li, Y. and Kowdley, K.V. (2012) 'Method for microRNA isolation from clinical serum samples', Analytical biochemistry, 431(1), pp. 69–75.

Li, Z., Chen, D., Wang, Q., Tian, H., Tan, M., Peng, D., Tan, Y., Zhu, J., Liang, W. and Zhang, L. (2021) 'mRNA and microRNA stability validation of blood samples under different environmental conditions', Forensic science international. Genetics, 55, p. 102567.

Llera-Oyola, J., Carceller, H., Andreu, Z., Hidalgo, M.R., Soler-Sáez, I., Gordillo, F., Gómez-Cabañes, B., Roson, B., de la Iglesia-Vayá, M., Mancuso, R., Guerini, F.R., Mizokami, A. and García-García, F. (2024) 'The role of microRNAs in understanding sex-based differences in Alzheimer's disease', Biology of sex differences, 15(1), p. 13.

Ludwig, N., Becker, M., Schumann, T., Speer, T., Fehlmann, T., Keller, A. and Meese, E. (2017) 'Bias in recent miRBase annotations potentially associated with RNA quality issues', Scientific reports, 7(1), p. 5162.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E. and Keller, A. (2016) 'Distribution of miRNA expression across human tissues', Nucleic acids research, 44(8), pp. 3865–3877.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R. and Golub, T.R. (2005) 'MicroRNA expression profiles classify human cancers', Nature, 435(7043), pp. 834–838.

Lu, T.X., Hartner, J., Lim, E.-J., Fabry, V., Mingler, M.K., Cole, E.T., Orkin, S.H., Aronow, B.J. and Rothenberg, M.E. (2011) 'MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity', Journal of immunology , 187(6), pp. 3362–3373.

Lynch, H.E., Sanchez, A.M., D'Souza, M.P., Rountree, W., Denny, T.N., Kalos, M. and Sempowski, G.D. (2014) 'Development and implementation of a proficiency testing program for Luminex bead-based cytokine assays', Journal of immunological methods, 409, pp. 62–71.

MacPherson, J.C., Comhair, S.A.A., Erzurum, S.C., Klein, D.F., Lipscomb, M.F., Kavuru, M.S., Samoszuk, M.K. and Hazen, S.L. (2001) 'Eosinophils are a major source of nitric oxide-derived oxidants in severe asthma: characterization of pathways available to eosinophils for generating reactive nitrogen species', The Journal of Immunology, 166(9), pp. 5763–5772.

Maes, T., Cobos, F.A., Schleich, F., Sorbello, V., Henket, M., De Preter, K., Bracke, K.R., Conickx, G., Mesnil, C., Vandesompele, J., Lahousse, L., Bureau, F., Mestdagh, P., Joos, G.F., Ricciardolo, F.L.M., Brusselle, G.G. and Louis, R. (2016) 'Asthma inflammatory phenotypes show differential microRNA expression in sputum', The Journal of allergy and clinical immunology, 137(5), pp. 1433–1446.

Mahmoud, A.A.F., Austen, K.F. and Simon, A.S. (1980) The Eosinophil in Health and Disease. Grune & Stratton.

Mao, Z., Shi, Y., Cao, Q., Chen, Y., Sun, Y., Liu, Z., Zhang, Q. and Huang, M. (2018) 'Transcriptional regulation on the gene expression signature in combined allergic rhinitis and asthma syndrome', Epigenomics, 10(2), pp. 119–131.

Maqueda, J.J., Giovanazzi, A., Rocha, A.M., Rocha, S., Silva, I., Saraiva, N., Bonito, N., Carvalho, J., Maia, L., Wauben, M.H.M. and Oliveira, C. (2023) 'Adapter dimer contamination in sRNA-sequencing datasets predicts sequencing failure and batch effects and hampers extracellular vesicle-sRNA analysis', Journal of extracellular biology, 2(6). Available at: https://doi.org/10.1002/jex2.91.

Marchand, E. and Cordier, J.-F. (2006) 'Idiopathic chronic eosinophilic pneumonia', Orphanet journal of rare diseases, 1, p. 11.

Martinez-Nunez, R.T., Rupani, H., Platé, M., Niranjan, M., Chambers, R.C., Howarth, P.H. and Sanchez-Elsner, T. (2018) 'Genome-Wide Posttranscriptional Dysregulation by MicroRNAs in Human Asthma as Revealed by Frac-seq', Journal of immunology , 201(1), pp. 251–263.

Ma, S., Kotar, A., Grote, S., Rouskin, S. and Keane, S.C. (2023) 'Structure of pre-miR-31 reveals an active role in Dicer processing', bioRxiv : the preprint server for biology [Preprint]. Available at: https://doi.org/10.1101/2023.01.03.519659.

Mattes, J., Collison, A., Plank, M., Phipps, S. and Foster, P.S. (2009) 'Antagonism of microRNA-126 suppresses the effector function of TH2 cells and the development of allergic airways disease', Proceedings of the National Academy of Sciences of the United States of America, 106(44), pp. 18704–18709.

Mattiuzzi, C. and Lippi, G. (2020) 'Worldwide asthma epidemiology: insights from the Global Health Data Exchange database', International forum of allergy & rhinology, 10(1), pp. 75–80.

Maurya, N., Awasthi, S. and Dixit, P. (2012) 'Association of CFTR gene mutation with bronchial asthma', The Indian journal of medical research, 135(4), pp. 469–478.

Ma, X., Lou, C., Pan, J., Zhou, C., Zhao, X., Li, N., Tian, H. and Meng, X. (2023) 'The diagnostic potential of a circRNA-miRNA network in non-small cell lung cancer', Journal of molecular medicine , 101(6), pp. 671–684.

Mc Brien, C.N. (2020) The Bacterial and Fungal Communities in the Airways of Adults with Asthma and Eosinophilic Lung Diseases. Edited by M.M.W.C. Menzies-Gow. PhD. Imperial College London.

Meijer, H.A., Smith, E.M. and Bushell, M. (2014) 'Regulation of miRNA strand selection: follow the leader?', Biochemical Society transactions, 42(4), pp. 1135–1140.

Michalek, R.D., Gerriets, V.A., Jacobs, S.R., Macintyre, A.N., MacIver, N.J., Mason, E.F., Sullivan, S.A., Nichols,

A.G. and Rathmell, J.C. (2011) 'Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets', The journal of immunology, 186(6), pp. 3299–3303.

Mitchell, M.I., Ben-Dov, I.Z., Ye, K., Liu, C., Shi, M., Sadoughi, A., Shah, C., Siddiqui, T., Okorozo, A., Gutierrez, M., Unawane, R., Biamonte, L., Parikh, K., Spivack, S. and Loudig, O. (2024) 'Exhaled breath condensate contains extracellular vesicles (EVs) that carry miRNA cargos of lung tissue origin that can be selectively purified and analyzed', Journal of extracellular vesicles, 13(4), p. e12440.

Moffatt, M.F. and Cookson, W.O. (2017) 'The lung microbiome in health and disease', Clinical medicine, 17(6), pp. 525–529.

Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., Heinzmann, A., Simma, B., Frischer, T., Willis-Owen, S.A.G., Wong, K.C.C., Illig, T., Vogelberg, C., Weiland, S.K., von Mutius, E., Abecasis, G.R., Farrall, M., Gut, I.G., Lathrop, G.M. and Cookson, W.O.C. (2007) 'Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma', Nature, 448(7152), pp. 470–473.

