UNIVERSITY OF Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Dr Hannah Burke (2021) "The targets and functions of lung-derived extracellular vesicle microRNA in Chronic Obstructive Pulmonary Disease", University of Southampton, Clinical and Experimental Sciences, Faculty of Medicine, PhD Thesis, p1-244.

University of Southampton

FACULTY OF MEDICINE

CLINICAL AND EXPERIMENTAL SCIENCES

THE TARGETS AND FUNCTIONS OF LUNG-DERIVED EXTRACELLULAR VESICLE MICRORNA IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

By

Dr Hannah Burke

BMBS, MRCP (UK)

ORCID ID: 0000-0003-3553-4590

Thesis for the degree of DOCTOR OF PHILOSOPHY

March 2021

University of Southampton

Abstract

Faculty of MEDICINE

CLINICAL AND EXPERIMENTAL SCIENCES

Doctor of Philosophy

THE TARGETS AND FUNCTIONS OF LUNG-DERIVED EXTRACELLULAR VESICLE MICRORNA IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

By

Dr Hannah Burke

Chronic obstructive pulmonary disease (COPD) is a heterogeneous disease characterised by chronic airway inflammation, which is dependent upon a complex network of intercellular communication between the damaged airway epithelium and the immune system. Extracellular vesicles (EVs) are novel major signalling mediators between cells, and shuttle cargo such as micro RNA (miRNA) in health and disease. This thesis examines the role of EV miRNA regulation of inflammatory signalling pathways in COPD.

EVs were isolated from bronchoalveolar lavage fluid (BALF), and epithelial brushings were taken from the same patients with COPD and healthy controls. RNA sequencing was used to quantify EV miRNA expression and gene expression in epithelial brushings. Negative binomial models were used to identify differentially expressed miRNA and genes in patients with COPD, and a combination of bioinformatic approaches were used to identify the biologically significant interactions between these miRNA and genes. Further analysis assessed the predictive ability of the differentially expressed miRNA in discriminating between health and disease and their relationship with inflammatory endotypes in COPD.

Differential miRNA expression was observed in BALF EVs from patients with COPD compared to healthy ex-smokers. Specifically, five miRNA were found to be up-regulated in COPD (miR-2110, miR-182-5p, miR-223-3p, miR-200b-5p and miR-625-3p) and three miRNA were down-regulated in COPD (miR-338-3p, miR-138-5p and miR-204-5p). *In silico* analysis demonstrated these differentially expressed miRNA targeted differentially expressed genes in epithelial brushings in the same patients. These miRNA-gene expression interactions may have a significant impact on key inflammatory pathways in COPD. In addition, lung-derived EV miRNA were a strong predictor of disease presence. Moreover, specific EV miRNA correlated with expression of inflammatory cells within the airways of COPD patients, which may provide novel insights into distinct inflammatory endotypic disease mechanisms.

..In summary, the findings from this thesis suggest that lung-derived EV miRNA may regulate gene expression in COPD leading to aberrant inflammatory signalling. Furthermore, EV miRNA may provide a novel diagnostic opportunity to detect early disease presence in COPD, and give further insights into the underlying endotypic mechanisms of disease, whereby EV miRNA could help future treatment stratification.

Table of Contents

TABLE OF CONTENTS ..... **I**

TABLE OF TABLES ..... **IX**

TABLE OF FIGURES ..... **XIII**

RESEARCH THESIS: DECLARATION OF AUTHORSHIP ..... **XVII**

ACKNOWLEDGEMENTS ..... **XIX**

DEFINITIONS AND ABBREVIATIONS ..... **XXI**

CHAPTER 1 **INTRODUCTION** **1**

1.1 CHRONIC OBSTRUCTIVE PULMONARY DISEASE 1

1.1.1 Overview of COPD *1*

1.1.2 Burden of COPD *1*

1.1.3 Risk factors for development and progression of disease *2*

1.1.3.1 . Genetic factors 2

1.1.3.2 . Ageing and COPD 3

1.1.3.3 . Lung growth and development 3

1.1.3.4 . Smoking & domestic fuel exposure 3

1.2 PATHOGENESIS AND PATHOPHYSIOLOGY OF COPD 4

1.2.1 Inflammatory cells *5*

1.2.1.1 . Neutrophils 5

1.2.1.2 . Macrophages 6

1.2.1.3 . T Lymphocytes 7

1.2.1.4 . Airway epithelial cells 7

1.2.2 Inflammatory mediators *8*

1.2.3 Oxidative stress *8*

1.2.4 Protease-anti-protease imbalance *9*

1.2.5 Airflow obstruction and hyperinflation *9*

1.2.6 Gas exchange abnormalities and pulmonary hypertension *10*

1.2.7 Systemic features *10*

1.3 DIAGNOSIS OF COPD 10

1.4 TREATMENT OF COPD 11

1.5 COPD ENDOTYPES 12

1.5.1 Neutrophilic COPD *13*

1.5.2 Eosinophilic COPD *13*

1.6 EXACERBATIONS IN COPD 14

Table of Contents

1.7 EXTRACELLULAR VESICLES 15
1.7.1 Definition 16
1.7.2 Subtypes and classification 16
1.7.2.1 .. Apoptotic bodies 16
1.7.2.2 .. Microvesicles 17
1.7.2.3 .. Exosomes 17
1.7.3 Extracellular vesicles in the lung 18
1.7.4 Extracellular vesicles in COPD 19
1.7.4.1 .. Extracellular vesicles as a diagnostic biomarker in COPD 22
1.7.4.2 .. Extracellular vesicles as a potential therapy in COPD 23
1.7.5 Extracellular vesicle isolation methods 23
1.7.5.1 .. Ultracentrifugation 24
1.7.5.2 .. Size exclusion chromatography 24
1.7.5.3 .. Filtration 24
1.7.5.4 .. Polymeric precipitation 24
1.7.5.5 .. Immunoaffinity isolation 25
1.7.5.6 .. Membrane affinity isolation 25
1.7.6 Characterisation of extracellular vesicles 25
1.8 MICRORNA 26
1.8.1 MicroRNA biogenesis and function 26
1.8.2 MicroRNA in COPD 27
1.9 HYPOTHESIS AND AIMS 31

CHAPTER 2 **METHODS 33**
2.1 ETHICS 33
2.2 MICA II STUDY DESIGN 34
2.3 STUDY POPULATION FOR BRONCHOALVEOLAR LAVAGE EXOSOMAL MIRNA ANALYSIS 34
2.4 SAMPLING PROTOCOLS 35
2.4.1 Spirometry 35
2.4.2 Transfer factor 35
2.4.3 Blood serum preparation 36
2.4.4 High resolution computer tomography 36
2.4.5 Fibreoptic bronchoscopy 36
2.4.5.1 .. Epithelial brushings for RNA isolation and sequencing 37
2.4.5.2 .. Bronchoalveolar lavage fluid analysis 37
2.4.5.3 .. Bronchoalveolar lavage cell counts 37
2.5 EXTRACELLULAR VESICLE ISOLATION 38
2.5.1 Ultrafiltration for sample purification and concentration 38
2.5.2 Size exclusion chromatography 38

2.5.2.1 . Size exclusion chromatography using PURE-EV™ columns 39
2.5.2.2 . Size exclusion chromatography using Exo-spin™ Midi-columns 39
2.5.3 Measurement of protein concentration of SEC-derived EV samples 40
2.5.3.1 . Protein concentration of size exclusion chromatography fractions 40
2.5.4 Summary of EV isolation from BALF 41
2.5.5 EV isolation from serum by filtration using ExoMir™ kit 42
2.6 EXTRACELLULAR VESICLE CHARACTERISATION 43
2.6.1 Quantification of BALF derived EVs using enzyme-linked immunosorbent assays 43
2.6.1.1 . CD63 enzyme-linked immunosorbent assay 43
2.6.1.2 . CD9 enzyme-linked immunosorbent assay 44
2.6.2 Transmission electron microscopy of BAL derived EVs to visualise characteristic size and shape 46
2.6.3 Characterisation of serum derived EVs 47
2.6.3.1 . SDS PAGE and Western blotting to determine presence of serum EV surface markers 47
2.6.3.2 . Transmission electron microscopy of serum derived EVs to visualise characteristic size and shape 48
2.7 RNA ISOLATION, CDNA SYNTHESIS AND REAL-TIME QPCR 50
2.7.1 BALF EV RNA isolation, quantification and quality control to assess SEC EV isolation suitability for downstream application (RNA seqencing) 50
2.7.2 BAL EV RNA quantification and quality control prior to library preparation & next generation sequencing – performed by Qiagen® Genomic Services 51
2.7.3 Serum EV RNA isolation, quantification and quality control 55
2.8 NEXT GENERATION MICRORNA SEQUENCING OF BRONCHOALVEOLAR LAVAGE EXTRACELLULAR VESICLE RNA 57
2.8.1 BALF EV RNA Library preparation – performed by Qiagen® Genomic Services 57
2.8.2 microRNA library pre-sequencing quality control and preparation 57
2.8.3 Sequencing run setup 58
2.9 MICRORNA SEQUENCING OUTPUT PROCESSING, QUALITY CONTROL, MAPPING AND ALIGNMENT 59
2.9.1 Trimming of adaptors and UMI correction – performed by Qiagen® Genomic Services 59
2.9.2 microRNA sequencing quality control 60
2.9.2.1 . Average read quality 60
2.9.2.2 . Per base sequence quality 61
2.9.2.3 . Sequence length distribution 62
2.9.2.4 . Additional FastQC quality metrics 62
2.9.2.5 . Summary of microRNA sequencing quality control 63
2.9.3 Mapping and aligning to reference genome – performed by Qiagen® Genomic Services 63
2.9.3.1 . Analysis of mapping and alignment to reference genome 64

2.10 MICRORNA SEQUENCING FILTERING, DATA ANALYSIS AND DIFFERENTIAL EXPRESSION ANALYSIS 68
2.10.1 Unsupervised filtering 68
2.10.1.1 Median log2-transformed CPM cut-off method 69
2.10.1.2 CPM>1 in a minimum of n samples, where n=size of the smallest group 70
2.10.1.3 CPM>10 in a minimum of n samples, where n=size of the smallest group 70
2.10.2 Exploratory data analysis for quality control 71
2.10.2.1 Basic quality control plots 71
2.10.2.2 Interquartile range versus median plot 75
2.10.2.3 Principle component analysis 77
2.10.2.4 Summary of exploratory data analysis 78
2.10.3 Normalisation methods 79
2.10.4 Negative binomial models 81
2.10.4.1 Biological coefficient of variation 81
2.10.5 Differential expression analysis between COPD subjects and healthy ex-smokers 82
2.11 REAL TIME-QPCR DATA ANALYSIS 84
2.11.1 RT-qPCR data quality control 84
2.11.2 Normalisation of RT-qPCR data 84
2.11.3 Differential expression analysis of RT-qPCR data 84
2.12 IDENTIFYING MIRNA TARGET GENES 86
2.12.1 miRNA target prediction in silico analysis using multiMiR 86
2.12.2 Next generation mRNA sequencing of epithelial brushings – performed by the Translational Science & Experimental Medicine team at AstraZeneca 87
2.12.2.1 mRNA sequencing data preparation - performed by the Bioinformatics team at AstraZeneca 88
2.12.3 Differential gene expression analysis of the epithelial brushing mRNA 89
2.12.4 Identify miRNA-mRNA putative interactions 91
2.13 MIRNA-MRNA INTERACTION ANALYSIS 92
2.13.1 Pairwise correlation analysis between miRNA and mRNA 92
2.13.2 Network analysis of miRNA-mRNA interaction network 93
2.13.3 miRNA-mRNA interaction network topology 95
2.13.3.1 Cluster analysis of networks 96
2.13.4 Gene Ontology enrichment analysis 96
2.13.4.1 The Biological Networks Gene Ontology tool (BiNGO) 96
2.13.4.2 Enrichment Map for gene-set enrichment visualisation and interpretation 97
2.14 STATISTICS 98

CHAPTER 3 **STUDY COHORT CHARACTERISTICS FOR EXTRACELLULAR VESICLE ISOLATION FROM BRONCHOALVEOLAR LAVAGE FLUID 99**
3.1 INTRODUCTION 99
3.2 CHARACTERISTICS OF THE SUBJECTS INCLUDED IN EV ISOLATION FROM BALF 99

3.2.1 Baseline & historic blood count 101
3.2.2 HRCT measurement 101
3.2.3 BALF count analysis 102
3.3 DISCUSSION 104
3.3.1 Subject characteristics 104
3.3.2 Blood eosinophilia 105
3.3.3 Airway inflammatory cell profile 106
3.3.4 Strength and Limitations 107
3.3.5 Summary 108
CHAPTER 4 **MICRORNA SEQUENCING OF BRONCHOALVEOLAR LAVAGE EXTRACELLULAR VESICLES AND VALIDATION OF THE RESULTS 109**
4.1 INTRODUCTION 109
4.1.1.1 . Characteristics of the subjects included in differential expression analysis of BALF EV miRNA 109
4.2 MAPPING AND ALIGNMENT RESULTS COMPARING PATIENTS WITH COPD AND HEALTHY EX-SMOKERS 110
4.2.1 Total number of reads 110
4.2.2 Proportion of miRNA and smallRNA mapped reads in COPD and healthy ex-smokers 111
4.3 DIFFERENTIAL EXPRESSION OF EV MIRNA BETWEEN COPD SUBJECTS AND HEALTHY EX-SMOKERS 113
4.4 VALIDATION OF DIFFERENTIALLY EXPRESSED MIRNA WITH RT-QPCR 114
4.4.1 Characteristics of subjects used for differential expressed EV miRNA validation by RT-qPCR, N=46 114
4.4.2 MiRNA chosen for validation by RT-qPCR 115
4.4.3 RT-qPCR data quality control 117
4.4.3.1 . Number of detected miRNAs 118
4.4.4 Differential miRNA expression analysis of RT-qPCR data 118
4.5 DISCUSSION 121
4.5.1 RNA mapping and alignment 121
4.5.2 Extracellular vesicle miRNA packaging 122
4.5.3 Differential expression of EV miRNA in COPD subjects compared with healthy ex-smokers 124
4.5.4 Strengths and limitations 128
4.5.5 Summary 129
CHAPTER 5 **IDENTIFICATION, VISUALISATION AND ANALYSIS OF MICRORNA-TARGET INTERACTION NETWORKS 131**
5.1 INTRODUCTION 131
5.2 IDENTIFYING MIRNA TARGET GENES 131

5.2.1 miRNA target identification in silico using multiMiR 131

5.2.2 Differential gene expression analysis of epithelial brushings in COPD and healthy ex-smokers 132

5.2.3 Identifying putative miRNA-mRNA interactions using multiMiR 135

5.2.3.1 .. Identification of the up-regulated EV miRNA gene targets in epithelial brushings 135

5.2.3.2 .. Identification of the down-regulated EV miRNA gene targets in epithelial brushings 137

5.2.4 miRNA-mRNA interaction analysis 140

5.2.4.1 .. Pairwise correlations between dysregulated miRNA and mRNA 140

5.2.4.2 .. Identifying the dominant miRNAs of the miRNA-mRNA interactions using miRMapper 143

5.3 miRNA-mRNA INTERACTION NETWORK TOPOLOGY 146

5.3.1 Cluster analysis of networks 149

5.4 GENE ONTOLOGY ENRICHMENT ANALYSIS 151

5.4.1 Gene ontology enrichment analysis for GO: biological process 151

5.4.2 Gene ontology enrichment analysis for GO: molecular function 153

5.5 DISCUSSION 156

5.5.1 Identifying differentially expressed EV miRNA target genes within epithelial brushings 156

5.5.1.1 .. Up-regulated genes in COPD epithelial brushings 157

5.5.1.2 .. Down-regulated genes in COPD epithelial brushings 158

5.5.2 Identifying putative miRNA-mRNA interactions 159

5.5.2.1 .. Identifying the most dominant miRNA-mRNA interactions 160

5.5.2.2 .. Visualising the miRNA-mRNA interactions in a network 162

5.5.3 GO enrichment analysis to identify key pathways regulated by miRNA-mRNA network 162

5.5.4 Strengths and limitations 163

5.5.5 Summary 164

CHAPTER 6 **DIAGNOSTIC USE OF EXTRACELLULAR VESICLE MIRNA AND THE RELATIONSHIP WITH COPD INFLAMMATORY ENDOTYPES 165**

6.1 INTRODUCTION 165

6.2 CHARACTERISTICS OF THE SUBJECT INCLUDED IN THE ANALYSIS EXPLORING THE PREDICTIVE ABILITY OF EV miRNA TO DIFFERENTIATE COPD FROM HEALTH AND ASSOCIATE WITH INFLAMMATORY ENDOTYPES 166

6.3 LUNG-DERIVED EV miRNA IN RELATION TO COPD CLINICAL CHARACTERISTICS 167

6.4 PREDICTIVE ABILITY OF miRNA TO DIFFERENTIATE BETWEEN COPD AND HEALTH 169

6.5 EV miRNA IN RELATION TO COPD INFLAMMATORY ENDOTYPES 171

6.5.1 Describing inflammatory endotypes in COPD subjects 171

6.5.2 Relationship between EV miRNA expression and levels of inflammatory cells in BAL 172

6.5.3 Using EV miRNA to predict inflammatory endotypes in COPD 173

6.6 DIFFERENCES IN SERUM EXTRACELLULAR VESICLE MIRNA EXPRESSION 176

6.6.1 Subject characteristics 176

6.6.2 Differential miRNA expression analysis from serum EV miRNA 178

6.7 DISCUSSION 179

6.7.1 Predictive ability of EV miRNA to differentiate between health and disease 179

6.7.2 Relationships between lung EVs and inflammatory endotypes in COPD 181

6.7.3 Strengths and limitations 183

6.7.4 Summary 184

CHAPTER 7 **SUMMARY DISCUSSION AND FUTURE WORK 185**

7.1 OVERVIEW 185

7.2 AIM 1: ISOLATE AND CHARACTERISE EVS FROM BRONCHOALVEOLAR LAVAGE FLUID OF COPD SUBJECTS AND HEALTHY EX-SMOKERS 187

7.3 AIMS 2&3: IDENTIFY DIFFERENTIALLY EXPRESSED LUNG-DERIVED EV MIRNA IN PATIENTS WITH COPD COMPARED WITH HEALTHY EX-SMOKERS AND THEIR BIOLOGICAL SIGNIFICANT TARGETS 189

7.4 AIM 4 EXPLORE THE DIAGNOSTIC USE OF THE LUNG-DERIVED EV MIRNA AND THEIR RELATIONSHIP WITH SPECIFIC COPD INFLAMMATORY ENDOTYPES 196

7.5 SUMMARY OF THE IMPLICATIONS OF STUDY FINDINGS 198

7.5.1 EV miRNA as a biomarker in COPD 198

7.5.2 Therapeutic potential of EVs in COPD 198

7.6 STRENGTHS & LIMITATIONS 199

7.7 FURTHER WORK 200

7.7.1 Ex vivo cell culture models 200

7.7.2 Explore EV miRNA-mRNA interactions in other cell types 200

7.7.3 Interrogate multi-omic readouts to identify downstream effects of EV miRNA 201

7.7.4 Study of EV miRNA signature in early COPD 201

7.7.5 Summary of future work 202

APPENDIX A .. SUPPLEMENTARY RESULTS FROM MICRORNA-MIRNA INTERACTION 203

A.1 DIFFERENTIALLY EXPRESSED GENES IN EPITHELIAL BRUSHINGS IN COPD 203

A.2 CORRELATION ANALYSIS OF MIRNA-MRNA INTERACTIONS 210

LIST OF REFERENCES 213

Table of Tables

Table 1.1 Classification of airflow limitation severity in COPD based on post-bronchodilator FEV1 11

Table 1.2 Summary of the EV studies in smokers, patients with COPD and murine models 20

Table 1.3 Expression pattern of miRNAs in COPD 28

Table 2.1 Inclusion and exclusion criteria for the MICA II study 34

Table 2.2 Summary of combined SEC fractions for PureEV™ columns 43

Table 2.3 Fifty-two QIAseq miRNA library QC spike-ins 53

Table 2.4 Number of reads after each step of the UMI correction process 63

Table 2.5 Types and proportions of smallRNA found in BALF EV miRNA (n=35) 67

Table 2.6 Summary of outliers identified from the boxplots and IQR/median plots for each unsupervised filtering method 76

Table 2.7 List of library sizes and normalisation factors for the "CPM filtered dataset" generated by TMM normalisation. 80

Table 3.1 Characteristics of subjects included in BALF EV RNA isolation for miRNA sequencing, n=35 100

Table 3.2 Correlation between HRCT measures of small airways disease and emphysema and physiology measures of disease 102

Table 3.3 Correlation between blood and BALF eosinophil expression 103

Table 4.1 Characteristics of subjects included in differential expression analysis of BALF EV miRNA, n=31 110

Table 4.2 Logistic regression of proportions of miRNA reads in COPD and healthy ex-smokers 112

Table 4.3 Top differentially expressed miRNA between COPD subjects and healthy ex-smokers 114

Table 4.4 Characteristics of subjects included in miRNA validation by RT-qPCR, n= 46 115

Table of Tables

Table 4.5 List of miRNA targets validated by RT-qPCR 116

Table 4.6 miRNA identified from NormFinder analysis of miRNA sequencing data as those most stably expressed across COPD and healthy ex-smoker samples 117

Table 4.7 Significantly differentially expressed miRNA measured by RT-qPCR between COPD subjects and healthy ex-smokers 119

Table 5.1 Summary results of *in silico* target prediction using multiMiR 132

Table 5.2 Top ten differentially expressed genes in epithelial brushings in COPD subjects 134

Table 5.3 Gene targets of the up-regulated EV miRNA in COPD, identified as down-regulated within the epithelial brushing transcriptome in COPD 136

Table 5.4 Gene targets of the down-regulated EV miRNA in COPD, identified as up-regulated in the epithelial brushing transcriptome in COPD 138

Table 5.5 "*Direct*" miRNA-mRNA interactions from pairwise correlation analyses 142

Table 5.6 Adjacency matrix for the gene targets with the greatest degree centrality 143

Table 5.7 miRNA impact on the differential gene expression 144

Table 5.8 Attributes of the five clusters identified from the miRNA-mRNA network cluster analysis 150

Table 5.9 Most significant GO terms generated from BiNGO output for GO biological process 153

Table 5.10 Most significant GO terms generated from BiNGO output for GO molecular function 155

Table 6.1 Characteristics of subjects included in the analysis exploring the diagnostic use of BALF EV miRNA and associations with inflammatory endotypes, N=44 166

Table 6.2 Correlations of up-regulated lung-derived EV miRNA expression with COPD phenotypic disease characteristics 168

Table 6.3 Correlations of down-regulated BAL EV miRNA expression with COPD phenotypic disease characteristics 169

Table 6.4 ROC analysis for predictive ability of up-regulated miRNA to differentiate between COPD and healthy ex-smokers 170

Table 6.5 ROC analysis to determine optimal combination of EV miRNA in differentiating between COPD and healthy ex-smokers 170

Table 6.6 Correlations between EV miRNA expression and immune cells proportions within BALF 172

Table 6.7 Definitions of inflammatory endotypes in COPD using %neutrophil and eosinophil pre-defined cut-offs 173

Table 6.8 ROC analyses to determine predictive ability of miRNA to differentiate between eosinophilic and non-eosinophilic subtypes in COPD 174

Table 6.9 ROC analysis to differentiate between pure eosinophilic COPD and paucigranulocytic or neutrophilic COPD 175

Table 6.10 ROC analyses to determine predictive ability of miRNA to differentiate between neutrophilic and non-neutrophilic subtypes in COPD 176

Table 6.11 Characteristics of subjects included in serum EV miRNA target validation by RT-qPCR, N=24 177

Table of Figures

Figure 1.1 An overview of the inflammatory and cellular interactions in COPD 5

Figure 1.2 Biogenesis and secretion of different subtypes of extracellular vesicles. 16

Figure 1.3 MiRNA biogenesis and function. 27

Figure 2.1 Size exclusion chromatography platform separating soluble proteins (yellow) from the extracellular vesicles (blue).....39

Figure 2.2 Protein concentration measured by BCA assay of the SEC fractions isolated from both Exo-spin™ (A) and PureEV™ (B) columns 41

Figure 2.3 EV abundance according to presence of CD63, measured by direct ELISA in combined SEC # 1-4 from PureEV™ columns 44

Figure 2.4 EV abundance (A) according to presence of CD9 and protein concentration (B) in combined SEC # 1-4 from PureEV™ columns 45

Figure 2.5 Whole mounted BALF-derived EVs isolated using PureEV™ SEC columns viewed by transmission electron microscopy from combined SEC # 2 (A) and # 3 (B) 47

Figure 2.6 Western blot analysis of isolated EVs from serum.....48

Figure 2.7 Whole mounted serum-derived EVs isolated using ExoMir™ kit viewed by transmission electron microscopy 49

Figure 2.8 Mean Cq values for miR-29a and miR-240 for combined SEC fractions from PureEV™ columns 51

Figure 2.9 Radar plot showing relative spike-in signal for each sample 54

Figure 2.10 Summary of library preparation process.....57

Figure 2.11 Overview of trimming of adapters and UMI correction process 60

Figure 2.12 Illustration of the principle behind UMI correction 60

Figure 2.13 Average read quality of miRNA sequencing data.....61

Figure 2.14 Box whisker plot showing the quality scores across all bases for a representative BALF sample. 61

Table of Figures

Figure 2.15 Read length distribution after filtering of adaptors for a representative BALF sample 62

Figure 2.16 Overview of RNA sequencing quality control and mapping. 64

Figure 2.17 Total number of reads for each sample. 65

Figure 2.18 Summary of mapping results of reads for each sample. 66

Figure 2.20 Overview of unsupervised filtering, exploratory data analysis and differential expression analysis methods ..... **Error! Bookmark not defined.**

Figure 2.21 Non-scaled Venn diagram showing the number of miRNA included in each filtered dataset and the overlap between each method..... **Error! Bookmark not defined.**

Figure 2.22 Boxplots (A-C) and histograms (D-F) showing the distribution and frequency of log-transformed CPM data across datasets for different filtering methods. ..... **Error! Bookmark not defined.**

Figure 2.23 Interquartile range/median plots of CPM data for different filtering methods. 75

Figure 2.24 Three-dimensional PCA plot showing the variation and clusters within the Limma filtered dataset..... 78

Figure 2.25 Biological coefficient of variation plot showing trended, common and miRNA ("Tagwise") specific estimates for CPM filtered TMM normalised data..... 82

Figure 2.26 An example of an MA plot showing differential expressed miRNA..... 83

Figure 2.27 An overview of the methods used to identify miRNA-mRNA interactions in this study 86

Figure 2.28 An overview of the methods used in this study for miRNA-mRNA interaction analysis 92

Figure 2.29 An overview of miRmapper outputs describing the miRNA-mRNA interaction network 95

Figure 3.1 Baseline (A) and historic (B) blood eosinophil counts for COPD subjects compared with healthy ex-smokers..... 101

Figure 3.2 BALF expression of immune cells in COPD subjects and healthy ex-smokers. 103

Figure 4.1 Total number of reads for COPD and healthy ex-smoker samples..... 111

Figure 4.2 Proportions of different types of reads in COPD and healthy ex-smoker samples111

Figure 4.3 Total number (A) and proportions (B) of smallRNA for COPD and healthy ex-smoker samples.....112

Figure 4.4 MA plot showing differentially expressed miRNA between COPD subjects and healthy ex-smokers 113

Figure 4.5 Raw Cq values for spike-in assays used to assess quality of cDNA synthesis (UniSp3) and reverse transcription reaction (UniSp6).....117

Figure 4.6 Number and expression level of miRNA detected by RT-qPCR for each sample ..118

Figure 4.7 Volcano plot showing relationship between P values and expression data 119

Figure 4.8 Normalised expression levels for significantly differentially expressed miRNA ...120

Figure 5.1 MA plot showing differentially expressed mRNA in epithelial brushings between COPD subjects and healthy ex-smokers 133

Figure 5.2 An overview of the methods used to identify putative miRNA-mRNA interactions135

Figure 5.3 Summary of miRNA-mRNA interaction analyses performed in this thesis. 140

Figure 5.4 Summary of pairwise correlation analyses to identify mRNA targets of miRNA ..141

Figure 5.5 Predicted miRNA impact on genes. 145

Figure 5.6 Dendrogram showing clustering of differentially expressed EV miRNA based on the similarity of the miRNAs' Jaccard index values to each other.....146

Figure 5.7 Up-regulated EV miRNA- mRNA interaction network in epithelial brushings in COPD 147

Figure 5.8 Down regulated EV miRNA-mRNA interaction network in epithelial brushings in COPD 148

Figure 5.9 Cluster analysis of miRNA-mRNA interaction network 149

Figure 5.10 Enrichment Map from BiNGO output with most significant GO terms for biological process and their interactions. 152

Figure 5.11 Enrichment Map from BiNGO output with most significant GO terms for molecular function and their interactions.....154

Table of Figures

Figure 6.1 ROC curve for miR-2110, miR-223-3p and miR-182-5p for the predicting the presence of COPD in the cohort 171

Figure 6.2 Venn diagram to describe the inflammatory endotypes in the COPD subjects based on pre-defined cut-offs 174

Figure 6.3 Normalised expression levels for miRNA in serum EVs 178

Figure 6.4 Relationship between Post FEV1% predicted and miR-2110 expression in serum.179

Research Thesis: Declaration of Authorship

Print name:	Hannah Burke

Title of thesis:	THE TARGETS AND FUNCTIONS OF LUNG-DERIVED EXTRACELLULAR VESICLE MICRORNA IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as conference abstracts:

Burke H, Heinson A, Freeman A, Ostridge K, Watson A, Staples K, et al., Late Breaking Abstract - Differentially expressed exosomal miRNAs target key inflammatory pathways in COPD. European Respiratory Journal. 2018; 52 (suppl 62):OA4922.

Burke H, Freeman A, Ostridge K, Watson AS, Spalluto C, Staples KJ et al., S13:Extracellular vesicle miRNA: a mechanism for chronic airway inflammation and macrophage dysfunction in COPD Thorax 2018;73:A9-A10

Signature:	Dr Hannah Burke	Date:	22nd March 2021

Acknowledgements

There are a number of people I wish to thank for their assistance and support over the last five years of study.

Firstly, I wish to express my sincere gratitude to my supervisors, Professor Tom Wilkinson, Dr Mirella Spalluto, Dr Karl Staples and Dr Tilman Sanchez-Elsner. Your support, patience and generous commitment of time have been greatly appreciated. Specifically I would like to thank Tom for giving me the courage to pursue a Wellcome Fellowship and for believing that I could 'deliver' despite setbacks along the way. Mirella, thank you for your enduring optimism even when times were tough, and most importantly for teaching me what 'cDNA' actually is! Karl, thank you for your wisdom and unwavering attention to detail. Tilman, thank you for interest and expertise in molecular biology, which helped craft the original idea.

My appreciation also goes Dr Anna Freeman, whose kindness and support made the finishing of this thesis possible in exceptional times.

I gratefully acknowledge the funding award received for my PhD from the Wellcome Trust Research Training Fellowship. Thanks also to the bioinformatics team at AstraZeneca, in particular Bastian Angermann for his help with the gene expression analysis.

I would also like to thank some very important people who are not directly associated with my research. Thank you Ryan for your belief and encouragement. Your love, support and confidence in my ability made this thesis possible. Thank you to my sister George, for the excellent example she sets, and her eternal support and friendship.

Finally, but most importantly, I would like to express my sincere appreciation to all patients, participants and professionals who were involved in this research. Without your generous commitment of time and effort, the investigations within this thesis would not have been possible.

Definitions and Abbreviations

AAT	Alpha-1 antitrypsin
AECOPD	Acute exacerbation of COPD
AGO2	Argonaute protein-2
Alix	Programmed cell death 6 interacting protein
AKR1B10	Aldo-Keto Reductase Family 1 Member B10
AKR1C2	Aldo-Keto Reductase Family 1 Member C2
AKT2	Protein kinase B2
ALOX15B	Arachidonate 15-lipoxygenase type B
ANP	Atrial natriuretic peptide
ARHGAP6	Rho GTPase activating protein 6
aSMase	Acid sphingomyelinase
ATG7	Autophagy related 7
ATS	American Thoracic Society
BALF	Bronchoalveolar lavage fluid
BCL	Binary base call
Bcl-2	B-cell lymphoma 2
BCV	Biological coefficient of variation
BDNF	Brain-derived neurotrophic factor
BiNGO	Biological Networks Gene Ontology tool
BMI	Body mass index
BSA	Bovine serum albumin
BTBD7	BTB Domain containing 7

Definitions and Abbreviations

cAMP	cyclic adenosine monophosphate
CCL	CC chemokine ligand
CD34	CD34 molecule
CD44	CD44 molecule
C/EBPα	CCAAT enhancer protein α
CEMIP	Cell migration inducing hyaluronidase
CFTR	Cystic fibrosis transmembrane conductance regulator
CLR	C-type lectin receptor
COL17A1	Collagen type XVII alpha 1 chain
COPD	Chronic Obstructive Pulmonary disease
COX	Cyclooxygenase
CPM	Counts per million
CREB5	Cyclic AMP-Responsive Element-binding Protein 5
CRIM1	Cysteine Rich Transmembrane BMP Regulator 1
CRP	C-reactive protein
CSE	Cigarette smoke extract
CXCL	CXC Chemokine ligand
CXCR	Chemokine receptor
CXR	Chest X-ray
CYP1B1	Cytochrome P450 1B1
DALY	Disability adjusted life years
DGRC8	DiGeorge syndrome critical region gene 8
DLCO	Diffusion capacity of the lung for carbon monoxide
DNA	Deoxyribonucleic acid

dNTP	Deoxynucleotide
DTT	Dithiothreitol
EBV	Epstein-Barr virus
edgeR	Empirical analysis of digital gene expression in R
EGFR	Epidermal growth factor receptor
ELAM-1	Endothelial-leucocyte adhesion molecule 1 or CD62E
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
EMT	Extracellular mesenchymal transition
EMV	Endothelial microvesicle
ERK	Extracellular regulated kinase-1/2
ERS	European Respiratory Society
ETS1	V-ets erythroblastosis virus E26 oncogene homolog 1
EV	Extracellular vesicle
FACs	Fluorescence-activated cell sorting
FBXL14	F-Box And Leucine Rich Repeat Protein 14
FC	Fold change
FDR	False discovery rate
FEF	Forced expiratory flow rate
FETUB	Fetuin B
FEV1	Forced expiratory volume in one second
FRCpleth	Functional residual capacity
FVC	Forced vital capacity
FXYD6	FXYD Domain Containing Ion Transport Regulator 6

GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GLI3	GLI Family Zinc Finger 3
GLP	Glucagon-like peptide
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GO	Gene ontology
GOLD	Global initiative for chronic obstructive lung disease
GVHD	Graft-vs-host-disease
HDAC	Histone deacetylase
Healthy-ES	Healthy ex-smoker
HITS-CLIP	high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation
HIF	Hypoxia-inducible factor
hnRNA	Pre-messenger RNA
hnRNP	heterogeneous nuclear ribonucleoprotein
HOMER1	Homer scaffold protein 1
HRCT	High-resolution computer tomography
HSP70	Heat shock protein 70
HT1	Hybridisation buffer
ICS	Inhaled corticosteroids
IFI30	Gamma-interferon-inducible lysosomal thiol reductase
IFN	Interferon
IGF	Insulin growth factor
IKKβ	Inhibitor of nuclear factor kappa-B kinase subunit beta
IL	Interleukin

ILC2	Type 2 innate lymphoid cells
IL1R2	Interleukin 1 receptor type 2
ING1	Inhibitor Of Growth Family Member 1
INSR	Insulin receptor
IQR	Interquartile range
IRAK4	Interleukin-1 receptor-associated kinase 4
ISEV	International Society for Extracellular Vesicles
ITGA2B	Integrin alpha-IIB or CD41
ITM2A	Integral membrane protein 2A
iTRAQ	Isobaric tags for relative and absolute quantitation
JADE1	Jade family PHD finger 1
KCO	Diffusion capacity of the lung of carbon monoxide per unit volume
KLF10	Kruppel Like Factor 10
LPS	Lipopolysaccharide
LT	Leukotriene
LTK	Leucocyte receptor tyrosine kinase
MACC1	Metastasis associated in colon cancer 1
MAPK	Mitogen-activated protein kinase
MCP-1	Monocyte chemoattractant protein 1
MDM	Monocyte derived macrophage
MHC	Major Histocompatibility complex
MICA	Microbiology and Immunology of the Chronically-inflamed Airway
miRISC	microRNA induced silencing complex
miRNA	MicroRNA

MMP	Matrix metalloproteinase
MOGSA	Multi omics data integrative clustering and gene set analysis
MPK-1	Mitogen-activated protein kinase-1
mRNA	Messenger RNA
MSC	Mesenchymal stem cells
MT1G	Metallothionein 1G
mTOR	Mammalian target of rapamycin
mtRNA	Mitochondrial RNA
MUC	Mucin
MVB	Multivesicular body
MV	Microvesicle
MYD88	Myeloid differentiation primary response 88
NaOH	Sodium hydroxide
NETs	Neutrophil extracellular traps
NF-κB	Nuclear factor kappa B
NFI-A	Nuclear factor I A
NLR	NOD-like receptor
NPR3	atrial Natriuretic Peptide Receptor 3
NSCLC	Non-small cell lung cancer
nSMase2	Neural sphingomyelinase 2
NTA	Nanoparticle tracking analysis
NTHi	Non-typeable *Haemophilus influenzae*
PAH	Polycyclic aromatic hydrocarbon
PARP	Poly(ADP-ribose) polymerase

PBS	Phosphate-buffered saline
PCA	Principle component analysis
PD	Programmed cell death
PDE4	Phosphodiesterase 4
PECAM-1	Platelet endothelial cell adhesion molecule 1 or CD31
PG	Prostaglandin
PI3K	Phosphoinositide 3-kinase
PIEZO2	Piezo Type Mechanosensitive Ion Channel Component 2
piRNA	Piwi-interacting RNA
PL	Phospholipase
PLAG1	Pleomorphic adenoma gene 1
Pred	Predicted
Pri-miRNA	Primary miRNA
PR	Pulmonary rehabilitation
PRR	Pattern recognition receptor
QC	Quality control
qCML	Quantile-adjusted conditional maximum likelihood
Q-score	Mean sequence quality score
RAETL1	Retinoic acid early transcript-1
RISC	RNA-induced silencing complex
RLR	Retinoic acid-inducible gene-I-like receptor
ROS	Reactive oxygen species
RNA	Ribonucleic acid
rRNA	Ribosomal RNA

RSV	Respiratory syncytial virus
RT	Reverse transcription
RT-qPCR	Real time quantitative polymerase chain reaction
RV	Residual volume
SCN2A	Sodium voltage-gated channel α subunit 2
SD	Standard deviation
SEC	Size exclusion chromatography
SERPINE1	Serpin family E member 1
SIRT1	Sirtuin 1
SLC39A14	Solute Carrier Family 39 Member 14
SLC45A3	Solute Carrier Family 45 Member 3
SLP1	Secretory leukocyte protease inhibitor
SMAD	Mothers against decapentaplegic homolog
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SOCS	Suppressor of cytokine signalling
SOP	Standard operating procedure
STAT	Signal transducer and activator of transcription protein
SUSD2	Sushi domain containing 2
SVC	Slow vital capacity
TBC1D3C	TBC1 Domain Family Member 3C
TE	Tris- Ethylenediaminetetraacetic acid
TEM	Transmission electron microscopy
TERT	Telomerase reverse transcriptase gene

TGF	Transforming growth factor
Th	T helper cell
TLC	Total lung capacity
TLL1	Tolloid-like 1
TLR	Toll-like receptor
Tm	Primer melting temperature
TMM	Trimmed mean of M-values
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
TRBP	Trans-activation response RNA-binding protein
tRNA	Transfer RNA
Tsg101	Tumour susceptibility gene-101
TSLP	Thymic stromal lymphopoietin
TSKU	Tsukushi
UMI	Unique molecular index
UTR	Untranslated region
VE	Vascular endothelial
VEGF	Vascular endothelial growth factor
VEGFR2	Vascular endothelial growth factor receptor 2
VTG	Volume of thoracic gas
WGCNA	Weighted correlation network analysis
WNK4	WNK lysine deficient protein kinase 4
WNT	Wingless/Integrase-1

Chapter 1 **Introduction**

1.1 Chronic obstructive pulmonary disease

1.1.1 Overview of COPD

Chronic obstructive pulmonary disease (COPD) affects 384 million people worldwide and is the third leading cause of death globally, claiming three million lives in 2016 (1, 2). The burden of COPD is predicted to increase over the next few decades due to continued exposure to COPD risk factors, such as tobacco smoke and an aging population (3). This is in marked contrast to other chronic diseases, such as heart disease and stroke, where there have been considerable decreases in mortality (4). Consequently, COPD represents an important public health challenge that is both preventable and treatable. COPD is characterised by irreversible airflow obstruction associated with emphysema, chronic inflammation and fibrosis, mucus gland hyperplasia and pulmonary arteriolar wall thickening and remodelling (5). In addition to pulmonary disease, COPD is also associated with several systemic complications, which contribute to overall morbidity and mortality (6, 7). COPD is therefore a complex and heterogeneous disease with many distinct clinical phenotypes and varied disease progression. This complexity has limited our understanding of disease and therefore has hindered the development of effective therapies.