Moncada, S., Palmer, R.M. and Higgs, E.A. (1991) 'Nitric oxide: physiology, pathophysiology, and pharmacology', Pharmacological reviews, 43(2), pp. 109–142.

Moretti, F., D'Antona, P., Finardi, E., Barbetta, M., Dominioni, L., Poli, A., Gini, E., Noonan, D.M., Imperatori, A., Rotolo, N., Cattoni, M. and Campomenosi, P. (2017) 'Systematic review and critique of circulating miRNAs as biomarkers of stage I-II non-small cell lung cancer', Oncotarget, 8(55), pp. 94980–94996.

Muhammad, S.A., Fatima, N., Syed, N.-I.-H., Wu, X., Yang, X.F. and Chen, J.Y. (2015) 'MicroRNA Expression Profiling of Human Respiratory Epithelium Affected by Invasive Candida Infection', PloS one, 10(8), p. e0136454.

Mukherjee, K., Amsel, D., Kalsy, M., Billion, A., Dobrindt, U. and Vilcinskas, A. (2020) 'MicroRNAs regulate innate immunity against uropathogenic and commensal-like Escherichia coli infections in the surrogate insect model Galleria mellonella', Scientific reports, 10(1), p. 2570.

Mukherjee, M., Sehmi, R. and Nair, P. (2014) 'Anti-IL5 therapy for asthma and beyond', The World Allergy Organization journal, 7(1), p. 32.

Mullard, A. (2022) 'FDA approves first-in-class TSLP-targeted antibody for severe asthma', Nature reviews. Drug discovery, 21(2), p. 89.

Munafo, D., Hoda, S., Ettwiller, L. and Langhorst, B. (2013) Automated Size Selection of NEBNext Small RNA Libraries with the Sage Pippin Prep. Available at: http://www.sagescience.com/wp-content/uploads/2014/04/Sage_AppNote_NEBNextRNALib_Final_12_20_13_3.pdf.

Nadi, E., Hajilooi, M., Babakhani, D. and Rafiei, A. (2012) 'Platelet endothelial cell adhesion molecule-1 polymorphism in patients with bronchial asthma', Iranian journal of allergy, asthma, and immunology, 11(4), pp. 276–281.

Nair, P. and O'Byrne, P.M. (2019) 'The interleukin-13 paradox in asthma: effective biology, ineffective biologicals', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology, 53(2), p. 1802250.

Newton, A.C. and Trotman, L.C. (2014) 'Turning off AKT: PHLPP as a drug target', Annual review of pharmacology and toxicology, 54, pp. 537–558.

Nilsen, T.W. (2013) 'The fundamentals of RNA purification', Cold Spring Harbor protocols, 2013(7), pp. 618–624.

Novosad, J., Krčmová, I., Bartoš, V., Drahošová, M., Vaník, P., Růžičková-Kirchnerová, O., Teřl, M. and Krejsek, J. (2020) 'Serum periostin levels in asthma patients in relation to omalizumab therapy and presence of chronic rhinosinusitis with nasal polyps', Postepy dermatologii i alergologii, 37(2), pp. 240–249.

O'Brien, J., Hayder, H., Zayed, Y. and Peng, C. (2018) 'Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation', Frontiers in endocrinology, 9, p. 402.

O'Connell, R.M., Rao, D.S., Chaudhuri, A.A. and Baltimore, D. (2010) 'Physiological and pathological roles for microRNAs in the immune system', Nature reviews. Immunology, 10(2), pp. 111–122.

Odgers, T. and Gow, A. (2014) 'P166 - Inducible nitric oxide synthase is responsible for macrophage recruitment and activation following bleomycin mediated lung injury', Nitric oxide: biology and chemistry / official journal of the Nitric Oxide Society, 42, pp. 135–136.

Oglesby, I.K., Bray, I.M., Chotirmall, S.H., Stallings, R.L., O'Neill, S.J., McElvaney, N.G. and Greene, C.M. (2010) 'miR-126 is downregulated in cystic fibrosis airway epithelial cells and regulates TOM1 expression', Journal of immunology , 184(4), pp. 1702–1709.

Oliphant, C.J., Barlow, J.L. and McKenzie, A.N.J. (2011) 'Insights into the initiation of type 2 immune responses: Initiation of type 2 responses', Immunology, 134(4), pp. 378–385.

Ono, S., Lam, S., Nagahara, M. and Hoon, D.S.B. (2015) 'Circulating microRNA Biomarkers as Liquid Biopsy for Cancer Patients: Pros and Cons of Current Assays', Journal of clinical medicine research, 4(10), pp. 1890–1907.

Pandey, G., Pandey, O.P., Rogers, A.J., Ahsen, M.E., Hoffman, G.E., Raby, B.A., Weiss, S.T., Schadt, E.E. and Bunyavanich, S. (2018) 'A Nasal Brush-based Classifier of Asthma Identified by Machine Learning Analysis of Nasal RNA Sequence Data', Scientific reports, 8(1), p. 8826.

Panek, M.G., Karbownik, M.S., Górski, K.M., Koćwin, M., Kardas, G., Marynowski, M. and Kuna, P. (2022) 'New insights into the regulation of TGF-β/Smad and MPK signaling pathway gene expressions by nasal allergen and methacholine challenge test in asthma', Clinical and translational allergy, 12(7), p. e12172.

Panganiban, R.P., Lambert, K.A., Hsu, M.-H., Laryea, Z. and Ishmael, F.T. (2019) 'Isolation and profiling of plasma microRNAs: Biomarkers for asthma and allergic rhinitis', Methods , 152, pp. 48–54.

Panganiban, R.P., Wang, Y., Howrylak, J., Chinchilli, V.M., Craig, T.J., August, A. and Ishmael, F.T. (2016) 'Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma', The Journal of allergy and clinical immunology, 137(5), pp. 1423–1432.

Papiris, S., Kotanidou, A., Malagari, K. and Roussos, C. (2002) 'Clinical review: severe asthma', Critical care / the Society of Critical Care Medicine, 6(1), pp. 30–44.

Park, J.H., Chang, H.S., Park, C.-S., Jang, A.-S., Park, B.L., Rhim, T.Y., Uh, S.-T., Kim, Y.H., Chung, I.Y. and Shin, H.D. (2007) 'Association analysis of CD40 polymorphisms with asthma and the level of serum total IgE', American journal of respiratory and critical care medicine, 175(8), pp. 775–782.

Pascoe, C.D., Obeidat, M. 'en, Arsenault, B.A., Nie, Y., Warner, S., Stefanowicz, D., Wadsworth, S.J., Hirota, J.A., Jasemine Yang, S., Dorscheid, D.R., Carlsten, C., Hackett, T.L., Seow, C.Y. and Paré, P.D. (2017) 'Gene expression analysis in asthma using a targeted multiplex array', BMC pulmonary medicine, 17(1), p. 189.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. and Ruvkun, G. (2000) 'Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA', Nature, 408(6808), pp. 86–89.