1.1.2 Burden of COPD

COPD is a leading cause of morbidity and mortality worldwide that generates an economic and social burden that is both substantial and increasing (8, 9).

Prevalence data is widely variable in COPD, due to differences in survey methods, diagnostic criteria, and analytical approaches (3). The lowest estimates of prevalence are those based on self-reporting of a doctor-diagnosis of COPD or an equivalent condition. For example, most national data show that less than six percent of the adult population have been told that they have COPD (10) and this likely reflects the widespread under-recognition and under-diagnosis of COPD (11). Large-scale epidemiological studies with more robust methodology, including standardised questionnaires and pre- and post-bronchodilator spirometry, have estimated the global prevalence of COPD to be 11.7% (8.4-15%, 95% confidence intervals) (1).

COPD is associated with significant economic burden. In the European Union, the total direct costs of respiratory disease are estimated to be about six percent of the total healthcare budget, with COPD accounting for 56% (38.6 billion Euros) of the cost of respiratory disease (12). In particular,

COPD exacerbations account for the greatest proportion of the total COPD burden on the healthcare system. Unsurprisingly, the health care costs rise exponentially with COPD disease severity and in ageing populations with advancing comorbidities, this rise in costs is multiplied.

The burden of COPD to a patient is high, both in terms of health-related quality of life and health status. Disability adjusted life years (DALY) is a composite measure of the burden of each health problem on a patient. COPD is an increasing contributor to disability and mortality, with COPD ranked as the fifth leading cause of DALYs lost globally. In addition, there are significant impacts on the patient's family, friends and carers, with added emotional, social and financial burdens as a result of the disease (13).

1.1.3 Risk factors for development and progression of disease

COPD results from a complex interplay of long-term cumulative exposure to noxious gases and particles, combined with a variety of host factors including genetics, airway hyper-responsiveness and poor lung growth during childhood (14-17).

Cigarette smoking is the leading environmental risk factor for COPD and is estimated to account for up to 80-90% of cases within the developed world. Yet even in heavy smokers, fewer than 50% develop COPD during their lifetime (18) suggesting additional complex interactions between genes and other environmental factors.

1.1.3.1 Genetic factors

The most well described genetic risk factor is Alpha-1 antitrypsin (AAT) deficiency. A hereditary deficiency of AAT leads to uninhibited action of serine proteases in the lung leading to massive tissue destruction and emphysema in people who smoke. This genetic abnormality is relatively rare and affects only a small proportion of the world's population (estimated prevalence 0.01-0.02%), with a prevalence of ~1% in patients with COPD (17, 19).

Another clearly defined genetic defect linked to COPD are mutations of the telomerase reverse transcriptase gene (*TERT*), which results in early onset emphysema in smokers, who are predominantly female and have increased incidence of pneumothorax. These mutations are also described in idiopathic pulmonary fibrosis and there may be a family history of pulmonary fibrosis. These *TERT* polymorphisms were found in approximately 1% of COPD patients in two small independent cohorts, thus similar prevalence to AAT deficiency (20). The mutations lead to a shortening of telomeres leading to cellular senescence, which is discussed in more detail in the next section.

Early familial aggregation and linkage analysis studies strongly suggest genetic contributions to COPD (21-25), and more recent genome-wide association studies have identified several genomic regions that are clearly related to COPD susceptibility (26, 27). However, despite research advances, much of the heritability of COPD remains unexplained. Therefore the genetic determinants of COPD are likely to compose of multiple genetic susceptibility variants acting together to create a diverse array of COPD-related phenotypes.

1.1.3.2 Ageing and COPD

There is accumulating evidence that ageing hallmarks are prominent features of COPD (28). The structural changes seen in ageing airways and parenchyma (e.g. enlarged alveolar spaces and loss of lung elasticity) are similar to the changes seen in COPD. Broadly, the ageing hallmarks in COPD can be divided into processes affecting gene transcription (e.g. genomic instability, telomere attrition and epigenetic alterations), cellular metabolism (e.g. loss of proteostasis, dysregulated nutrient sensing and mitochondrial dysfunction) and other cellular processes (e.g. cellular senescence, stem cell exhaustion and altered intracellular communication) (29-31). The strongest evidence of the role of abnormal ageing in COPD come from studies demonstrating increased cellular senescence (mediated by the phosphoinositide 3-kinase (PI3K) -mammalian target of rapamycin (mTOR) pathway) (32, 33), however more work needs to be done to integrate findings and assess how age-related changes affect tissue repair and contribute causally to COPD.

1.1.3.3 Lung growth and development

Problems occurring during gestation, birth, and exposures during childhood and adolescence may all affect lung growth (34, 35), and have the potential to increase an individual's risk of developing COPD. For example, low birth weight has been associated with lower FEV1 (forced expiratory volume in one second) in adulthood (36) and several studies have found a link between early childhood lung infections and COPD (37, 38).

1.1.3.4 Smoking & domestic fuel exposure

Cigarette smoking is the most common risk factor for COPD worldwide. Cigarette smokers have a higher prevalence of respiratory symptoms, greater lung function abnormalities, a faster decline in FEV1 and greater COPD mortality rate than non-smokers (39). Passive smoke exposure is also an important contributor to the development of COPD (40), as well as smoking during pregnancy affecting lung growth and development in utero (39, 41, 42).

Globally, indoor air pollution is an underappreciated, but important risk factor for COPD. Almost three billion people worldwide use biomass as their main source of energy for cooking and

heating, which in poorly ventilated dwellings can accelerate lung damage (43-45). In these communities, indoor air pollution is responsible for a greater fraction of COPD risk than smoking.

1.2 Pathogenesis and pathophysiology of COPD

Inhalation of cigarette smoke or other noxious particles, such as smoke from biomass fuels, causes lung inflammation (46). The normal protective response to inhaled toxins is amplified in COPD leading to a chronic inflammatory response (47). This inflammation is thought to be responsible for parenchymal tissue destruction (resulting in emphysema), disruption of normal repair and defence mechanisms (resulting in small airway fibrosis), and goblet cell hyperplasia with mucus gland hypertrophy, as well as loss of cilia and reduced mucociliary function (resulting in chronic bronchitis). These pathological changes lead to gas trapping and progressive airflow limitation. The mechanisms for this amplified inflammation are not yet understood, but may at least in part, be genetically determined (48), and importantly persists despite smoking cessation. In support of the latter, studies have shown that levels of airway inflammatory cells are similar in current and ex-smokers (49) and markers of oxidative stress are found persistently elevated in the airways post-smoking cessation (50). One of the mechanisms postulated is increased phosphorylation of extracellular regulated kinase-1/2 (ERK) within emphysematous lungs driving matrix metalloproteinases (MMP)-1 induction leading to airway inflammation and matrix remodelling (51). Furthermore autoimmunity (via antielastin antibodies) (52), embedded particles/heavy metals from smoking (53) and changes in the lung microbiome have all been implicated in persistent airway inflammation in COPD (52, 54). However, novel mechanisms are emerging by which cells adapt to environmental cues such as smoking, which include changes in DNA methylation, histone modifications and regulation of transcription and translation by noncoding RNAs. These epigenetic mechanisms may be in key in driving persistent inflammation as a consequence of prior cigarette smoke exposure.

The inflammatory and cellular interactions in COPD are summarised in Figure 1.1, and will be discussed in more detail in the next section.

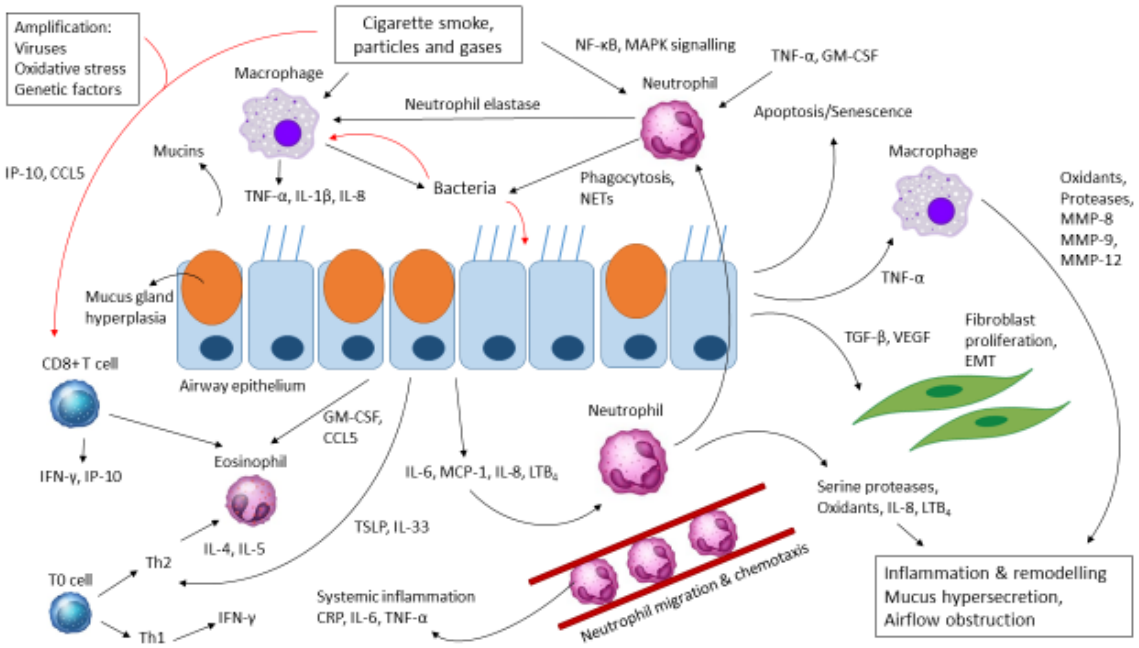


Figure 1.1 An overview of the inflammatory and cellular interactions in COPD

Cigarette smoke induced activation of immune cells leads to the release of cytokine, chemokines and proteases. Amplification signals (e.g. viruses, oxidative stress) are important in augmenting these inflammatory responses. CCL, CC chemokine ligand; CRP, C-reactive protein; CXCL, CXC chemokine ligand; EMT, epithelial mesenchymal transition; GM-CSF, granulocyte-macrophage colony stimulating factor; IFN, interferon; IL, interleukin; IP, interferon (IFN)-γ-inducible protein; LT, Leukotriene, MCP, monocyte chemotactic protein; MMP, Matrix metalloproteinase; NETs, neutrophil extracellular traps; NF-κB, nuclear factor kappa B; TGF, transforming growth factor; Th, T-helper cell; TNF, tumour necrosis factor; TSLP, thymic stromal lymphoprotein; VEGF, vascular endothelial growth factor. Adapted from Chung et al., 2008 (55)

1.2.1 Inflammatory cells

COPD is characterised by increased numbers of neutrophils, macrophages and lymphocytes (mainly CD8+ T lymphocytes) in bronchial biopsies (56), the small airways and lung parenchyma (57).

1.2.1.1 Neutrophils

Neutrophils are the front line of defence for the immune system. As summarised in Figure 1.1, they are a source of proteases and oxidants as well as inflammatory cytokines and antibacterial peptides (58, 59). They are, in part, responsible for both goblet cell hyperplasia in chronic bronchitis and tissue destruction in emphysema.

Neutrophils are present in the lung tissue in steady state conditions in small numbers, but rapidly migrate to the sites of infection or injury in response to chemoattractants released by epithelial cells and alveolar macrophages (60). Cigarette smoke, oxidative stress, bacteria and viruses all activate neutrophilic inflammation via the transcription factor nuclear factor kappa B (NF-kB), and p38 mitogen-activated protein kinase (MAPK) signalling in airway epithelial cells, resulting in the release of neutrophilic mediators, including neutrophil chemoattractants CXC chemokine ligand

(CXCL)1, interleukin (IL)-8 and leukotriene (LT) B_4 (61). Neutrophils are maintained in the airway by tumour necrosis factor (TNF)-α and granulocyte-macrophage colony stimulating factor (GM-CSF), and secrete serine proteases (e.g. neutrophil elastase) and MMP-8 and -9, which may contribute to alveolar destruction (62). Furthermore, neutrophil elastase secretion can activate epidermal growth factor receptors (EGFRs) leading to mucus hypersecretion and hyperplasia of goblet cells and submucosal glands (63). Neutrophils also release granule proteins and chromatin that together form extracellular fibres that bind bacteria. These neutrophil extracellular traps (NETs) degrade virulence factors and kill bacteria (64). Furthermore, neutrophils generate reactive oxygen species (ROS) causing oxidative stress, which further activates inflammation and induces corticosteroid resistance (65). Neutrophils from patients with COPD show marked abnormalities in chemotactic response, with increased migration but reduced accuracy (66), which may lead to increased damage to lung tissue and emphysema.

1.2.1.2 Macrophages

Macrophages are the dominant immune cell type in the healthy airways (67). Similar to neutrophils, macrophages release oxidants, inflammatory cytokines, growth factors, chemotactic factors and an array of MMPs when activated. They are critical to host defence and display a high degree of plasticity by changing their functional phenotype depending on the cytokine environment (68). In COPD, there are increased numbers of macrophages in the airways (69), they have a dysregulated response to virus (70), defective phagocytic ability (71) and reduced cytokine response (72), all of which may contribute to microbial dysbiosis and susceptibility to exacerbation.

Alveolar macrophages have an important role in host immune defence and regulate both innate and adaptive immunity (73). They are significantly increased in the lungs of smokers and in patients with COPD (74), where they accumulate in the areas of emphysema (75, 76). The increase in numbers of macrophages in COPD may be as a result of increased recruitment from peripheral blood (77), local proliferation and/or local survival of macrophages (78, 79). In addition, evidence suggests that the effects of acute and chronic cigarette smoke exposure on macrophage survival may differ (80).

Macrophages are a heterogeneous population of innate immune cells and can therefore perform different roles within the lung. Alveolar macrophages are primed to interact with inhaled pathogens and particles, whereas interstitial macrophages have an important role in regulating extracellular matrix proteins (81-83). Alveolar macrophages release more inflammatory mediators (e.g. TNF-α, IL-1β and IL-8), ROS and nitrogen intermediates, and show increased chemotaxis in response to cigarette smoke exposure (84-87). As a result, this activation implicates alveolar

macrophages in the airway inflammation and remodelling characteristic of COPD pathogenesis. MMP-9 is the predominant protease secreted by alveolar macrophages in COPD, and amongst other inflammatory proteins (e.g. TNF-α), is up-regulated by the transcription factor NF-κB, which is activated in alveolar macrophages in COPD patients, particularly during exacerbations (88).

In addition, in response to cigarette smoke, alveolar macrophages also recruit neutrophils to the lungs via release of IL-8 (89), which further enhances the inflammatory cascade. This inflammation persists even after smoking cessation, suggesting there are self-perpetuating mechanisms, although these are not well understood (90).

1.2.1.3 T Lymphocytes

T lymphocytes are also increased in the airways and lung parenchyma, with an increased CD4:CD8 ratio (91). T cell activation in COPD typically results in a T helper (Th)-1 mediated immune response and production of interferon (IFN)-γ (92). The mechanisms by which CD8+ T cells may cause COPD are unclear, however both CD8+ cells and IFN-γ levels correlate with airflow limitation and severity of emphysema, suggesting a role in pathogenesis (93). Furthermore, CD8+ T cells release perforin, granzyme B and cathepsins, which can contribute to alveolar cell apoptosis and therefore the development of emphysema (62).

1.2.1.4 Airway epithelial cells

The primary function of the airway epithelial cells is to form a physical and biochemical barrier between pathogens and harmful inhaled toxins. In COPD, this barrier function is compromised due to changes in the airway epithelial architecture, such as squamous metaplasia resulting in ciliary dysfunction, goblet cell hyperplasia and mucosal gland hypertrophy (94, 95). These changes result in reduced mucociliary clearance and chronic colonisation with respiratory pathogens (96, 97).

An important secondary function of these cells is to stimulate the innate immune response on recognition of an airway pathogen. Microbial presence is detected by pattern recognition receptors (PRRs) such as toll-like receptors (TLRs), C-type lectin receptors (CLRs), cytoplasmic retinoic acid-inducible gene-I-like receptors (RLRs) and NOD-like receptors (NLRs). Oxidative stress and viral infection can trigger epithelial cell production of IL-6 and IL-8 (98-100), which in turn leads to recruitment of innate immune cells, such as neutrophils, to surrounding tissues. In health these are protective mechanisms, however in COPD excessive or sustained recruitment and/or activation of immune cells can lead to tissue damage. Furthermore, IL-8 release from epithelial cells can alter structural cell phenotypes resulting in squamous metaplasia and impaired barrier function (101).

1.2.2 Inflammatory mediators

Inflammatory mediators, such as chemokines, cytokines and growth factors, are increased in COPD and their role in COPD pathogenesis has been extensively studied (102). Increased levels of IL-1β, TNF-α, IL-6 and IL-8 have all been measured in the sputum of COPD patients, with further increases during COPD exacerbations (103, 104).

TNF-α up-regulates adhesion molecules and facilitates migration of leucocytes into the bronchial mucosa during exacerbations by inducing IL-8 expression, and stimulating neutrophil degranulation and superoxide production (105). TNF-α activates NF-κB driven pro-inflammatory cytokine production, via activation of TNF receptor 2 (TNFR2) (106). Similarly, TNF-α activates p38 MAPK, which in turn may activate a similar array of genes and interact with the NF-κB pathway (107). This network of inflammatory signalling suggests a role for TNF-α in amplification of inflammation in COPD, resulting in activation of neutrophils, macrophages, the epithelium, mucus secretion, and destruction of lung parenchyma through release of proteinases.

IL-1β has similar actions to TNF-α and is also a potent activator of alveolar macrophages from COPD patients. Together, they both stimulate macrophages to produce MMP-9 (108).

The airway epithelium over-expresses monocyte chemotactic protein (MCP)-1 and IL-8 in response to cigarette smoke (109). MCP-1 is a chemoattractant of monocytes and may therefore be involved in the recruitment of macrophages in COPD. Whereas IL-8, along with LTB_4, are potent chemoattractants of neutrophils, and are the major drivers for neutrophilic recruitment to the airways in COPD (110).

Additionally, growth factors such as transforming growth factor (TGF)-β are also increased in the airway epithelium and lung tissue in COPD (111, 112). TGF-β is a potent regulatory cytokine with many effects on cell proliferation and differentiation, and on initiation and resolution of inflammatory responses (113). TGF-β is also implicated in the development small airway fibrosis and remodelling in COPD, by inducing epithelial mesenchymal transition (EMT) (114). EMT is a process whereby fully differentiated epithelial cells undergo transition to a mesenchymal phenotype giving rise to fibroblasts and myofibroblasts, and is increasingly recognized as playing an important role in aberrant repair and fibrosis in the airways in COPD (115).

1.2.3 Oxidative stress

In addition to chronic inflammation, an imbalance between oxidants and antioxidants (defined as oxidative stress) also occurs in the lungs of COPD patients (116). Sources of oxidants include inhaled cigarette smoke and ROS and nitrogen species released from inflammatory cells, such as

neutrophils. Biomarkers of oxidative stress (e.g. hydrogen peroxide, 8-isoprostane) are increased in stable COPD and further increased during exacerbations (117). Oxidative stress can lead to inactivation of anti-proteases and stimulation of mucus production. There may also be a reduction in antioxidants through activation of transcription factors such as NF-κB, which regulates many anti-oxidants (118).

Furthermore, studies have demonstrated a direct relationship between oxidative stress and cyclooxygenase (COX)-2 (119, 120). COX-2 is one of the rate-limiting enzymes in the metabolic pathway that transforms arachidonic acid into prostaglandins (PGs) and ROS, both of which promote inflammation via p53 signalling pathway (119). Moreover, the expression of COX-2 is increased by cigarette smoke exposure in lung epithelial cells (121), and activation of COX-2-PG signalling is thought to be central in promoting cigarette smoke induced-airway inflammation (120).

1.2.4 Protease-anti-protease imbalance

Both cigarette smoke exposure and chronic inflammation drive oxidative stress, which primes several inflammatory cells to release a combination of proteases and inactivates several anti-proteases such as AAT, secretory leucoprotease inhibitor (SLPI) and tissue inhibitors of MMPs (122). The main inflammatory cell types involved are neutrophils releasing proteases such as serine proteases, elastases, cathepsin G and protease 3, and macrophages releasing cysteine proteases, cathespins E, A, L and S, and various MMPs (e.g. MMP-8, MMP-9 and MMP-12). Protease mediated destruction of elastin, a major tissue component in lung parenchyma, is thought to lead to alveolar wall destruction and emphysema (123).

1.2.5 Airflow obstruction and hyperinflation

The main site of airflow obstruction occurs in the smaller airways (< 2mm in diameter) due to inflammation, airway remodelling and inflammatory exudates driven by the above processes (57). This peripheral airway limitation progressively traps gas during expiration, resulting in hyperinflation. Evidence suggests that hyperinflation occurs early in disease, and is the main mechanism for exertional dyspnoea in COPD (124). Exertional dyspnoea occurs as a consequence of dynamic hyperinflation, where loss of inspiratory capacity at rest leads to a reduction in functional residual capacity during exercise (125).

Other factors contributing to airflow obstruction include loss of elastic recoil (due to destruction of elastin fibres by proteases) and destruction of alveolar support (from alveolar attachments) (126). These features are the hallmarks of emphysema.

1.2.6 Gas exchange abnormalities and pulmonary hypertension

Both gas exchange abnormalities and pulmonary hypertension occur in late disease, due to advanced structural changes (e.g. emphysema) in the COPD lung. Loss of alveolar surface area and a reduction in the pulmonary vascular bed lead to a ventilation: perfusion mismatch (127). Chronic hypoxia leads to vasoconstriction of the small pulmonary arteries, which in turn leads to endothelial dysfunction, remodelling of the pulmonary arteries with smooth muscle hypertrophy and hyperplasia (128). Persistent pulmonary hypertension results in right ventricular hypertrophy and dysfunction (known as cor pulmonale) (129).

1.2.7 Systemic features

Comorbidities are common in COPD, are often linked to the same risk factors (i.e. smoking, ageing, and inactivity), and have a major impact on morbidity and mortality (130). Cachexia, skeletal muscle wasting, cardiovascular disease, osteoporosis, anaemia and metabolic syndromes (e.g. diabetes) are all recognised complications of COPD. The mechanisms underlying these systemic effects are likely multifactorial, including inactivity, systemic inflammation, tissue hypoxia and oxidative stress (131).

The above pathogenic mechanisms result in the pathological changes found in COPD. These in turn result in pathophysiological abnormalities – mucus hypersecretion and ciliary dysfunction, airflow obstruction and hyperinflation, gas exchange abnormalities, pulmonary hypertension and systemic effects.

1.3 Diagnosis of COPD

COPD should be considered in any patient who has dyspnoea, chronic cough or sputum production, and/or a history of exposure to risk factors for the disease. In this clinical context, post-bronchodilator spirometry is the current test used to diagnose COPD by demonstrating airflow obstruction (4). It is a physiological test that measures the volume of air an individual can expel from his lungs after a maximum inspiration. The standard spirometry manoeuvre is maximal exhalation after a maximum deep inspiration. A number of indices can be derived from this:

- Forced vital capacity (FVC): The volume delivered during expiration made as forcefully and completely as possible, starting from full inspiration.
- FEV1: Maximum volume of air exhaled in the first second of a forced expiration.
- FEV1/FVC ratio: Airflow obstruction is defined as, an FEV1/FVC ratio of under 0.7.
- FEF 25%–75%: The mean forced expiratory flow rate between 25% and 75% of the FVC.

Predicted values for spirometry have been calculated from large sample groups and vary with age, height, sex and race (132). A diagnosis of airflow obstruction in COPD is made when the ratio of FEV1/FVC is less than 0.7 on post-bronchodilator spirometry. However, in certain populations this absolute cut-off may be unreliable as it can over diagnose in the elderly and under diagnose in the young (133, 134). There are therefore advocates of using the lower limit of normal value for these populations, which involves taking the bottom 5% as abnormal (135). The FEV1 can also be used to severity grade COPD (see Table 1.1).

Table 1.1 Classification of airflow limitation severity in COPD based on post-bronchodilator FEV1

GOLD classification	Severity	FEV1
GOLD 1	Mild	FEV1 ≥ 80% predicted
GOLD 2	Moderate	50% ≤ FEV1 < 80%
GOLD 3	Severe	30% ≤ FEV1 < 50%
GOLD 4	Very severe	FEV1 < 30%

GOLD, Global initiative for Chronic Obstructive Lung Disease; FEV1, Forced expiratory volume in one second. Adapted from GOLD guidelines, 2019 report (4).

Spirometry is cheap, easy to perform and highly reproducible in most populations and has therefore been utilised as the investigation of choice for diagnosing and severity grading COPD. In early studies, reduced FEV1 was associated with higher mortality (136, 137). However in the large National Emphysema Treatment trial, FEV1 was not associated with mortality in multivariate analysis (138) and other studies have also not found this link with mortality (139). There are also only weak correlations between FEV1 and change of quality of life and health status (140). There are also concerns that spirometry is an effort dependent test and can be insensitive to early disease (141). FEV1 provides little information about the distal airways and so FEF 25%–75% is often cited as a measure of small airways pathology. However FEF 25-75% measure is not very sensitive and has poor reproducibility (142). Given these issues, efforts should be made to find more specific and sensitive measures of early disease, which would allow earlier interventions (i.e. smoking cessation) to alter disease course.

1.4 Treatment of COPD

Smoking cessation is key and has the greatest capacity to influence the natural history of COPD. Pharmacotherapy and nicotine replacement reliably increase long-term smoking abstinence rates (143). Furthermore, effective tobacco control policies (i.e. smoking ban legislation) have been shown increase quit rates and reduce harm from second hand smoke exposure (4).

Pulmonary rehabilitation (PR) is defined as a comprehensive intervention tailored to patient's needs, and includes exercise training, education and self-management (144). PR should be

considered in all COPD patients as it has been shown to be the most effective therapeutic strategy to improve breathlessness, increase exercise capacity, and improve quality of life in COPD patients (144, 145).

Vaccination is also an integral part of COPD management, with the annual influenza vaccine reducing lower respiratory tract infections requiring hospitalisation and death in COPD patients (146).The pneumococcal vaccine also provides significant protection against community acquired pneumonia (147).

Pharmacological therapy can reduce COPD symptoms, reduce the frequency and severity of exacerbations, and improve health status and exercise tolerance (4). However, to date, there is no conclusive clinical trial evidence that any existing medications for COPD modify the long-term decline in lung function (4). Medications used to treat stable COPD encompass bronchodilators and anti-inflammatory agents, such as inhaled corticosteroids (ICS). *In vitro* evidence suggests that COPD-associated inflammation has limited responsiveness to corticosteroids, however, in combination with bronchodilators, improvements in lung function, health status and exacerbation frequency are shown (4). In addition, regular treatment with ICS increases risk of pneumonia, especially in those with severe disease (148). More recently, a number of studies have shown that blood eosinophil counts predict the magnitude of the effect of ICS treatment in preventing future exacerbations (149, 150). The issues around using blood eosinophils to guide treatment are discussed in the next section, however a blood eosinophil count > 300 cells/µL can be used to identify patients with the greatest likelihood of treatment benefit with ICS (4).

In the current paradigm, treatment options for COPD are used relatively indiscriminately, without consideration of specific phenotypes and endotypes. This arbitrary treatment strategy is largely due to limits in our understanding of the pathophysiology in COPD and the disease heterogeneity, which is a challenge in itself when designing clinical trials of future treatments. The following section, will discuss the concept of endotypes in COPD and their role in defining patients that may benefit from specific targeted therapy.

1.5 COPD Endotypes

Although, COPD is defined by characteristic symptoms and fixed airflow obstruction, patients may show different clinical features, rates of progression, frequency of exacerbations and associated diseases (co-morbidities), suggesting different clinical phenotypes. The classically defined phenotypes of chronic bronchitis and emphysema in COPD are based on the two major pathologies in COPD, yet despite these different pathologies, the clinical presentation in both these phenotypes is very similar, with progressive dyspnoea on exertion and reduced exercise

tolerance as the major presenting symptoms. Furthermore, identification of frequent exacerbations in a patient (151) (so called frequent exacerbator phenotype – discussed in section 1.6) may influence the choice of therapy, however there is no evidence that exacerbation frequency is related to differences in underlying mechanisms of COPD (4).

Endotypes describe a distinct pathophysiological mechanism at a cellular and molecular level leading to a clinical phenotype of disease. Despite similar clinical symptoms, patients may respond very differently to the same therapeutic intervention and therefore precision medicine is used to describe treatment targeted at specific patient endotypes. Different inflammatory patterns have been described in COPD and are referred to as "*inflammatory endotypes*", however the true molecular mechanisms underlying these remain uncertain.

1.5.1 Neutrophilic COPD

Neutrophilic inflammation is one of these inflammatory patterns described in COPD, with sputum neutrophilia characteristic of the disease (152). Neutrophilic inflammation in COPD is unresponsive to corticosteroids, even in high doses (153) and this may reflect the marked reduction in histone deacetylase 2 (HDAC2) seen in COPD lungs, which is secondary to oxidative stress (154). This resistance to corticosteroids is reflected by the lack of effect of high-dose ICS on mortality and progression of COPD (155) and thus indicates the need for more specific anti-neutrophilic therapies in COPD. However although a CXC chemokine receptor-2 (CXCR2) antagonist (navarixin) that blocks the chemotactic effect of the neutrophil chemoattractant IL-8 and related chemokines does reduce sputum neutrophils in COPD patients, it had no clinical benefit on lung function, symptoms or exacerbations (156). Other therapies directed towards neutrophilic inflammation, including antibodies against TNF-α and IL-1β, and p38 MAPK have also been largely clinically ineffective (157-160). The phosphodiesterase (PDE)-4 inhibitor roflumilast, which reduces neutrophilic (and eosinophilic) inflammation, can reduce exacerbations in carefully selected patients (161). However, its dose is limited by side effects, such as diarrhoea, headaches and nausea (162). Patients with frequent exacerbations, FEV_1<50% predicted and chronic bronchitis appear to show the best responses to roflumilast in terms of reduced exacerbations (163), which may indicate a neutrophilic endotype given neutrophils drive mucus hypersecretion in COPD.

1.5.2 Eosinophilic COPD

Some COPD patients may have more reversibility to bronchodilators and a better response to corticosteroid therapy. These patients show increased sputum eosinophils, increased FeNO

(marker of eosinophilic airway inflammation) (164), and have been classified as eosinophilic COPD. The prevalence of eosinophilic COPD is uncertain as different studies have used different criteria and eosinophil measurements, reports vary from a prevalence of 15% (165), to 60% of patients with COPD (166). There is evidence that eosinophilic COPD patients have more frequent exacerbations (167, 168) and are more responsive to corticosteroid treatment (149).

Defining eosinophilic COPD has been challenging since although blood eosinophils are easy to measure in clinical practice, it is uncertain whether they closely reflect sputum or tissue eosinophilia. Increased blood eosinophils predict an increase in sputum eosinophils (>3%) in around 70% of patients (169) but there is only a weak relationship between blood and sputum or bronchial tissue eosinophils in COPD patients (170, 171). Furthermore, blood eosinophils counts are variable over time and even within the same day, so that it is important to make repeated measurements (172).

The mechanism of eosinophilia in COPD is not yet certain. An increase in sputum IL-5, a key mediator of eosinophil proliferation, differentiation and maturation, has been reported in COPD patients with eosinophilia (173). Furthermore, there is an increase in granulocyte-macrophage GM-CSF, which is also important for maintaining eosinophil survival in lungs, and CC chemokine ligand 5 (CCL5), a recruiter of eosinophils, in the sputum of these patients (174). Both these cytokines are released by airway epithelial cells in addition to the upstream cytokines thymic stromal lymphopoietin (TSLP) and IL-33 in response to cigarette smoke and viral infection. These latter stimuli recruit Th2 and type 2 innate lymphoid cells (ILC2), which secrete IL-5, resulting in eosinophilic inflammation (175-177). However, it is not understood why only a minority of COPD patients have significant eosinophilia, but it may be linked to associated allergy or asthma.

Treatments targeting eosinophilic inflammation, such as the anti-IL-5 antibody mepolizumab, have only a minor effect in reducing exacerbations in eosinophilic COPD patients with a history of exacerbations (178). Furthermore, benralizumab, which blocks the IL-5 receptor-α even more effectively and removes eosinophils from the airways, does not seem to provide a significant clinical benefit in COPD patients (179). Importantly, resident eosinophils have been described in healthy human lungs that are independent of IL-5 regulation (180), and therefore this IL-5 independent sub-population may contribute to lung eosinophilia in COPD and thus may require alternative targets to achieve a treatment response in this disease.

1.6 Exacerbations in COPD

The clinical course of COPD is often characterised by “exacerbations”, defined as episodes of worsening symptoms such as breathlessness, cough and an increase in sputum production and/or

purulence (181). These exacerbations have a detrimental effect on patients' quality of life (182), accelerate disease progression and are associated with increased hospital admission and death (183, 184). Up to 70% of exacerbations are due to respiratory infections, including bacterial and respiratory viral pathogens (185). Other causes include environmental pollution, which can depend on season and geographical placement (186), and up to 30% are of unknown aetiology (187). Exacerbations are 50% more likely in winter, probably due to improved survival of respiratory viruses in colder weather, the crowding together of people indoors during the winter and a possible reduced immunological response (188).

Exacerbation frequency increases with COPD severity (189). Although, a sub-group of patients have frequent exacerbations independent of lung function and this has been recognized as a clinical phenotype, which are often referred to as "frequent exacerbators" (151). These patients have a poorer quality of life (182), faster decline in lung function (184), higher readmission rates to hospital (190, 191) and increased mortality (192). Given these worsening outcomes, the ability to predict and prevent exacerbations would have a significant impact on survival in this patient group. However so far, studies have been unsuccessful in identifying markers of exacerbation risk (193-196). Therefore new insights into the defective immune functions in COPD are required to identify key cells, proteins or mediators of disease, which increase susceptibility to exacerbation.

1.7 Extracellular vesicles

Defective innate immunity is key to COPD pathogenesis and susceptibility to exacerbations. Although some of the mechanisms are well described (outlined in section 1.2), much is still uncertain and a clearer understanding of the different interactions between the main cellular players may result in new mechanistic insights. Aberrant intercellular communication has been demonstrated so far through investigation of classical pro-inflammatory cytokine pathways. However, more recently, a novel mechanism of intercellular signalling has been identified via extracellular vesicles (EVs) and this may provide new insights for disease pathogenesis.

Intercellular communication is fundamental for organisms to respond and adapt to changes in the environment. It is the cornerstone of the immune response and is typically defined as either direct cell to cell contact or via secretion of soluble mediators (e.g. cytokines) (197). EVs have been shown to "shuttle" cargo between cells and have undergone intensive investigation in the last two decades (198). The study of EVs in COPD may provide novel insights into the underlying mechanisms driving defective innate immunity and chronic airway inflammation.

1.7.1 Definition

Almost all cell types release EVs, and the term EVs encompasses apoptotic bodies, microvesicles (MVs) and exosomes, which are differentiated by size, origin and biochemical composition (Figure 1.2). Each subset of vesicles are enriched for a subset of diverse proteins reflecting that of the parent cell, and include adhesion molecules, membrane trafficking molecules, cytoskeleton molecules, heat shock proteins, signal transduction proteins, cytoplasmic enzymes, cytokines and chemokines. In addition, EVs can contain messenger ribonucleic acid (mRNA), non-coding RNAs (e.g. microRNA (miRNA)) (199) and extra-chromosomal deoxyribonucleic acid (DNA) (200), all of which are derived from the parent cell.