Patterson, R., Greenberger, P.A., Lee, T.M., Liotta, J.L., O'Neill, E.A., Roberts, M. and Sommers, H. (1987) 'Prolonged evaluation of patients with corticosteroid-dependent asthma stage of allergic bronchopulmonary aspergillosis', The Journal of allergy and clinical immunology, 80(5), pp. 663–668.

Pavlidis, P., Jensen, J.D., Stephan, W. and Stamatakis, A. (2012) 'A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans', Molecular biology and evolution, 29(10), pp. 3237–3248.

Pavord, I.D., Hoyte, F.C.L., Lindsley, A.W., Ambrose, C.S., Spahn, J.D., Roseti, S.L., Cook, B., Griffiths, J.M., Hellqvist, Å., Martin, N., Llanos, J.-P., Martin, N., Colice, G. and Corren, J. (2023) 'Tezepelumab reduces exacerbations across all seasons in patients with severe, uncontrolled asthma (NAVIGATOR)', Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology, 131(5), pp. 587–597.e3.

Peake, H.L., Currie, A.J., Stewart, G.A. and McWilliam, A.S. (2003) 'Nitric oxide production by alveolar macrophages in response to house dust mite fecal pellets and the mite allergens, Der p 1 and Der p 2', The Journal of allergy and clinical immunology, 112(3), pp. 531–537.

Pelaia, C., Paoletti, G., Puggioni, F., Racca, F., Pelaia, G., Canonica, G.W. and Heffler, E. (2019) 'Interleukin-5 in the Pathophysiology of Severe Asthma', Frontiers in Physiology, 10. Available at: https://doi.org/10.3389/fphys.2019.01514.

Peng, J., Wang, H., Lu, J., Hui, W., Wang, Y. and Shang, X. (2017) 'Identifying term relations cross different gene ontology categories', BMC bioinformatics, 18(Suppl 16), p. 573.

Peters, U., Dixon, A.E. and Forno, E. (2018) 'Obesity and asthma', The Journal of allergy and clinical immunology, 141(4), pp. 1169–1179.

Pfaffl, M.W., Tichopad, A., Prgomet, C. and Neuvians, T.P. (2004) 'Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations', Biotechnology letters, 26(6), pp. 509–515.

Phipps, S., Lam, C.E., Kaiko, G.E., Foo, S.Y., Collison, A., Mattes, J., Barry, J., Davidson, S., Oreo, K., Smith, L. and Others (2009) 'Toll-interleukin-1 signalling is critical for house dust mite-specific Th2 and Th17 responses', American journal of respiratory and critical care medicine, 179, pp. 883–893.

Piatek, M.J. and Werner, A. (2014) 'Endogenous siRNAs: regulators of internal affairs', Biochemical Society transactions, 42(4), pp. 1174–1179.

Picado, C. (1992) 'Early and late-phase asthmatic reactions: a hypothesis', Allergy, 47(4 Pt 1), pp. 331–333.

Picado, C., Deulofeu, R., Lleonart, R., Agustí, M., Casals, E., Quintó, L. and Mullol, J. (1999) 'Lipid and protein metabolism in asthma. Effects of diet and corticosteroid therapy', Allergy, 54(6), pp. 569–575.

Porcu, E., Sadler, M.C., Lepik, K., Auwerx, C., Wood, A.R., Weihs, A., Sleiman, M.S.B., Ribeiro, D.M., Bandinelli, S., Tanaka, T., Nauck, M., Völker, U., Delaneau, O., Metspalu, A., Teumer, A., Frayling, T., Santoni, F.A., Reymond, A. and Kutalik, Z. (2021) 'Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome', Nature communications, 12(1), p. 5647.

Prabahar, A. and Natarajan, J. (2017) 'MicroRNA mediated network motifs in autoimmune diseases and its crosstalk between genes, functions and pathways', Journal of immunological methods, 440, pp. 19–26.

Pre-Analytix (2020) PAXgene® Blood RNA Kit Instructions for Use (Handbook). Qiagen. Available at: https://www.qiagen.com/us/resources/download.aspx?id=9a4cdd13-8b72-4b33-877c-ba60a83b87be&lang=en.

Prechl, J., Papp, K. and Erdei, A. (2010) 'Antigen microarrays: descriptive chemistry or functional immunomics?', Trends in immunology, 31(4), pp. 133–137.

Prinz, F., Jonas, K., Balihodzic, A., Klec, C., Reicher, A., Barth, D.A., Riedl, J., Gerger, A., Kiesslich, T., Mayr, C., Rinner, B., Kargl, J. and Pichler, M. (2022) 'MicroRNA mimics can distort physiological microRNA effects on immune checkpoints by triggering an antiviral interferon response', RNA biology, 19(1), pp. 1305–1315.

Pua, H.H. and Ansel, K.M. (2015) 'MicroRNA regulation of allergic inflammation and asthma', Current opinion in immunology, 36, pp. 101–108.

Puddicombe, S.M., Torres-Lozano, C., Richter, A., Bucchieri, F., Lordan, J.L., Howarth, P.H., Vrugt, B., Albers, R., Djukanovic, R., Holgate, S.T., Wilson, S.J. and Davies, D.E. (2003) 'Increased expression of p21(waf) cyclin-dependent kinase inhibitor in asthmatic bronchial epithelium', American journal of respiratory cell and molecular biology, 28(1), pp. 61–68.

Qin, Z., Chen, Y., Wang, Y., Xu, Y., Liu, T., Mu, Q. and Huang, C. (2024) 'Immunometabolism in the pathogenesis of asthma', Immunology, 171(1), pp. 1–17.

Qi, P., Cheng, S.-Q., Wang, H., Li, N., Chen, Y.-F. and Gao, C.-F. (2011) 'Serum microRNAs as biomarkers for hepatocellular carcinoma in Chinese patients with chronic hepatitis B virus infection', PloS one, 6(12), p. e28486.

qPCR Reference Gene Selection Protocol (no date). Available at: https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/genomics/qpcr/qpcr-reference-gene-selection (Accessed: 12 September 2024).

Quail, M.A., Gu, Y., Swerdlow, H. and Mayho, M. (2012) 'Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation', Electrophoresis, 33(23), pp. 3521–3528.

Ramshankar, V. and Krishnamurthy, A. (2013) 'Lung cancer detection by screening - presenting circulating miRNAs as a promising next generation biomarker breakthrough', Asian Pacific journal of cancer prevention: APJCP, 14(4), pp. 2167–2172.

Raponi, M., Dossey, L., Jatkoe, T., Wu, X., Chen, G., Fan, H. and Beer, D.G. (2009) 'MicroRNA classifiers for predicting prognosis of squamous cell lung cancer', Cancer research, 69(14), pp. 5776–5783.

Razzak, R., Bédard, E.L.R., Kim, J.O., Gazala, S., Guo, L., Ghosh, S., Joy, A., Nijjar, T., Wong, E. and Roa, W.H. (2016) 'MicroRNA expression profiling of sputum for the detection of early and locally advanced non-small-cell lung cancer: a prospective case-control study', Current oncology , 23(2), pp. e86–94.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D. and Bader, G.D. (2019) 'Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap', Nature protocols, 14(2), pp. 482–517.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) 'The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans', Nature,

403(6772), pp. 901–906.