EVs have an important role in immune function, by regulating both inflammatory and immunosuppressive pathways, and therefore are implicated in driving processes responsible for many inflammatory, autoimmune and infectious diseases (201-204). There is growing interest into their use as diagnostic biomarkers in these diseases, but also as therapeutics by targeted manipulation of their cargo for modulating the immune response in these diseases (198).

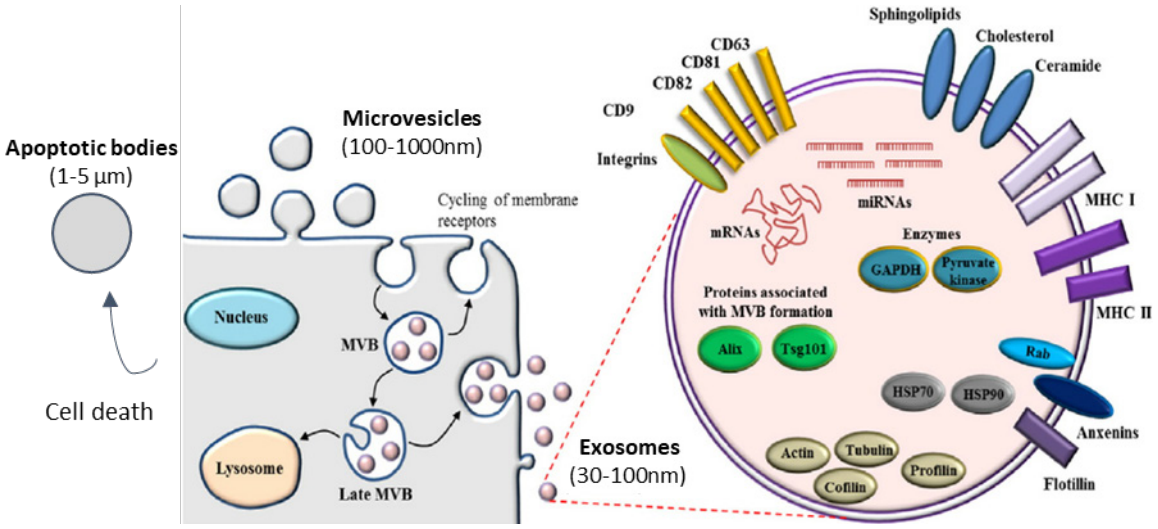


Figure 1.2 Biogenesis and secretion of different subtypes of extracellular vesicles.
MVB, multivesicular body; MHC, Major histocompatibility complex; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; HSP70 Heat shock protein 70; Tsg101, Tumour susceptibility gene-101. Adapted from Giau et al., 2016 (205)

1.7.2 Subtypes and classification

1.7.2.1 Apoptotic bodies

Apoptotic bodies, the largest of the EVs, are 1-5 µm in diameter and are released indiscriminately from the plasma membrane during cellular apoptosis (206, 207) (see Figure 1.2). They can transfer DNA (208), specifically oncogenes (209), and are also capable of antigen presentation (210, 211).

1.7.2.2 Microvesicles

MVs (also known as microparticles or ectosomes) are 100-1000nm in diameter and are released from the plasma membrane by budding or blebbing (212-215) (see Figure 1.2). As a consequence they express the same surface antigens found on the parent cells. For example, vascular endothelial (VE)-cadherin (CD144), platelet endothelial cell adhesion molecule 1 (PECAM-1 or CD31) and endothelial-leucocyte adhesion molecule 1 (ELAM-1 or CD62E) are expressed on endothelial cell-derived MVs (216), whereas P-selectin (CD62P) and integrin alpha-IIb (ITGA2B or CD41) are expressed on platelet-derived MVs (217). These surface markers enable efficient characterisation of these vesicles using antigen detection with fluorescence-activated cell sorting (FACs) (218). MVs bind via specific receptors to the surface of target cells leading to endocytic uptake of the vesicle and direct delivery of its contents (214). MVs have a number of diverse functions including pro-coagulant activity (219), secretion of IL-1β (220) and a role in tumour progression (221), in addition to several immunoregulatory functions (201, 222, 223).

1.7.2.3 Exosomes

Exosomes, the smallest of the EVs (50-100nm), were first discovered over 30 years ago, but considered little more than “cell junk” whose job it was to discard unwanted cellular components. Exosomes were originally discovered by researchers using transmission electron microscopy (TEM) to investigate the loss of the transferrin surface receptor from reticulocytes (224, 225). They noted that radioactive-labelled transferrin bound to its receptor made its way from the cell surface into the cell via the endocytic pathway and travelled through different cellular compartments to multivesicular bodies (MVBs). The transferrin, still bound to its receptor, was sequestered in small vesicles that formed inside the larger MVB. Unexpectedly, these MVBs fused with the plasma membrane and released these small vesicles – now known as exosomes, to the outside of the cell.

The next major exosome discovery implicated their role in adaptive immunity. In 1996, Raposa et al. showed that Epstein-Barr virus (EBV)-transformed B-lymphocytes secreted exosomes that bore major histocompatibility complex (MHC) II bound to antigenic peptides which could induce antigen-specific MHC class II T cell responses (226).

Subsequently, major advances in large-scale protein analysis techniques allowed further characterisation of exosome composition and studies showed that exosomes contained specific proteins originating mainly from the endosomal compartment (e.g. tetraspanins; CD9, CD63 and CD81), the endocytic pathway (e.g. Tumour susceptibility gene 101 (Tsg101) and programmed cell death 6 interacting protein (Alix)), and the cytosol (e.g. Actin and Tubulin) (207, 227).

Simultaneously, studies showed that exosomes secreted from one cell could transfer information to a recipient cell, such as MHC-peptide complexes inducing immune responses (228-230) and apoptotic signals initiating cell death (231).

More recently Valadi et al. discovered the presence of mRNA and miRNA inside exosomes, suggesting that exosomes can transfer genetic information and alter gene expression in recipient cells (199). This discovery not only indicated a new form of intercellular communication, but also suggested that exosomes could behave similarly to viruses, in that they could deliver genetic material, which was then processed by the recipient cells machinery.

1.7.3 Extracellular vesicles in the lung

EVs have been implicated in the pathogenesis of a number of lung diseases. Adymre et al. found that exosomes isolated from healthy human bronchoalveolar lavage fluid (BALF) express MHC class I and II, CD54 and CD63 and the co-stimulatory molecule CD86. Based on these findings, they concluded that these exosomes were likely to originate from antigen presenting cells (232). However, subsequent analysis of exosomes isolated from BALF in asthmatic mice originated predominantly from epithelial cells and exosome production was enhanced by IL-13. In addition, these exosomes were able to induce proliferation and chemotaxis of undifferentiated macrophages (233). *In vitro* work by Kesimer et al. showed that epithelial cells secreted exosomes enriched for mucins (MUC1, MUC4 and MUC16), which reduced the ability of the human influenza virus to infect epithelial cells by up to 85-90% (234). Surfactant proteins have also been detected in EVs derived from lung epithelial cells (235).

Epithelial cell injury secondary to cigarette smoke is also known to stimulate exosome production. Moon et al. showed that cigarette smoke extract (CSE) increased the release of exosomes from lung epithelial cells, mediated by the induction of RAB27a expression (an important regulator of exosome production) (236). These exosomes were shown to encapsulate full length cysteine-rich angiogenic protein 61 (CCN1) that induced epithelial cell IL-8 and vascular endothelial growth factor (VEGF) secretion, both of which are implicated in the pathogenesis of COPD. Further work has also shown that CSE triggers epithelial cells to produce EVs containing miRNA-210, which promotes myofibroblast differentiation in lung fibroblasts via autophagy related 7 (ATG7) silencing (a regulator of autophagy) (237). Consequently, EV miRNA autophagic regulation of myofibroblast differentiation could be a critical determinant of airway fibrotic remodelling in COPD pathogenesis.

As well as CSE inducing EV release from epithelial cells, macrophages also release EVs. In response to CSE, macrophage derived exosomes are both pro-thrombotic and proteolytic (containing MMP-

14) and therefore may contribute to the instability of atherosclerotic plaques and the destruction of pulmonary connective tissue in emphysema (238, 239). Cordazzo et al. demonstrated the effects of CSE-induced MVs from monocytes increasing production of pro-inflammatory and pro-thrombotic mediators from human lung epithelial cells (240). In addition, Cerri et al. demonstrated EVs from human macrophages have pro-inflammatory potential in human airways through their ability to up-regulate IL-8, CD54 and MCP-1 synthesis by bronchial and alveolar epithelial cells. Finally, Soni at el. demonstrated that alveolar macrophage MVs mediate acute lung injury in response to lipopolysaccharide (LPS) stimulus by transporting biologically active TNF-α. When instilled intra-tracheally into mice, these MVs induced neutrophil influx and CD54 expression in epithelial cells (241). All of these mechanisms may contribute to the persistent airway inflammation recognised in COPD.

In contrast, alveolar macrophage derived EVs have also been shown to have anti-inflammatory effects. Bourdonnay et al. demonstrated that EVs derived from alveolar macrophages were a source of suppressor of cytokine signalling (SOCS) 1 and 3 proteins. When delivered to airway epithelial cells these EVs inhibited cytokine-induced signal transducer and activator of transcription protein (STAT) activation and expression of MCP-1 (a STAT-dependent gene product). Furthermore they found reduced levels of SOCS1 and 3 in BALF of smokers compared with non-smokers, suggesting that EV delivery of SOCS proteins may be dysregulated by smoke exposure as a result of persistent inflammation (242). In addition, alveolar macrophages have been shown to secrete EVs which inhibit influenza infection of nearby airway epithelial cells. However, when exposed to cigarette smoke this inhibitory mechanism is lost (243). These findings may have important consequences for patients with COPD who demonstrate increased susceptibility to viral infections.

In summary, EVs have been implicated in both driving chronic airway inflammation and altering cellular function of important immune cells within the lung. Their release and function is altered by cigarette smoke exposure and therefore it is plausible that EVs play a key role in COPD pathogenesis.

1.7.4 Extracellular vesicles in COPD

To date, no study has directly sampled and characterised EVs from the lungs of patients with COPD. However, a number of studies have investigated EVs in smokers, murine models, cell culture models and circulating EVs in patients with COPD and these are summarised in Table 1.2.

Table 1.2 Summary of the EV studies in smokers, patients with COPD and murine models

EV type	Subject cohort/Cell type	Main findings	Ref
Bronchoalveolar lavage EVs			
	Smokers (n=10) and non-smokers (n=10)	MiRNAs let-7e, let-7g and miR-26b were significantly down-regulated in smokers. MiRNAs let-7e and let-7g expression was also found to be reduced in human bronchial epithelial cells (BEAS-2B) exposed to smokers EVs in comparison with non-smoker EVs.	(244)
	Smoke exposed mice (n=8), LPS challenged mice (n=8), saline & air control mice (n=16)	Findings support that infectious agents (bacterial or viral) can trigger EV release in the airways and that ATP drives EVs to release IL-1β and IL-18 via a P2X receptor-7/caspase-dependent mechanism.	(245)
Lung tissue EVs *(isolated using UC from surgically resected lung tissue)*			
	Non-smokers (n=13), healthy smokers (n=13), COPD (n=13)	EVs showed greater biodiversity (more operational taxonomic units) than lung tissue. Firmicutes were highly present in the EVs of the COPD group compared with other samples or groups.	(246)
Sputum microparticles (MPs)			
	Male, COPD patients (n=18)	MPs were identified in the sputum of COPD patients. CD31- MPs (i.e. those not associated with the endothelium) correlated negatively with FEV1, suggesting relationship with disease severity.	(247)
Plasma endothelial microparticles (EMPs)			
	Healthy non-smokers (n=32), healthy smokers with normal spirometry and DLCO (n=42), and healthy smokers with normal spirometry and low DLCO (n=19)	Smokers with low DLCO (evidence of emphysema) have increased circulating EMPs, suggesting that measurement of plasma EMPs may be helpful in identifying early disease.	(248)
	Stable COPD patients (n=80), exacerbating COPD patients (n=27), healthy controls (n=20)	Levels of CD144+, CD31+ and CD62E+ EMPs were significantly higher in stable COPD than healthy controls and even higher in exacerbating patients when compared to stable COPD. Suggesting EMPs may be a biomarker for exacerbations.	(249)

EV type	Subject cohort/Cell type	Main findings	Ref
Plasma endothelial microparticles (EMPs) continued...			
	Stable COPD patients (n=104), healthy controls (n=74)	Levels of CD31+ EMPs were significantly higher in mild COPD consistent with endothelial apoptosis CD31+ EMPs were also associated with percent of emphysema and correlated with reductions in pulmonary microvascular perfusion. Levels of CD63E+ EMPs were elevated in severe COPD and patients with hyperinflation	(250)
	Non-smokers (n=28), healthy smokers (n=61), COPD smokers (n=49)	COPD and healthy smokers had elevated plasma EMPs, which remained elevated over 12 months, but returned to non-smoker levels in healthy smokers only who quit.	(251)
	Healthy non-smokers (n=8), COPD subjects (n=17). Primary human lung microvascular endothelial cell (HLMVEC) model exposed to CS.	Levels of EMPs significantly increased in COPD. Furthermore, miRNAs let-7d, miR-191, miR-126 and miR-125a were significantly enriched in EMPs released by HLMVEC exposed to CS. These EMPs were ceramide-rich and required aSMase for their release, an enzyme found to exhibit significantly higher activity in the plasma of COPD patients.	(252)
	Smoke exposed rats (n=30), sham-smoke exposed rats (n=30)	Plasma CD42b-/CD31+ EMPs were significantly increased in rats exposed to cigarette smoke and associated with lung function decline. These EMP markers are suggestive of pulmonary endothelial apoptosis	(253)
Plasma exosomes			
	Stable COPD patients (n=20), exacerbating COPD patients (n=20), healthy controls (n=20)	Circulating plasma exosome levels (CD9+) were significantly higher in COPD patients both at stable state and acute exacerbation. Levels of exosomes correlated with CRP, sTNFR1 and IL-6 in plasma suggesting a role in exacerbations in COPD.	(254)
Serum EVs			
	Stable COPD patients (n=5), 5 healthy controls (n=5).	Levels of miR-21 increased in the serum EVs of COPD patients and suggested a role for EV miR-21 in macrophage polarisation.	(255)

aSMase, acid sphingomyelinase; CRP, C-reactive protein; CS, cigarette smoke; DLCO, diffusing capacity of the lungs for carbon monoxide; EMPs, endothelial microparticles; EV, extracellular vesicle; HLMVEC, human lung microvascular endothelial cell; LPS, lipopolysaccharide; IL, interleukin, sTNFR1, soluble tumour necrosis factor receptor 1

1.7.4.1 Extracellular vesicles as a diagnostic biomarker in COPD

Over the past decade, there have been numerous studies focused on discovery and assessment of biomarkers in relation to acute exacerbations of COPD (AECOPD). Owing to their heterogeneity and the lack of available diagnostic laboratory tests, AECOPD are often based on clinical suspicion alone, which is subjective and variable within and across physicians. The most well studied biomarkers are C-reactive protein (CRP), IL-6, and TNF-α. However these are rather non-specific as they are raised by many inflammatory conditions and had variable statistical significance and results (256). EVs could be used as potential biomarkers for diagnosis and monitoring of COPD patients. Indeed, numerous EV proteins and miRNAs have already been identified as potentially useful biomarkers for various diseases, especially cancer detection (257-259)

Recently, Tan et al. assessed the expression levels of exosomes in the plasma of patients with AECOPD (n=20) versus patients with stable COPD (n=20) and healthy non-smokers (n=20). They showed that plasma exosome levels were highest in patients with AECOPD and stable COPD compared with healthy non-smokers. In addition, exosome levels correlated with plasma levels of CRP, soluble TNFR1 and IL-6 (254).

Circulating endothelial microvesicles (EMVs) have also been analysed to evaluate the endothelial damage in COPD patients and several studies have reported that some types of EMVs could be used as potential new biomarkers. Thomashaw et al. reported that CD31+ EMVs, reflecting endothelial apoptosis, were elevated in mild COPD and emphysema. In contrast, CD62E+ EMVs, indicating endothelial activation, were elevated in severe COPD and hyperinflation (250). Takahashi et al. showed that CD144+ (the most specific marker for endothelial cells), CD31+ and CD62E+ EMVs were significantly higher in patients with stable COPD than in the healthy non-COPD subjects (260). In addition, Lacedonia et al. reported a negative correlation between the number of EMVs in the sputum and FEV1 (247).

As well as overall levels of circulating EVs used in distinguishing disease, EV miRNA cargo can also be used as a potential biomarker. Furthermore, various miRNAs have been implicated in the development and progression of COPD (see section 1.4.3) (261). To date, studies in sarcoidosis and asthma have shown the potential of EV miRNA as biomarkers of disease. Levanen et al. demonstrated that BALF-derived exosomal miRNA are different in asthmatic patients compared with healthy volunteers. The effect of these exosomes to promote leukotriene and IL-8 release from bronchial epithelial cells was significantly increased in asthmatic patients, giving an additional functional role to this potential biomarker (262). In a recent study, the profile of exosomal miRNA in BALF and serum of patients with sarcoidosis has been investigated. In this

study, exosomal miR-21, miR-150 and miR-146a expression was increased in patients with chest X-ray (CXR) stage II sarcoid compared with CXR stage I disease (263).

1.7.4.2 Extracellular vesicles as a potential therapy in COPD

The utility of EVs as a potential for therapeutic intervention is currently under intense investigation. One of the considerations for developing an EV therapeutic is the cellular source. Mesenchymal stem cells (MSC), known for their anti-inflammatory and regenerative properties have been trialled as possible therapies for COPD (264, 265). However there is still no evidence for clinically relevant effects and further study of MSC-base treatments for COPD patients is needed (266). EVs derived from MSC have recently been proposed as having the potential for tissue repair, wound healing and lung tissue regeneration (267, 268). Furthermore, MSC-EVs and dendritic cell EVs are already being studied in on-going phase 1 and 2 trials for cancer and graft versus host disease in leukaemia (269-272). Issues such as culture conditions, yield and manufacturability have been addressed by research groups who have reported using EVs derived from plant-based systems, as they are able to be loaded with smaller cargo, which can mediate a therapeutic effect in animal models (273, 274).

Importantly, with reference to EV miRNA use in therapy, Zhang et al. used a novel protocol to manipulate EV miRNAs with high efficiency and deliver them to recipient cells (275). However, in order for EV-mediated miRNA delivery to be used as therapy, several issues still need to be addressed. For example, studies have reported that the load of miRNA from tumour cells are low in individual EVs from patient plasma, and that host-derived EVs are unable to deliver sufficient copies of desired miRNAs (276). In addition, EV-mediated miRNA therapy may need to be delivered at regular intervals as the effects are likely to be short lasting. However, emerging technology has seen the development of EVs containing adeno-associated vectors (AAVs), which can provide sustained transgene expression *in vivo* and may overcome some of the challenges seen with the EV-miRNA or AAVs not contained within an EV system (277).

1.7.5 Extracellular vesicle isolation methods

A major challenge in the field of EV research is to improve and standardise the methods for EV isolation and analysis. The International Society for Extracellular Vesicles (ISEV) has attempted to address this in their 2018 statement on the minimal information required for studies of EV sample collection, isolation and analysis (278), which represents the majority viewpoint of over 70 experts in the field.

1.7.5.1 Ultracentrifugation

Differential ultracentrifugation (with or without density gradient separation) is the method most often reported in the literature (226, 279). Sequential centrifugation at higher speeds separates out larger apoptotic bodies from smaller exosomes, the latter usually pellet at 100,000-120,000 g. This process is time consuming, (up to 48 hours to complete purification), requires a large sample volume and may lead to low yield of desired EV population. In addition, there is a suggestion that the centrifugal forces may cause vesicle aggregation (affecting separation) and physical damage leading to impaired structure and function (280, 281).

1.7.5.2 Size exclusion chromatography

Size exclusion chromatography (SEC) separates EVs based on size and passage through a suspension, whereby different particles move at different rates (282). Used alone, this method often results in co-elution of multiple EV classes and therefore can be combined with filtration to remove larger cell fragments and/or EVs (e.g. apoptotic bodies and MVs) to ensure enrichment of desired EV population (279, 283). SEC is ideally performed under gravity to prevent EVs being deformed by forcing particles through filter pores. Studies have shown the resultant EV fraction has minimal soluble protein contamination (282, 284, 285).

1.7.5.3 Filtration

The currently available commercial membrane filters (e.g. ExoMir™ kit from Bioo Scientific, Austin, USA) have pores of various diameters with a narrow range of pore size distribution, which simplifies isolation of the particles with a specified size. When isolating EVs by filtration, larger particles are removed first (by filters with pore diameters 0.8 and 0.45 μm) and the particles with a size smaller than the target EVs are separated from the filtrate at the next stage (by filters with pore diameters 0.22 and 0.1 μm). This method is simple and requires no additional equipment, however EVs may bind to the membrane even when using materials with low affinity for proteins resulting in lower yield (286). Furthermore, the pressure that is used to “push” the specimen through the membrane may result in contamination or deformation of the desired EV population (286, 287). However, this method has been used to successfully isolate EVs from cell culture media (288), serum and BALF (289).

1.7.5.4 Polymeric precipitation

Polymeric precipitation kits are commercially available (e.g. ExoQuick™ from System Biosciences®, Cambridge, UK) and typically use polyethylene glycol to precipitate EVs. This method is quick, technically easy to reproduce and result in typically high yields of EVs (290). However, there are

concerns around the purity of the resultant EV fraction with protein and extra-vesicular RNA contamination being common. In addition, polyethylene glycol may affect downstream analysis, such as mass spectrometry platforms (291).

1.7.5.5 Immunoaffinity isolation

Antibodies to characteristic surface markers present on certain EV classes (e.g. tetraspanins CD9, CD63 and CD81 present on exosome surface) can be used to select desired EV populations (immune-enrichment) or exclude unwanted EV fractions (immune-depletion) (279, 292). These antibodies are typically bound to beads, facilitating separation by low-speed centrifugation or magnetic techniques (293, 294). This selectivity allows high specificity, however concomitantly results in lower yields and will exclude subpopulations of EVs, given that some markers are not represented on all vesicles within a given class (279).

1.7.5.6 Membrane affinity isolation

More recently, a new membrane affinity spin column method for the isolation of highly pure EVs from biofluids was released (exoEasy kit from Qiagen®) (295). Briefly, prefiltered biofluid (e.g. plasma) is mixed with a binding buffer and added to the exoEasy membrane affinity column to bind the EVs to the membrane. After centrifugation, the flow-through is discarded and a wash buffer is added to the column to wash off non-specifically retained material. After another centrifugation and discarding of the flow-through, the EVs are then lysed by adding QIAzol to the spin column, and the lysate is collected by centrifugation. This new procedure is reported to capture nearly 100% of mRNA from plasma samples and is equal to or better than ultracentrifugation in mRNA yield (295). The method also allows for intact vesicles to be eluted from the column material for further characterisation.

1.7.6 Characterisation of extracellular vesicles

Given the complexities in EV isolation, detailed characterisation of the resultant EV fraction of interest is recommended by 2 or more techniques (291). These include electron microscopy (EM) to visualise the size and structure (226); nanoparticle tracking analysis (NTA), a commercial method to obtain size and concentration of EVs (296, 297); flow cytometry, for characterisation of larger EVs (>500nm) (298); and conventional western blotting or enzyme linked immunosorbent assay (ELISA) to demonstrate presence of characteristic protein markers (e.g. tetraspanins CD9, CD63 and CD81) (279, 299).

In summary, EVs are recognised as an emerging novel intercellular communication tool in numerous physiological and pathological processes. Investigating the role of EVs is an emerging

and rapidly progressing area of research, especially in lung disease. The above section highlights that EVs may play a pivotal role in COPD pathogenesis and further work could contribute to the understanding of the disease pathogenesis as well as the development of novel therapies. Of particular importance is the role of EVs as carriers of miRNA, which may provide novel insights into the persistent inflammation and defective innate immunity described in COPD.

1.8 MicroRNA

Protein-coding genes only represent about 2% of the human genome (300); the rest are transcribed into non-coding RNA, of which the most widely studied is miRNA. MiRNAs are small RNA molecules (approximately 21-25 nucleotides in length) that negatively regulate gene expression post-transcriptionally, by degrading mRNA or by blocking translation (301). They are predicted to regulate more than 60% of the human genome (302) and are involved in controlling cellular proliferation, apoptosis and differentiation (303). Dysregulated miRNA expression is reported to be involved in the pathogenesis of many diseases (304-308), including COPD (309).

1.8.1 MicroRNA biogenesis and function

Briefly, miRNA biogenesis begins with the cleavage of primary miRNA (pri-miRNA) into pre-miRNA in the nucleus, which is mediated by a nuclear RNase III enzyme, Drosha (310). The resulting 70 nucleotide pre-miRNA is then actively transported into the cytoplasm and processed into a 22 nucleotide double-stranded miRNA by a cytoplasmic RNase III enzyme, Dicer (311). One strand of this duplex is degraded, whilst the other functions as mature miRNA, and is incorporated into a RNA-induced silencing complex (RISC) (312). This mature miRNA-RISC complex then binds to mRNA depending on sequence complementarity between the miRNA 5’ region (the seed sequence) and the mRNA 3’ untranslated region (UTR) (the biogenesis and function of miRNA is summarised in Figure 1.2). This matching does not need to be perfect and therefore a single miRNA may regulate several hundred mRNA (313). Moreover, by targeting transcription factors or genes involved in epigenetic regulation, miRNAs can alter expression of hundreds of genes at once. Guo et al. found that miRNA regulation of target mRNA accounted for ~84% of protein repression (314).

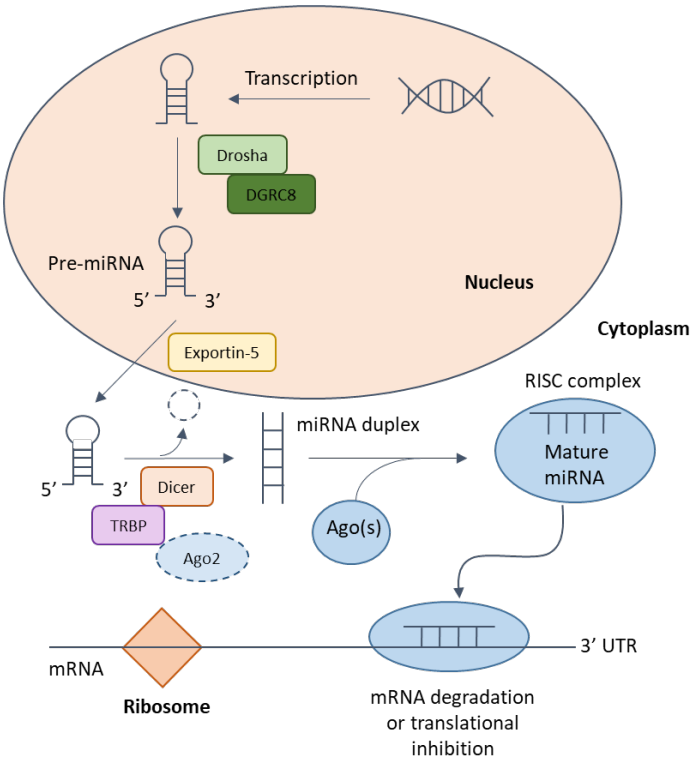


Figure 1.3 MiRNA biogenesis and function.
Ago, Argonaute protein; DGRC8, DiGeorge syndrome critical region gene 8; Pri-miRNA, Primary miRNA; TRBP, trans-activation response RNA-binding protein; RISC: RNA-induced silencing complex; UTR: untranslated region. Adapted from Rupani et al., 2013 (315).

1.8.2 MicroRNA in COPD

As discussed in section 1.1.3, cigarette smoke exposure is the most important risk factor for the development of COPD and has been shown to dysregulate miRNA expression. Schembri et al. found that miR-218 was down-regulated in smokers relative to never smokers in human bronchial epithelium (316). In addition, 34 miRNA were found to be differentially expressed in induced sputum from never-smokers compared with current smokers (309). It is known that changes in gene expression were found to persist in the bronchial epithelium of ex-smokers years after smoking cessation (317). In line with this, cigarette smoke induced changes in miRNA expression in murine lung are irreversible and depend on the duration and dose of smoke exposure (318). Taken together, these findings support the theory that persistent changes in gene expression found in the lungs of ex-smokers, might be attributable to changes in miRNA expression induced by prolonged (dose dependant) exposure to cigarette smoke.

More in-depth expression profiling of miRNA in COPD subjects compared with non-COPD smokers has shown 70 miRNAs differentially expressed in lung tissue (319). In particular, miR-146a, miR-15b, miR-223 and miR-1274a were up-regulated in COPD samples. In addition, Ezzie et al. demonstrated down-regulation of a number of target genes (including mothers against decapentaplegic homolog (*SMAD)-7* gene, a target of miR-15b) involved in the TGF-β signalling

pathway in COPD. Thus, miRNA regulation of fibrotic pathways could be key to understanding airway remodelling in COPD.

Increased COX-2 protein is one of the hallmarks of chronic inflammation in COPD. Sato et al. reported reduced expression of miR-146a in lung fibroblasts from COPD following stimulation by inflammatory cytokines, IL-1β and TNF-α. They suggested that this decrease in miR-146a resulted in reduced degradation of COX-2 mRNA (a predicted target of miR-146a), and in turn, over-expression of COX-2 protein (320). Furthermore, miR-146a has been implicated in other inflammatory diseases, such as rheumatoid arthritis (321, 322).

Analysis of miRNA expression in induced sputum showed expression of let-7c was reduced in COPD patients compared with smoking and non-smoking controls (309). Moreover, let-7c was shown to be down-regulated in response to cigarette smoke exposure in murine models (318, 323). TNFR2 is a predicted target of let-7c and levels inversely correlated with let-7c in the same COPD patients (309). Importantly, TNFR2 plays a vital role in COPD pathogenesis, demonstrated in a TNFR2 knock-out mouse, which protects against cigarette-induced inflammation and emphysema (324). In addition, let-7c has also been investigated as a therapeutic target in cancer (325).

Several other miRNAs have been implicated in the pathogenesis of COPD and these are summarised in Table 1.3.

Table 1.3 Expression pattern of miRNAs in COPD

miRNA(s)	Expression in COPD	Tissue/cell type	Target Gene(s)	Reference
let-7c	Down-regulated	Induced sputum	TNFR2 (predicted)	(309)
miR-218, miR-146a, miR-125, miR-34b	Down-regulated	Induced sputum		
miR-34c	Down-regulated	Induced sputum		(326)
miR-34c, miR-30e-3p miR-218, let-7c	Down-regulated	Induced sputum		(327)
miR-31, miR135b, miR-148a, miR-155, miR-191	Up-regulated	Lung tissue		(327)
miR-149	Down-regulated			
miR-34c	Down-regulated	Lung tissue	*SERPINE1* (validated)	(328)
miR-34a	Up-regulated	Peripheral lung samples	SIRT1 (validated)	(329)
miR-34a, miR-199a-5p	Up-regulated	Lung tissue	HIF-1α (validated)	(330)
miR-101, miR-144	Up-regulated	Lung tissue	CFTR (validated), MPK-1 (predicted)	(331)
miR-15b	Up-regulated	Lung tissue	*SMAD7* (validated)	(319)
miR-223, miR-1274a, miR-424	Up-regulated	Lung tissue		(319)

miRNA(s)	Expression in COPD	Tissue/cell type	Target Gene(s)	Reference
miR-638	Up-regulated	Lung tissue		(332)
miR-212-5p	Up-regulated	Lung tissue	Genes enriched in the nuclear lumen and transcription initiation process	(333)
miR-28-3p, miR-374a-3p, miR-181d, miR-151a-3p, miR-30a-5p, miR181b-5p, miR-30-2-3p, miR-143-3p, miR-30a-3p, miR-378f, miR-361-5p	Down-regulated			
miR-320d	Up-regulated in response to ICS	Bronchial biopsies	CXCL8	(334)
miR-146a	Down-regulated	Lung fibroblasts	COX-2 (predicted)	(320)
miR-31, miR-155, miR-218, let-7c				
miR-1	Down-regulated	Quadriceps muscle	IGF-1 (validated)	(335)
miR-1, miR-499, miR-206, miR-133	Up-regulated	Plasma		(336)
miR-328, miR-21	Down-regulated	Exhaled breath condensate		(337)
miR-29b, miR-483-5p, miR-152, miR-629, miR-26b, miR-101, miR-106b, miR-532-5p, miR-133b	Down-regulated	Plasma		(338)
miR-223	Up-regulated	*In vitro* (human endothelial cells), *in vivo* (mouse)	HDAC2 (validated)	(339)
miR-20a, miR-28-3p, miR-34c-5p, mir-100	Down-regulated	Serum		(340)
miR-7	Up-regulated			
miR-145-5p, miR-338-3p	Down-regulated	Plasma	IFI30 (predicted); LTK (predicted), TNF2 (predicted)	(341)
miR-132, miR-212	Up-regulated	Cell fraction of BALF	TLR-2, -4, -5, Myc, MYD88, IRAK4, IL-6, BDNF, AAT (predicted)	(342)
miR-1, miR-133, miR-206	Down-regulated	Diaphragmatic muscle	HDAC4, Med2c (validated)	(343)
miR-21	Up-regulated	Serum		(344)
miR-181	Down-regulated			
miR-34a-5p, miR-374a-5p	Down-regulated in COPD exposed to biomass smoke	Serum		(345)
miR-150-5p, miR-191-5p, miR-223-3p	Up-regulated in COPD exposed to biomass smoke			

AAT, Alpha-1 antitrypsin; BDNF, Brain-derived neurotrophic factor; CFTR, Cystic fibrosis transmembrane conductance regulator; COX2, Cycloxygenase-2; HDAC, Histone deacetylase; HIF, Hypoxia-inducible factor; IFI30, Gamma-interferon-inducible lysosomal thiol reductase; IGF, Insulin growth factor; IL, Interleukin; IRAK4, Interleukin-1 receptor-associated kinase 4; LTK, Leucocyte receptor tyrosine kinase; MPK-1, Mitogen-activated protein kinase-1; MYD88, Myeloid differentiation primary response 88; SERPINE1, Serpin family E member 1; SIRT1, Sirtuin 1; SMAD7, Mothers against decapentaplegic homolog 7; TLR, Toll-like receptor; TNF, Tumour necrosis factor; TNFR2, Tumour necrosis factor receptor 2.

As shown in Table 1.3, much of the work has focused on miRNA as a potential biomarker in COPD. In addition, miRNAs have potential as novel therapeutic targets, for example by either miRNA replacement therapy using miRNA mimics or inhibition of miRNA function by antagomiRs (346). The application of miRNA as therapy in COPD however faces several challenges, including mode of delivery, stability of miRNA in biofluids and tissue specificity. In the context of lung disease, aerosolisation is a strategy for enhancing drug delivery and reducing side effects. However, free miRNAs are rapidly degraded by nucleases present in extracellular fluids. EVs may provide a solution to this, by encapsulating the miRNAs in a protective package and allowing tissue specific delivery (198, 347)

1.9 Hypothesis and aims

COPD is characterised by chronic airway inflammation and defective innate immunity (5), which is dependant upon a complex network of intercellular communication between the damaged airway epithelium and the immune system (93). This complexity and the disease heterogeneity has limited our understanding of the pathogenesis of COPD and therefore hindered the development of novel and effective disease modifying therapies.

As outlined in section 1.7, EVs are a unique method of intercellular communication (199); they are released in response to tissue damage (236, 237) and can alter the phenotype of recipient cells (348). In addition, miRNA are powerful regulators of post-transcriptional gene expression and can have far reaching effects on downstream cellular function (313, 314). So far, no studies have directly sampled and characterised EV miRNA from the lungs of COPD patients and investigated their impact on COPD pathogenesis. Therefore my study focus and hypothesis is:

MicroRNA is differentially expressed in extracellular vesicles in the airways of patients with COPD, and leads to differential gene expression, which drives chronic inflammation in COPD.

Overall study aims:

1. To isolate EVs from the BALF of COPD subjects and healthy ex-smokers.
2. To identify differentially expressed miRNA in lung-derived EVs in COPD subjects compared with healthy ex-smokers.
3. To identify the biologically significant targets of these differentially expressed miRNA in the airway epithelium
4. To investigate the diagnostic use of the lung-derived EV miRNA and explore their relationship with specific COPD inflammatory endotypes.

Chapter 2 **Methods**

This section will detail the methods for patient sample collection, extracellular vesicle (EV) isolation and characterisation, RNA extraction and library preparation, microRNA (miRNA) and smallRNA sequencing, bioinformatics analysis of sequencing results and the validation steps using real time quantitative polymerase chain reaction (RT-qPCR). It will also cover the methods used to explore the miRNA – messenger RNA (mRNA) interactions to identify those most relevant to the pathophysiology of Chronic Obstructive Pulmonary Disease (COPD).

Due to the extensive and advanced methodology used in this thesis, some of the methods were performed by external contributors. The following methods were performed by the author, Dr Hannah Burke: patient and healthy volunteer recruitment to the MICA II study (outlined in section 2.2), bronchoscopy including epithelial brushings and bronchoalveolar lavage, EV isolation and characterisation, and RNA isolation, cDNA synthesis and RT-qPCR for optomisation of EV methods. Furthermore, all the analysis of the miRNA sequencing data including unsupervised filtering, exploratory data analysis, and differential expression analysis was peformed by the author. Analysis of the epithelial brushing transcriptome was also performed by Dr Hannah Burke as well as the miRNA-mRNA interaction analysis and gene ontology enrichment analysis. Dr Hannah Burke performed all of the statistical analysis included in the thesis. The wider MICA II research team helped with the subject phenotyping including the physiological measurements, high-resolution computer tomography (HRCT) analysis and BAL inflammatory cell counts. Qiagen® Genomic services performed the final EV RNA isolation, library preparation and next generation miRNA sequencing protocols, including mapping and alignment to the reference genome. All other quality control measures of the miRNA sequencing data was performed by the author. RNA isolation and next generation mRNA sequencing of the epithelial brushings was performed by the Translational Science and Experimental Medicine team at AstraZeneca. The mRNA sequencing data preparation was performed by the Bioinformatics team at AstraZeneca, with Dr Hannah Burke performing the final differential gene expression analysis.

2.1 Ethics

All subjects gave written informed consent for the study RHM MED1277 Microbiology and Immunology of the Chronically-Inflamed Airway (MICA) II. This study was approved by the South Central - Oxford C Research Ethics Committee (15/SC/0528).

2.2 MICA II study design

MICA II is a longitudinal cohort study in which patients with mild-moderate COPD, healthy ex-smokers and healthy non-smokers underwent full lung function, blood and sputum sampling, high-resolution computer tomography (HRCT) and bronchoscopy. The study principle outcomes were to investigate the microbiology, inflammation and immunology within the airways of COPD patients. Detailed inclusion and exclusion criteria are shown in Table 2.1.

Table 2.1 Inclusion and exclusion criteria for the MICA II study

Inclusion Criteria	Exclusion Criteria
Written informed consent obtained from the participant	A confirmed diagnosis of asthma, cystic fibrosis, pneumonia risk factors or other respiratory disorders (e.g., tuberculosis, lung cancer).
Patients with COPD and healthy ex-smokers with a history of ≥10 pack-years of cigarette smoking	History of lung surgery.
Male or female aged 40–85 years	AAT deficiency as underlying cause of COPD.
COPD subjects must have a confirmed diagnosis of mild/moderate COPD based on post bronchodilator spirometry with FEV1 >50% of predicted normal and FEV1/FVC <0.7.	Moderate or severe COPD exacerbation not resolved at least 1 month prior to enrolment and less than 30 days following the last dose of oral corticosteroids and/or antibiotics.
Healthy subjects must have an FEV1/FVC >0.7.	Long-term corticosteroid or antibiotic therapy. Use of any antibacterial, antiviral or respiratory investigational drug or vaccine within 30 days of the enrolment visit
Subjects must be fit to undergo bronchoscopy	Presence of other conditions that the principal investigator judges may interfere with the study findings.
	Evidence of alcohol or drug abuse.

AAT, Alpha-1 antitrypsin; COPD, Chronic obstructive pulmonary disease; FEV1, Forced expiratory volume in one second; FVC, Forced vital capacity.

2.3 Study population for bronchoalveolar lavage exosomal miRNA analysis

Bronchoalveolar lavage fluid (BALF) samples from 20 subjects with stable mild and moderate COPD and 15 healthy ex-smokers were used from the MICA II cohort for EV isolation, RNA extraction and miRNA sequencing. For COPD patients, post-bronchodilator spirometry was used to assess airflow obstruction with a Forced expiratory volume in one second (FEV1)/Forced vital capacity (FVC) ratio of <0.7 and an FEV1 of ≥50% predicted value required for enrolment.

In addition, samples from a further six COPD subjects and five healthy ex-smokers from the MICA II cohort were included in the validation cohort. BALF samples from these patients underwent EV isolation, RNA extraction and RT-qPCR for direct target miRNA identification.

All of the above patients had epithelial brushing samples processed for mRNA sequencing, however only 14 subjects with COPD and 10 healthy ex-smokers from the study population who had lung EV isolation, had serum processed for EV miRNA analysis due to sample availability. The details of this sub-group will be explored in detail in Chapter 6, section 6.6.1.

2.4 Sampling protocols

2.4.1 Spirometry

Spirometry was performed with a MicroLab™ spirometer (CareFusion®, San Diego, US) in accordance with the American Thoracic Society (ATS)/European Respiratory Society (ERS) guidelines (349), by a trained individual following the standard operating procedure (SOP). Briefly, the subject was seated and a nose clip was applied during the procedure. The patient was instructed to take a deep breath in filling their lungs to total lung capacity (TLC), making a good seal around the mouthpiece and blowing out as hard and as fast as possible, and for as long as possible. This continued until their lungs were completely empty and they were encouraged through the process. The blows were repeated until 3 technically good blows were obtained with the highest two FEV1 and FVC measurements being within 150 mL of each other. Spirometry results recorded were FEV1, FEV1% predicted, FVC, FVC% predicted, FEV1/FVC ratio, FEF 25%–75% and FEF 25%–75% % predicted.

2.4.2 Transfer factor

Gas transfer was performed using a body plethysmograph HDpft™4000 (nSpire™ Healthcare Ltd, Hertford, UK), by a trained member of staff following the SOP. Briefly, tidal breathing was traced to ensure there are no irregularities. The patient was then instructed to exhale slowly to residual volume (RV), at this point the patient was instructed to immediately take a deep and full breath into TLC. The patient was then asked to hold their breath for 10 s until the red sample line reaches the vertical lines at which point they were asked to breathe out with medium force until RV (where the red sample line touches/plateaus with the RV dotted blue line). Four min were given between tests, allowing adequate washout of the gases from the lungs. Tests were acceptable if complete inspiratory breaths were within 2 s, breath holds were between 8 and 12 s and complete exhalation was within four s (349). A maximum of 5 tests were performed and two diffusion capacity of the lung for carbon monoxide (DLCO) results needed to be within 10% of each other or 1 $mmol \cdot min^{-1} \cdot kPa^{-1}$ of each other, and diffusion capacity of the lung of carbon monoxide per unit volume (KCO) measurement within 0.1 $mmol.min^{-1} \times kPa^{-1} \times l^{-1}$.

2.4.3 Blood serum preparation

Blood samples were collected in 2 x 9 mL plain (no anticoagulant additive) tubes and placed upright for 30-60 min to allow a clot to form. Tubes were then centrifuged at 1500 g, room temperature for 15 min to pellet clotted cells. The supernatant was aspirated into fresh tubes and underwent further centrifugation at 1500 g, room temperature for 15 min to remove any residual cells and cellular debris. Purified serum was stored at -80^{o}C for future EV isolation.

2.4.4 High resolution computer tomography

HRCT scanning was undertaken on a Siemens Sensation 64 CT scanner (Siemens Medical Solutions®, Erlangen, Germany). The scans comprised of a helical scan in inspiration and one in held expiration. During the scan each subject was given strict instructions about position, which consisted of having the arms raised above their heads. The imaging protocol consisted of; slice thickness 0.75 mm, slice separation 0.5 mm, tube voltage 120 KV, effective milliamps 90mAs (using dose modulation), collimation 0.6 mm and a pitch of 1.

The scan data was anonymised to a study number on the CT scanner. All CT images were reviewed by a thoracic radiologist to determine if there were any clinically important abnormalities, which were then managed as per local clinical practice. Emphysema was quantified by the percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units ($\%LAA_{<-950}$) (350, 351). A surrogate marker for small airways disease was measured using the ratio of mean lung attenuation on expiratory and inspiratory scans (E/I MLD), which has previously been validated (352). Quantitative analysis of lung attenuation values derived from the CT images were used to determine the amount of small airways disease and emphysema. This was performed using Apollo™ pulmonary analysis software version 2.0 (VIDA Diagnostics®, Iowa, USA), and used to assess regional variability of disease in COPD and guide bronchoscopic sampling from disease and less-diseased lobes.

2.4.5 Fibreoptic bronchoscopy

All subjects underwent a fibreoptic bronchoscopy, performed on an outpatient basis. Subjects were asked not to eat and drink four hours prior to the procedure. All subjects were given 2.5 mg of nebulised salbutamol and underwent spirometry. A combination of intravenous alfentynyl (100-500 µg) and midazolam (1-10 mg) were given as sedation. Lignocaine spray and gel were used as local anaesthetic to the nose and pharynx. Ten millilitres of 1% lignocaine was used as local anaesthetic for the larynx and bronchial tree and were given via the bronchoscope channel.

The fibreoptic bronchoscope was passed either through the nose or mouth and two lobes were sampled in all subjects.

Two unprotected bronchial epithelial brushes were taken for RNA isolation and sequencing from the same lobe location. BAL was performed by advancing and wedging the bronchoscope into a segmental bronchus. One hundred millilitres of pre-warmed 0.9% sodium chloride in 20 mL aliquots were introduced into each lobe and recovered by gentle aspiration. Samples were processed within two hours of the procedure. Finally, two bronchial biopsies were taken from each lobe for RNA isolation and sequencing. All samples were processed within 2 hours of the procedure.

2.4.5.1 Epithelial brushings for RNA isolation and sequencing

Two unprotected brushes from the same lobe were collected in 5 mL of 1X Phosphate-buffered saline (PBS) and centrifuged at 400 g, room temperature for 5 min. The pellet was re-suspended in 600 µL of RNAprotect® Cell Reagent (Qiagen®, Manchester, UK) and incubated at 4^{o}C overnight and then stored at -80^{o}C for RNA isolation and sequencing.

2.4.5.2 Bronchoalveolar lavage fluid analysis

BAL fluid was poured through 100 µm cell strainer to remove mucus and cells were removed by centrifugation at 400 g, 4^{o}C for 10 min. The cell-free supernatant was stored at -80^{o}C prior to EV isolation, while the resulting cell pellet was resuspended in 10 mL hypotonic lysis buffer for 2 min to remove any red blood cell contamination. Ten millilitres of hypertonic recovery was then added with 1X PBS to make-up the volume to 40 mL. The sample was centrifuged at 400 *g* for 10 min. The pellet was resuspended in 1 mL 1X PBS and cell counts were performed using the Trypan blue exclusion method. The cell solution was then adjusted to 0.5 x 10^{6} cells/mL, and 75 µL of this solution was loaded onto cytospin funnels. Cells were centrifuged at 350 *g* for 6 min and collected on Poly-L-Lysine slides. Any leftover cell solution in the polypropylene tube was spun at 400 *g,* 4^{o}C for 10 min, and the resultant pellet was resuspended in 500 µL of TRIzol® (ThermoFisher Scientific®, Basingstoke, UK) and stored at -80^{o}C.

2.4.5.3 Bronchoalveolar lavage cell counts

Cell slides were left to air-dry overnight, and stained the next day using Rapid Romanowsky A-B-C kit (TCS Biosciences, Buckingham, UK). A differential cell count was performed by counting 500 cells using light microscopy.at X40 magnification. Numbers and relative percentages of eosinophils, neutrophils, macrophages, lymphocytes, bronchial epithelial cells and squamous cells were calculated.

2.5 Extracellular vesicle isolation

For the purposes of this study, EVs refers to an **"enriched" exosome population**. Given it is currently technically challenging to obtain a pure exosome fraction (i.e. free from MVs and other non-vesicular components), it is more widely accepted (and scientifically correct) to refer to an isolated vesicle fraction as EVs rather than exosomes (353).

After trialling both polymeric precipitation and ultracentrifugation methods I chose to proceed with a combination of ultrafiltration and size exclusion chromatography (SEC) to isolate EVs from BALF. Although polymeric precipitation was an established technique in my research laboratory, I had concerns over the purity of the resultant EV sample based on the size of the EV pellet isolated (very large and difficult to resuspend) and the transmission electron microscopy (TEM) images, which demonstrated large amounts of debris (e.g. soluble protein, cell fragments) surrounding the vesicles. Furthermore, I also trialled ultracentrifugation (the most widely used technique in the field), but due to the low initial concentration of EVs in BALF, my sample yield was too low for adequate RNA isolation. I therefore proceeded with SEC isolation of EVs, using ultrafiltration as a method of purifying and concentrating the BALF supernatant to prior to this technique.

2.5.1 Ultrafiltration for sample purification and concentration

BALF samples were purified and concentrated using Amicon® Ultra centrifugal filters (Merck Millipore®, Watford, UK) using an adapted manufacturer's protocol. Briefly, the 15 mL BALF was filtered through a 0.22 µm PVDF, 33 mm gamma sterilised filter (Merck Millipore®), to remove any larger particles (e.g. apoptotic bodies). The sample was then loaded onto an equilibrated Amicon® Ultra-15 (10,000 MWCO) spin filter and centrifuged at 4000 g, 4°C for 15 min. The filter device was washed with 14 mL of 1X PBS and centrifuged at 4000 g, 4°C for a further 15 min. The EV containing sample was recovered from the filter device using a Gilson pipette, by pipetting up and down vigorously to ensure all EVs recovered.

2.5.2 Size exclusion chromatography

Separation of EV based on size and passage through physical filters can be achieved through the use of SEC. In addition, SEC is considered an effective method for isolating EVs from soluble proteins and is not thought to affect the original shape or functionality of the vesicles (Figure 2.1). I trialled two different commercial SEC columns for EV isolation: PURE-EVs™ columns (HansaBioMed Life-Sciences®, Tallinn, Estonia) and Exo-spin™ Midi-Columns (CELL guidance systems®, Cambridge, UK).

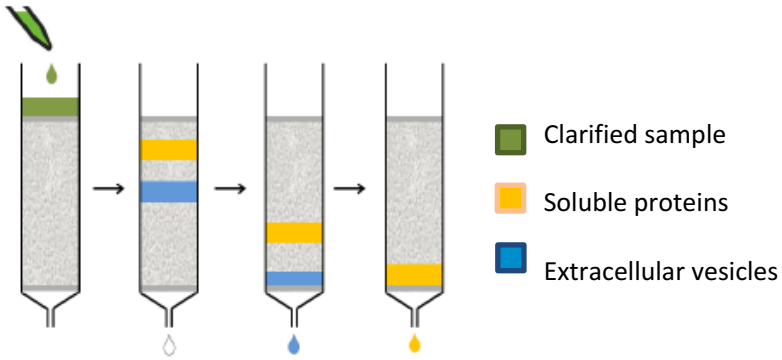


Figure 2.1 Size exclusion chromatography platform separating soluble proteins (yellow) from the extracellular vesicles (blue)

2.5.2.1 Size exclusion chromatography using PURE-EV™ columns

Columns were prepared by allowing the preservative buffer to flow through and washed with three 10 mL volumes of 1X PBS to eliminate any preservative buffer residues. Up to 2 mL of the ultrafiltrate generated in section 2.5.1 was then placed in the column and allowed to pass through under gravity. The column was topped up with 1X PBS regularly to ensure the column did not dry out. 24 x 500 µL fractions were collected, and once satisfied with the purity of the EV containing fractions, fractions 1-5, 6-11, 14-17, 18-24 were collated. After all fractions were collected the column was washed with up to 30 mL of 1X PBS to remove any sample residue. Columns were stored at 4°C and only reused for the same sample to ensure no cross contamination. The pore size of the agarose beads within the resin is between 40-60 nm to ensure an exosome-enriched population of EVs is eluted.

If the EV fraction (fractions 6-11) was to be used for RNA isolation, then the sample was concentrated using an equilibrated Amicon® Ultra-4 (10,000 MWCO) spin filter, centrifuged at 4000 g, 4°C for 25 min. The EV containing sample was recovered from the filter device using a Gilson pipette, by pipetting up and down vigorously to ensure all EVs recovered.

2.5.2.2 Size exclusion chromatography using Exo-spin™ Midi-columns

Columns were prepared by discarding the preservative buffer and equilibrated using two 10 mL volumes of 1X PBS. When no buffer remained on the surface of the column, up to 1 mL of the ultrafiltrate generated in section 2.5.1 was added to the column and allowed to pass through under gravity. The column was topped up with 1X PBS regularly to ensure the column did not dry out. Twenty-four fractions of 500 µL were collected ready for downstream analysis. According to the manufacturer's protocol, the majority of EVs eluted between fractions 7 and 12. The pore size

within the resin is approximatively 30 nm to attain a highly pure exosome elution. Columns could not be reused and therefore were discarded after use.

2.5.3 Measurement of protein concentration of SEC-derived EV samples

To compare the performance of these two commercial SEC columns, I analysed the protein concentration of the 24 fractions from both columns using the Pierce BCA Protein Assay kit (ThermoFisher Scientific®) according to the manufacturer's instructions. Briefly, diluted bovine serum albumin (BSA) standards were prepared according to the protocol dilution scheme. Ten microlitres of each standard or unknown sample replicate was added to an appropriately labelled well of a 96-well, flat-bottomed plate (Nunclon®, Sigma-Aldrich Company Ltd, Gillingham, UK). Two hundred microlitres of prepared working reagent was added to each well to achieve a sample to working reagent ratio of 1:20. The plate was covered and incubated at 37°C for 1 h. The plate was cooled on ice and absorbance measured using a Softmax® microtitre plate spectrophometer at 550 nm. The average 550 nm absorbance measurement of the Blank standard replicates was subtracted from the 550 nm measurement of all the other individual standard and unknown sample replicates. A standard curve was plotted using the Blank-corrected 550 nm measurement for each BSA standard versus its concentration in μg/mL. This was then used to determine the concentration of each unknown sample.

In addition, protein concentrations were also determined by NanoDrop 1000 (ThermoFisher Scientific®), using the A_{280} protein method for both the BSA protein standards and EV samples. A standard curve was generated from the BSA standards and was used to determine the concentration of each unknown sample.

2.5.3.1 Protein concentration of size exclusion chromatography fractions

To compare the performance of these two commercial SEC columns, I analysed the protein concentration of the 24 fractions from both columns using the Pierce BCA Protein Assay kit (ThermoFisher Scientific®). The majority of the protein was eluted in fractions 13-24 in both columns, demonstrating the separation of soluble proteins from EVs. However, the protein concentrations were more consistent and reproducible using the PureEV™ columns, which also showed a small rise in protein concentration in fractions 6-11, corresponding to the EV-dominant fractions (Figure 2.2-B). Based on this and the reproducibility of the protein concentrations, all proceeding EV isolation from BALF was performed using the PureEV™ columns.

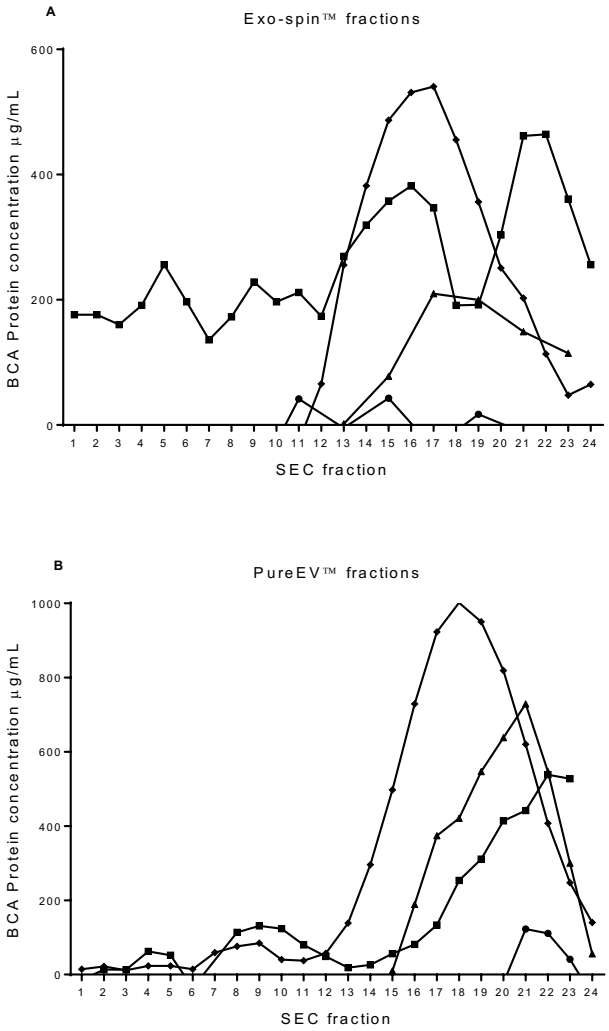


Figure 2.2 Protein concentration measured by BCA assay of the SEC fractions isolated from both Exo-spin™ (A) and PureEV™ (B) columns

EV-dominant fraction for Exo-spin™ (A) are SEC fractions 7-12 and for PureEV™ (B) are SEC fractions 6-11. Individual data points plotted plus connecting line. Each line represents a different BALF sample, N=4.

2.5.4 Summary of EV isolation from BALF

To my knowledge, this is the first study demonstrating EV isolation from BALF using SEC. All other published studies have used ultracentrifugation as their method of choice (232, 233, 354-360). Of those, the studies performed on human samples are more than seven years old (232, 262, 357, 359) and EV isolation using SEC has only been published in the last six years (284). Interestingly, in a 2016 survey on EV isolation techniques used worldwide, although ultracentrifugation was the most widely used isolation method (81%), it was mainly for isolation of EVs from conditioned cell culture media. While those analysing more complex biofluids tended to use a combination of methods, and in this setting, SEC was used by up to 15% of respondents (similar to density gradient centrifugation and filtration) (361).

As discussed previously (section 1.3.5.1), ultracentrifugation can result in low EV yield; and vesicle aggregation and/or damage affecting structure and function (280, 281). The International Society for Extracellular Vesicles (ISEV) agree there is no "gold standard" for EV isolation, but accept that the technique used will rely on the original sample input volume (for my study 15 mL of BALF) and the subsequent required degree of EV purity and concentration for on-going analysis (291). Although I trialled ultracentrifugation, the EV yield was poor and I could not isolate enough RNA for sequencing purposes. Therefore, I employed the combination of ultrafiltration, to purify and concentrate my input sample, and SEC to separate the EV fraction from the soluble proteins. This latter step was particularly important for my downstream RNA analysis, as extra-vesicular RNA can be bound to soluble protein complexes, such as AGO2 (362), which would have contaminated my final EV RNA sample.

Two commercial SEC platforms were available for EV isolation (PureEV™ and ExoSpin™), however these were not validated for use in BALF samples, with the only prior studies being done in urine and plasma (282, 284, 285). Therefore, I initially confirmed separation of the EV fraction from the soluble protein outlined in sections 2.5.3.1. These analyses confirmed the EV dominant fractions and demonstrated the higher reproducibility of the PureEV™ column. Therefore a combination of ultrafiltration and SEC using the PureEV™ columns was used as the method for subsequent BALF EV isolation including characterisation of these EVs (outlined in section 2.6).

BALF samples from 20 COPD subjects and 15 healthy ex-smokers in the MICA II research cohort were initially processed for EV isolation and RNA sequencing. A further 6 COPD subjects and 5 healthy ex-smokers BALF samples were processed for EV RNA isolation for the RT-qPCR validation study. This gave a total of 46 BALF samples processed (26 COPD and 20 healthy ex-smoker samples) for EV isolation and miRNA quantification. A pragmatic sample size for this study was driven by access to patient bronchoscopy samples and adequate volume of BALF available for EV isolation. Furthermore, comparison with previous studies investigating EV miRNA in BALF in other inflammatory lung diseases (262, 358, 360) suggest that this sample size would be adequate to detect a significant difference in EV miRNA signatures between COPD and healthy controls.

2.5.5 EV isolation from serum by filtration using ExoMir™ kit

The ExoMir™ kit uses filters of different sizes to capture EVs and larger membrane-bound particles (*e.g.* apoptotic bodies and microvesicles). Filters are then flushed with an RNA extraction reagent to lyse captured particles and release their contents. This method was chosen due to the previous experience of the laboratory group in using this method and the abundance of EV RNA recovered from serum using this method.

The kit was used according to the manufacturer's instructions. Briefly, 100 µL of Proteinase K was added to 4 mL of serum and incubated at 37°C for 30 min, to minimise non-specific signal and prevent filter clogging. The sample was loaded into a 10 mL syringe, which was pre-connected to a filter stack. The top filter had a pore size of ~200 nm to capture the larger particles and the bottom filter had a smaller pore size (~20 nm) to capture the EVs. The plunger was gently depressed at a rate of 2-3 drops per second. After sample filtration, residual fluid was completely removed from the filter disks using an air filled syringe. Each filter was then flushed with 1 mL BiooPure™-MP (a single-phase RNA extraction reagent containing guanidinium thiocyanate and phenol) to lyse captured particles.

2.6 Extracellular vesicle characterisation

2.6.1 Quantification of BALF derived EVs using enzyme-linked immunosorbent assays

To quantify the BALF EVs isolated using the PureEV™ columns and ensure these were present in the correct fractions; I used two types of ELISA: a CD63 direct ELISA and a CD9 double sandwich ELISA I performed these experiments on **combined** SEC fractions (see Table 2.2) based on the manufacturer's instructions and the results from the protein concentration experiments (section 2.5.3.1).

Table 2.2 Summary of combined SEC fractions for PureEV™ columns

Combined SEC fraction	SEC fractions included	Total final volume (mL)
SEC # 1	1-5	2.5
SEC # 2*	**6-11***	3
SEC # 3	12-17	3
SEC # 4	18-24	3.5

*Corresponds to the EV-dominant fraction SEC, Size exclusion chromatography; #, fraction.

2.6.1.1 CD63 enzyme-linked immunosorbent assay

CD63 is one of the tetraspanins found on the surface of EVs and can be used to confirm the presence of EVs, but also quantified to determine relative abundance. EV abundance was quantified in the SEC fractions by direct ELISA for CD63, using the ExoELISA-ULTRA™ CD63 assay (System Biosciences®, Cambridge, UK). EV samples required pelleting and resuspending in Coating Buffer, therefore after SEC they were prepared using ExoQuick-TC™ kit (System Biosciences®). Briefly, ExoQuick-TC™ was added to the collated SEC fractions in a 1:5 ratio (i.e. 500 µL of ExoQuick-TC™ to 2500 µL SEC fraction) and incubated overnight (12-16 h) at 4°C. The EVs were pelleted by centrifugation twice at 1500 g, 4°C for 30 min. The EV pellet was resuspended in 120 µL Coating Buffer.

Diluted ExoELISA-ULTRA™ protein standards were prepared according to the manufacturer's dilution scheme. 50 µL of each standard or EV sample replicate was bound to a high protein binding microtitre plate (supplied by System Biosciences®). The wells were incubated with an anti-CD63 primary antibody, and a horseradish peroxidase enzyme linked secondary antibody was used for signal amplification. A colorimetric substrate (extra-sensitive TMB) was then used for the assay read out. The results were quantitated using a Softmax® microtitre plate spectrophometer at 450 nm absorbance.

The average 450 nm absorbance measurement of the Blank standard replicates was subtracted from the 450 nm measurement of all the other individual standard and unknown sample replicates. A standard curve was plotted using the Blank-corrected 450 nm measurement for each ExoELISA-ULTRA™ standard versus its concentration in µg/mL. This was then used to determine the EV abundance of each unknown sample.

The results of this ELISA are shown in Figure 2.3, where CD63 was not present in any of the combined fractions, apart from in combined SEC # 4 in one sample.

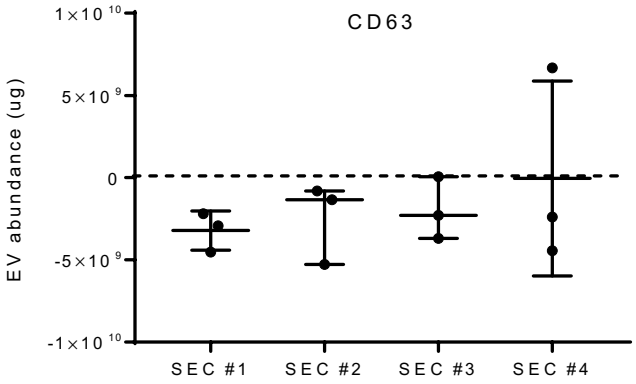


Figure 2.3 EV abundance according to presence of CD63, measured by direct ELISA in combined SEC # 1-4 from PureEV™ columns
Median and 95% confidence intervals shown. Dotted line shows zero level of EV abundance. BALF samples, N=3.

2.6.1.2 CD9 enzyme-linked immunosorbent assay

EV populations have different expression levels of the tetraspanins, and given I had poor results using the CD63 ELISA, I used an alternative, double sandwich ELISA for CD9 (ExoTest™, HansaBioMed® Life-Sciences) which could be performed directly on the combined SEC fractions.

Diluted lyophilized exosome standards were prepared according to the manufacturer's dilution scheme. One hundred microlitres of each standard or EV sample replicate was added to a pre-coated ELISA plate with primary antibodies against CD9. The sample was incubated overnight (12-20 h) at 4°C. The wells were incubated with an anti-CD9 mouse primary antibody, and a

horseradish peroxidase enzyme conjugated anti-mouse Ig secondary antibody was used for signal amplification. A colorimetric substrate (100 µL per well of the Substrate Chromogenic solution provided by the manufacturer HansaBioMed® Life-Sciences) was then used for the assay read out. The results were quantitated using a Softmax® microtitre plate spectrophotometer at 450 nm and 550 nm absorbance.

The 550 nm absorbance measurement was subtracted from the 450 nm measurement for all the samples. The average 450 - 550 nm absorbance measurement of the Blank standard replicates was subtracted from the 450 - 550 nm measurement of all the other individual standard and unknown sample replicates. A standard curve was plotted using the corrected 450 - 550 nm measurement for each ExoTest™ standard versus its concentration in µg/mL and used to determine the EV abundance of each unknown sample.

These results showed significant EV abundance, (quantified by the presence of CD9) in combined SEC # 2 compared with the other SEC fractions (Figure 2.4-A). This is despite a far higher protein concentration in combined SEC #3 and #4 (Figure 2.4-B).

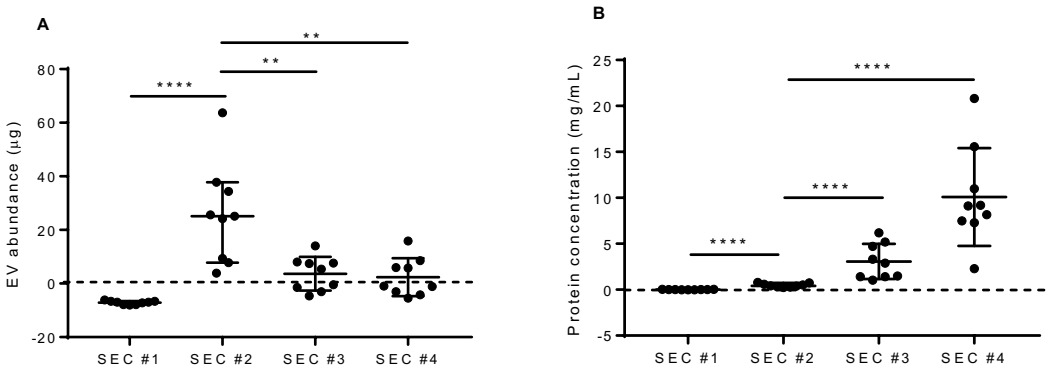


Figure 2.4 EV abundance (A) according to presence of CD9 and protein concentration (B) in combined SEC # 1-4 from PureEV™ columns

Median and 95% confidence intervals shown. Dotted line shows zero level of EV abundance/protein concentration. Mann-Whitney U test, ****p<0.0001, **p<0.01. BALF samples, N=9.

These results confirmed the presence of CD9 positive EVs in combined SEC # 2 and demonstrated the separation of soluble protein from the EV fraction.

It is worth noting that this method of EV quantification will only capture a sub-population of EVs (i.e. the CD9 positive EVs). Overall EV particle number can be measured by light scattering technologies, such as nanoparticle tracking analysis (NTA) (297). This approach could be used to determine if there is a difference in EV number between health and COPD. This may be important as an increase in EV particle number in either health or disease may confer additional miRNA carrying capacity. Therefore, EV particle number may be important when considering normalisation of RNA sequencing results. In this thesis, normalisation was performed in four

ways, firstly the starting volume of BALF was fixed at 15 mL for all samples, the same concentration of final library preparation was used prior to sequencing for each samples (see section 2.8.2), filtered reads underwent normalisation (see section 2.10.3) and miRNA expression data was normalised to stably expressed miRNA across health and disease (see section 2.11.2). There are inherent challenges of normalisation strategies in this field and EV particle number may affect RNA content, however correcting for this variable may be an important signal which is a relevant to the biology between health and disease. Therefore using the same starting volume of 15 mL was considered to be a more biologically relevant strategy in this thesis.

2.6.2 Transmission electron microscopy of BAL derived EVs to visualise characteristic size and shape

TEM techniques are well established and proven useful in EV characterisation, providing direct evidence of the characteristic vesicular structures. In addition, the use of heavy metal stains such as ammonium molybdate in TEM enables visualisation of the bi-lipid layer. Briefly, 10 µL of EVs in 1X PBS were fixed with 1 µL 25% (v/v) glutaraldehyde. Five microlitres of the preparation was layered onto individual formvar-carbon coated 200 mesh copper grids (Agar Scientific Ltd, Stansted, UK) and dried at room temperature. The samples were then contrasted in a solution of 3% ammonium molybdate in 0.1 M ammonium acetate buffer pH 7.0 for 10 s, then blotted and allowed to dry (363). The samples were examined under an electron microscope (Hitachi®, High-Technologies Ltd, High Wickham, UK) with a 16-megapixel side mounted camera (Morada® G3, EMSIS Ltd, Muenster, Germany). The EM images were analysed using RADIUS™ 2.0 software (EMSIS Ltd).

TEM was used to characterise the EVs isolated from BALF using the PureEV™ SEC columns. Figure 2.5 shows the EV sample morphology in combined SEC # 2 (A) compared with combined SEC # 3 (B). In image A, there is little surrounding debris and the characteristic EV cup-shaped morphology and size between 30-150 nm is illustrated. In contrast, image B shows some larger vesicles, which are difficult to distinguish from the surrounding debris, which is likely to comprise largely of soluble protein aggregates.

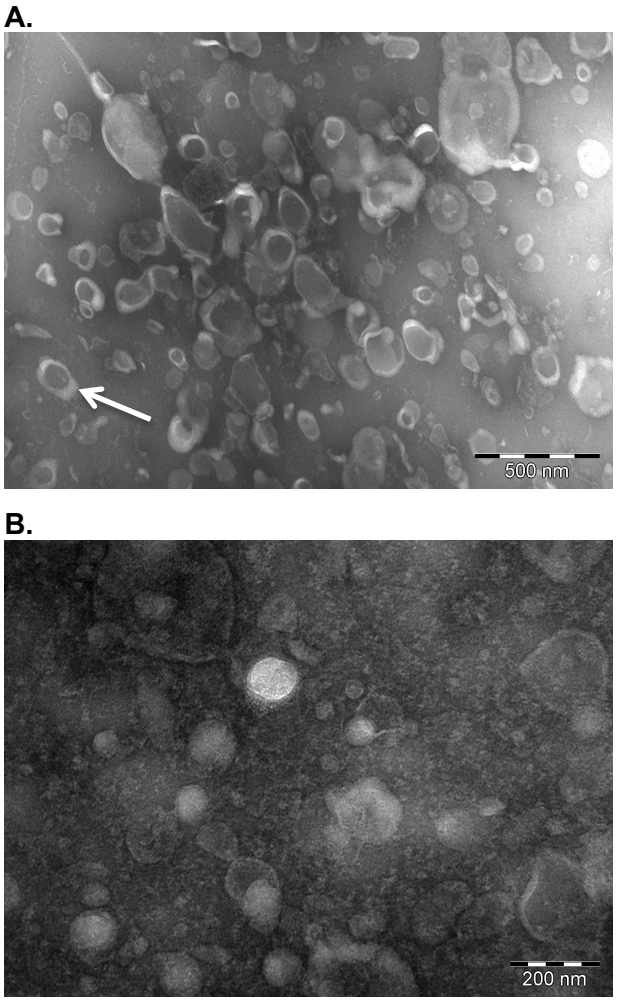


Figure 2.5 Whole mounted BALF-derived EVs isolated using PureEV™ SEC columns viewed by transmission electron microscopy from combined SEC # 2 (A) and # 3 (B)
White arrow points to a characteristic EV with cup-shaped morphology and size between 30 – 150 nm Scale bar shown in bottom right hand corner. BALF, bronchoalveolar lavage fluid; EV, extracellular vesicles; SEC, Size exclusion chromatography

2.6.3 Characterisation of serum derived EVs

Isolation of serum EVs differed from the BALF as it was performed using the ExoMir kit. This method was chosen due to the previous experience of the laboratory group and the abundance of EV RNA recovered from serum using this method. Characterisation of the serum EVs using this method was performed by western blot (section 2.6.3.1) and TEM (section 2.6.3.2).

2.6.3.1 SDS PAGE and Western blotting to determine presence of serum EV surface markers

Western blotting can be used to demonstrate the presence of proteins reportedly associated with EV or EV subgroups. Markers can include tetraspanins (CD9, CD63, CD81), major histocompatibility complex (MHC) molecules and cytosolic proteins such as certain stress proteins

(e.g. Tumour susceptibility gene 101 (Tsg101) and programmed cell death 6 interacting protein (Alix)) or cytoskeletal proteins (e.g. actin, tubulin) (207, 227).

Briefly EVs were lysed and denatured in NuPAGE™ LDS Sample buffer 4X with 12% B-Mercaptoethanol (ThermoFisher Scientific®) at 70°C for 10 min and resolved in NuPAGE™ 4-12% Bis-Tris Protein gels (ThermoFisher Scientific®). Expression of CD63 (anti-rabbit polyclonal antibody, Atlas Antibodies, Bromma, Sweden) and calnexin (anti-rabbit monocolonal antibody, Cell Signaling Technology, Danvers, US) were detected by specific antibodies.

Figure 2.6 demonstrates the presence of CD63 (a tetraspanin, known to be present on EV surface) and the absence of the endoplasmin reticulum protein calnexin from EVs isolated using ExoMir™ method from serum.

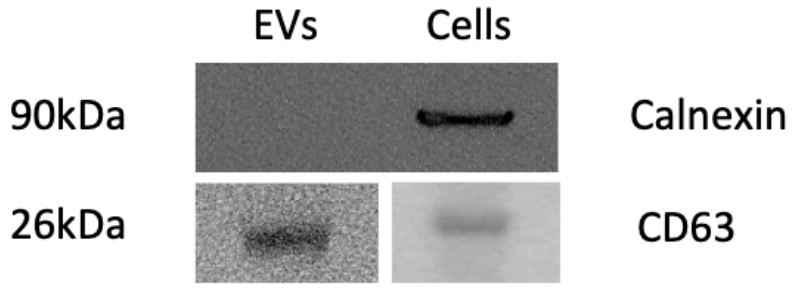


Figure 2.6 Western blot analysis of isolated EVs from serum
30ug of protein was used for Western blot analysis of isolated EVs. Images captured from different WB for CD63 due to issues with running gel non-reducing conditions. Best image shown for both EVs and Cells.

Of note, Western blot analysis was attempted on EVs isolated using SEC from BALF, but the antibodies did not bind and therefore an alternative, more sensitive approach was used with the high-sensitivity ELISAs (see section 2.6.1).

2.6.3.2 Transmission electron microscopy of serum derived EVs to visualise characteristic size and shape

TEM was used to characterise the EVs isolated from serum using the ExoMir™ kits. Figure 2.7 shows the EV sample morphology with bilayer membranes visible, the characteristic EV cup-shaped morphology and size between 30-150 nm is illustrated.