Resztak, J.A., Choe, J., Nirmalan, S., Wei, J., Bruinsma, J., Houpt, R., Alazizi, A., Mair-Meijers, H.E., Wen, X., Slatcher, R.B., Zilioli, S., Pique-Regi, R. and Luca, F. (2023) 'Analysis of transcriptional changes in the immune system associated with pubertal development in a longitudinal cohort of children with asthma', Nature communications, 14(1), p. 230.

Rick, E.-M., Woolnough, K.F., Seear, P.J., Fairs, A., Satchwell, J., Richardson, M., Monteiro, W.R., Craner, M., Bourne, M., Wardlaw, A.J. and Pashley, C.H. (2020) 'The airway fungal microbiome in asthma', Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology, 50(12), pp. 1325–1341.

Rijavec, M., Korošec, P., Žavbi, M., Kern, I. and Malovrh, M.M. (2014) 'Let-7a is differentially expressed in bronchial biopsies of patients with severe asthma', Scientific reports, 4, p. 6103.

Robinson, M.D. and Oshlack, A. (2010) 'A scaling normalization method for differential expression analysis of RNA-seq data', Genome biology, 11(3), p. R25.

Roche, N., Aggarwal, B., Boucot, I., Mittal, L., Martin, A. and Chrystyn, H. (2022) 'The impact of inhaler technique on clinical outcomes in adolescents and adults with asthma: A systematic review', Respiratory medicine, 202, p. 106949.

Roffel, M.P., Boudewijn, I.M., van Nijnatten, J.L.L., Faiz, A., Vermeulen, C.J., van Oosterhout, A.J., Affleck, K., Timens, W., Bracke, K.R., Maes, T., Heijink, I.H., Brandsma, C.-A. and van den Berge, M. (2021) 'Identification of asthma associated microRNAs in bronchial biopsies', The European respiratory journal: official journal of the European Society for Clinical Respiratory Physiology [Preprint]. Available at: https://doi.org/10.1183/13993003.01294-2021.

Ronchetti, D., Manzoni, M., Todoerti, K., Neri, A. and Agnelli, L. (2016) 'In Silico Characterization of miRNA and Long Non-Coding RNA Interplay in Multiple Myeloma', Genes, 7(12). Available at: https://doi.org/10.3390/genes7120107.

Rosenberg, M., Patterson, R., Mintzer, R., Cooper, B.J., Roberts, M. and Harris, K.E. (1977) 'Clinical and immunologic criteria for the diagnosis of allergic bronchopulmonary aspergillosis', Annals of internal medicine, 86(4), pp. 405–414.

Rothenberg, M.E. and Hogan, S.P. (2006) 'The eosinophil', Annual review of immunology, 24, pp. 147–174.

Rotival, M., Siddle, K.J., Silvert, M., Pothlichet, J., Quach, H. and Quintana-Murci, L. (2020) 'Population variation in miRNAs and isomiRs and their impact on human immunity to infection', Genome biology, 21(1), p. 187.

Rozowsky, J., Kitchen, R.R., Park, J.J., Galeev, T.R., Diao, J., Warrell, J., Thistlethwaite, W., Subramanian, S.L., Milosavljevic, A. and Gerstein, M. (2019) 'exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling', Cell systems, 8(4), pp. 352–357.e3.

Ruby, J.G., Jan, C.H. and Bartel, D.P. (2007) 'Intronic microRNA precursors that bypass Drosha processing', Nature, 448(7149), pp. 83–86.

Sada, K.-E., Amano, K., Uehara, R., Yamamura, M., Arimura, Y., Nakamura, Y., Makino, H. and Research Committee on Intractable Vasculitides, the Ministry of Health, Labour, Welfare of Japan (2014) 'A nationwide survey on the epidemiology and clinical features of eosinophilic granulomatosis with polyangiitis (Churg-Strauss) in Japan', Modern rheumatology / the Japan Rheumatism Association, 24(4), pp. 640–644.

Saeedavi, M., Goudarzi, M., Mehrzadi, S., Basir, Z., Hasanvand, A. and Hosseinzadeh, A. (2022) 'Sinapic acid ameliorates airway inflammation in murine ovalbumin-induced allergic asthma by reducing Th2 cytokine production', Life sciences, 307, p. 120858.

Samant, M., Ziemniak, J. and Paolini, J.F. (2023) 'First-in-Human Phase 1 Randomized Trial with the Anti-CD40 Monoclonal Antibody KPL-404: Safety, Tolerability, Receptor Occupancy, and Suppression of T-Cell-Dependent Antibody Response', The Journal of pharmacology and experimental therapeutics, 387(3), pp. 306–314.

Sarkis, E., Patel, S., Burns, K., Batarseh, H. and Mador, M.J. (2020) 'Anti-interleukin (IL)-5 as a steroid-sparing agent in chronic eosinophilic pneumonia', The Journal of asthma: official journal of the Association for the Care of Asthma, 57(1), pp. 82–86.

Sastre, I., Blanco, J., Mata, H. and García, F. (2012) 'A case of allergic bronchopulmonary aspergillosis treated with omalizumab', Journal of investigational allergology & clinical immunology: official organ of the International Association of Asthmology and Sociedad Latinoamericana de Alergia e Inmunologia, 22(2), pp. 145–147.

Sevignani, C., Calin, G.A., Siracusa, L.D. and Croce, C.M. (2006) 'Mammalian microRNAs: a small world for fine-tuning gene expression', Mammalian genome: official journal of the International Mammalian Genome Society, 17(3), pp. 189–202.

Shah, A. (2007) 'Allergic bronchopulmonary aspergillosis: an Indian perspective', Current opinion in pulmonary medicine, 13(1), pp. 72–80.

Shah, A., Bhagat, R. and Panchal, N. (1993) 'Allergic bronchopulmonary aspergillosis with clubbing and cavitation', Indian pediatrics, 30(2), pp. 248–251.

Shang, L., Du, Y., Zhao, Y., Zhang, Y. and Liu, C. (2023) 'The interaction of OTUB1 and TRAF3 mediates NLRP3 inflammasome activity to regulate TGF-β1-induced BEAS-2B cell injury', Applied biochemistry and biotechnology, 195(11), pp. 7060–7074.

Shell, S., Park, S.-M., Radjabi, A.R., Schickel, R., Kistner, E.O., Jewell, D.A., Feig, C., Lengyel, E. and Peter, M.E. (2007) 'Let-7 expression defines two differentiation stages of cancer', Proceedings of the National Academy of Sciences of the United States of America, 104(27), pp. 11400–11405.

Shi, J., Chen, M., Ouyang, L., Wang, Q., Guo, Y., Huang, L. and Jiang, S. (2020) 'miR-142-5p and miR-130a-3p regulate pulmonary macrophage polarization and asthma airway remodeling', Immunology and cell biology, 98(9), pp. 715–725.

Shirshova, A.N., Shamovskaya, D.A., Boyarskikh, U.A., Kushlinskii, N.E. and Filipenko, M.L. (2018) 'One-phase phenol-free method for microRNA isolation from blood plasma', MethodsX, 5, pp. 737–743.

Silver, N., Best, S., Jiang, J. and Thein, S.L. (2006) 'Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR', BMC molecular biology, 7(1), p. 33.

Simpson, J.L., Scott, R.J., Boyle, M.J. and Gibson, P.G. (2005) 'Differential proteolytic enzyme activity in eosinophilic and neutrophilic asthma', American journal of respiratory and critical care medicine, 172(5), pp. 559–565.