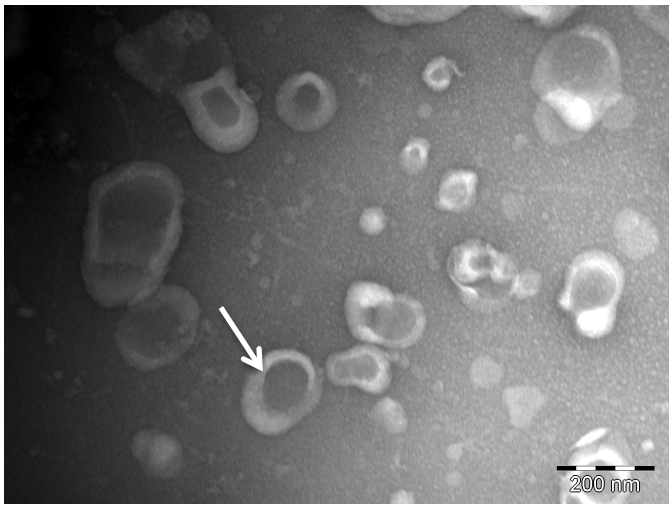


Figure 2.7 Whole mounted serum-derived EVs isolated using ExoMir™ kit viewed by transmission electron microscopy
White arrow points to a characteristic EV with cup-shaped morphology and size between 30 – 150 nm
Scale bar shown in bottom right hand corner. EV, extracellular vesicles

Although difficult to make direct comparison due to the different isolation methodologies and samples types, there appears to be less characteristic vesicles observed in the EM images from the serum samples (Figure 2.7) than the BALF EV samples (Figure 2.5 A), and there appears to be more contaminating material in the serum EV sample.

2.7 RNA isolation, cDNA synthesis and real-time qPCR

2.7.1 BALF EV RNA isolation, quantification and quality control to assess SEC EV isolation suitability for downstream application (RNA seqencing)

This method was used for RNA isolation, cDNA synthesis and RT-qPCR of RNA extracted during optimisation of BALF EV isolation to ensure adequate RNA quantity and quality prior to formal sequencing.

RNA from BALF EVs was isolated using miRNeasy Micro kit (Qiagen®) according to the manufacturer's instructions. Briefly, EVs were lysed with 1 mL of QIAzol Lysis Reagent. Chloroform was then added, mixed and incubated at room temperature for 2-3 min. Following centrifugation at 12,000 *g*, 4°C for 15 min, RNA was isolated by removing the aqueous phase which was then mixed with 1.5X volume of 100% ethanol. The total sample was transferred to an RNeasy MinElute spin column and centrifuged at 10,000 *g* for 15 s at room temperature followed by addition of Buffer RWT to the spin column. After further centrifugation at 10,000 *g* for 15 s, the flow-through was discarded and spin column washed with 80% ethanol. The column was left to air dry and the resultant RNA eluted in 14 µL of pre-heated (60°C) nuclease-free water. Concentrations of RNA were determined by NanoDrop 1000 (ThermoFisher Scientific®).

Custom primer pools were made up for the reverse transcription and pre-amplification reactions using specific Taqman™ miRNA assays (5X RT Primer and 20X mix of forward and reverse primers respectively) for the miRNA of interest. These were diluted with 1X Tris-Ethylenediaminetetraacetic acid (TE) Buffer and stored at -20°C for up to 2 months, according to the manufacturer's instructions.

Reverse transcription was carried out in 15 µL reactions; 45 ng of RNA was added to 1X RT buffer, 2X RT primer pool, 2mM deoxynucleotide (dNTP) mix, 150 U MultiScribe™ Reverse Transcriptase and 3.8 U RNase Inhibitor (ThermoFisher Scientific®). Remaining volume was made up to 15 µL with nuclease-free water.

Following cDNA synthesis, a preamplification reaction was performed in 25 µL reactions, as recommended by the manufacturer for a starting total RNA of 1-350 ng. Briefly, 2.5 µL of cDNA was added to 3X PreAmp Primer Pool (both forward and reverse primers) and 1X TaqMan® PreAmp Master Mix. Remaining volume was made up to 25 µL with nuclease-free water. Following thermal-cycling, the amplified cDNA (PreAmp product) was diluted with 0.0875X TE Buffer, pH 8.0 to a total volume of 200 µL and stored at -20°C for up to 1 week.

Quantitative PCR was carried out in duplicate 10 µL reactions, where 1X Taqman® Universal Master Mix II, No AmpErase® UNG was added to 1X Taqman® miRNA primer. Remaining volume was made up to 10 µL with nuclease-free water. This master mix cocktail was added to appropriate wells on a 384 well PCR plate before addition of 1 µL of diluted PreAmp product. Gene expression was normalized to stably expressed miRNA (varied per sample) and presented as either ΔCq or ΔΔCq to show fold induction.

To ensure RNA was of sufficient quantity to be analysed by sequencing, I performed RNA isolation, cDNA synthesis with pre-amplification and RT-qPCR on the combined SEC fractions from the PureEV™ columns. Primers for miR-29a and miR-340 were used as these miRNA have been used as normalisers in previous work I have performed using BALF EV samples (364). Figure 2.8 shows there was no RNA detected in either combined SEC # 1 or the blank sample, defined as a Cq value of above 36 (represented by the dotted line). There was however sufficient quantities of RNA detected in the other three combined SEC fractions (2-4), with a mean Cq value of 29.1 (SD±3.2, N=6) for combined SEC #2 (EV-dominant fraction). Importantly, RNA was detected in higher quantities (corresponding to a lower Cq value) in both combined SEC # 3 and 4, demonstrating the ability of soluble protein to carry RNA. It was therefore important to ensure this was separated from the EV-dominant fraction prior to RNA isolation to prevent any non-vesicular RNA contamination.

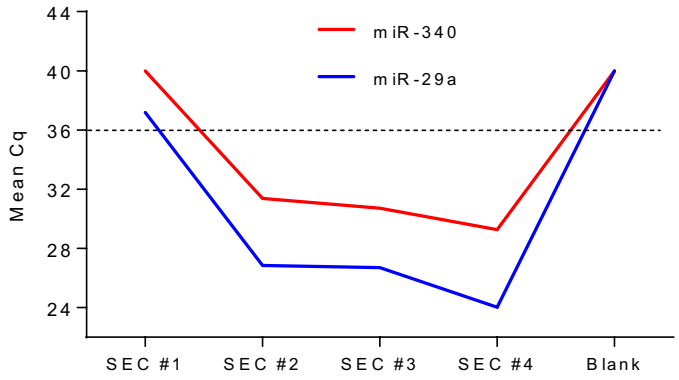


Figure 2.8 Mean Cq values for miR-29a and miR-240 for combined SEC fractions from PureEV™ columns

BALF samples, N=6. Combined SEC fractions 1-4 from PureEV™ columns used. Dotted line represents cut-off for RNA detection, where a Cq>36 shows no miRNA detected. SEC, size exclusion chromatography.

2.7.2 BAL EV RNA quantification and quality control prior to library preparation & next generation sequencing – performed by Qiagen® Genomic Services

Qiagen® Genomic Services performed the following methods as part of the next generation sequencing protocol. Prior to RNA isolation, EVs (suspended in up to 200 µL of 1X PBS) were lysed

using Buffer RPL at room temperature for 3 min. In order to assess the quality of RNA isolation across samples, Qiaseq miRNA Library Quality control (QC) Spike-Ins solution (contents listed in Table 2.3) was added to each of the lysed EV samples. RNA from EVs was then isolated by Qiagen® using the miRNeasy® Serum/Plasma Advanced kit (Qiagen®) following the manufacturer's instructions.

Table 2.3 Fifty-two QIAseq miRNA library QC spike-ins

NGS Spike-In name	Sequence	NGS Spike-In name	Sequence
UniSp100	uugauucccaauccaagcaag	UniSp126	acaaacaccuuggauguucuu
UniSp101	uaccaaccuuucaucguuccc	UniSp128	uaguccgguuuuggauacgug
UniSp102	ucccaaauguagacaaagca	UniSp102	ucccaaauguagacaaagca
UniSp103	ugaagcugccagcaugaucua	UniSp129	uuagaugaccaucaacaaacu
UniSp104	cagccaaggaugacuugccgg	UniSp130	ucuugcuuaaaugaguauucca
UniSp105	uccggcaaguugaccuuggcu	UniSp131	agcucugauaccaaaugauggaau
UniSp106	agaaucuugaugaugcugcau	UniSp132	ugaucucuucguacucuucuug
UniSp107	uuggcauucuguccaccucc	UniSp133	cgaaacuggugucgaccgaca
UniSp108	uuggcauucuguccaccucc	UniSp134	uucuugcauauguucuuuauc
UniSp109	cgaaacuggugucgaccgaca	UniSp135	uccuguguuuccuuugaugcgugg
UniSp110	uucgaggccuauuaaaccucug	UniSp136	aucaguuucuuguucguuuca
UniSp111	uagaaugcuauuguaauccag	UniSp137	ucauggucagauccgucaucc
UniSp112	gguucguacguacacuguuca	UniSp138	ucauggucagauccgucaucc
UniSp113	uaaacuaaucacggaaaugca	UniSp139	uugaauugaagugcuugaauu
UniSp114	uuuuggaaauuuguccuuacg	UniSp140	ugacaugggacugccuaagcua
UniSp115	ugagccucuguggguagcccuca	UniSp141	uaacuaaacauugguguagua
UniSp116	uuugcuuccagcuuuugucuc	UniSp142	uaagauccggacuacaacaaag
UniSp117	uugguuacccauauggccauc	UniSp143	uaauccuaccaauaacuucagc
UniSp118	uucgaugucuagcagugcca	UniSp144	gauggauaugucuucaaggac
UniSp119	ucuaagucuucuauugauguu	UniSp145	ccuuggagaaauaugcgucaa
UniSp120	uacgcauugaguuucguugcuu	UniSp146	uuaugucuuguugaucucaau
UniSp121	uggcuugguuuauguacaccg	UniSp147	uaaagucaauaauaccuugaag
UniSp122	uucugcuauguugcugcucau	UniSp148	uuuuuccucaaauuuauccaa
UniSp123	ugauuggaaauuucguugacu	UniSp149	augaauuuggaucuaauugag
UniSp124	ucuagcagcuguugagcaggu	UniSp150	auugguucaauucugguguug
UniSp125	uucuucgugaauaucuggcau	UniSp151	uaauuugguguuucuucgauc

These are 5' phosphorylated miRNAs with sequence length in the range 20-24 nucleotides. The sequences are of plant origin and bear no significant homology to miRNAs from the following species: human (hsa), mouse (mmu), rat (rno), rhesus monkey (mml), orangutan (ppy), chimpanzee (ptr) or pig (ssc).

Reverse transcription was performed in 10 µL reactions using the miRCURY LNA RT kit (Qiagen®), with an artificial RNA spike-in (UniSp6) to assess the quality of the reverse transcriptase reaction.

Following cDNA synthesis, quantitative PCR was performed by Qiagen in a LightCycler® 480 Real-Time PCR System (Roche®, Welwyn Garden City, UK) in 384 well plates. Primers used were: miR-23a, miR-30c, miR-103, miR-142-3p and miR-451, as well as the primers for the 52 RNA spike-ins (listed in Table 2.3) and the primer for the artificial RNA spike-in (UniSp6). Negative controls excluding the template from the reverse transcription reaction was performed and profiled like the samples. Amplification curves were analysed using the Roche® LC software, both for determination of Cq (by the 2nd derivative method) and for melting curve analysis.

The ISEV position paper on "*EV RNA analysis and bioinformatics*" recommends the use of external spike-in RNA to evaluate the sensitivity, accuracy and comparability of RNA-sequencing experiments (365). Pearson correlation analysis of the 52 RNA-spike-in Cq values was performed across the samples. The following radar plot demonstrates excellent correlation of counts corresponding to the spike-ins between the samples (Figure 2.9). The actual R^2 values were between 0.94 and 1.0 for all samples.

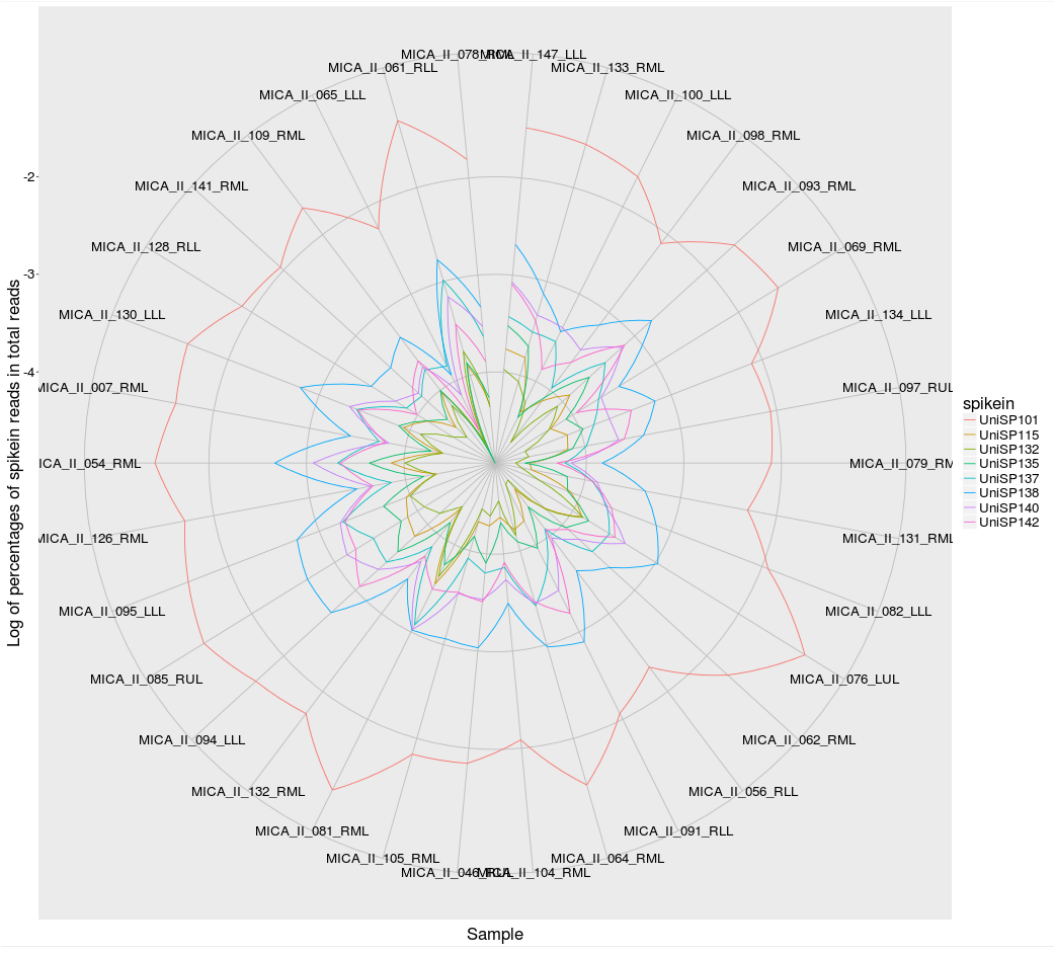


Figure 2.9 Radar plot showing relative spike-in signal for each sample
Figure adapted from Qiagen® sequencing data report. N=35; COPD, n=20.

Importantly, no sample outliers were identified at this point, which is key to ensure the quality and reliability of the library preparations.

The methods for RNA isolation, cDNA synthesis and RT-qPCR for validation of the sequencing results were the same as above, and again performed by Qiagen® Genomic Services. The only differences were the RNA spike-ins used for QC. Instead of the 52 RNA spike-Ins listed in Table 2.3, the following spike-ins were used to assess RNA isolation efficiency: UniSp2, UniSp4 and UniSp5. In addition to this, a DNA spike-in (UniSp3) was also added. This DNA spike-in consists of a premixed combination of DNA template and primers. Deviations in this reaction would indicate if there had been inhibitions at the qPCR level. For cDNA synthesis control, the same RNA spike-In (UniSp6) was used during the reverse transcription reaction.

2.7.3 Serum EV RNA isolation, quantification and quality control

Serum EVs were isolated using the ExoMir™ kit (see section 2.5.5) and were processed for RNA isolation using the following method. 200 µL of chloroform was added to the lysed particles that had been flushed from the ~20 nm filter (smaller, bottom filter) with 1 mL BiooPure™-MP. The resultant sample was vortexed for 20 sec and left at room temperature for 5 min. Samples were centrifuged 12,000 g, 4^{o}C for 15 min. RNA was isolated by removing the aqueous phase, adding 3 µL of co-precipitant (linear acrylamide) and incubating at room temperature for 5 min. To precipitate the RNA, 550 µL of isopropanol was added to the sample and then left at 20^{o}C for at least 1 hour. RNA was recovered by centrifugation at 12,000 *g*, 4^{o}C for 15 min. The resultant RNA pellet was washed with 900 µL 75% ethanol, re-suspended in 25 µL of nuclease-free water. Prior to storage at -80^{o}C, concentrations of RNA were determined by NanoDrop 1000 (ThermoFisher Scientific®). Custom primer pools were made up for the reverse transcription and pre-amplification reactions using specific Taqman™ miRNA assays (5X RT Primer and 20X mix of forward and reverse primers respectively) for the miRNA of interest. These were diluted with 1X Tris- Ethylenediaminetetraacetic acid (TE) Buffer and stored at -20^{o}C for up to 2 months, according to the manufacturer's instructions.

Reverse transcription was carried out in 15 µL reactions; 45 ng of RNA was added to 1X RT buffer, 2X RT primer pool, 2mM deoxynucleotide (dNTP) mix, 150 U MultiScribe™ Reverse Transcriptase and 3.8 U RNase Inhibitor (all applied biosystems® by ThermoFisher Scientific®). Remaining volume was made up to 15 µL with nuclease-free water.

Following cDNA synthesis, a preamplification reaction was performed in 25 µL reactions, as recommended by the manufacturer for a starting total RNA of 1-350 ng. Briefly, 2.5 µL of cDNA was added to 3X PreAmp Primer Pool (both forward and reverse primers) and 1X TaqMan®

PreAmp Master Mix. Remaining volume was made up to 25 µL with nuclease-free water. Following thermal-cycling, the amplified cDNA (PreAmp product) was diluted with 0.0875X TE Buffer, pH 8.0 to a total volume of 200 µL and stored at -20^{o}C for up to 1 week.

Quantitative PCR was carried out in duplicate 10 µL reactions, where 1X Taqman® Universal Master Mix II, No AmpErase® UNG was added to 1X Taqman® miRNA primer. Remaining volume was made up to 10 µL with nuclease-free water. This master mix cocktail was added to appropriate wells on a 384 well PCR plate before addition of 1 µL of diluted PreAmp product. Gene expression was normalized to stably expressed miRNA (varied per sample) and presented as either ΔCq or ΔΔCq to show fold induction.

2.8 Next generation microRNA sequencing of bronchoalveolar lavage extracellular vesicle RNA

2.8.1 BALF EV RNA Library preparation – performed by Qiagen® Genomic Services

Library preparation was performed by Qiagen® Genomic Services using the methods outlined in the Qiaseq™ miRNA Library Kit Handbook – precision small RNA library prep for Illumina® NGS system (version November 2016). Figure 2.10 summarises the methods for library preparation.

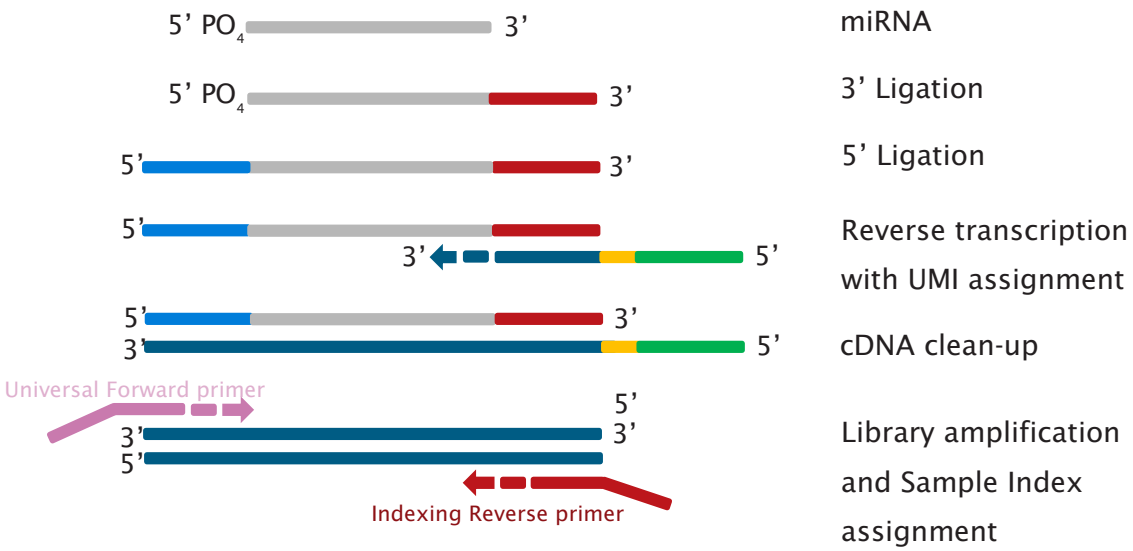


Figure 2.10 Summary of library preparation process
UMI, unique molecular index

2.8.2 microRNA library pre-sequencing quality control and preparation

MiRNA library pre-sequencing quality control was performed by Qiagen® Genomic Services on an Agilent® Bioanalyser 2100 using a High Sensitivity DNA chip.

The concentration of the miRNA sequencing library was performed by Qiagen® on a Qubit™ Fluorimeter (Invitrogen™). The molarity of each sample (in nM) was determined using the following equation:

$$(X \text{ ng/µl})(10^6)/(112450) = Y \text{ nM}$$

The individual libraries were diluted with nuclease-free water to 4 nM. The libraries were combined in equimolar amounts ready for sequencing.

Prior to sequencing, Qiagen® Genomic Services performed the "Standard Normalisation Method" outlined in the NextSeq System Denature and Dilute Libraries Guide (Document # 15048776 v5). A

final concentration of 1.8 pM of the denatured, diluted library solution was used for loading of the Flow cell.

2.8.3 Sequencing run setup

Next generation sequencing methods were performed by Qiagen® Genomic Services using the NextSeq500 instrument (Illumina®, Chesterford, UK). The high output Flow cell was prepared and the following sequencing parameters were used:

Read type - Single-end read

Number of sequencing cycles (read length) - 75 nucleotides

Average number of reads - 10 million reads/sample

2.9 MicroRNA sequencing output processing, quality control, mapping and alignment

2.9.1 Trimming of adaptors and UMI correction – performed by Qiagen® Genomic Services

The NextSeq500 sequencing system generates raw data files in binary base call (BCL) format. Qiagen® Genomic Services used bcl2fastq conversion software v2.20 (Illumina®) to demultiplex data and convert BCL files to standard FASTQ file format for downstream analysis. FASTQ is a text based sequencing data file format that stores both raw sequence data and quality scores. Next, cutadapt (1.11) is used to remove low quality bases, and identify the adapter and UMIs applied during library preparation (366). The output from cutadapt is used to remove adapter sequences and to collapse reads by unique molecular index (UMI) with in-house script (summarised in Figure 2.11). According to the experiment protocol, each raw read is expected to contain (starting from the 5' end): an insert sequence, the adapter sequence, 12nt-long UMI sequence, and other ligated sequence (see Figure 2.12). Depending on the read length and insert length, not all parts are present on all reads. To correct PCR bias with UMI information, raw reads are processed as follows:

1. Use cutadapt on raw reads with provided adapter sequence to acquire output with information about the presence of adapter for each read.
2. Parse cutadapt output and keep only reads that fulfil all of the following:
    a. Reads contain adapters
    b. Insert sequences should be equal or larger than minimal insert length (default 16 nucleotides)
    c. UMI sequence should be equal or longer than minimal UMI length (default 10 nucleotides)
3. Extract insert sequences from reads which do not contain full length UMI sequence from step 2 output as "partial-UMI reads".
4. Examine the reads with full length UMI in step 2 output and identify all unique insert + UMI combinations. Extract insert sequences from unique inset + UMI combinations as "full-UMI reads".
5. Combine "partial-UMI reads" and "full-UMI reads" as output of UMI correction.

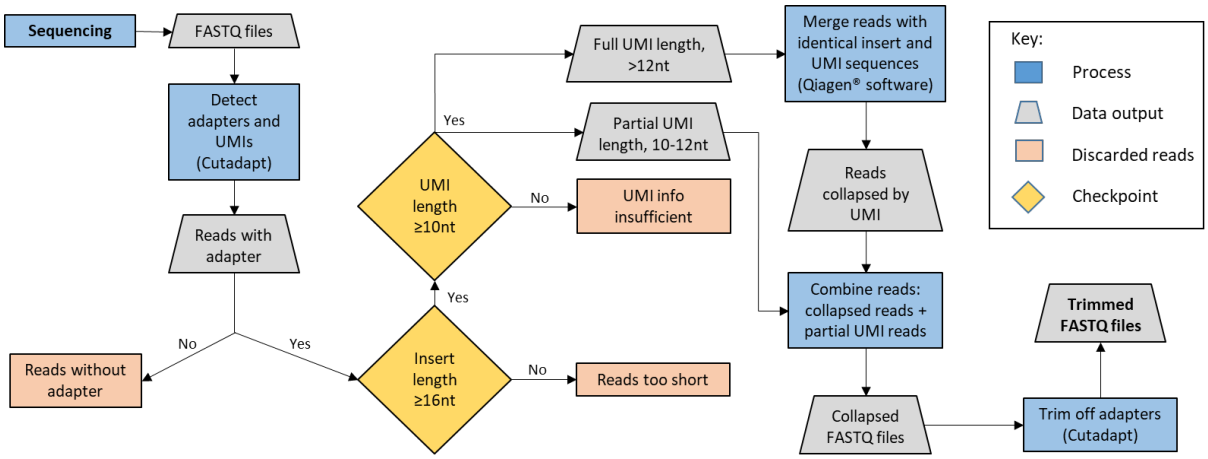


Figure 2.11 Overview of trimming of adapters and UMI correction process

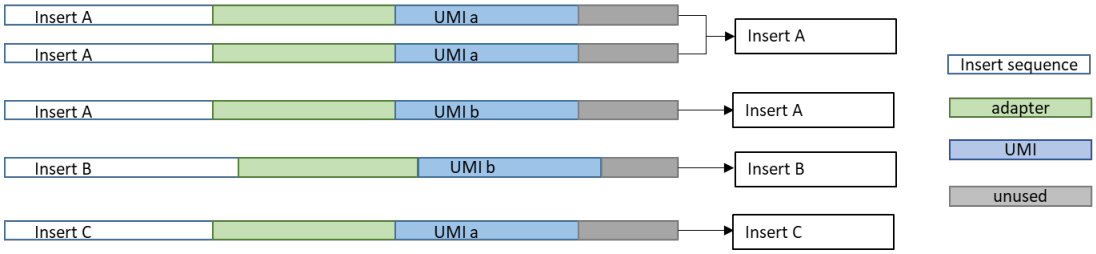


Figure 2.12 Illustration of the principle behind UMI correction
In order to correct for amplification biases, sequences containing identical insert and UMI are collapsed into a single sequence.

2.9.2 microRNA sequencing quality control

Trimmed FASTQ files were analysed using FastQC (367) a quality control tool for high throughput sequence data. This provided an overview on average read quality, average base quality and highlighted any potential outliers. The FastQC report provides basic statistics, per base sequence quality, per tile sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences and adapter content metrics. Qiagen® generated the FastQC reports and I then performed the quality control analysis.

2.9.2.1 Average read quality

Using the trimmed FASTQ files, FastQC generates a per sequence quality report, which identifies if a subset of sequences have universally low quality values. If a significant proportion of the sequences in a run have overall low quality then this could indicate a systematic problem (e.g. a problem with the flowcell). All of my samples passed this module, with the majority of the reads having a mean sequence quality score (Q-score) of above 30, which is consistent with high quality

data. An example of the average read quality is shown in Figure 2.13. The rest of the samples had similar plots, with the majority of reads having a Q score of greater than 30.

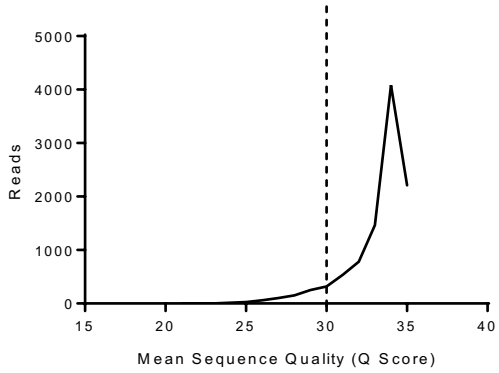


Figure 2.13 Average read quality of miRNA sequencing data
The mean sequence quality (Q score) is plotted against the number of reads. A Q-score above 30 is considered high quality.

2.9.2.2 Per base sequence quality

The per base sequence quality metric gives an overview of the range of quality scores across all bases at each position in the FASTQ file. All of my samples passed this module, with Q-scores above 30 (>99.9% correct), and therefore indicative of high quality data. An example of the quality scores across all bases is shown in Figure 2.14. All other samples in my dataset had simple profiles. Of note, the quality of calls on most platforms will degrade as the run progresses, so it is common to see base calls falling into the orange area towards the end of a run.

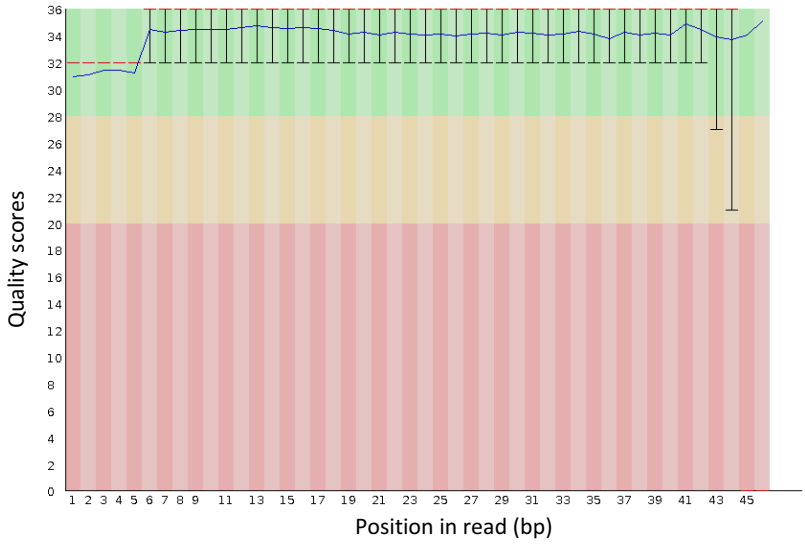


Figure 2.14 Box whisker plot showing the quality scores across all bases for a representative BALF sample.
Blue line represents the mean quality. Red dotted line represents the median. Whiskers represent IQR. Scores within the green area are considered high quality. Figure adapted from FASTQ file.

2.9.2.3 Sequence length distribution

The data generated from BAL EVs show a peak in sequence length around 21-22 base pairs, which represents miRNA (Figure 2.15). Longer sequences are of other origin (i.e. rRNA), transfer RNA (tRNA), mRNA and Y-RNA fragments) and have a length of ~30-50 nucleotides.

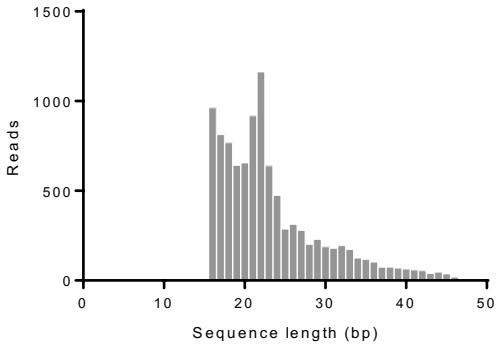


Figure 2.15 Read length distribution after filtering of adaptors for a representative BALF sample
MiRNA appear as a peak at 21-22 nucleotides. The other samples had very similar sequence length profiles corresponding to miRNA.

2.9.2.4 Additional FastQC quality metrics

In addition to the average read quality, per base sequence quality and sequence length distribution, the FastQC report generates a number of other outputs. This includes per base sequence content, which measures the proportion of each base (e.g. A, T, C, and G) position in a file. In a random library there is very little difference between bases in a sequencing run, however libraries produced by priming using random hexamers (i.e. RNA sequencing libraries) inherit an intrinsic bias in the positions at which reads start. Given my data is from an RNA sequencing library, the FastQC reported an error in this module, highlighting the difference between A and T, or C and G was greater than 10-20% in any position. Whilst this is considered true technical bias, the literature suggests it does not seem to affect downstream analysis.

Sequence duplication levels and overrepresented sequences are also presented in the quality control report. Diverse libraries will have very low levels of sequence duplication and overrepresented sequences. However, in RNA-sequencing libraries different transcripts are present at very different levels in the starting population. Therefore, in order to observe lowly expressed transcripts, it is common to greatly over-sequence highly expressed transcripts resulting in sequencing duplicates and overrepresented sequences. The FastQC reported an error in both sequence duplication and overrepresented sequences for this reason. However by using the UMI correction (outlined in section 2.9.1), I accounted for this technical duplication of sequences by collapsing identical insert sequences and UMI sequences (insert-UMI pairs) into a

single read. Table 2.4 summarises the number of reads after each step of the UMI correction process.

Table 2.4 Number of reads after each step of the UMI correction process

	Raw reads	Reads with adapter	Reads after length filtering	Reads with full or partial UMI	Collapsed and Partial reads
Mean (±SD)	15,911,755 ($\pm 1.27 \times 10^6$)	15,626,191 ($\pm 1.27 \times 10^6$)	3,760,379 ($\pm 9.54 \times 10^5$)	3,627,757 ($\pm 9.18 \times 10^5$)	2,233,781 ($\pm 4.75 \times 10^5$)
% of original raw reads	100	98	24	23	14

2.9.2.5 Summary of microRNA sequencing quality control

The quality metrics presented in section 2.9.2 are consistent with high quality data as demonstrated by the Q-scores of above 30. In addition, the sequence length distribution indicates that the majority of the reads correspond to an RNA sequence length of 18-22 nucleotides, representative of miRNA, the RNA of interest in my study. At this stage no outliers were identified in the sample cohort.

2.9.3 Mapping and aligning to reference genome – performed by Qiagen® Genomic Services

Qiagen® Genomic Services used bowtie2 (v 2.2.2) tool to align sequencing reads to the reference genome, Genome Reference Consortium Human Build 37 (GRCh37/hg19), and miRNA to the miRNA database, miRBase (version mirbase_20) (368). The reads had to have a perfect match to the reference sequences to be included for aligning reads to spike-ins, abundant sequences and miRBase. For mapping to the genome, 1 mismatch was allowed in the first 32 bases of the read. No indels were allowed in mapping. MiRNA counts were generated by "in-house" software by Qiagen® (mapping summarised in Figure 2.16). Mapped reads were classified into the following classes:

- **Outmapped**: Reads aligning to poly(A) and poly(C) homopolymers as well as abundant ribosomal RNA (rRNA) and mitochondrial RNA (mtRNA)
- **Unmapped reads**: No alignment to reference genome possible.
- **Genome**: Reads align to reference genome (GRCh37), but not to smallRNA or miRNA.
- **MiRNA**: Reads aligning to miRNA in miRBase (version mirbase_20).
- **SmallRNA**: Reads map to smallRNA database (compiled by Exiqon).
- **Predicted miRNA**: Reads map to a sequence found in miRBase in another organism ("predicted putative") or reads that don't match any known miRNAs in miRBase, but have the structural properties (i.e. read count distribution and secondary structure) of the genome in specific locations that resemble known miRNAs ("predicted miRBase").

An overview of the RNA sequencing quality control and mapping process performed by Qiagen® is summarised in Figure 2.16.

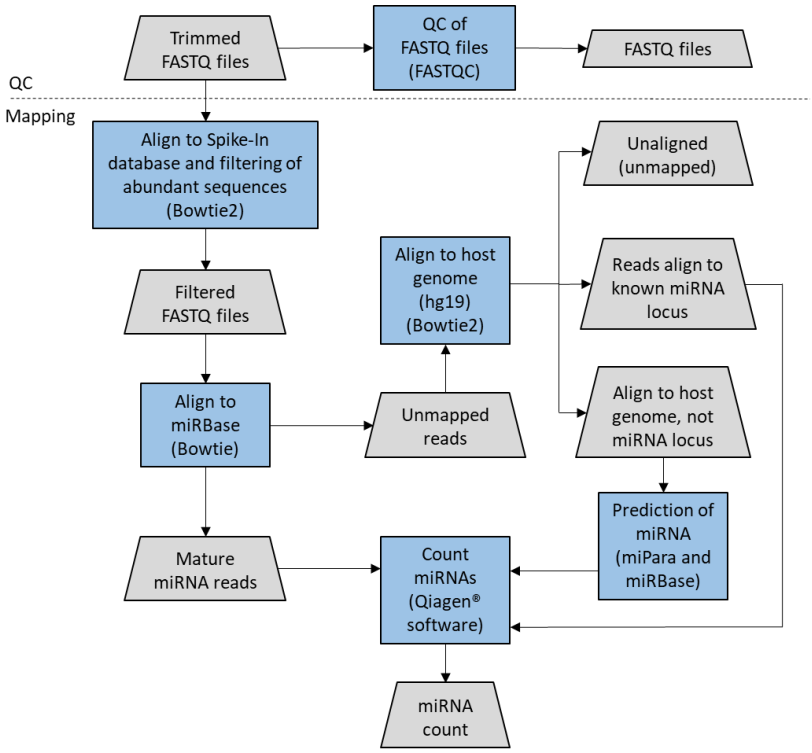


Figure 2.16 Overview of RNA sequencing quality control and mapping.
Blue rectangle indicates a process, with the software/tool specified in parenthesis. Grey trapezium indicates data output.

Given that the EV RNA was isolated from BALF, a non-sterile biofluid with possible bacterial and fungal contamination (both of which can release EVs), it is worth noting that during the mapping process, Qiagen® Genomic services also aligned sequenced reads to several bacterial and fungal reference genomes, however found no positive alignment. Therefore only sequenced reads aligning to the human genome were taken forward for further analysis.