Sinyor, B. and Perez, L.C. (2023) Pathophysiology Of Asthma. StatPearls Publishing.

Skeparnias, I., Anastasakis, D., Grafanaki, K., Kyriakopoulos, G., Alexopoulos, P., Dougenis, D., Scorilas, A., Kontos, C.K. and Stathopoulos, C. (2020) 'Contribution of miRNAs, tRNAs and tRFs to Aberrant Signaling

and Translation Deregulation in Lung Cancer', Cancers, 12(10). Available at: https://doi.org/10.3390/cancers12103056.

Smulders, T., Van Der Schee, M.P., Maitland-Van Der Zee, A.H., Dikkers, F.G. and Van Drunen, C.M. (2024) 'Influence of the gut and airway microbiome on asthma development and disease', Pediatric allergy and immunology: official publication of the European Society of Pediatric Allergy and Immunology, 35(3), p. e14095.

Sohel, M.H. (2016) 'Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges', Achievements in the Life Sciences, 10(2), pp. 175–186.

Sokol, K.C., Sharma, G., Lin, Y.-L. and Goldblum, R.M. (2015) 'Choosing wisely: adherence by physicians to recommended use of spirometry in the diagnosis and management of adult asthma', The American journal of medicine, 128(5), pp. 502–508.

Solberg, O.D., Ostrin, E.J., Love, M.I., Peng, J.C., Bhakta, N.R., Hou, L., Nguyen, C., Solon, M., Nguyen, C., Barczak, A.J., Zlock, L.T., Blagev, D.P., Finkbeiner, W.E., Ansel, K.M., Arron, J.R., Erle, D.J. and Woodruff, P.G. (2012) 'Airway epithelial miRNA expression is altered in asthma', American journal of respiratory and critical care medicine, 186(10), pp. 965–974.

Soliman, M.M. (2018) COMPARISON OF RNA PRESERVATION TECHNIQUES. Edited by D.T. Harris. University of Arizona, Tucson, USA.

Specjalski, K. and Niedoszytko, M. (2020) 'MicroRNAs: future biomarkers and targets of therapy in asthma?', Current opinion in pulmonary medicine, 26(3), pp. 285–292.

Steele, A.M., Meuret, A.E., Millard, M.W. and Ritz, T. (2012) 'Discrepancies between lung function and asthma control: asthma perception and association with demographics and anxiety', Allergy and asthma proceedings: the official journal of regional and state allergy societies, 33(6), pp. 500–507.

Sterne-Weiler, T., Martinez-Nunez, R.T., Howard, J.M., Cvitovik, I., Katzman, S., Tariq, M.A., Pourmand, N. and Sanford, J.R. (2013) 'Frac-seq reveals isoform-specific recruitment to polyribosomes', Genome research, 23(10), pp. 1615–1623.

Sunadome, H., Matsumoto, H., Izuhara, Y., Nagasaki, T., Kanemitsu, Y., Ishiyama, Y., Morimoto, C., Oguma, T., Ito, I., Murase, K., Muro, S., Kawaguchi, T., Tabara, Y., Chin, K., Matsuda, F. and Hirai, T. (2020) 'Correlation between eosinophil count, its genetic background and body mass index: The Nagahama Study', Allergology international: official journal of the Japanese Society of Allergology, 69(1), pp. 46–52.

Sun, J., Nishiyama, T., Shimizu, K. and Kadota, K. (2013) 'TCC: an R package for comparing tag count data with robust normalization strategies', BMC bioinformatics, 14, p. 219.

Suyundikov, A., Stevens, J.R., Corcoran, C., Herrick, J., Wolff, R.K. and Slattery, M.L. (2015) 'Incorporation of subject-level covariates in quantile normalization of miRNA data', BMC genomics, 16, p. 1045.

Suzuki, Y. and Suda, T. (2018) 'Long-term management and persistent impairment of pulmonary function in chronic eosinophilic pneumonia: A review of the previous literature', Allergology international: official journal of the Japanese Society of Allergology, 67(3), pp. 334–340.

Takayama, G., Arima, K., Kanaji, T., Toda, S., Tanaka, H., Shoji, S., McKenzie, A.N.J., Nagai, H., Hotokebuchi, T. and Izuhara, K. (2006) 'Periostin: a novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals', The journal of allergy and clinical immunology, 118(1), pp. 98–104.

Tam, S., Tsao, M.-S. and McPherson, J.D. (2015) 'Optimization of miRNA-seq data preprocessing', Briefings

in bioinformatics, 16(6), pp. 950–963.

Tang, M., Sun, J., Shimizu, K. and Kadota, K. (2015) 'Evaluation of methods for differential expression analysis on multi-group RNA-seq count data', BMC bioinformatics, 16, p. 361.

Tay, J.W., James, I., Hughes, Q.W., Tiao, J.Y. and Baker, R.I. (2017) 'Identification of reference miRNAs in plasma useful for the study of oestrogen-responsive miRNAs associated with acquired Protein S deficiency in pregnancy', BMC research notes, 10(1), p. 312.

Thomeer, M.J., Costabe, U., Rizzato, G., Poletti, V. and Demedts, M. (2001) 'Comparison of registries of interstitial lung diseases in three European countries', The European respiratory journal. Supplement, 32, p. 114s–118s.

Thwaites, R.S., Ito, K., Chingono, J.M.S., Coates, M., Jarvis, H.C., Tunstall, T., Anderson-Dring, L., Cass, L., Rapeport, G., Openshaw, P.J., Nadel, S. and Hansel, T.T. (2017) 'Nasosorption as a Minimally Invasive Sampling Procedure: Mucosal Viral Load and Inflammation in Primary RSV Bronchiolitis', The Journal of infectious diseases, 215(8), pp. 1240–1244.

Thwaites, R.S., Jarvis, H.C., Singh, N., Jha, A., Pritchard, A., Fan, H., Tunstall, T., Nanan, J., Nadel, S., Kon, O.M., Openshaw, P.J. and Hansel, T.T. (2018) 'Absorption of Nasal and Bronchial Fluids: Precision Sampling of the Human Respiratory Mucosa and Laboratory Processing of Samples', Journal of visualized experiments: JoVE [Preprint], (131). Available at: https://doi.org/10.3791/56413.

Tian, M., Ji, Y., Wang, T., Zhang, W., Zhou, Y. and Cui, Y. (2018) 'Changes in circulating microRNA-126 levels are associated with immune imbalance in children with acute asthma', International journal of immunopathology and pharmacology, 32, p. 2058738418779243.

Tillie-Leblond, I., Germaud, P., Leroyer, C., Tétu, L., Girard, F., Devouassoux, G., Grignet, J.-P., Prudhomme, A., Dusser, D. and Wallaert, B. (2011) 'Allergic bronchopulmonary aspergillosis and omalizumab', Allergy, 66(9), pp. 1254–1256.

Tiwari, A., Li, J., Kho, A.T., Sun, M., Lu, Q., Weiss, S.T., Tantisira, K.G. and McGeachie, M.J. (2021) 'COPD-associated miR-145-5p is downregulated in early-decline FEV1 trajectories in childhood asthma', The journal of allergy and clinical immunology, 147(6), pp. 2181–2190.