2.9.3.1 Analysis of mapping and alignment to reference genome

I performed the analysis of the alignment and mapping results. The total number of reads mapped after full UMI correction for each sample is shown in Figure 2.17. The mean total number of reads mapped for the dataset was 2,804,969 (range: $1.6x10^6$ – $4.3x10^6$, SD±$7.07x10^5$) and the average genome-mapping rate was 53.4%.

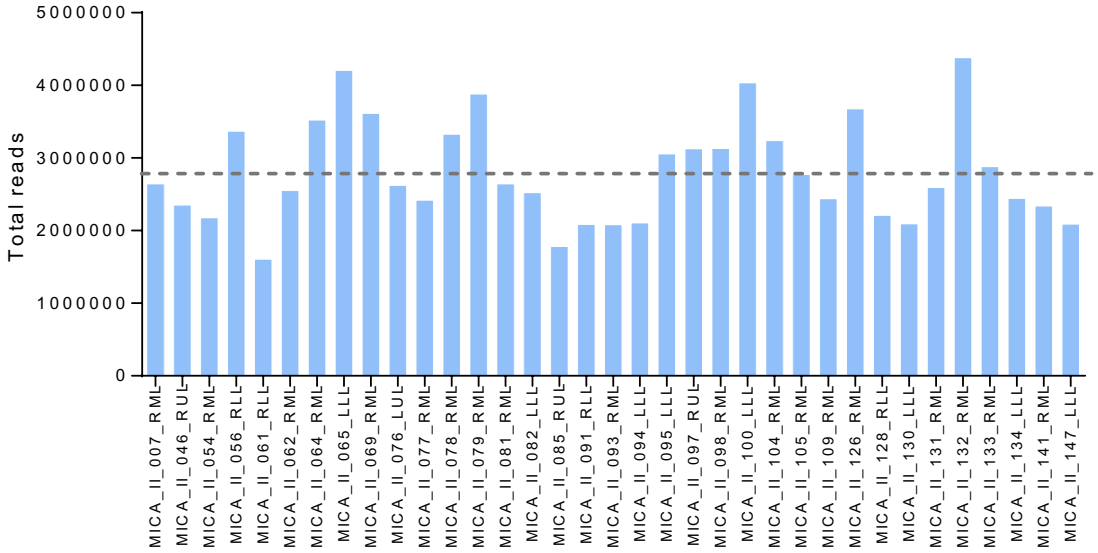


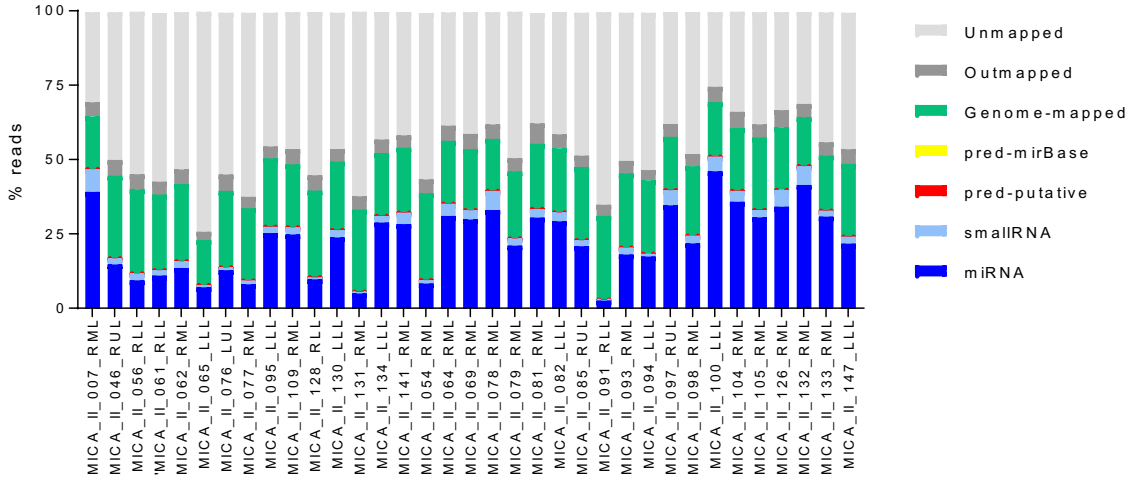
Figure 2.17 Total number of reads for each sample.

The dotted line represents in the mean total number of reads obtained ~ $2.8x10^6$. N=35; COPD, n=20.

2.9.3.1.1 Classification of mapped reads

As described in section 2.9.3, mapped reads were classified into the following types outmapped, unmapped, genome mapped, miRNA mapped, smallRNA mapped or predicted. The proportions of each type were calculated for each sample (Figure 2.18 A). After excluding the unmapped reads, (Figure 2.18 B) the most predominant mapped reads were genome-mapped, (mean 45.3%, SD±14.7%) and miRNA (mean 40.5%, SD±14.3%). SmallRNA, predicted RNA and outmapped RNA only accounted for a mean of 4.9%, 0.3% and 8.9% reads respectively.

A



B

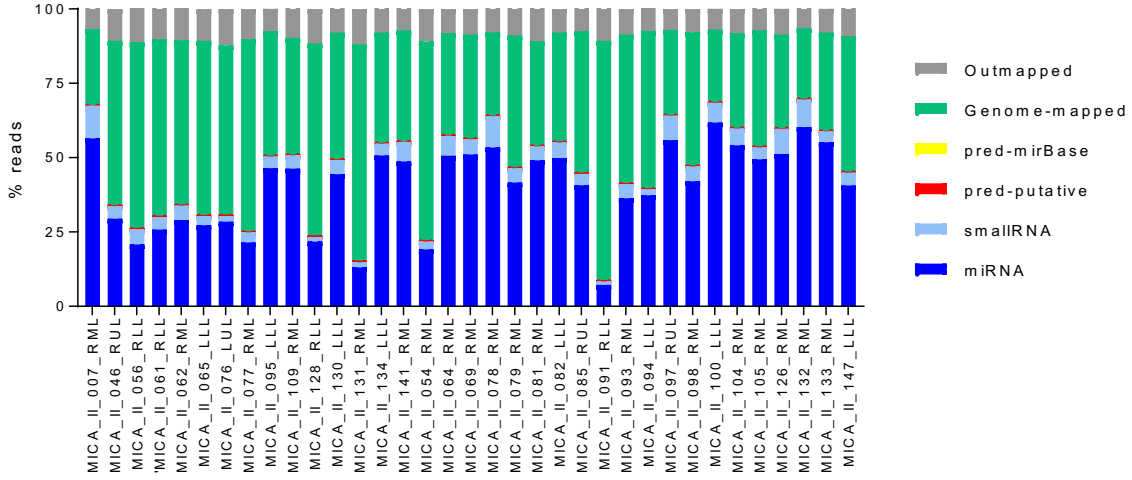


Figure 2.18 Summary of mapping results of reads for each sample.

Each sample consists of reads that can be classified into the following categories: miRNA, smallRNA, genome-mapped, outmapped or high abundance (e.g. rRNA, polyA, polyC, mtRNA) and reads which did not align to anything (unmapped). Figure 4.6 A unmapped reads included, B.unmapped reads excluded. N=35; COPD, n=20.

Of note, although several samples showed lower mapping rates when including both mapped and unmapped reads (Figure 2.18 A), MICA_II_091_RLL appears to have significantly lower mapping rates for both miRNA and smallRNA. This may well influence downstream analysis and will be considered further in the exploratory data analysis (see section 2.10.2).

2.9.3.1.2 Classification of smallRNA mapped reads

Five different types of smallRNA were found, transfer RNA (tRNA), piwi-interacting RNA (piRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA) and Y-RNA. tRNA was the most abundant (92.2%) smallRNA found in my EV samples (Table 2.5). It is integral to protein synthesis

by helping decode mRNA into protein. piRNA, is the largest class of small non-coding RNA molecules expressed in animal cells (369). This is reflected in my dataset, as they had the highest number of different RNA (2314) detected across all samples (Table 2.5). They form RNA-protein complexes through interactions with piwi proteins and have been linked to both epigenetic and post-transcriptional gene silencing, particularly in germ line cells (370). snoRNA are mainly involved in the modification of other RNAs, mainly rRNA, tRNA and snRNA. They were the smallest proportion (0.008%) of smallRNA found in both COPD and healthy ex-smoker EVs (Table 2.5). snRNA are one of many small RNA species confined to the nucleus and responsible for the processing of pre-messenger RNA (hnRNA) in the nucleus (371). The final type of smallRNA found are Y-RNAs, which are components of the Ro60 ribonucleoprotein particle. They are necessary for RNA replication through interactions with chromatin and initiation proteins (372).

Table 2.5 Types and proportions of smallRNA found in BALF EV miRNA (n=35)

Class of small RNA	Number of different types of small RNA found across samples	Proportion of small RNA type found across samples (%)
piRNA	2314	0.008
snoRNA	1612	0.8
snRNA	2065	1.9
tRNA	624	92.2
Y-RNA	866	5.1

piRNA, piwi-interacting RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; tRNA, transfer RNA.

2.10 MicroRNA sequencing filtering, data analysis and differential expression analysis

Unsupervised filtering, data analysis and differential expression analysis was performed in RStudio®, an open source software for using the R statistical computing environment, using R (v 3.8.2). The methods were adapted from the Bioconductor package, "Empirical analysis of digital gene expression in R" (edgeR) (373) and are summarised in Figure 2.19.

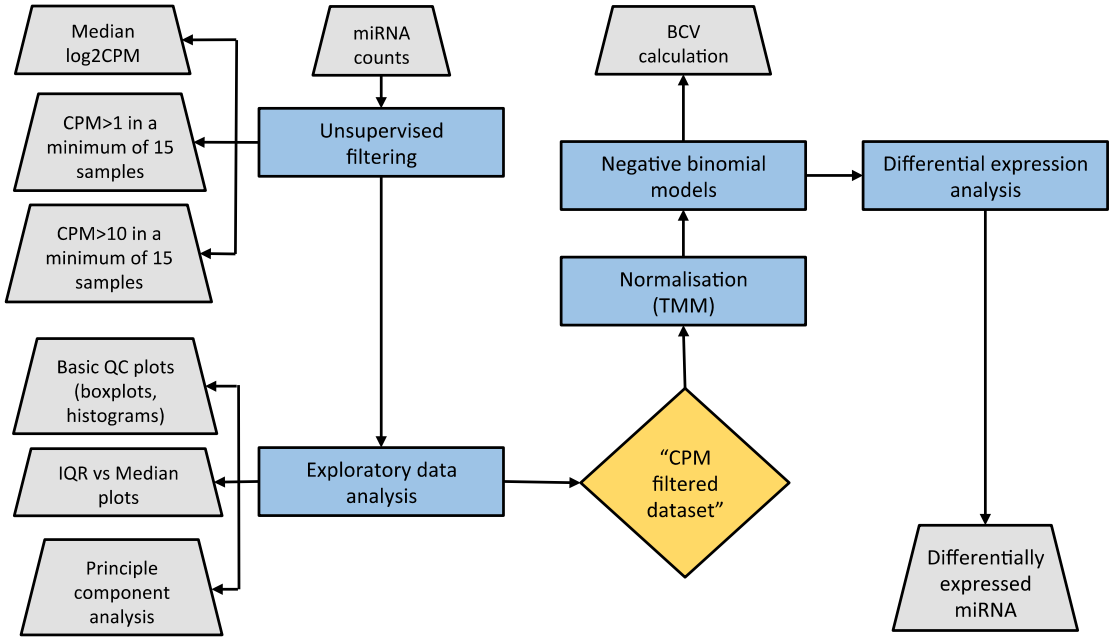


Figure 2.19 Overview of unsupervised filtering, exploratory data analysis and differential expression analysis methods
Blue rectangle indicates a process, grey trapezium indicate a data output, yellow diamond is an important checkpoint. BCV, Biological coefficient of variation; CPM, counts per million; IQR, Interquartile range; miRNA, microRNA; QC, Quality control; TMM, Trimmed mean of M values.

2.10.1 Unsupervised filtering

MiRNA with very low counts across all libraries provide little evidence for differential expression (373). From a biological standpoint, a miRNA must be expressed at some minimal level before it is likely to have a downstream effect on gene expression. In addition, the notable discreteness of these low counts interferes with some of the statistical approximations used later in the analysis.

Therefore, miRNA can be excluded from the dataset if it cannot be expressed in all samples for any of the conditions (i.e. in healthy ex-smokers or COPD). For a miRNA to be considered "expressed" in a library, it is usually required to have a raw read count of 5-10. However, it is also recommended, that filtering should be performed on count-per-million (CPM) data rather than on

raw counts directly, as the latter does not account for differences in library sizes between samples.

I found three examples of 3 different cut-offs for unsupervised filtering of RNA sequencing data:

I. Median log2-transformed CPM cut-off (326).
II. CPM>1 in a minimum of n samples, where n=size of the smallest group (373).
III. CPM>10 in a minimum of n samples, where n=size of the smallest group (374).

After removing the lowly expressed miRNA the library sizes were recalculated, for each filtered dataset, although the differences are usually negligible.

2.10.1.1 Median log2-transformed CPM cut-off method

This method uses a cut-off based on the median log2-transformed CPM for each miRNA (326). Briefly, the raw count data is converted into CPM and then log transformed (defined as median_log2_cpm in R). The median of this dataset is then used as a cut-off and every miRNA with a log2-transformed CPM of less than the median is discarded. The following code will perform this filtering process in R, where the raw count data is defined as data_clean_1:

```
> cpm_log <- cpm(data_clean_1), log = TRUE)
> median_log2_cpm <- apply(cpm_log, 1, median)
> hist(median_log2_cpm)
> expr_cutoff <- median_log2_cpm
> abline(v = expr_cutoff, col = "red", lwd = 3)
> sum(cpm_log > expr_cutoff)
> keep_1 <- data_clean_1[cpm_log > expr_cutoff,]
```

Library sizes were recalculated using the following code, where groups is defined by a vector that assigns each of the samples to a subject group, either COPD or healthy:

```
> group <- read.table("Experimental_Design_File.csv", header = TRUE, row.names = 1,
sep = ",")
> groups <- group$Disease
> keep_1 <- DGEList(counts = keep_1, genes = row.names(keep_1), group = groups)
```

2.10.1.2 CPM>1 in a minimum of n samples, where n=size of the smallest group

This method uses a CPM cut-off of greater than 1 in a minimum of 15 or more of the samples (373). Here, a CPM of 1 corresponds to a raw read count of 6-7 in the smallest sample. A requirement for expression in 15 or more libraries is used, as the minimum number of samples in each group is 15 (i.e. there are 15 healthy controls and 20 COPD subjects, therefore the minimum number in each group is 15). The following code will perform this filtering process using the edgeR statistical package in R (v 3.8.2), where the raw count data is contained within a simple list-based data object called DGEList (defined as data_clean_2):

```
> keep_2 <- rowSums(cpm(data_clean_2)>1) >= 15
```

Library sizes were recalculated using the following code:

```
> keep_2 <-eds[keep_2, , keep.lib.sizes=FALSE]
```

2.10.1.3 CPM>10 in a minimum of n samples, where n=size of the smallest group

Again, this uses the edgeR statistical package in R (v 3.8.2), but utilises the function filterByExpr, which has a pre-set CPM cut-off of greater than 10 in a minimum of 15 or more samples (374). This is therefore, the most stringent cut-off and can be performed using the following code, where the raw count data is contained within the DGEList, data_clean_3:

```
> keep_3 <- filterByExpr(data_clean_3, design)
```

Library sizes were recalculated using the following code:

```
> keep_3 <-eds[keep_3, , keep.lib.sizes=FALSE]
```

Each filtering method resulted in a different number of miRNA being included in the filtered dataset (Figure 2.18). The median cut-off method resulted in 527 included miRNA, whereas the CPM>10 in a minimum number of samples (n=15) was the most stringent cut-off and resulted in only 275 included miRNA. This method included all of the miRNA from the other two filtered datasets.

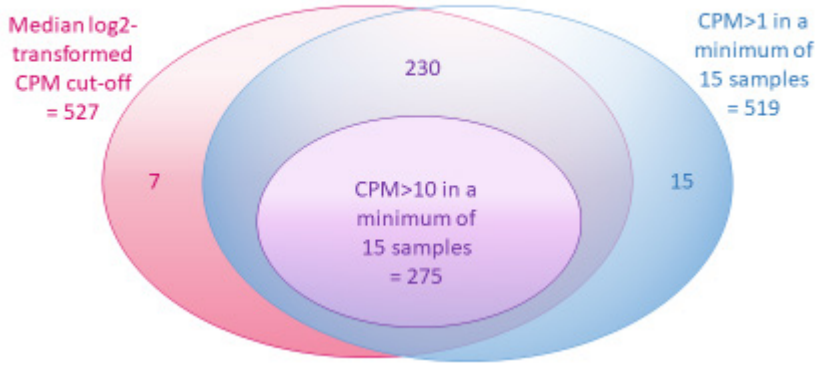


Figure 2.20 Non-scaled Venn diagram showing the number of miRNA included in each filtered dataset and the overlap between each method

For each of these three cut-offs for unsupervised filtering, I used several analytical and graphical approaches to determine which cut-off was the most robust for my dataset and to identify any outliers within the sample dataset. The next section will explored these quality control measures.

2.10.2 Exploratory data analysis for quality control

In order to build familiarity with the sequencing data, determine overall quality and identify possible outliers, which could bias further analysis, it is important to visualise and summarize aspects of the data. I used a number of methods to do this: basic quality control plots (*e.g.* boxplots, histograms), interquartile range (IQR) versus median plots and principle component analysis.

2.10.2.1 Basic quality control plots

Boxplots and histograms were used to visualise the distribution and density respectively, of read counts across each of the sample libraries. These were generated in R (v 3.8.2) using the following code, where "cpm_log" is defined as the log-transformed CPM value for each miRNA, in each sample:

```
> boxplot(cpm_log, outlines=FALSE, las=2, cex.axis=0.5, main="cpm_log data miRNA", col="gray79")

> hist(cpm_log, main="cpm_log data miRNA", xlab = "CPM_log counts")
```

The boxplots for each filtered dataset are shown in Figure 2.21, A-C. The median log2CPM cut-off (A) and CPM>1 in a minimum of 15 samples (B) methods show the highest variation in data (highlighted in red circles) suggesting these methods are still including miRNA with very low counts. Low counts are also highlighted in the histograms (Figure 2.21, D-F), which show higher

frequencies in the 0 log-transformed CPM counts for the median logCPM cut-off (D) and CPM>1 in a minimum of 15 samples (E) filtered data (highlighted in red circles).

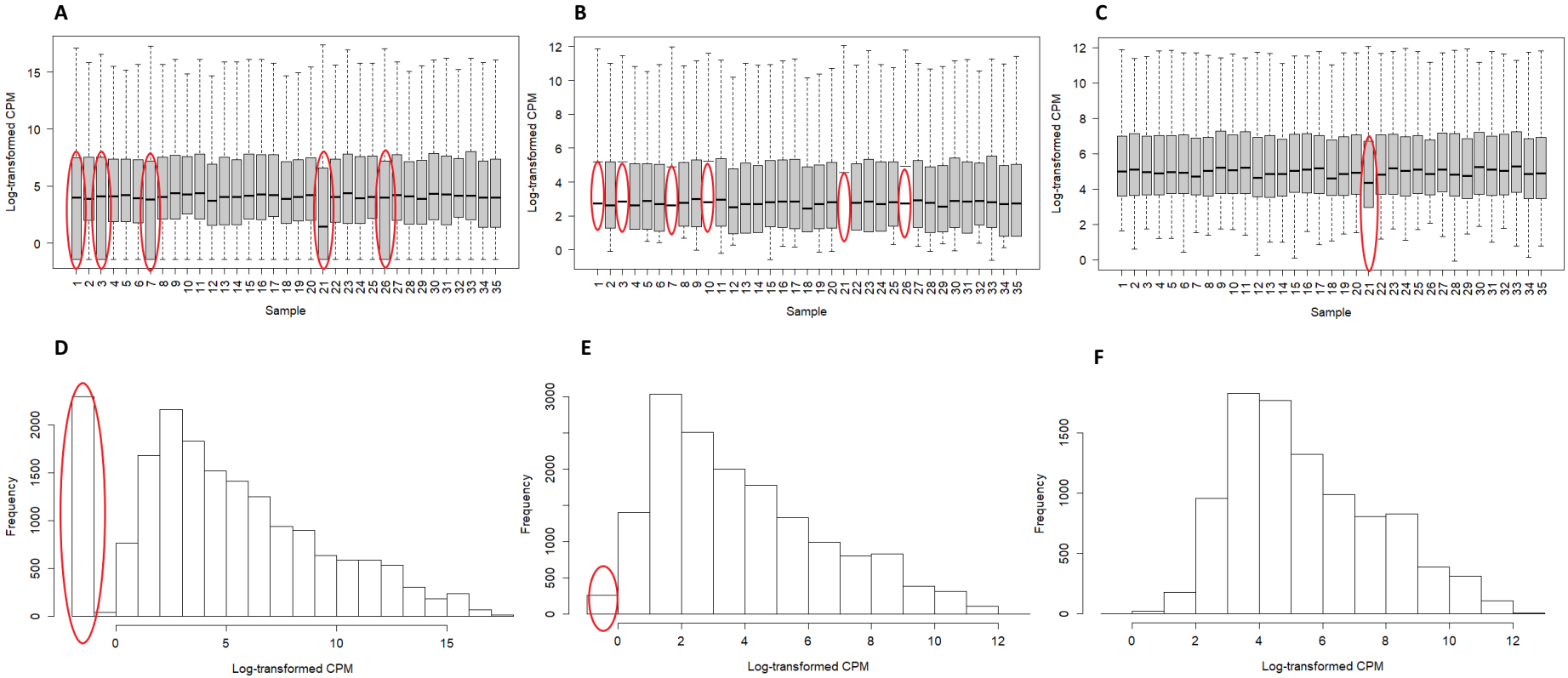


Figure 2.21 Boxplots (A-C) and histograms (D-F) showing the distribution and frequency of log-transformed CPM data across datasets for different filtering methods.

Boxplots: Median log2CPM cut-off (A), CPM>1 in a minimum of 15 samples (B), CPM>10 in a minimum of 15 samples (C). Black lines represent medians, grey boxes represent IQRs, whiskers represent the range except for "outliers" that are more than ±1.5 times the IQR larger or smaller than the median. Histograms: Median log2CPM cut-off (D), CPM>1 in a minimum of 15 samples (E), CPM>10 in a minimum of 15 samples (F) Red circles highlight large variation in data or frequency of miRNA included with zero log-transformed CPM. N=35, COPD n= 20. CPM, counts per million.

2.10.2.2 Interquartile range versus median plot

For each of the filtering methods, the IQR and median was calculated, using Microsoft® Excel® (2011), from the CPM miRNA expression data for each sample library. The values were plotted on an XY graph along with reference points for 1 standard deviation (SD) of the mean and 2 SDs of the mean. This allows visualisation of the spread or variance in data across the samples. It also highlights potential outliers, by identifying those with very large variance and therefore outside 2 SDs of the mean (Figure 2.22).

CPM>1 and CPM>10 in a minimum of 15 samples filtering identified similar outliers (defined as outside 2 SD of IQR/Median CPM) with both MICA_II_100_LLL and MICA_II_091_RLL showing a high degree of variance. In addition, MICA_II_007_RML was also identified as a possible outlier in the CPM>10 in a minimum of 15 samples filtered dataset. In contrast, the median log2CPM filtered dataset identified several different outliers, and except for MICA_II_091_RLL, all of which were samples from healthy ex-smokers (MICA_II_077_RML, MICA_II_061_RLL, MICA_II_128_RLL and MICA_II_131_RML).

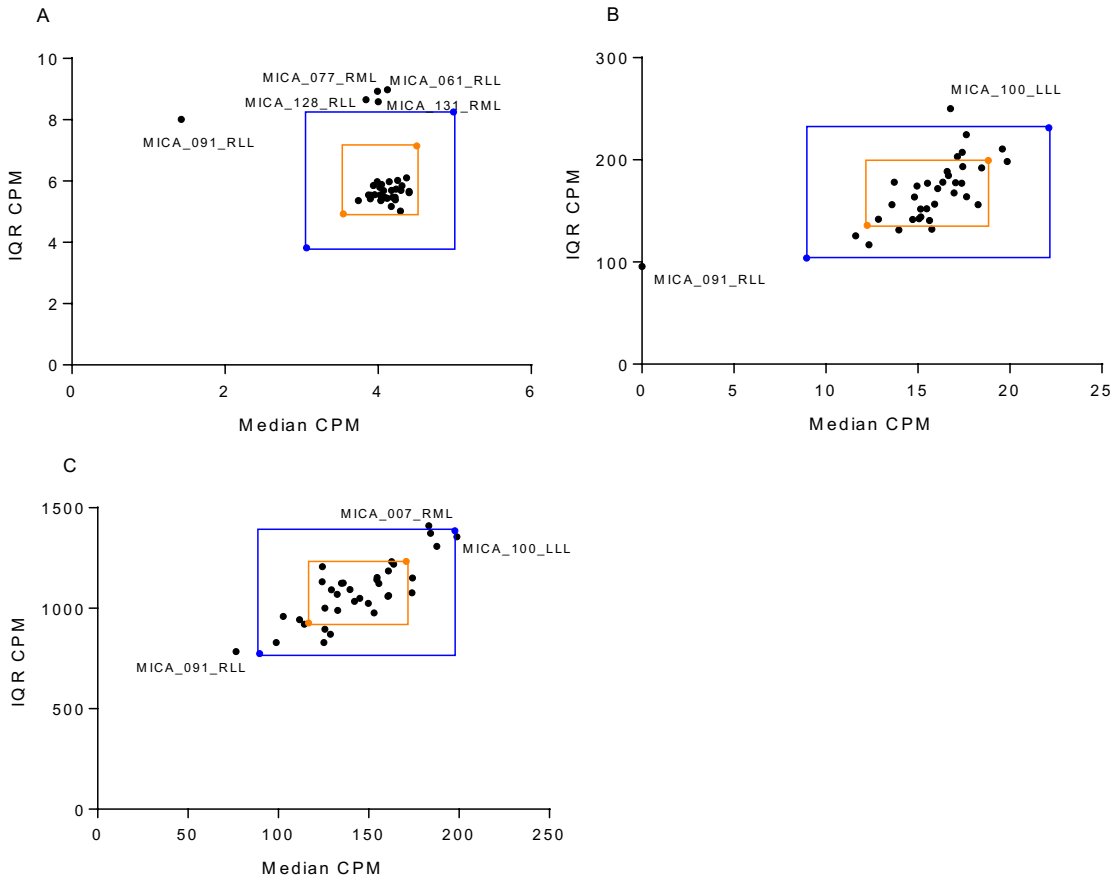


Figure 2.22 Interquartile range/median plots of CPM data for different filtering methods.
Median log2CPM cut-off (A), CPM>1 in a minimum of 15 samples (B), CPM>10 in a minimum of 15 samples (C) datasets. Black dots represent samples (n=35, COPD n = 20). Blue dots and perimeter represent ±2SD from the mean. Orange dots and perimeter represent ±1SD from the mean. Outlier samples outside ± 2SD from the mean labelled.

The outliers identified from the boxplots and IQR/Median plots are summarised in Table 2.6 for each of the different filtering methods. The sample identified most commonly as an outlier was MICA_II_091_RLL. This was identified by all three filtering methods in both boxplots and IQR/Median plots. The median log2CPM cut-off method included the highest number of miRNA, however this meant lowly expressed miRNA with high variance were included. I concluded that the most robust data for further analysis was the CPM>10 in a minimum of 15 samples filtered data. Although, this method used the most stringent cut-off (resulting in the smallest number of miRNA, 275), when visualising the data and examining the variance using the above techniques it included the least lowly expressed miRNA with the least variance. Therefore, this dataset is likely to contain more biologically significant miRNA, which is an important consideration for differential expression analysis.

Table 2.6 Summary of outliers identified from the boxplots and IQR/median plots for each unsupervised filtering method

Unsupervised filtering method	**Number of miRNA included in filtered dataset**	**Sample outliers identified on boxplots of CPM filtered data**	**Sample outliers identified on IQR/Median plots of CPM filtered data**
Median log2CPM cut-off	527	MICA_II_077_RML MICA_II_061_RLL MICA_II_128_RLL ***MICA_II_091_RLL** MICA_II_131_RML	MICA_II_077_RML MICA_II_061_RLL MICA_II_128_RLL ***MICA_II_091_RLL** MICA_II_131_RML
CPM>1 in a minimum of 15 samples	519	MICA_II_077_RML MICA_II_061_RLL MICA_II_128_RLL MICA_II_054_RML ***MICA_II_091_RLL** MICA_II_131_RML	MICA_II_100_LLL ***MICA_II_091_RLL**
CPM>10 in a minimum of 15 samples	275	***MICA_II_091_RLL**	MICA_II_ 007_RML MICA_II_100_LLL ***MICA_II_091_RLL**

*Sample identified as an outlier in every filtering method, in all both analysis. CPM, counts per million; IQR, interquartile range.

Interestingly, MICA_II_091_RLL was identified as an outlier in all three unsupervised filtering methods and in all graphical analysis of the data (Table 2.6). Furthermore, as highlighted in section 2.9.3.1.1, this sample appeared to have significantly lower mapping rates for both miRNA and smallRNA. Therefore, as these anomalies are likely to affect downstream differential expression analysis, the sample was excluded from the final analysis (see section 2.10.2.4).

2.10.2.3 Principle component analysis

Principle component analysis (PCA) was used to transform the multi-dimensional sequencing dataset to smaller, discrete sets of orthogonal principle components. The first principle component specifies the direction with the largest variability in the data, the second component is the direction with the second largest variation and so on. The PCA and plots were generated using the rgl package in R (v 3.8.2) using the function "prcomp" on the filtered CPM, log-transformed dataset, using the following code (375):

```
> pca <- prcomp(t(cpm_log), scale. = TRUE)

> PC1 <- pcs$x[ ,1]

> PC2 <- pcs$x[ ,2]

> PC3 <- pcs$x[ ,3]

> PCA_details <- cbind(PC1, PC2, PC3)

> write.table (PCA_details, "PCA_details.txt", sep = "/t')

## Open PCA_details.txt file in excel and shift column titles to the right, add "Sample_ID" heading and Disease column (i.e. assign each sample to COPD ("dodgerblue" color) or Healthy ("firebrick" color) group) and save.

> pca <- read.table ("PCA_details.txt", sep="/t")

## Plot 3D PCA plot

> library (rgl)

> p3d <- plot3d (pca$PC1, pca$PC2, pca$PC3, x lab = "Comp 1", y lab = "Comp 2", z lab = "Comp 3", col = (pca$Disease), box = FALSE, size = 0.5, type = 's')

> text3d (PC1, PC2, PC3, text = pca$Sample_ID, font = 1, cex = 0.6)
```

Outliers were identified on the PCA plot by capturing those within 1 SD of the mean using the following code:

```
> mean.vec <- c(mean(PC1), mean(PC2), mean(PC3))

> all_comp <- cbind (PC1, PC2, PC3)

> sigma <- cov(all_comp)

> plot3d(ellipse3d(x = sigma, centre= mean.vec, scale = c(1,1,1), col = "PeachPuff", alpha = 0.5, add = TRUE, level = 0.95, smooth = TRUE)
```

This generated the following three-dimensional PCA plots (Figure 2.23-A and -B). The outliers identified from the previous methods (e.g. Boxplots and IQR/Median plots) are labelled. In addition, Figure 2.23-B shows the data points within one SD of the mean (contained within the ellipsoid), MICA_II_054_RML is the only sample outside this cut-off showing high variation within the sample.

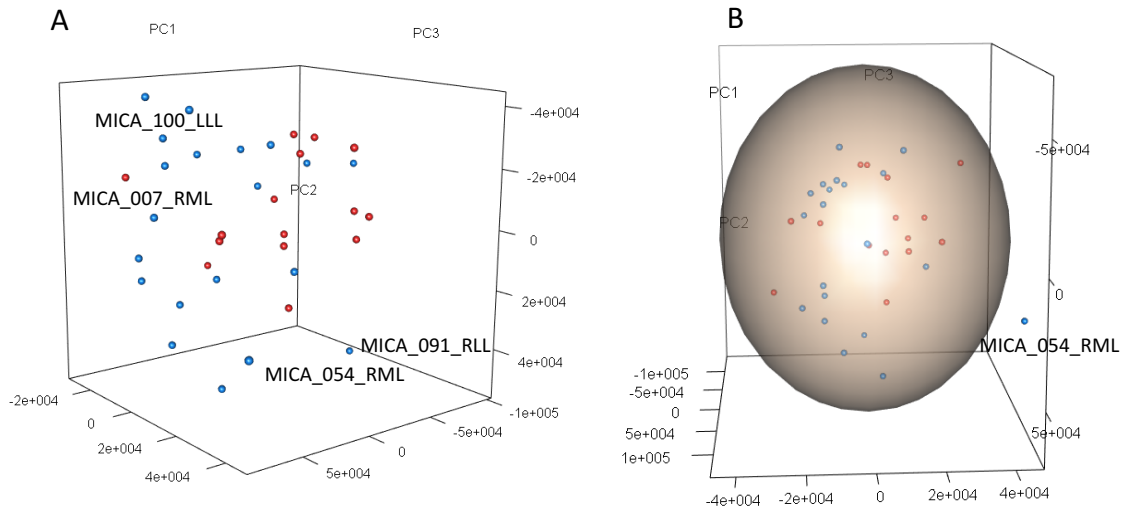


Figure 2.23 Three-dimensional PCA plot showing the variation and clusters within the Limma filtered dataset

One standard deviation from the mean is plotted as an ellipsoid in Figure 4.11-B. Red dots represent healthy ex-smokers (n=15), blue dots represent COPD subjects (n=20). Potential sample outliers labelled.

2.10.2.4 Summary of exploratory data analysis

There is no consensus as to which unsupervised filtering method to use for processing raw RNA count data (365) and there is no previous work directly comparing one strategy against another. Therefore, I tested three cut-off options (median log2CPM; CPM>1 and CPM>10 – both in a minimum of 15 samples) and used a number of graphical outputs (e.g. boxplots, histograms, IQR/median plots and PCA) to analyse the data. The choice of methodology came from multiple sources including: attending an international EV conference ("*Extracellular Vesicles 2017*"; Cambridge, UK; September 2017), supervised analysis on a bioinformatics research methods course ("*Introduction to RNA-seq analysis";* University of Cambridge, UK; October 2017), collaborations with researchers in the field (Professor Mark Lindsay; University of Bath, UK) and the literature (373, 374). This process allowed me to build familiarity with my dataset and identify sample outliers, which would bias differential miRNA expression analysis.

Although the median log2CPM and CPM>1 in a minimum of 15 samples methods resulted similar numbers of miRNA (527 and 519 respectively), the CPM>1 and CPM>10 in a minimum of 15 samples methods were more similar with respect to outlier identification, particularly on the

IQR/median plots (Figure 2.19). In this analysis, the median log2CPM method identified a group of healthy ex-smokers (MICA_II_077_RML, MICA_II_061_RLL, MICA_II_128_RLL and MICA_II_131_RML), who did not flag up on any of the additional outputs leading to the assumption that this was a possible spurious result. Although, the CPM>10 in a minimum of 15 samples method resulted in the smallest number of included miRNA (275 compared with 519 and 527), the lowly expressed miRNA included in the other methods are likely to introduce increased variance and therefore bias any future differential expression testing. Given this, I proceeded with the **CPM>10 in a minimum of 15 samples filtered dataset**, with **four outliers identified above removed** (MICA_II_007_RML, MICA_II_054_RML, MICA_II_091_RLL and MICA_II_100_LLL), leaving **n=31** (COPD, n=17 and Healthy ex-smoker, n=14). This is known as the **"CPM filtered dataset"** and was used for all onward analysis.

2.10.3 Normalisation methods

CPM is a unit to measure expression levels in next generation sequencing experiments and is calculated by the following equation:

CPM = (Number of reads mapped to specific miRNA/ Total number of reads in the library) x 1 million

This is a simple normalisation procedure that corrects only for sequencing depth and provides a measure of quantity for each miRNA. However, this method does not account for the most important technical influence on differential expression, RNA composition. MiRNA sequencing provides a measure of the relative abundance of each miRNA in each RNA sample, but does not provide any measure of the total RNA output on a per-cell basis. This becomes important when a small number of miRNA are very highly expressed in one sample, but not in another. These highly expressed miRNA can constitute a substantial proportion of the total library size, causing the remaining miRNA to be under-sampled in that library. Therefore, unless these highly expressed miRNA are adjusted for, the remaining miRNA may falsely appear to be downregulated in that sample.

The calcNormFactors function in the edgeR package in R normalises for RNA composition by finding a set of scaling/normalisation factors for the library sizes that minimise the log-fold changes between the samples for most miRNA. The default method for computing these scale factors uses a trimmed mean of M-values (TMM) between each pair of samples (376). The following code performs the TMM normalisation and generates a list of normalisation factors for each sample, where keep is the log-transformed "**CPM filtered dataset**":

```
> TMM_normalised <- calcNormFactors(keep)

> TMM_normalised$samples
```

The normalisation factors of all the libraries multiply to unity. A normalisation factor below one indicates that a small number of high count miRNA are dominating the sequencing, causing the counts for the other miRNA to be lower than would be usual given the library size. Conversely, a factor above one scales up the library size, analogous to downscaling the counts. The recalculated library sizes and normalisation factors are listed in Table 2.7.

Table 2.7 List of library sizes and normalisation factors for the "CPM filtered dataset" generated by TMM normalisation.