Tonga, K.O., King, G.G., Farah, C.S., Thamrin, C., Tang, F.S., Santos, J., Sharma, P., Chapman, D.G. and Oliver, B.G. (2018) 'Steroid insensitive fixed airflow obstruction is not related to airway inflammation in older non-smokers with asthma', Respiratory research, 19(1), p. 176.

Turchinovich, A. and Burwinkel, B. (2012) 'Distinct AGO1 and AGO2 associated miRNA profiles in human cells and blood plasma', RNA biology, 9(8), pp. 1066–1075.

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. and Johnson, R. (2018) 'Towards a complete map of the human long non-coding RNA transcriptome', Nature reviews. Genetics, 19(9), pp. 535–548.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', Genome biology, 3(7), p. RESEARCH0034.

Van Goethem, A., Yigit, N., Everaert, C., Moreno-Smith, M., Mus, L.M., Barbieri, E., Speleman, F., Mestdagh, P., Shohet, J., Van Maerken, T. and Vandesompele, J. (2016) 'Depletion of tRNA-halves enables effective small RNA sequencing of low-input murine serum samples', Scientific reports, 6, p. 37876.

Vernon, M.K., Wiklund, I., Bell, J.A., Dale, P. and Chapman, K.R. (2012) 'What do we know about asthma

triggers? a review of the literature', The Journal of asthma: official journal of the Association for the Care of Asthma, 49(10), pp. 991–998.

Veryaskina, Y.A., Titov, S.E., Ivanov, M.K., Ruzankin, P.S., Tarasenko, A.S., Shevchenko, S.P., Kovynev, I.B., Stupak, E.V., Pospelova, T.I. and Zhimulev, I.F. (2022) 'Selection of reference genes for quantitative analysis of microRNA expression in three different types of cancer', PloS one, 17(2), p. e0254304.

Vigneron, N., Meryet-Figuière, M., Guttin, A., Issartel, J.-P., Lambert, B., Briand, M., Louis, M.-H., Vernon, M., Lebailly, P., Lecluse, Y., Joly, F., Krieger, S., Lheureux, S., Clarisse, B., Leconte, A., Gauduchon, P., Poulain, L. and Denoyelle, C. (2016) 'Towards a new standardized method for circulating miRNAs profiling in clinical studies: Interest of the exogenous normalization to improve miRNA signature accuracy', Molecular oncology, 10(7), pp. 981–992.

Walczak-Drzewiecka, A., Wyczólkowska, J. and Dastych, J. (2003) 'Environmentally relevant metal and transition metal ions enhance Fc epsilon RI-mediated mast cell activation', Environmental health perspectives, 111(5), pp. 708–713.

Wang, J.-S., Lin, H.-Y., Cheng, M.-L. and Wong, M.-K. (2007) 'Chronic intermittent hypoxia modulates eosinophil- and neutrophil-platelet aggregation and inflammatory cytokine secretion caused by strenuous exercise in men', Journal of applied physiology, 103(1), pp. 305–314.

Wang, K., Yuan, Y., Cho, J.-H., McClarty, S., Baxter, D. and Galas, D.J. (2012) 'Comparing the MicroRNA spectrum between serum and plasma', PloS one, 7(7), p. e41561.

Wang, L., Yang, X., Li, W., Song, X. and Kang, S. (2019) 'MiR-202-5p/MATN2 are associated with regulatory T-cells differentiation and function in allergic rhinitis', Human cell, 32(4), pp. 411–417.

Wang, Z. (2011) 'The guideline of the design and validation of MiRNA mimics', Methods in molecular biology (Clifton, N.J.), 676, pp. 211–223.

Wang, Z., Pianosi, P.T., Keogh, K.A., Zaiem, F., Alsawas, M., Alahdab, F., Almasri, J., Mohammed, K., Larrea-Mantilla, L., Farah, W., Daraz, L., Barrionuevo, P., Morrow, A.S., Prokop, L.J. and Murad, M.H. (2018) 'The Diagnostic Accuracy of Fractional Exhaled Nitric Oxide Testing in Asthma: A Systematic Review and Meta-analyses', Mayo Clinic proceedings. Mayo Clinic, 93(2), pp. 191–198.

Ward, C., Pais, M., Bish, R., Reid, D., Feltis, B., Johns, D. and Walters, E.H. (2002) 'Airway inflammation, basement membrane thickening and bronchial hyperresponsiveness in asthma', Thorax, 57(4), pp. 309–316.

Wardzyńska, A., Pawełczyk, M., Rywaniak, J., Makowska, J., Jamroz-Brzeska, J. and Kowalski, M.L. (2021) 'Circulating miRNA expression in asthmatics is age-related and associated with clinical asthma parameters, respiratory function and systemic inflammation', Respiratory research, 22(1), p. 177.

Wenzel, S.E. (2012) 'Asthma phenotypes: the evolution from clinical to molecular approaches', Nature medicine, 18(5), pp. 716–725.

Wenzel, S., Ford, L., Pearlman, D., Spector, S., Sher, L., Skobieranda, F., Wang, L., Kirkesseli, S., Rocklin, R., Bock, B., Hamilton, J., Ming, J.E., Radin, A., Stahl, N., Yancopoulos, G.D., Graham, N. and Pirozzi, G. (2013) 'Dupilumab in persistent asthma with elevated eosinophil levels', The New England journal of medicine, 368(26), pp. 2455–2466.

Williams, A.E., Larner-Svensson, H., Perry, M.M., Campbell, G.A., Herrick, S.E., Adcock, I.M., Erjefalt, J.S., Chung, K.F. and Lindsay, M.A. (2009) 'MicroRNA expression profiling in mild asthmatic human airways and effect of corticosteroid therapy', PloS one, 4(6), p. e5889.

Williams, Z., Ben-Dov, I.Z., Elias, R., Mihailovic, A., Brown, M., Rosenwaks, Z. and Tuschl, T. (2013) 'Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations', Proceedings of the National Academy of Sciences of the United States of America, 110(11), pp. 4255–4260.

Willis-Owen, S.A.G., Cookson, W.O.C. and Moffatt, M.F. (2018) 'The genetics and genomics of asthma', Annual review of genomics and human genetics, 19(1), pp. 223–246.

Wong LLK, Lee K, Russell I, Chen C. (2010) Endogenous Controls for Real-Time Quantitation of miRNA UsingTaqMan MicroRNA Assays. Applied Biosystems Application Note. Available at: https://assets.thermofisher.com/TFS-Assets/LSG/Application-Notes/cms_044972.pdf.

Wong, M., Forno, E. and Celedón, J.C. (2022) 'Asthma interactions between obesity and other risk factors', Annals of allergy, asthma & immunology: official publication of the American College of Allergy, Asthma, & Immunology, 129(3), pp. 301–306.

Wu, W.-S., Brown, J.S., Chen, T.-T., Chu, Y.-H., Huang, W.-C., Tu, S. and Lee, H.-C. (2019) 'piRTarBase: a database of piRNA targeting sites and their roles in gene regulation', Nucleic acids research, 47(D1), pp. D181–D187.

Wynn, T.A. (2015) 'Type 2 cytokines: mechanisms and therapeutic strategies', Nature reviews. Immunology, 15(5), pp. 271–282.