Sample ID	Sample cohort	Adjusted library size	Normalisation factor
MICA_II_077_RML	Healthy	193077	0.996
MICA_II_078_RML	COPD	1092789	1.000
MICA_II_061_RLL	Healthy	175842	1.046
MICA_II_065_LLL	Healthy	294122	0.977
MICA_II_109_RML	Healthy	601252	0.985
MICA_II_141_RML	Healthy	659567	0.985
MICA_II_128_RLL	Healthy	214688	0.970
MICA_II_130_LLL	Healthy	496871	1.004
MICA_II_126_RML	COPD	1248508	1.106
MICA_II_095_LLL	Healthy	768741	0.922
MICA_II_085_RUL	COPD	371065	0.919
MICA_II_094_LLL	COPD	364761	0.901
MICA_II_132_RML	COPD	1805946	1.007
MICA_II_081_RML	COPD	801747	1.128
MICA_II_105_RML	COPD	844511	0.969
MICA_II_046_RUL	Healthy	343816	0.899
MICA_II_104_RML	COPD	1152597	0.990
MICA_II_064_RML	COPD	1089079	0.974
MICA_II_056_RLL	Healthy	313940	1.034
MICA_II_062_RML	Healthy	344192	1.067
MICA_II_076_LUL	Healthy	333900	1.008
MICA_II_082_LLL	COPD	732867	1.007
MICA_II_131_RML	Healthy	127769	1.017
MICA_II_079_RML	COPD	810943	1.072
MICA_II_097_RUL	COPD	1077755	0.974
MICA_II_134_LLL	Healthy	699110	0.961
MICA_II_069_RML	COPD	1076087	1.161
MICA_II_093_RML	COPD	373396	0.996
MICA_II_098_RML	COPD	678973	1.046
MICA_II_133_RML	COPD	882149	0.935
MICA_II_147_LLL	COPD	451583	0.998

CPM filtered dataset, N=31. COPD, n=17.

The TMM-normalised dataset (n=31) was then used for the differential expression analysis between patients with COPD and healthy ex-smokers. By using the TMM normalisation values, the statistical tests used in differential expression analysis were less skewed and the false positive rate was reduced.

2.10.4 Negative binomial models

Original methods for modelling RNA-sequencing count data used the Poisson distribution. However, the Poisson assumes the mean and variance are identical and in RNA-sequencing measurements, the variance of miRNA expression is larger than the mean (termed "overdispersion"). Therefore the negative binomial distribution is used which has a dispersion parameter for modelling the increase in variance from a Poisson process.

2.10.4.1 Biological coefficient of variation

The package edgeR in R (v 3.8.2) was used to perform the statistical methodology (based on negative binomial models) to analyse the differentially expressed miRNA. Firstly, edgeR shares information across miRNA to determine a common dispersion. It then extends this to a trended dispersion to model the mean-variance relationship (lowly expressed genes are typically more noisy). Lastly, it calculates a dispersion estimate per miRNA and shrinks it towards the trended dispersion, referred to as the "biological coefficient of variation" (BCV). The BCV is normally ~0.4 in human studies (373). The miRNA-specific (referred to in edgeR as tagwise) dispersion estimates are used in the test for differential expression. The following code can be used to calculate dispersion estimates and visualise them in a BCV plot (see Figure 2.24); where "TMM_normalised" is the normalised miRNA dataset and "design" is a model matrix based on the experimental design of the study (i.e. samples either assigned to COPD or Healthy).

```
> Condition <- factor(group[ ,"Disease"], levels=c("Healthy", "COPD")
> design <- model.matrix(~Condition)
> y <- estimateDisp(TMM_normalised,design)
## The square root of the common dispersion gives the BCV
> sqrt(y$common.dispersion)
# The dispersion estimates can be view in a BCV plot
> plotBCV(y)
```

The package edgeR applies an empirical Bayes strategy for squeezing the tagwise dispersions towards a global dispersion trend or towards a common dispersion value. The amount of squeeze is determined by the weight given to the global value on one hand and the precision of the tagwise estimates on the other. The relative weights given to the two are determined the prior and residual degrees of freedom. By default, the prior degrees of freedom, which determines the amount of empirical Bayes moderation, is estimated by examining the heteroskedasticity of the data (377).

Negative binomial distribution methods were used to model the CPM filtered-TMM normalised-dataset. The common dispersion estimate was calculated as 0.398 (~0.4 is usual for biological studies). Trended dispersion estimates and miRNA specific estimates (referred to as “Tagwise” in edgeR) were plotted (Figure 2.24) and used in testing for differential expression.

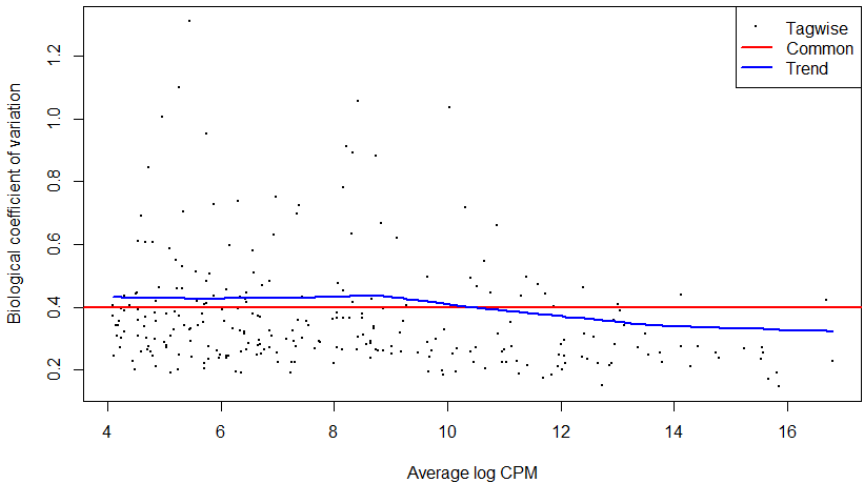


Figure 2.24 Biological coefficient of variation plot showing trended, common and miRNA ("Tagwise") specific estimates for CPM filtered TMM normalised data

2.10.5 Differential expression analysis between COPD subjects and healthy ex-smokers

Once negative binomial models were fitted and dispersion estimates obtained, edgeR determined differential miRNA expression using the exact test.(373)

The exact test is based on quantile-adjusted conditional maximum likelihood (qCML) method, which is commonly used for RNA-sequencing experiments with a single factor (e.g. COPD versus health). Knowing the conditional distribution for the sum of counts in a group, (i.e. calculated by the dispersion estimates), p-values were computed by summing over all sums of counts that have a probability less than the probability under the null hypothesis of the observed sum of counts. The exact test for negative binomial distribution has strong parallels with Fisher’s exact test (373).

The following code performs this function in R, where "y" is the dispersion estimates calculated above:

```
>  et <- exactTest(y)

## To give the top 10 differentially expressed miRNA

> topTags(et)

## To list the miRNA differentially expressed at a false discovery rate (FDR) of 5%

> results_edgeR <- topTags (et, n= nrow(data_clean), sort.by = "none")

> sum(results_edgeR$table$FDR<0.05)

## To visualise the data on an MA plot, showing the log2 fold change on y axis versus
average log 2 CPM on x axis for differentially expressed miRNA, with miRNA with an
FDR<0.05 in red.

> plotSmear (et, de.tags  = rownames (results_edgeR) [results_edgeR$table
$FDR<0.05], pch=16, cex=1)

## Additional information can be added to the MA plot e.g. blue line representing two-
fold change in expression

> abline(h=c(-1,1), col = "blue")
```

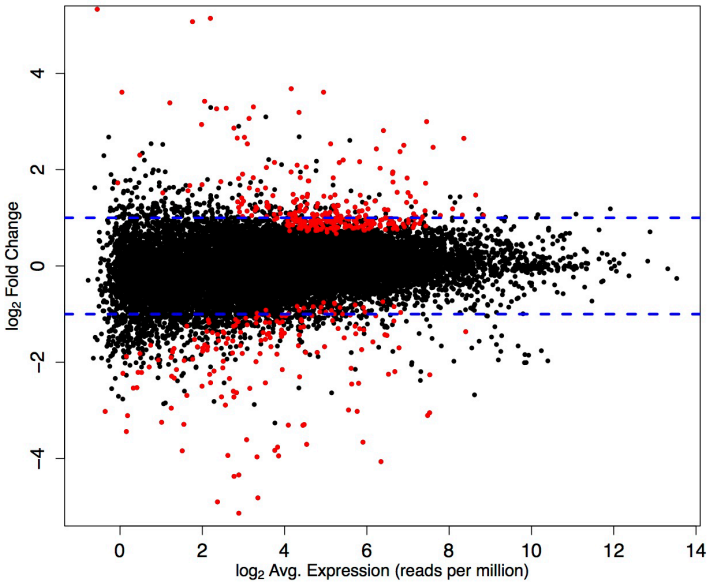


Figure 2.25 An example of an MA plot showing differential expressed miRNA
An MA plot is an application of Bland-Altman plot. Red dots represent miRNA with a FDR <0.05. Blue lines represent a twofold change in expression.

2.11 Real time-qPCR data analysis

2.11.1 RT-qPCR data quality control

The amplification efficiency was calculated using algorithms by Qiagen® proprietary software (378). All assays were inspected for distinct melting curves and the primer melting temperature (Tm) was checked to be within known specifications for the assay. Furthermore, assays must be detected with five Cq less than the negative control, and with Cq<37 to be included in the data analysis. Data that did not pass these criteria were omitted from any further analysis. Cq was calculated as the 2nd derivative.

Analysis of the expression levels of the RNA spike-in controls were performed to assess the quality of the RNA extraction (UniSp2, UniSp4 and UniSp5), reverse transcriptase reaction (UniSp6) and qPCR steps (UniSp3). These assays are not used for normalisation.

2.11.2 Normalisation of RT-qPCR data

NormFinder software in Microsoft® Office Excel was used to identify miRNA that were most stably expressed across all samples from the miRNA sequencing results (379). These miRNA were then used as normalisers for the qPCR validation study. Although all of these "normaliser" miRNA were measured by RT-qPCR, only "normaliser" miRNA detected in all samples were used for normalisation of Cq data. The mean Cq for all the universally expressed "normaliser" miRNA was calculated to give a Geomean Cq. Then, the following formula was used to calculate the normalized Cq values:

$$\text{Normalized Cq of miRNA of interest } (\Delta Cq) = \text{Geomean Cq} - \text{miRNA of interest Cq}$$

A higher value thus indicates that the miRNA is more abundant in the particular sample. Values were then presented as $2^{\Delta Cq}$ to represent fold change.

2.11.3 Differential expression analysis of RT-qPCR data

To compare the differences between COPD and healthy ex-smokers, the SD, and the average ΔCq was calculated for the two groups. Based on this, the ΔΔCq was calculated for each miRNA. The distribution of the data was assessed by a Shapiro-Wilk normality test, with a $p>0.05$ suggesting data is normally distributed.

An unpaired t test with Welch's correction (assuming the SD was not equal for groups) was performed. Raw P values were then adjusted for multiple testing using the Benjamini-Hochberg correction (380).

Volcano plots were generated in R (v 3.8.2) using the following code:

```
>qPCR_results <- read.table("MA_data_qPCR.csv", header = TRUE, row.names = 1, sep = ",")
>with(qPCR_results, plot(log2FC, -log10(Pvalue), pch=20, main="Volcano plot", xlim=c(-2.5,2)))
>with(subset(qPCR_results, FDR<0.05), points(log2FC, -log10(Pvalue), pch=20, cex=2,col="red"))
```

2.12 Identifying miRNA target genes

An overview of the methods to identify of miRNA target genes is summarised in Figure 2.26. Each of the following sections will cover the processes outlined in the blue boxes.

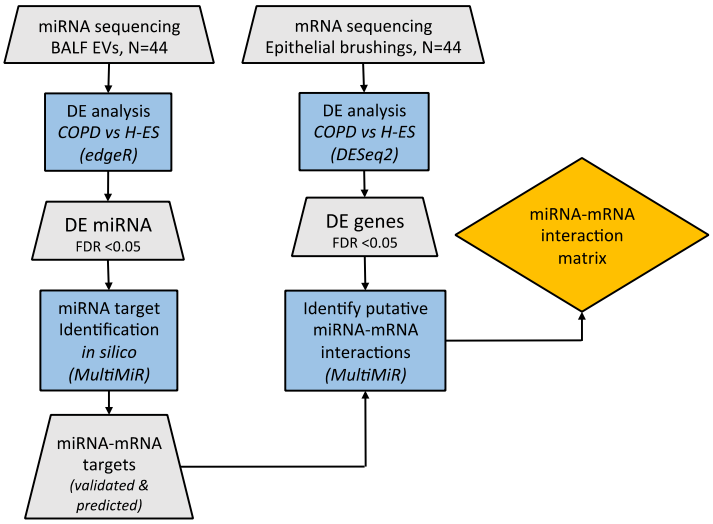


Figure 2.26 An overview of the methods used to identify miRNA-mRNA interactions in this study
Blue rectangle indicates a process with the software/tool specified in parenthesis, grey trapezium indicate a data output, yellow diamond is an important checkpoint. BALF, bronchoalveolar lavage fluid; COPD, Chronic obstructive pulmonary disease; DE, differential expression; EVs, extracellular vesicles; FDR, false discovery rate; H-ES, healthy ex-smoker; miRNA, microRNA; mRNA, messenger RNA

2.12.1 miRNA target prediction *in silico* analysis using multiMiR

MicroRNA targets sites are catalogued in databases based on experimental validation and computational prediction using a variety of algorithms. Several online resources provide collections of multiple databases such as miRBase (381), TargetScan 7.2 (303) and the multiMiR package in R (382). The latter is a comprehensive collection of nearly 50 million predicted and validated miRNA-target interactions and their associations with diseases and drugs. MultiMiR was chosen for this analysis over other databases as it includes several novel features:

1. Compilation of 14 different databases, more than any other collection
2. Expansion of databases to those based on disease annotation and drug response, in addition to many experimental and computational databases.
3. User-defined cut-offs for predicted binding strength to provide the most confident selection.

Lists of validated and predicted mRNA targets for the differentially expressed lung EV miRNA identified from the differential expression analysis (Chapter 4) were generated in using multiMiR package (v.2.1.1.) in R (v 4.0.2) using the following code (where UP_mRNA corresponds to the up-regulated miRNA in COPD):

```
>library(multiMiR)

##List of validated targets of up-regulated miRNA in COPD
>UP_multimir_results <- get_multimir(org     = 'hsa',
                                    mirna   = UP_miRNA,
                                    table   = 'validated',
                                    summary = TRUE)
>table(UP_multimir_results@data$mature_mirna_id)

##List of validated results for a specific miRNA (e.g. miR-2110)
>miR_2110 <- UP_multimir_results@data[grep("hsa-miR-2110", UP_multimir_results@data[,
                "mature_mirna_id"]), ]
>miR_2110 <- as_tibble(miR_2110)
>miR_2110 <- miR_2110 %>% distinct(target_ensembl, .keep_all = TRUE)

##List of predicted results for all up-regulated miRNA in COPD, with the top 20% of targets
considered.
> UP_multimir_results_pred <- get_multimir(org     = 'hsa',

                                    mirna   = UP_miRNA,

                                    table   = 'predicted',

                                    predicted.cutoff.type = 'p',

                                    predicted.cutoff = 20,

                                    use.tibble = TRUE,

                                    summary = TRUE)

> table(UP_multimir_results_pred@data$mature_mirna_id)
```

The analysis was completed for both up and down-regulated miRNA identified from the differential expression analysis comparing COPD with healthy ex-smokers.

2.12.2 Next generation mRNA sequencing of epithelial brushings – performed by the Translational Science & Experimental Medicine team at AstraZeneca

Paired epithelial brushings (processed as outlined in section 2.4.5.1) from the same study subjects and lung lobe location as the BAL EV were isolated from, were processed for mRNA sequencing by the Translational Science & Experimental Medicine team from AstraZeneca (Gothenburg, Sweden). The following section outlines the methods of the library preparation and sequencing.

Briefly, total RNA was extracted from epithelial brushings using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen®). The quantity and quality of RNA samples were determined using the standard RNA analyzer kit on a 96-channel Fragment analyzer (Agilent® Technologies, Stockport, UK). Extracted samples with a yield concentration >25 ng/µl total RNA, and a DV_{200} value (percentage of RNA fragments >200nucleotides) >=30% were deemed to be of sufficient quantity and quality for TotalRNA-seq analysis. Samples were diluted to 25 ng/µl using a Tecan Fluent liquid handling automation system (Tecan, Männedorf, Switzerland). Library preparation was done in four separate runs, one 96 well plate per run. The Kapa RNA HyperPrep Kit with RiboErase was used for reverse transcription, generation of double stranded cDNA and subsequent library preparation and indexing to facilitate multiplexing (Roche, Basel, Switzerland), all of which was performed through automation on a Tecan fluent. The libraries were quantified with the 96-channel Fragment Analyzer using the standard sensitivity NGS kit (Agilent® Technologies). Samples from each preparation plate were pooled and the final pools (4 in total) were quantified using a Qubit™ instrument for concentration determination with the DNA High Sensitivity kit (ThermoFisher Scientific®). Fragment size was determined using the Fragment Analyzer, standard sensitivity NGS kit (Agilent® Technologies). Three of four library pools were further diluted to 1 nM and sequenced on a NovaSeq 6000 (Illumina®) using NovaSeq 6000 S4 Reagent Kit, 2x76 cycles. The remaining library pool was diluted to 1.9 nM and sequenced on NovaSeq 6000 (Illumina®) using 2 NovaSeq 6000 SP S1 Reagent Kits, 2x51 cyclers. Average reads per sample were 52.6 million.

2.12.2.1 mRNA sequencing data preparation - performed by the Bioinformatics team at AstraZeneca

The epithelial brushing mRNA sequencing output processing, quality control, mapping and alignment was performed by the AstraZeneca bioinformatics team. Briefly, Fastq files from paired-end sequencing libraries were collected and read quality for all libraries was accessed using FastQC (v 0.11.7) (383), Qualimap (v 2.2.2c) (384) and samtools stats (v 1.9) (385). QC metrics for Qualimap were based on a STAR (v 2.7.2b) (386) alignment against the human genome (GRCh38, Ensembl v99). Next, QC metrics were summarized using MultiQC (v 1.7) (387). Two libraries were excluded; one due to a low mapping rate (57% vs [79%-97%]) and another due to low sequencing throughput (210k reads vs [20M-86M]), leaving 118 epithelial brushings for analysis. Sequencing adapters were then trimmed from the remaining libraries using NGmerge (v 0.3) (388). A human transcriptome index consisting of cDNA and ncRNA entries from Ensembl (v 99) was generated and reads were mapped to the index using Salmon (v1.1.0) (389). The bioinformatics workflow was organized using Nextflow workflow management system (v 19.07) (390) and Bioconda software management tool (391).

2.12.3 Differential gene expression analysis of the epithelial brushing mRNA

I performed the differential gene expression analysis of the epithelial brushing mRNA using DESeq2 (v 1.26.0) (392), using apeglm (v 1.8.0) (393) for fold change shrinkage, all in R (v 4.0.2). In the model for differential expression effects from a technical batch-effect (library preparation plate) were taken into account. Estimated counts were used as input for the DESeq2 with lowly expressed genes excluded (only genes with at least 10 counts in at least 20 samples were kept, n=27,229).

Differential gene expression analysis was performed in R (v 4.0.2) using the following code:

```
## Read in total RNA-seq data (provided by AstraZeneca, where a pre-prepared dds object is
  available with two assays:
1. counts: estimated raw counts (from Stargazer pipeline (counts from tximport, Salmon
  quantification)
2. vst_batch: vst normalized expression from batch-corrected counts where the Lane/Plate
  effect has been removed.
```

The dds object also has all metadata available (coldata).

```
> dds < - readRDS(file.path(params$dir_data, params$data_version))

## Subset dds object to sample type of interest, epithelial brushings
> dds_brush <- dds[ , dds$NGS.Sample.type %in% c(params$sample_type)]

## Subset dds object to only include Groups of interest, COPD (P_FE, P_IE) and healthy ex-
smokers (HV_ES).
>dds_brush <- dds_brush[ , dds_brush$Group %in% c("P_FE","P_IE","HV_ES")]

## Subset dds object to only include matched samples of interest (i.e. those with matched EV
miRNA from the same lobe location, N=44)
>rownames <- rownames(dds_brush_16g@colData)
>dds_brush_N44 <- dds_brush_16g[,rownames %in%
    c("200716V4007RNCRLLA1","111116V4034RNCRMLA1","C1111820013",
    "030816V4056RNCRLLA1","111116V4062RNCRLLA1","111116V4061RNCRMLA1",
    "020916V4065RNCLLLA1","111116V4073RNCRLLA1","111116V4076RNCLULA1","111116V4077R
    NCRMLA1","040518V4095RNCLLLA1","040518V4109RNCLLLA1","040518V4128RNCRLLA1","
    040518V4130RNCLLLA1","040518V4131RNCRMLA1","040518V4134RNCLLLA1","040518V413
```

```
5RNCLLLA1","040518V4140RNCRMLA1","040518V4141RNCRMLA1","040518V4150RNCLLLA
1","111116V4078RNCRLLA1","C1111820001","C1111820006","C1111820009","040518V4097RNC
RULA1","040518V4100RNCLULA1","040518V4105RNCLLLA1","040518V4145RNCRMLA1","04
0518V4148RNCRLLA1","040518V4151RNCRMLA1","210616V4054RNCRMLA1","020916V406
4RNCRMLA1","020916V4069RNCRMLA1","111116V4079RNCRMLA1","231116V4081RNCRML
A1","040518V4093RNCRMLA1","040518V4094RNCLLLA1","040518V4104RNCRMLA1","0405
18V4126RNCRMLA1","040518V4132RNCRMLA1","040518V4133RNCRMLA1","040518V4139
RNCRULA1","040518V4144RNCRMLA1")]
```

Set-up “design” & “reference level” to reflect experimental design of study (i.e. samples either assigned to COPD or Healthy) and adjust for potential bias.

```
>design(dds_brush_N44) <- formula(~Lane.Plate + Disease)
>dds_brush_N44$Disease <- relevel(dds_brush_N44$Disease, ref = "HV")
```

Drop levels prior to performing DESeq2 analysis:

```
>dds_brush_N44$Disease <- droplevels(dds_brush_N44$Disease)
```

Run DESeq2 analysis for differential expression between COPD and healthy ex-smokers.

```
>dds_brush_N44 <- DESeq(dds_brush_N44)
```

Perform log fold change-shrinkage using apeglm

```
>res_COPDvsES_g <- lfcShrink(dds_brush_N44, coef="Disease_COPD_vs_HV", type="apeglm")
```

To list the mRNA with differentially expressed at a FDR of 5%

```
>sum(res_COPDvsES_g$padj < 0.05, na.rm=TRUE)
```

To list the mRNA with fold change of greater than 2.

```
>sum((res_COPDvsES_g$padj < 0.05 & abs(res_COPDvsES_g$log2FoldChange) > 1),
na.rm=TRUE)
```

To visualise the data on an MA-plots, showing the log2 fold change on y axis versus vs expression-level (Mean of Normalised counts) on x axis for differentially expressed miRNA, with miRNA with an FDR<0.05 in red.

```
>plotMA(res_COPDvsES_g, ylim=c(-4,4), colSig ="red", alpha = 0.05, abline(h=c(-1,1), col
= "blue", lty = 2), ylab = "log2 fold change"
```

2.12.4 Identify miRNA-mRNA putative interactions

In addition to retrieving all validated and predicted target genes of given miRNA or set of miRNA, the multiMiR package within R (382) has functionality to retrieve interactions between miRNAs and a specific set of differentially expressed genes. Given than miRNA lead to a down-regulation of their gene targets (an inverse correlation relationship), only down-regulated genes in COPD were selected as possible targets for up-regulated miRNA in COPD and vice versa. Interactions between the up-regulated miRNA and down-regulated genes identified from the differential expression analysis were identified by the multiMiR (v 2.1.1.) in R (v 4.0.2) using the following code:

```
##Run multiMiR package for human organism, up-regulated miRNA, down-regulated
genes, for both validated and predicted targets, considering only the top 20% of
predicted interactions.
>UP_multiMir <- get_multimir(org     = "hsa",
                             mirna   = DE.miRNA.up,
                             target  = DE.entrez.dn,
                             table   = "all",
                             summary = TRUE,
                             predicted.cutoff.type = "p",
                             predicted.cutoff      = 20,
                             use.tibble = TRUE)
```

The analysis was also performed for the down-regulated miRNA and up-regulated genes. Any duplicates found from multiple databases were removed from the resultant miRNA-mRNA interaction matrix.

2.13 miRNA-mRNA interaction analysis

An overview of the methods to identify the miRNA-mRNA most important to COPD biology is summarised in Figure 2.27. Each of the following sections will cover the processes outlined in the blue boxes on the right under the miRNA-mRNA interaction analysis heading.

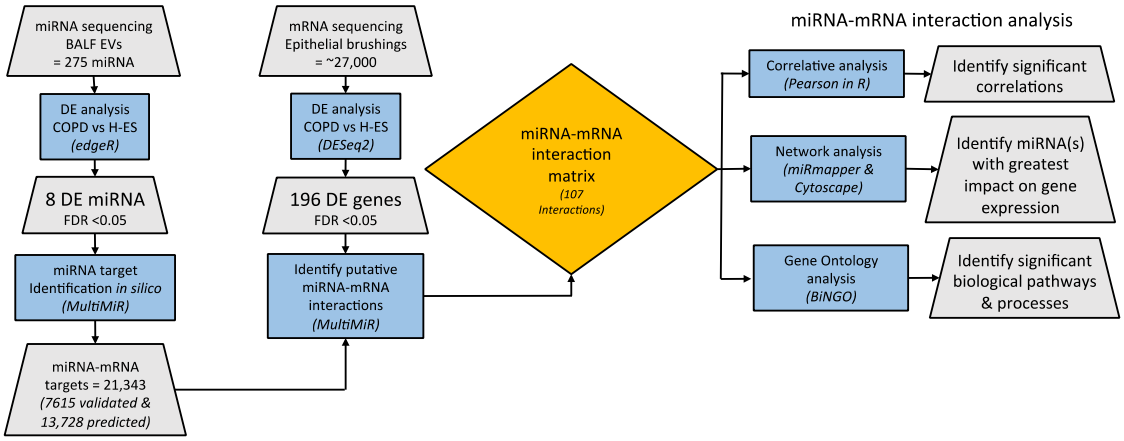


Figure 2.27 An overview of the methods used in this study for miRNA-mRNA interaction analysis
Blue rectangle indicates a process with the software/tool specified in parenthesis, grey trapezium indicate a data output, yellow diamond is an important checkpoint. BALF, bronchoalveolar lavage fluid; COPD, Chronic obstructive pulmonary disease; DE, differential expression; EVs, extracellular vesicles; FDR, false discovery rate; H-ES, healthy ex-smoker; miRNA, microRNA; mRNA, messenger RNA

2.13.1 Pairwise correlation analysis between miRNA and mRNA

Due to its simplicity and intuitive interpretation, Pearson correlation is widely used to analyse the relationships between miRNAs and mRNAs (394, 395). Therefore, Pearson correlations were generated using the Hmisc package in R (v 4.0.2) to determine the relationship between the normalised expression data of the differentially expressed miRNA and mRNA. The following code generated the correlation coefficients with adjusted P values using the Benjamini-Hochberg method.

The object mRNA_vs_miRNA contained the normalised expression data for miRNA and mRNA for each of the 44 matched patient samples.

```
>mRNA_vs_miRNA_rcorr <- rcorr(as.matrix(mRNA_vs_miRNA), type = c("pearson"))

##For R values
>mRNA_vs_miRNA_coeff <- mRNA_vs_miRNA_rcorr$r

##For p values
>mRNA_vs_miRNA_p <- mRNA_vs_miRNA_rcorr$P
```

```
##Flatten correlation matrix in order to visualise
>res2 <- rcorr(as.matrix(mRNA_vs_miRNA))
>flattenCorrMatrix <- function(cormat, pmat) {
        ut <- upper.tri(cormat)
        data.frame(
                row = rownames(cormat)[row(cormat)[ut]],
                column = rownames(cormat)[col(cormat)[ut]],
                cor  =(cormat)[ut],
                p = pmat[ut]
        )
}required to have an FEV1 % predicted of >50% to be included for a research rb
>correlationresults <- flattenCorrMatrix(mRNA_vs_miRNA_coeff, mRNA_vs_miRNA_p)
>P_values <- correlationresults[,4]

##Adjust P values using FDR correction
>Adj_P_values <- p.adjust(P_values, method = c("fdr"), n = length(P_values))
>Adj_P_values_DF <- as.data.frame(Adj_P_values)
>correlationresults_adj_P_values <- cbind(correlationresults,Adj_P_values_DF)

##Subset those results with a FDR<0.05
>sign_P <- subset(correlationresults_adj_P_values, Adj_P_values < 0.05, select = c(row,
column, cor, p, Adj_P_values))

##Filtering out correlations between miRNA-miRNA or mRNA-mRNA
>HSA_ALL <- correlationresults_adj_P_values[grep("HSA",
      correlationresults_adj_P_values$row), ]
>HSA_ALL <- HSA_ALL[grep("ENSG", HSA_ALL$column), ]
>HSA_ALL_sign_P <- subset(HSA_ALL, p < 0.05, select = c(row, column, cor, p, Adj_P_values))
```

2.13.2 Network analysis of miRNA-mRNA interaction network

The package miRmapper in R uses miRNA-mRNA predictions and a list of differentially expressed mRNAs to identify the most dominant miRNAs in the miRNA-mRNA interaction network and recognise the similarities between miRNA based on commonly regulated mRNA (396).

The following code was used in R (v 4.0.2) for this analysis, where input_1 is the miRNA-mRNA interaction matrix (generated by multiMiR – see section 2.12.4) and input_2 are the differentially expressed genes in the epithelial brushings (identified by DESeq2 analysis – see section 2.12.3).

```
>library(miRmapper)

>miRm <- miRmapper(interactions = input_1, DEgenes = input_2)

##To generate an adjacency matrix of the interaction network (SUM of miRNA targeting a specific gene)

>adjMat <- adjMat(miRm)

##To generate metrics that measure the predicted impact each miRNA has on the differentially expressed genes.

>impact <- getImpact(miRm)

##To depict the impact of each miRNA on the set differentially expressed gene targets.

>barPlot(miRm)

##To explore the similarity between the miRNA based on shared targets

>dendrogram(miRm)

>identityPlot(miRm)
```

The inputs and outputs of this workflow are summarised in Figure 2.29.

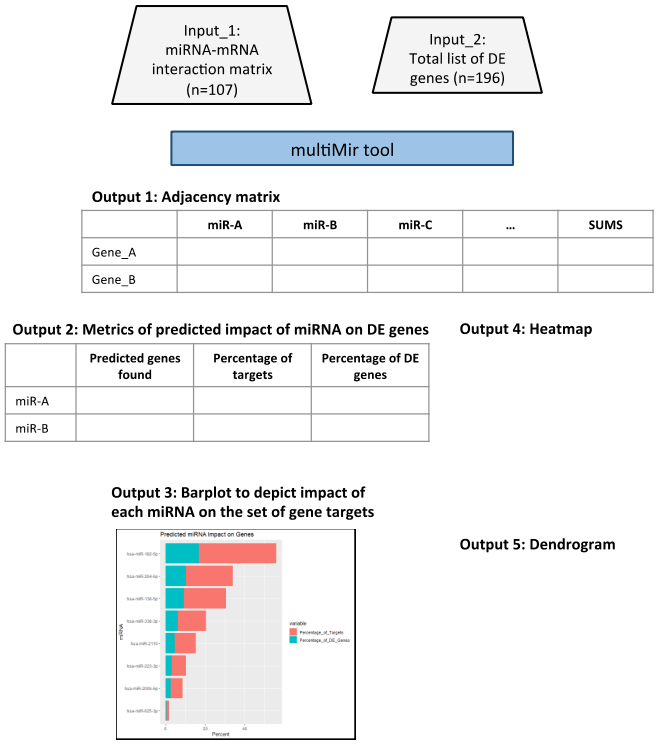


Figure 2.28 An overview of miRmapper outputs describing the miRNA-mRNA interaction network
DE, differentially expressed; miRNA, microRNA; mRNA, messenger RNA

2.13.3 miRNA-mRNA interaction network topology

Networks generated by the miRNA-mRNA interactions generated by the analysis in section 2.12.4 can be analysed further to identify a small number of miRNA-mRNA interactions that could be biologically important or related to a process in the study. Network topology analysis is a powerful way to prioritise nodes that can be important for gene network function. The methods used for this analyses are outlined by León and Calligaris in the book MicroRNA Profiling: Methods and Protocols (397).

Cytoscape is an open source software platform for visualising complex networks and integrating these with any type of attribute data (e.g. expression values) (398). The platform includes built-in tools that give basic network statistics such as:

- Node degree (number of edges incidents to the node)
- Betweeness centrality (indicator of a node's centrality in a network)
- Cluster coefficient (measure of the degree to which nodes in the network tend to cluster together).

Cytoscape was used to visual the miRNA-mRNA interaction networks for both up and down-regulated miRNA. The following node attributes were added to the network to integrate all of the analytical steps performed so far:

i. The circular nodes correspond to the differentially expressed genes identified from the methods outlined in section 2.12.3. The size of the node is proportional to the differential expression in COPD, i.e. for the up-regulated miRNA network, the larger the node, the more down-regulated the gene was in COPD epithelial brushings.

ii. The edges represent an interaction between a miRNA and mRNA. If the edge is a solid line, this represents a validated mRNA target, whereas a dotted line represents a predicted mRNA target (these targets were identified using the methods outlined in 2.12.1).

iii. The nodes coloured in orange are those identified as significant (FDR<0.05) from the correlative analysis (methods in section 2.13.1).

iv. The nodes coloured purple are those mRNA regulated by more than one miRNA and therefore have a degree centrality >1 (identified by methods outlined in section 2.13.2).

The nodes coloured yellow satisfy both iii. and iv. and therefore may be of greater significance.

2.13.3.1 Cluster analysis of networks

Cluster analysis of biological networks is one of the most important approaches for identifying functional modules and predicting downstream protein functions. I used ClusterViz, a Cytoscape plugin, to identify clusters (highly connected nodes) within the miRNA-mRNA interaction network (399). Within this tool I used the EAGLE algorithm (400), which has previously been used to analyse networks of this type (401). This approach allows the discovery of interconnected miRNA and genes, which may therefore identify those interactions with greater biological significance.

2.13.4 Gene Ontology enrichment analysis

The gene ontology (GO) project maintains a controlled hierarchical vocabulary of terms along with logical definitions to describe molecular functions, biological processes and cellular components (402). This knowledge can be applied to a given list of genes (referred to as a 'gene-set') to explore the GO terms annotating the genes and split them into functional groups ('annotation analysis'). In addition, 'enrichment analysis' can be performed by only focusing on terms significantly over-represented in the gene-set.

2.13.4.1 The Biological Networks Gene Ontology tool (BiNGO)

The Biological Networks Gene Ontology tool (BiNGO) is an open source Java-based tool to determine which GO categories are statistically overrepresented in a gene-set (403). It is implemented as a plugin for Cytoscape and maps the predominant functional themes of the gene-

set on the GO hierarchy, and outputs this mapping as a Cytoscape graph. It has been used previously to analyse functional enrichment of miRNA-mRNA networks (404). I performed the enrichment analysis for both GO biological processes and molecular functions within BiNGO using the hypergeometric statistical test, which is recommended for differentially expressed genes (405), followed by a multiple hypothesis correction by FDR ($p = 0.05$).

2.13.4.2 Enrichment Map for gene-set enrichment visualisation and interpretation

Although GO enrichment analysis is a helpful technique for high-throughput data interpretation, enrichment results are often characterized by lots of redundancy and inter-dependencies between gene-sets representing functional categories. Therefore a typical enrichment analysis can output up to 300 hundred different functional categories, which can be difficult to prioritise for further exploration. To address this, I used 'Enrichment Map', a network-based visualisation method within Cytoscape for organising and displaying the major enriched functional themes from the GO enrichment results (406).

2.14 Statistics

Statistical analyses were performed in GraphPad® Prism version 7.05 (GraphPad Software®, San Diego, USA), unless otherwise specified.

Baseline subject characteristics were summarised using standard descriptive statistics, with number and percentages for binary and categorical outcomes and appropriate measures for continuous outcomes – means and SD for normally distributed variables and medians and IQR for skewed distributions. The distribution of the data was assessed by the Shapiro-Wilk normality test, with a $p>0.05$ suggesting data is normally distributed. Welch two-sample t tests (for normally distributed data) and Mann Whitney U tests (for skewed data) were used to test whether there were significant differences in baseline subject characteristics between COPD subjects and healthy controls.

Logistic regression models were used to explore the relationship between co-variables (age, gender, smoking pack year history and lobe sampled) on the proportion of miRNA reads in COPD compared with healthy ex-smoker samples.

Receiver operative characteristic (ROC) curves were generated using the miRNA normalised expression data in SPSS® to investigate the predict ability of the differentially expressed miRNA to differentiate between health and disease.

Further specific statistical tests are defined either in the corresponding methods sub-section or where presented in the results.

Results were determined to be significant with a value of at least $P<0.05$; except when making multiple comparisons (e.g. differential expression analysis) where results were determined to be significant given a $FDR<0.05$ using the Benjamini-Hochberg correction (380).

Chapter 3 Study cohort characteristics for extracellular vesicle isolation from bronchoalveolar lavage fluid

3.1 Introduction

This chapter describes the characteristics of the study cohort used for extracellular vesicle (EV) isolation from bronchoalveolar lavage fluid (BALF). This study cohort comprises 20 subjects with Chronic Obstructive Pulmonary Disease (COPD) and 15 healthy ex-smokers and is a sub-group from the larger MICA II cohort study in which patients underwent full lung function, blood and sputum sampling, high-resolution computer tomography (HRCT) and bronchoscopy. This chapter will cover the phenotypic characteristics of this sub-group and highlight any differences that may be important to consider prior to any further analysis exploring underlying disease mechanisms.

3.2 Characteristics of the subjects included in EV isolation from BALF

BALF samples from 20 COPD subjects and 15 healthy ex-smokers in the MICA II research cohort were processed for EV isolation using size exclusion chromatography (SEC). The baseline characteristics for this group are summarised in Table 3.1.