Xie, F., Wang, J. and Zhang, B. (2023) 'RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes', Functional & integrative genomics, 23(2), p. 125.

Xue, F., You, W., Liu, Y. and Zhong, L. (2023) 'Identification of the key miRNA-mRNA regulatory network in lung adenocarcinoma', Journal of thoracic disease, 15(4), pp. 2037–2050.

Xu, Y., Wang, Z., Li, W., Zhu, Q. and Liang, Z. (2022) 'Clinical characteristics of patients with chronic eosinophilic pneumonia in a Chinese tertiary-care hospital: A 6-year retrospective study', The clinical respiratory journal, 16(1), pp. 35–42.

Yaghoubi, M., Adibi, A., Safari, A., FitzGerald, J.M. and Sadatsafavi, M. (2019) 'The Projected Economic and Health Burden of Uncontrolled Asthma in the United States', American journal of respiratory and critical care medicine, 200(9), pp. 1102–1112.

Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D. and Lai, E.C. (2010) 'Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis', Proceedings of the National Academy of Sciences of the United States of America, 107(34), pp. 15163–15168.

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q. and Tomari, Y. (2010) 'ATP-dependent human RISC assembly pathways', Nature structural & molecular biology, 17(1), pp. 17–23.

Yssel, Abbal, Pène and Bousquet (1998) 'The role of IgE in asthma', Clinical & Experimental Allergy, 28, pp. 104–109.

Zhang, H.-G., Liu, L., Song, Z.-P. and Zhang, D.-Y. (2020) 'Bioinformatics Analysis of the Regulatory lncRNA-miRNA-mRNA Network in Patients with Neuropathic Pain', Research Square. Research Square. Available at: https://doi.org/10.21203/rs.3.rs-54799/v1.

Zhang, K., Zhang, X., Cai, Z., Zhou, J., Cao, R., Zhao, Y., Chen, Z., Wang, D., Ruan, W., Zhao, Q., Liu, G., Xue, Y., Qin, Y., Zhou, B., Wu, L., Nilsen, T., Zhou, Y. and Fu, X.-D. (2018) 'A novel class of microRNA-recognition

elements that function only within open reading frames', Nature structural & molecular biology [Preprint]. Available at: https://doi.org/10.1038/s41594-018-0136-3.

Zhang, S., Laryea, Z., Panganiban, R.P.L., Lambert, K.A. and Ishmael, F. (2018) 'Plasma microRNA profiles identify distinct clinical phenotypes in human asthmatics'. Available at: https://doi.org/10.20517/jtgg.2018.22.

Zhang, X.-Y., Simpson, J.L., Powell, H., Yang, I.A., Upham, J.W., Reynolds, P.N., Hodge, S., James, A.L., Jenkins, C., Peters, M.J., Lin, J.-T. and Gibson, P.G. (2014) 'Full blood count parameters for the detection of asthma inflammatory phenotypes', Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology, 44(9), pp. 1137–1145.

Zhang, Y., Jing, Y., Qiao, J., Luan, B., Wang, X., Wang, L. and Song, Z. (2017) 'Activation of the mTOR signaling pathway is required for asthma onset', Scientific reports, 7(1), p. 4532.

Zhang, Y., Willis-Owen, S.A.G., Spiegel, S., Lloyd, C.M., Moffatt, M.F. and Cookson, W.O.C.M. (2019) 'The ORMDL3 Asthma Gene Regulates ICAM1 and Has Multiple Effects on Cellular Inflammation', American journal of respiratory and critical care medicine, 199(4), pp. 478–488.

Zhang, Z., Li, C., Zhang, J., Chen, F., Gong, Y., Li, Y., Su, Y., Wei, Y. and Zhao, Y. (2020) 'Selection of the Reference Gene for Expression Normalization in Papaver somniferum L. under Abiotic Stress and Hormone Treatment', Genes, 11(2). Available at: https://doi.org/10.3390/genes11020124.

Zhao, S., Gordon, W., Du, S., Zhang, C., He, W., Xi, L., Mathur, S., Agostino, M., Paradis, T., von Schack, D., Vincent, M. and Zhang, B. (2017) 'QuickMIRSeq: a pipeline for quick and accurate quantification of both known miRNAs and isomiRs by jointly processing multiple samples from microRNA sequencing', BMC bioinformatics, 18(1), p. 180.

Zheng, D., Haddadin, S., Wang, Y., Gu, L.-Q., Perry, M.C., Freter, C.E. and Wang, M.X. (2011) 'Plasma microRNAs as novel biomarkers for early detection of lung cancer', International journal of clinical and experimental pathology, 4(6), pp. 575–586.

Zhu, W., Zhou, K., Zha, Y., Chen, D., He, J., Ma, H., Liu, X., Le, H. and Zhang, Y. (2016) 'Diagnostic Value of Serum miR-182, miR-183, miR-210, and miR-126 Levels in Patients with Early-Stage Non-Small Cell Lung Cancer', PloS one, 11(4), p. e0153046.

Zhu, Z., Hasegawa, K., Camargo, C.A., Jr and Liang, L. (2021) 'Investigating asthma heterogeneity through shared and distinct genetics: Insights from genome-wide cross-trait analysis', The Journal of allergy and clinical immunology, 147(3), pp. 796–807.

Zoorob, R.J. and Cender, D. (1998) 'A different look at corticosteroids', American family physician, 58(2), pp. 443–450.

9. Appendices

Appendix Figure 1: Permission for Figure 1.3

From: **Sam Chesworth** samchesworth@dovepress.co.uk
Subject: ACI242222 - Eosinophilic Lung Diseases/Crowe - Permission Granted
Date: 30 August 2024 at 13:05
To: Headley-Morris, Lauren A l.headley17@imperial.ac.uk

SC

Dear Lauren Headley,

Many thanks for your attachment, we have reviewed the document and can confirm that this has now been approved and permission has been granted.

Permission is granted on a one-time basis and any other request for re-use in the future will require separate permission.

We look forward to working with you again shortly.

Thank you.

Best regards,
Sam

Sam Chesworth
Sales Administrator
Dove Medical Press Limited

Beechfield House,
Winterton Way,
Macclesfield SK11 0LP
United Kingdom
www.dovepress.com

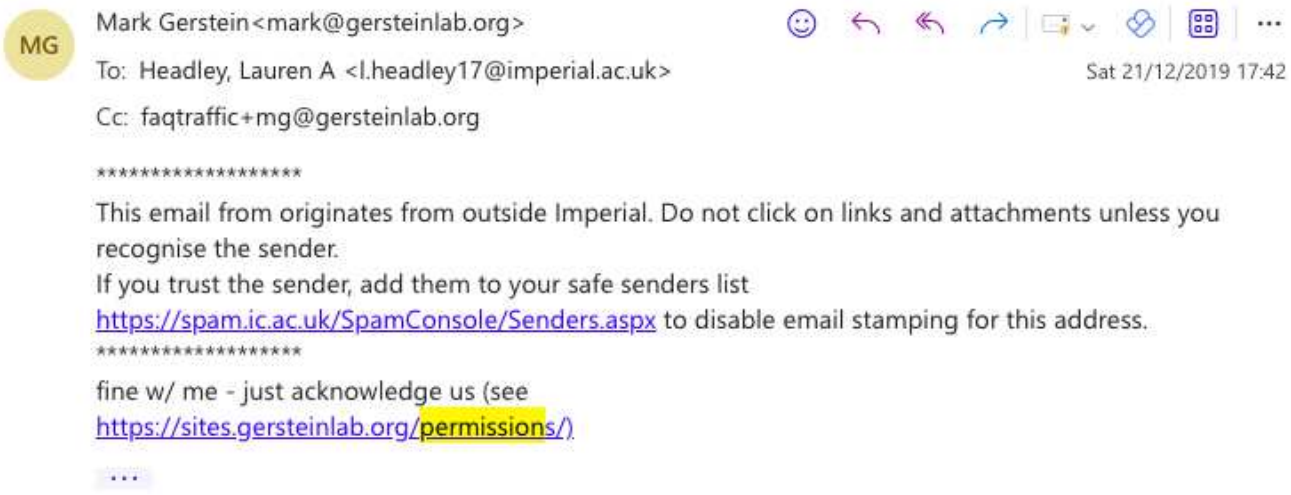
Company registration number: 4967656
Registered in England and Wales. Registered address 5 Howick Place, London, SW1P 1WG.
VAT Number: GB 365 4626 36

Dove Medical Press is part of Taylor & Francis Group, the Academic Publishing Division of Informa PLC

Direct: +44 (0)20 8052 2857
T: +44 (0)1625 509134
F: +44 (0)1625 617933
Email: samchesworth@dovepress.co.uk

Dove Journals Impact Factors:
6.922 J Inflamm Res; 6.4 Int J Nanomedicine; 5.828 J Hepatocell Carcinoma; 5.346 Nat Sci Sleep; 4.79 Clin Epidemiol; 4.458 Clin Interv Aging; 4.258 J Asthma Allergy; 4.162 Drug Des Devel Ther; 4.147 Onco Targets Ther; 4.003 Infect Drug Resist; 3.989 Cancer Manag Res; 3.912 Pharmgenomics Pers Med; 3.355 Int J Chron Obstruct Pulmon Dis; 3.2 Ris Manag Healthc Policy; 3.168 Diabetes Metab Syndr Obes; 3.133 J Pain Res; 2.945 Psychol Res Behav Manag; 2.773 Int J Womens Health; 2.711 Patient Prefer Adherence; 2.57 Neuropsychiatr Dis Treat; 2.489 Clin Cosmet Investig Dermatol; 2.466 Int J Gen Med; 2.423 Ther Clin Risk Manag; 2.404 J Multidiscip Healthc

Appendix Figure 2: Permission for Figure 2.19.



MG

Mark Gerstein<mark@gersteinlab.org>

To: Headley, Lauren A <l.headley17@imperial.ac.uk>

Sat 21/12/2019 17:42

Cc: faqtraffic+mg@gersteinlab.org

****

This email from originates from outside Imperial. Do not click on links and attachments unless you recognise the sender.
If you trust the sender, add them to your safe senders list https://spam.ic.ac.uk/SpamConsole/Senders.aspx to disable email stamping for this address.
****

fine w/ me - just acknowledge us (see https://sites.gersteinlab.org/permissions/)

...

Appendix Figure 3: Celtic Fire (Asthma) Recruitment Criteria

The Celtic Fire Study

Ethical approval received from: London - Stanmore Research Ethics Committee (london-stanmore@nhs.net

Study title: Heterogeneity of Asthma: a pilot study to investigate its microbiome and how it relates to asthma severity

REC reference: 14/LO/2063

IRAS project ID: 163724

Exclusion criteria

Asthmatic subjects must be non-smokers or ex-smokers with < 5 pack-years smoking

No morbid obesity (BMI>35)

No diagnosis of RA, ABPA or Churg-Strauss syndrome

No confounding drugs (beta-blockers, ACE inhibitors, anti-asthma immune modulators) other than steroids

No antibiotics within 4 weeks

No acute exacerbations of asthma within past 4 weeks

No history of an upper or lower respiratory infection (including common cold) within 4 weeks of baseline assessments

No confounding occupations (such as baking)

No significant vocal cord disorder (VCD)

Selection Criteria

Subjects will be 18 or more years of age

Subject groups with approximate target numbers will include

1. Healthy subjects: non-smoker (20 subjects)
2. Healthy subjects: current smokers (20 subjects)
3. Asthmatic patients taking short-acting beta agonists only (BTS Step 1 – 30 subjects)
4. Asthmatic patients on moderate dose of inhaled corticosteroid (ICS) (up to 800 µg/day of beclomethasone propionate (BDP equivalent)± long-acting β-agonist LABA (BTS Step 2/3 – 30 subjects)
5. Asthmatic patients on high dose ICS (ICS dose >=1600 µg/day) + LABA ± other controllers (theophyllines, LTRA, LAMA) (BTS Step 4 – 30 subjects)
6. Asthmatic patients on high dose ICS (ICS dose >=1600 µg/day) + LABA ± other controllers + oral prednisolone ± anti-IgE treatment (BTS Step 5 – 30 subjects)

Subjects will be admitted to the study irrespective of atopic status.

Appendix Figure 4: BEAM (ELD) Recruitment Criteria

Eosinophilic Lung Disease (BEAM) Cohort

Ethical approval received from: London - Brent Research ethics Committee brent.rec@hra.nhs.uk
Study Title: Bronchoscopic evaluation of the Eosinophilic Airway Microbiome (BEAM)
REC reference: 17/LO/0288
IRAS project ID: 215130

Inclusion criteria

General inclusion criteria

- Age 18-65 years
- Non-smoker or ex-smoker for at least the past 12 months with < 10 pack-years smoking history

Subjects with eosinophilic lung diseases

- Diagnosis of ELD confirmed by specialist respiratory assessment including, for CEP, characteristic appearances on computed tomography (CT) imaging
- FEV1 ≥ 60 % predicted

Healthy control subjects

- Healthy individuals, free of significant disease
- No history of asthma/ atopy/ rhinitis/ ELD
- Not taking inhaled therapies/ steroid treatment/ other immunosuppressive therapies
- Pre-bronchodilator FEV1 ≥ 80 % predicted

Exclusion criteria

The below exclusion criteria were applied:

- Unable to give informed consent
- History of an upper or lower respiratory infection (including common cold) within 4 weeks of baseline assessments
- Acute exacerbations of chronic lung disease (defined as a clinical episode requiring treatment with systemic corticosteroids, or requiring an increase in the dosage of systemic corticosteroids if the patient is on long term corticosteroid treatment) within the past 4 weeks
- Long-term antibiotics or any course(s) of antibiotics within the past 4 weeks
- Non-adherent to therapy
- Significant Vocal Cord Disorder (VCD)
- Diagnosis of Rheumatoid Arthritis (RA)
- Confounding drugs (beta-blockers, ACE inhibitors, anti-asthma immune modulators) other than steroids, omalizumab, azathioprine and methotrexate
- Body Mass Index (BMI) > 35 kg/m2
- Confounding occupations (such as baking)
- Previous bronchial thermoplasty
- Female subjects who are pregnant/ breastfeeding/ less than 6 weeks post-partum at the time of enrolment, or who fulfil any of these criteria at any point prior to the final investigation (bronchoscopy).