Table 3.1 Characteristics of subjects included in BALF EV RNA isolation for miRNA sequencing, n=35

Subject/sample characteristics	COPD (n=20)	Healthy ex-smoker (n=15)	P value
Age, mean ±SD	70.4 ±6.8	66.9 ±7.8	0.18
Male, n (%)	17 (85)	9 (60)	0.13
Smoking pack years, median (IQR)	40 (46.4)	25 (19)	0.16
BMI, mean ±SD	29.3 ±4.3	28.4 ±4.4	0.12
Lung Physiology			
FEV1 (% predicted), mean ±SD	79.9 ±13.9	99.5 ±14.2	**0.0003**
FVC (% predicted), mean ±SD	99.7 ±14.9	104.5 ±16.6	0.34
FEV1/FVC%, mean ±SD	58.6 ±8.0	77.3 ±3.2	**<0.0001**
FEF 25-75 (% predicted), mean ±SD	44 ±16.4	98.1 ±21	**<0.0001**
DLCO (% predicted), mean ±SD	75.8 ±14.9	91.1 ±12.6	**0.004**
COPD status, GOLD stage, n (%)			0.45
Mild	9 (45)	NA	
Moderate	11 (55)	NA	
Baseline & historic blood counts			
Total blood leucocytes (10^9/L), mean ±SD	7.3 ±1.4	6.7 ±1.3	0.13
Absolute neutrophil count (10^9/L), mean ±SD	4.6 ±1.2	3.9 ±1.0	0.06
Absolute eosinophil count (10^9/L), median (IQR)	0.2 (0.2)	0.1 (0.1)	0.1
Historic eosinophils (10^9/L), median (IQR)	0.35 (0.2)	0.1 (0.1)	**<0.0001**
HRCT measurements			
E/I MLD, mean ±SD	0.8 ±0.04	0.81 ±0.05	**0.02**
$\%LAA_{<-950,}$ mean ±SD	10.4 ±5.1	7.3 ±4.9	0.08

Fisher's exact test was performed for Gender given small sample size. Chi-squared test used for COPD status.
Shapiro-Wilk test for normality was performed for all continuous variables.
Welch two sample t test was performed for normally distributed data; Age, BMI, FEV1, FVC, FEV1/FVC and FEF 25-75, TLCO, RV/TLC SR, total blood leucocytes**,** absolute neutrophil count, E/I MLD and $\%LAA_{<-950}$.
Mann-Whitney U test was performed for skewed data; smoking pack years and eosinophil blood counts.
BMI, body mass index; FEV1, forced expiratory volume in 1 sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; DLCO, diffusion capacity of the lung for carbon monoxide; E/I MLD, ratio of mean lung attenuation on expiratory and inspiratory scans; Historic eosinophil count, highest ever recorded eosinophil measurement; HRCT, high resolution computer tomography; %LAA<-950, percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units; IQR, interquartile range; NA, non-applicable; SD, standard deviation.

The subjects were matched for age, sex, smoking history and body mass index (BMI) (Table 3.1). There were however more males in both groups. As expected post-bronchodilator forced expiratory volume in one second (FEV1) % predicted, FEV1/forced vital capacity (FVC), forced expiratory flow rate (FEF) 25-75 % predicted and diffusion capacity of the lung for carbon monoxide (DLCO) % predicted were all significantly reduced in the COPD group. The COPD subjects were balanced between mild (45%) and moderate (55%) GOLD status.

3.2.1 Baseline & historic blood count

Full blood count with cell differential were analysed for all subjects prior to bronchoscopy. Blood eosinophils are an important biomarker in COPD as they are seen as a useful marker of future exacerbation risk and have a role in predicting steroid responsiveness (407). Although baseline levels of blood eosinophils were not different between COPD subjects and healthy ex-smokers (median 0.2 and 0.1 respectively), historic blood eosinophil counts were significantly higher in the COPD subjects than the healthy controls (Table 3.1, p <0.0001), where historic blood eosinophil count refers to the highest ever-recorded eosinophil measurement.

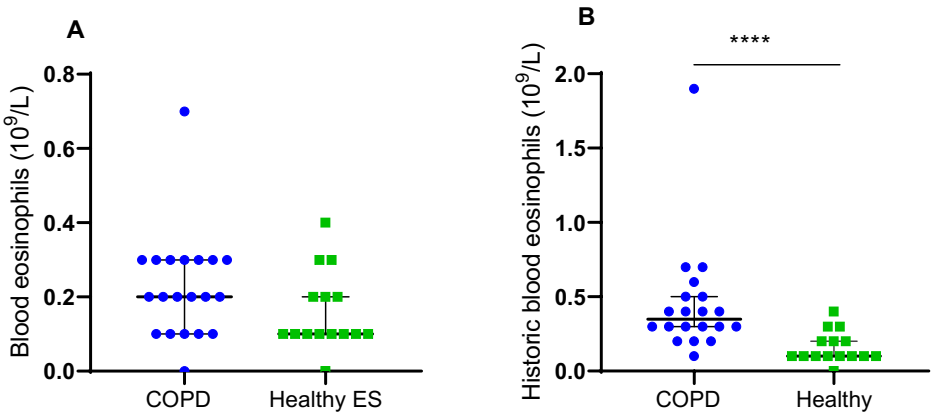


Figure 3.1 Baseline (A) and historic (B) blood eosinophil counts for COPD subjects compared with healthy ex-smokers
Median and IQR presented as skewed data. ****p<0.0001 using Mann Whitney test

When examining the variability of eosinophil counts across the COPD subjects, analysis showed one subject with COPD with higher levels of eosinophils both at baseline (absolute eosinophil count 0.7 10^9/L, Figure 3.1 A) and in historic blood eosinophil levels (absolute eosinophil count 1.9 10^9/L, Figure 3.1 B). Interestingly, this was a different COPD subject in each case, demonstrating the variability of eosinophil levels in the blood over time. After excluding these as possible outliers, the significance remained when comparing historic blood eosinophil expression in COPD subjects with health ex-smokers (p<0.001).

There were no differences in total blood leucocyte count or absolute neutrophil count between the two groups (see Table 3.1).

3.2.2 HRCT measurement

Quantitative assessment of gas trapping and emphysema were performed using HRCT analysis. The ratio of mean lung attenuation on expiratory and inspiratory scans (E/I MLD), a surrogate

marker for gas trapping (as discussed in section 2.4.4), was found to be significantly lower in COPD subjects compared with healthy controls (mean 0.8 (SD±0.04), 0.81 (SD±0.05) respectively, p = 0.02). E/I MLD also significantly negatively correlated with spirometric markers of airflow obstruction (FEV1/FVC, r = -0.43, p = 0.008), small airways disease (FEF 25-75%, r = -0.47, p = 0.007) and disease severity (FEV1%, r = -0.48, p = 0.006) across the whole cohort, but not in COPD subjects alone (see Table 3.2). The measure percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units ($\%LAA_{<-950}$) did not differ between the two groups (mean 10.4 (SD±5.1) and 7.3 (SD±4.9) in COPD and health ex-smokers respectively), and only correlated with FEV1/FVC (r = -0.39, p = 0.03) and FEF 25-75% predicted (r = -0.36, p = 0.03) across the whole study cohort. Similar to E/I MLD, $\%LAA_{<-950}$ did not correlate with any spirometric markers or DLCO in COPD subjects alone (Table 3.2).

Table 3.2 Correlation between HRCT measures of small airways disease and emphysema and physiology measures of disease

	FEV1 %	FVC %	FEV1/FVC	FEF 25-75%	DLCO
Whole cohort, n= 35					
E/I MLD	**-0.48****	-0.13	**-0.43****	**-0.47****	-0.26
$\%LAA_{<-950}$	-0.21	0.24	**-0.39***	**-0.36***	-0.23
COPD Subjects, n= 20					
E/I MLD	-0.3	0.07	-0.43	-0.37	-0.18
$\%LAA_{<-950}$	-0.1	0.41	-0.32	-0.24	-0.27

Spearman's correlation coefficient. N=35, whole cohort. N=20, COPD subjects. *p<0.05, **p<0.01
FEV1, forced expiratory volume in 1 sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; DLCO, diffusion capacity of the lung for carbon monoxide; E/I MLD, ratio of mean lung attenuation on expiratory and inspiratory scans; %LAA<-950, percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units

3.2.3 BALF count analysis

BALF differential cell counts were analysed in all 20 COPD subjects and 15 healthy ex-smokers (Figure 3.2). For this analysis the differential cell count from the same lobe as was sampled for the EVs was used. There were significantly increased neutrophils and eosinophils in the BALF of COPD subjects compared with the healthy ex-smokers (p = 0.02 and p = 0.04 respectively – Figure 3.2 A and C). As expected macrophages were the dominant cell type in the BALF (median 67.4% across the whole cohort), however there were no differences in proportions between COPD subjects and healthy ex-smokers (Figure 3.2 B). Relative numbers of lymphocytes in BALF were low and there were no significant differences between groups (Figure 3.2 D).

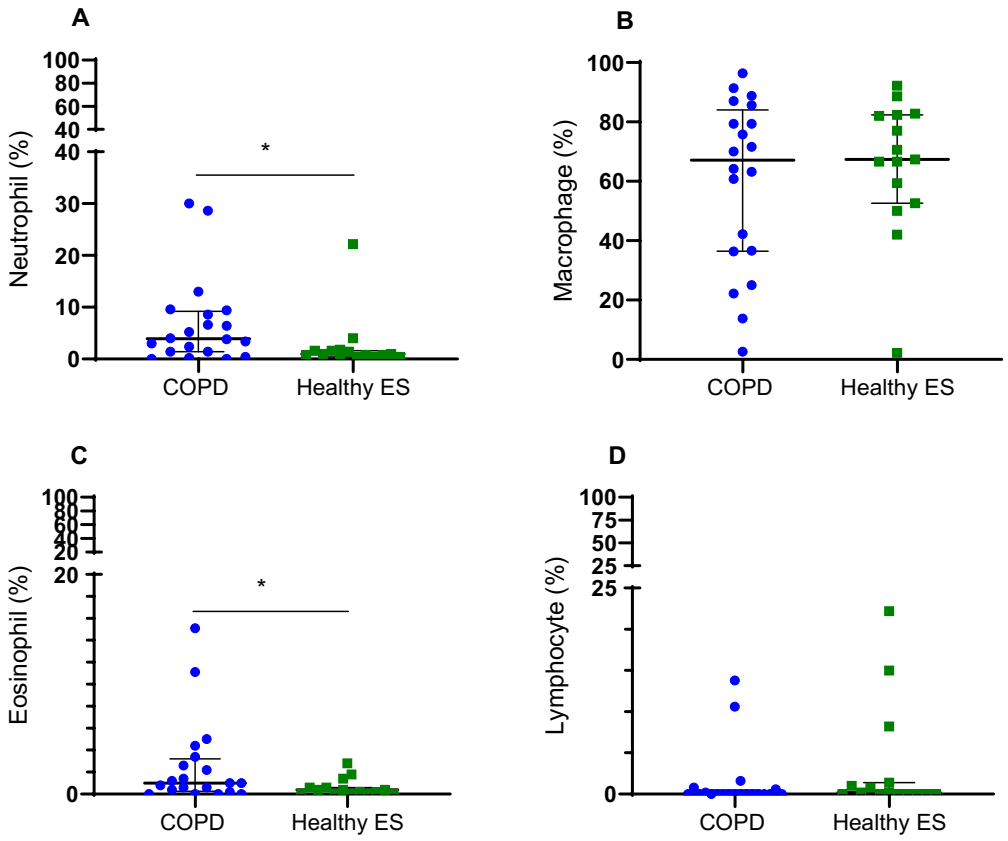


Figure 3.2 BALF expression of immune cells in COPD subjects and healthy ex-smokers.
(A) Neutrophil, (B) Macrophage, (C) Eosinophil, (D) Lymphocyte. Data represents median with interquartile range. Each dot represents BALF concentration of individual value in a specific patient. N=35; COPD, n=20. *p<0.05 using Mann-Whitney U test. COPD, Chronic Obstructive Pulmonary Disease; Healthy ES, healthy ex-smoker.

Given the differences in both blood and BALF eosinophil expression in COPD compared with healthy ex-smokers, I further investigated whether these two measures correlated across the whole cohort and in COPD subjects alone (Table 3.3).

Table 3.3 Correlation between blood and BALF eosinophil expression

	Absolute blood eosinophil count (10^9/L)	Historic blood eosinophil count (10^9/L)
Whole cohort, n=35		
BALF % eosinophils	0.2	0.29
COPD subjects, n = 20		
BALF % eosinophils	-0.17	-0.13

Spearman's correlation coefficient. N=35, whole cohort. N=20, COPD subjects. BALF, bronchoalveolar lavage fluid; COPD, Chronic Obstructive Pulmonary Disease.

None of the blood eosinophil counts correlated with BALF eosinophil expression in either the whole cohort or in COPD subjects alone. Although historic blood eosinophil count showed a possible weak correlation with BALF % eosinophils approaching significance ($r = 0.29$, $p = 0.08$).

3.3 Discussion

In this chapter I have outlined the subject characteristics of my study cohort, which underwent BALF EV isolation. Subjects enrolled in the study were fully phenotypically characterised with measures of inflammation, small airways disease and emphysema analysed.

3.3.1 Subject characteristics

The COPD patients sampled in this study had predominantly mild disease as determined by Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage (4), with a mean FEV1 % predicted of 79.9% (SD±13.9). This was primarily as a result of included subjects undergoing a fibreoptic bronchoscopy, whereby the safety FEV1 % predicted cut-off was set at 50%. A further reflection of the relatively mild cohort of COPD patients sampled in this study are the data from the HRCT analysis, which showed very little established emphysema, with no significant difference in $\%LAA_{<-950}$ in COPD subjects compared with healthy controls. These phenotypic characteristics are important to note, as any differential BALF EV microRNA (miRNA) expression discovered during subsequent analysis, would therefore reflect relatively mild disease changes. The lack of HRCT-defined emphysema in these patients has implications when determining the mechanistic effects of these dysregulated miRNA, as early epigenetic regulation of key inflammatory pathways may lead to novel understanding of the temporal nature of COPD pathogenesis. Furthermore, any differences in EV miRNA expression in early disease could have a future diagnostic role, particularly given the issues with spirometry as a relatively insensitive test in early or pre-clinical disease (142).

Despite not showing any significant difference levels of emphysema (as measured by $\%LAA_{<-950}$), there was a significant difference in presence of small airways disease in subjects with COPD with a reduced FEF 25-75 % predicted and E/I MLD ratio in subjects with COPD. As discussed in section 1.2, small airways disease is a cardinal feature of COPD and there is evidence that it occurs early in the natural history of COPD, with the narrowing and destruction of small airways appearing to precede the development of emphysema (408-410). Therefore targeting treatment towards small airways disease has the potential to treat both the progression of airway and parenchymal disease, although this remains to be proven. Therefore, by investigating mechanisms of disease in

this relatively mild cohort of patients with COPD, it may be possible to uncover new insights into the role of EV miRNA in the development of small airways disease and may identify targets that prevent disease progression.

As previously discussed in section 1.7.4, there are no published studies studying BALF EVs in the lungs of COPD patients. Therefore, it is difficult to directly compare my sample cohort to others in the literature. Those that have studied endothelial microvesicles (EMVs) in blood and sputum have included more severe COPD patients, with a mean FEV1 % predicted range of 52.4 – 71.6% (SD±8.51 - 40.9%) (247, 250, 260). These findings are applicable to a wider COPD cohort, however less applicable mechanistically given their broader range of included subject phenotypes. Similarly to my study, the subjects included in these previous studies were predominantly male (73.5%), which historically, reflects the global increased prevalence of COPD in men (2). Importantly, my groups were matched for gender (p=0.13) and therefore the male predominance should not affect the differential expression analysis of EV miRNA between the two groups.

All subjects included in my study were ex-smokers, defined as smoking cessation for at least six months prior to enrolment. This is contrary to previous EV studies, where cohorts have comprised a mixture of current, ex- and non-smokers (247, 250, 254, 260). My study deliberately excluded current smokers and non-smokers as my main aim was to investigate the effect of COPD on differential EV miRNA expression, rather than smoke exposure. Importantly, all subjects had at least a 10 pack-year history of smoking and again, groups were matched (p=0.16) so as not to effect differential EV miRNA expression analysis.

3.3.2 Blood eosinophilia

Historic blood eosinophil counts (referring to highest ever-recorded eosinophil measurement) were significantly higher in COPD subjects than healthy controls. This finding is consistent with the literature that describes around 60% of COPD patients have blood eosinophil counts of ≥2% (150). However, this difference was not shown at baseline absolute eosinophil measurement (Table 3.1, p = 0.1). This is in-keeping with the concept that blood eosinophils counts are variable over time and therefore is important to make repeated measurements (172).

Recognising eosinophilia in COPD has been shown to be important in predicting future exacerbation risk and those who respond to corticosteroid treatment (149, 168, 407). However, defining eosinophilia in COPD is challenging. As discussed in section 1.5.2, although blood eosinophils are easy to measure, there is conflicting evidence as to whether they correlate with eosinophil levels within the lungs of patients with COPD (169-171, 411). Data modelling suggests

that there is a continuous relationship between blood eosinophil counts and inhaled corticosteroid (ICS) effects, with ICS treatment having little or no effect at a blood eosinophil count of < 100 cells/µL, and maximal effect at a blood eosinophil count of >300 cells/µL (407). Therefore these thresholds are used as a guide in clinical practice to help estimate the likely beneficial preventive response to the addition of ICS to regular bronchodilator treatment (4).

Furthermore, the mechanism of eosinophilia in COPD is not yet certain and targeted eosinophilic treatments (such as anti-IL-5 therapies) have had limited success (178, 179). Therefore, by investigating the role of EVs in this context, whereby eosinophils may be a target or source of EVs, may have significant implications on their role in COPD pathophysiology and uncover new treatment targets.

3.3.3 Airway inflammatory cell profile

As discussed previously (section 1.7.3), epithelial cells have been shown to be one of the main EV sources in the lung (233, 412). However, studies have shown that all innate immune cells can release, and be regulated by EVs (413). Specifically, alveolar macrophages are known to release pro-inflammatory EVs in response to cigarette smoke (238-241) and these may have important consequences in driving persistent airway inflammation and vulnerability to infection in COPD (243). Furthermore, eosinophil-derived EVs have been implicated in asthma (414), and direct transport of neutrophil EVs are key in the innate immune response to eliminate bacterial infection (415). Therefore, when determining the role of EV miRNA in COPD pathogenesis, it is important to consider the role and abundance of these different cell types within the airways of patients with COPD, given they could be a key target or source of EVs and their cargo.

Proportions of neutrophils and eosinophils were significantly raised in COPD subjects compared with healthy controls, which is in keeping with previous studies (416, 417). This likely reflects the increased airway inflammation and tissue damage associated with disease (5). Although BAL eosinophils were elevated in COPD subjects, levels did not correlate with blood eosinophilia (Table 3.3). These findings reflect current literature, which shows only a weak relationship between blood and sputum or bronchial eosinophils in COPD patients (170, 171). Given this poor relationship, when investigating the impact of EV miRNA on specific inflammatory pathways in COPD, it may be important to distinguish between those patients with COPD who demonstrate peripheral eosinophilia compared with those with airway eosinophilia, as the mechanisms underlying the two are likely to be different.

Interestingly, proportions of macrophages were similar between the two groups, which is contrary to previous studies (74). This similarity in macrophage proportions may, in part, be due to the increased numbers of neutrophils and eosinophils in the COPD samples leading to a relative reduction in the proportion of macrophages in the COPD subjects. In addition, as highlighted above, this study cohort included patients with COPD with relatively mild disease (mean FEV1 % predicted 79.9%) and numbers of macrophages in the airways are known to negatively correlate with disease severity (56). Consistent with previous work, macrophages were the predominant cell type in the airways (median proportion 67.4%, IQR 36.2) (73). However the IQR for the proportion of macrophages found in COPD BALF samples were wide (Figure 3.6 B), suggesting a large degree of heterogeneity.

3.3.4 Strength and Limitations

As mentioned above, the 35 subjects (20 patients with COPD and 15 healthy ex-smokers) included in this study of BAL EV miRNA characterisation were part of a larger study investigating the inflammatory mechanisms within COPD (the MICA II study). By using patients from this larger study I was able to access and utilise the extensive phenotyping data for my analysis that included physiological, HRCT and inflammatory measurements. Importantly, this in-depth profiling gives unique insights into specific disease characteristics present in this cohort. For example, this analysis identified the presence of small airways disease (as defined by low FEF 25-75% predicted and low E/I MLD), but not emphysema in the subjects with COPD. Furthermore, it has provided insight into the heterogeneity of inflammation within this group, with particular reference to both airway and blood eosinophilia. Using this information will be key in determining the impact of any differential EV miRNA expression on disease relevant pathways.

One of the limitations of this study is that the cohort is relatively small (n=35), particularly when considering the heterogeneous nature of COPD. Ideally, I would have included all those recruited to the MICA II study (n=51), however I was limited by the volume of BALF available for EV isolation. As outlined in my methods (section 2.5), I required a minimum of 15 mL of BALF per subject, which was not possible for all, particularly in the case of patients with COPD, where recovery of BALF was more difficult due to airway closure (likely as a result of small airways disease). To try and mitigate this, I was able to validate the findings from my initial microRNA sequencing in a slightly larger group of subjects (n=46 – see section 4.4).

A further limitation is the sampling method used in this study. Bronchoscopy is an invasive procedure, with limits on sample availability and on subject inclusion, with COPD subjects

requiring an FEV1 % predicted of greater than 50% due to safety concerns. COPD is a disease characterised by pulmonary inflammation and therefore sampling the airways to look for novel mechanisms of disease seems logical. However if EVs and/or their contents are to be used as diagnostic or therapeutic biomarkers (as suggested in section 1.7.4), these findings will need translating into more readily available biofluids (such as blood), and this will be key to determining EVs utility in this context.

3.3.5 Summary

In summary, to my knowledge this is the first study to isolate and characterise EVs from BALF of COPD patients compared with healthy ex-smokers. Subjects across the two groups were well-matched for non-disease defining characteristics, which is an important consideration for the differential expression analysis used for the comparison of EV miRNA expression in COPD compared with healthy ex-smokers in the next chapter. In addition, the in-depth phenotypic characterisation of the COPD subjects identifying disease specific characteristics, such as presence of small airways disease and eosinophilia, will be important when exploring the involvement of the differentially expressed miRNA in the regulation of known/novel disease mechanisms given the heterogeneity of the disease.

Chapter 4 MicroRNA sequencing of bronchoalveolar lavage extracellular vesicles and validation of the results

4.1 Introduction

Extracellular vesicles (EVs) were isolated from bronchoalveolar lavage fluid (BALF) from patients with Chronic Obstructive Pulmonary Disease (COPD) (n=20) and healthy ex-smoker controls (n=15), using ultrafiltration and size exclusion chromatography (SEC) as described in section 2.5. RNA was isolated from the EV fraction and prepared for sequencing, as described in sections 2.7.2 and 2.8.1 respectively. Next generation microRNA (miRNA) sequencing was performed on the NextSeq500 instrument (illumina®) (described in section 2.8.3) and the resulting FASTQ files were trimmed and corrected using the unique molecular index (UMI) methods described in section 2.9.1. Raw reads were aligned and mapped to the reference genome, Genome Reference Consortium Human Build 37 (section 2.9.3). Unsupervised filtering and exploratory data analysis (outlined in section 2.10.1 and 2.10.2 respectively) resulted in the ***"counts per million (CPM) filtered dataset***" (n=31, with outliers removed) that was used for differential expression analysis of EV miRNA between COPD and healthy ex-smokers.

Chapter 4 describes the subject characteristics of the refined cohort (n=31) and miRNA sequencing mapping and alignment results for the ***CPM filtered dataset***. The main findings covered in this chapter are the differential expression analysis of the BALF EV miRNA comparing COPD with healthy ex-smokers and the results from the validation study using real-time quantitative polymerase chain reaction (RT-qPCR) in a larger cohort (n=46).

4.1.1.1 Characteristics of the subjects included in differential expression analysis of BALF EV miRNA

After exploratory data analysis identified a number of outliers (see section 2.10.2), a final **CPM filtered dataset** was used for the differential expression analysis consisting of 17 COPD subjects and 14 healthy ex-smokers. The subject characteristics are summarised in Table 4.1.

Table 4.1 Characteristics of subjects included in differential expression analysis of BALF EV miRNA, n=31

Subject/sample characteristics	COPD (n=17)	Healthy ex-smoker (n=14)	P value
Age, mean ±SD	69.6 ±7.2	66.2 ±7.6	0.18
Male, n (%)	15 (88)	9 (57)	0.13
Smoking pack years, mean ±SD	45.8 ±25.8	25.8 ±13.7	0.07
BMI, mean ±SD	29.8 ±4.1	28.5 ±4.6	0.39
FEV1 (% predicted), mean ±SD	78.9 ±14.4	100.6 ±14	0.0007
FVC (% predicted), mean ±SD	102.2 ±14.8	101.3 ±14.1	0.87
FEV1/FVC%, mean ±SD	59.1 ±8.2	76.9 ±2.9	<0.0001
FEF 25-75 (% predicted), mean+/-SD	44.1 ±17.2	97.7 ±22.2	<0.0001
COPD status, GOLD stage, n (%)			0.3
Mild	7 (41)	NA	
Moderate	10 (59)	NA	

Fisher's exact test for Gender given small sample size. Chi-squared test used for COPD status. Shapiro-Wilk test for normality was performed for all continuous variables. Welch two sample t test was performed for normally distributed data; Age, BMI, FEV1, FVC, FEV1/FVC and FEF 25-75. Mann-Whitney U test was performed for skewed data; smoking pack years. BMI, body mass index; FEV1, forced expiratory volume in one sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; NA, non-applicable; SD, standard deviation.

The subjects were matched for age, sex, smoking history and body mass index (BMI) (Table 4.1). There were however more males in both groups. As expected forced expiratory volume in one sec (FEV1) % predicted, FEV1/Forced vital capacity (FVC) and forced expiratory flow rate (FEF) 25-75% predicted were significantly reduced in the COPD group. The COPD subjects were balanced between mild (41%) and moderate (59%) GOLD status.

Given the groups were well-matched apart from disease defining characteristics (i.e. FEV1 % predicted), differential miRNA expression analysis was performed comparing between COPD and healthy ex-smokers without adjustment for multiple variables.

4.2 Mapping and alignment results comparing patients with COPD and healthy ex-smokers

4.2.1 Total number of reads

The mean total number of reads mapped for the dataset is 2,800,633 (SD±695,277) and the average genome mapping rate was 52.7%. When comparing COPD against healthy ex-smoker samples, there was no statistical difference between the total numbers of reads sequenced (Figure 4.1).

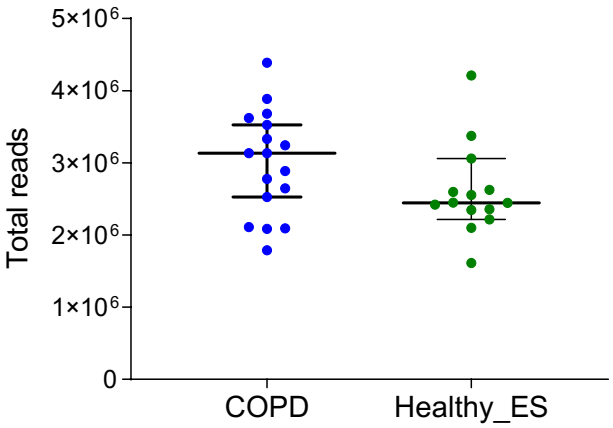


Figure 4.1 Total number of reads for COPD and healthy ex-smoker samples
Median and 95% confidence intervals shown. N=31; COPD, n=17. Mann Whitney U test showed no significant difference. COPD, Chronic obstructive pulmonary disease; Health-ES, healthy ex-smoker.

4.2.2 Proportion of miRNA and smallRNA mapped reads in COPD and healthy ex-smokers

When comparing different types of reads mapped in COPD and healthy ex-smoker samples, there is a higher proportion of miRNA in COPD samples than in healthy ex-smoker samples, both when including unmapped reads (p=0.03) (Figure 4.2 A) and without unmapped reads (p=0.02) (Figure 4.2 B).

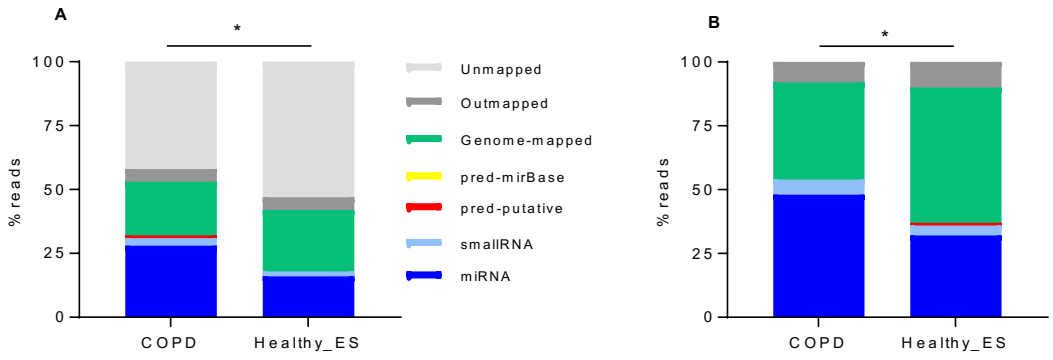


Figure 4.2 Proportions of different types of reads in COPD and healthy ex-smoker samples
A includes unmapped reads and B excludes unmapped reads. N=31; COPD, n=17. Chi-squared test performed on proportion of miRNA present in two groups, *p<0.05. COPD, Chronic obstructive pulmonary disease; Health-ES, healthy ex-smoker; miRNA, microRNA; Pred, predicted

Logistic regression was used to look at the effect of co-variables (age, gender, smoking pack year history and lobe sampled) on the proportion of miRNA reads in COPD compared with healthy ex-smoker samples. The model explained 47-63% (Cox&Snell R^2 model - Nagelkerke R^2 model) of the variance in COPD and correctly classified 83.9% of cases. Higher miRNA read % was the only variable significantly associated with the presence of COPD (p=0.02) (Table 4.2).

Table 4.2 Logistic regression of proportions of miRNA reads in COPD and healthy ex-smokers

Variable	OR (95% CI)*	P Value
Age in years	1.05 (0.9-1.2)	0.5
Smoking pack years	1.03 (0.9-1.1)	0.3
Gender	2.7 (0.02-30)	0.4
Lobe sampled	1.8 (0.03-2.4)	0.2
miRNA read %	**1.2 (1.02-1.3)**	**0.02***

*Without unmapped reads included. Calculated in SPSS. CI, confidence interval; miRNA, microRNA; OR, Odds ratio

In addition, the proportion of smallRNA mapped reads was significantly higher in COPD than healthy ex-smokers (p=0.01) (Figure 4.3 A). However there was no difference in the different types of smallRNA between health and disease (Figure 4.3 B). As previously discussed in section 2.9.3.1, tRNA was the most abundant (92.2%) smallRNA found in my EV samples and is integral to protein synthesis by helping decode mRNA into protein.

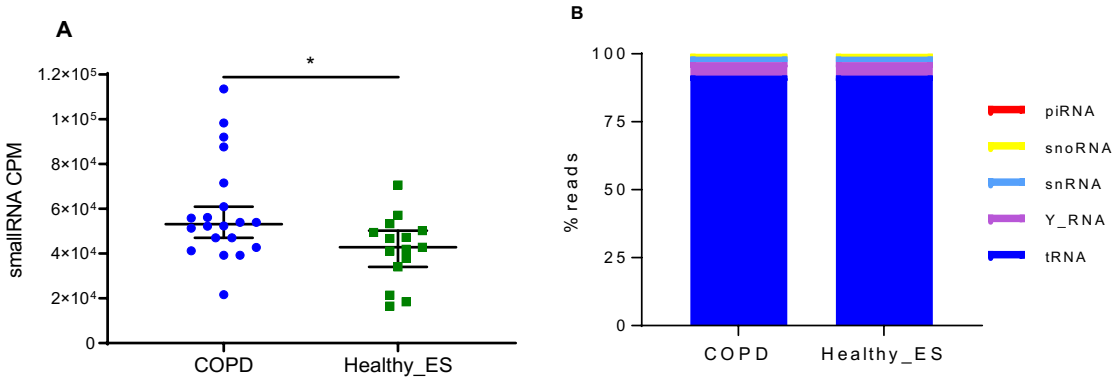


Figure 4.3 Total number (A) and proportions (B) of smallRNA for COPD and healthy ex-smoker samples

A. Median and 95% confidence intervals shown. N=31; COPD, n=17. Mann Whitney U test performed, *p <0.05. CPM, counts per million; COPD, Chronic obstructive pulmonary disease; Health-ES, healthy ex-smoker.

B. Proportions of types of smallRNA. N=31; COPD, n=17. Chi-squared test showed no significant difference. piRNA, piwi-interacting RNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; tRNA, transfer RNA.

Together these results suggest that EVs in COPD package a higher proportion of miRNA (Figure 4.2 A, p = 0.02) and smallRNA (Figure 4.3 A, p = 0.01) compared with healthy ex-smokers.

4.3 Differential expression of EV miRNA between COPD subjects and healthy ex-smokers

Differential expression analysis was performed using methods outlined in section 2.10.5. Given the comparator groups were well-matched apart from disease defining characteristics (i.e. FEV1 % predicted), an unadjusted analysis was performed comparing differential miRNA expression between COPD and healthy ex-smokers. The results are summarised in the MA plot in Figure 4.4, showing 54 significantly differentially expressed miRNA with p value adjustment using fold discovery rate (FDR) correction, whereby by significance is demonstrated with an FDR of less than 0.05.

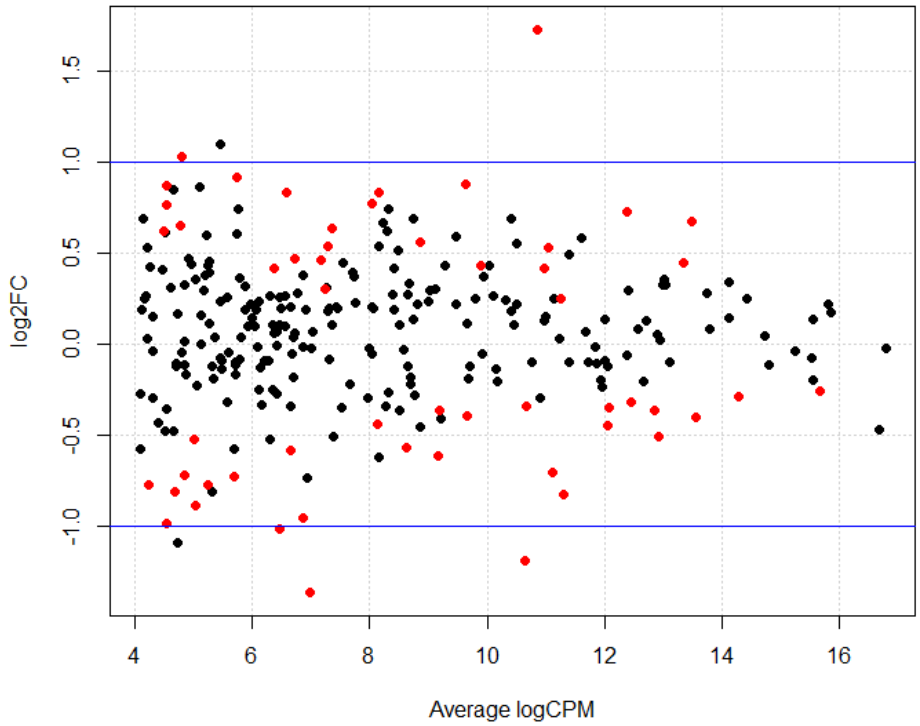


Figure 4.4 MA plot showing differentially expressed miRNA between COPD subjects and healthy ex-smokers
Red dots represent miRNA with an FDR <0.05. Blue lines represent a twofold change in expression. CPM, counts per million; FC, fold change

Of the 54 significantly differentially expressed miRNA, two miRNA were upregulated in COPD (miR-223-3p and miR-200b-5p) and three miRNA were downregulated in COPD (miR-138-5p, mIR-338-3p and miR-204-5p) by a log2 fold change (log2FC) of greater than one when compared with healthy ex-smokers (Table 4.3).

Table 4.3 Top differentially expressed miRNA between COPD subjects and healthy ex-smokers

miRNA	Expression in COPD	Log2FC	Average logCPM	P value	FDR
miR-223-3p	Up	1.73	10.87	<0.0001	0.006
miR-200b-5p	Up	1.03	4.79	<0.01	0.048
miR-138-5p	Down	-1.02	6.45	<0.001	0.007
miR-338-3p	Down	-1.18	10.65	<0.001	0.007
miR-204-5p	Down	-1.36	6.97	<0.01	0.019

CPM, counts per million; FC, fold change; FDR, false discovery rate; miRNA, microRNA;

4.4 Validation of differentially expressed miRNA with RT-qPCR

In order to confirm that the differentially expressed miRNA identified from miRNA sequencing were truly up or down-regulated in COPD, a second method was employed to confirm miRNA expression levels. This is an important step, especially when there are no biological replicates available (i.e. only one sample per patient analysed). Ideally this validation step would be performed in a separate cohort of subjects, however, due to sample availability this was not possible. However, an additional six COPD and five healthy ex-smoker BALF samples were used from the MICA II cohort to increase the total sample size, giving a total of 46 samples.

4.4.1 Characteristics of subjects used for differential expressed EV miRNA validation by RT-qPCR, N=46

A total of twenty-six COPD subjects and twenty healthy ex-smokers were included in the miRNA target validation by RT-qPCR. The subjects were matched for age, sex, smoking history and BMI (Table 4.4). There were more males in both groups. As expected FEV1% predicted, FEV1/FVC and forced FEF 25-75% predicted was significantly reduced in the COPD group. The COPD subjects were balanced between mild (41%) and moderate (59%) GOLD status.

Table 4.4 Characteristics of subjects included in miRNA validation by RT-qPCR, n= 46

Subject/sample characteristics	COPD (n=26)	Healthy ex-smoker (n=20)	P value
Age, mean ±SD	70.3 ±6.8	68 ±7.3	0.28
Male, n (%)	22 (85)	11 (55)	0.06
Smoking pack years, mean ±SD	46 ±28.8	27.8 ±13	0.06
BMI, mean ±SD	29.1 ±4.4	28.4 ±4	0.57
FEV1 (% predicted), mean ±SD	77.4 ±14.4	101.8 ±14.6	<0.0001
FVC (% predicted), mean ±SD	103.1 ±15.6	100.6 ±16.4	0.59
FEV1/FVC%, mean ±SD	57.4 ±8.6	78.2 ±4.2	<0.0001
FEF 25-75 (% predicted), mean ±SD	41 ±16.7	106 ±25.4	<0.0001
COPD status, GOLD stage, n (%)			0.27
Mild	11 (42)	NA	
Moderate	15 (57)	NA	

Fisher's exact test for Gender given small sample size. Chi-squared test used for COPD status. Shapiro-Wilk test for normality was performed for all continuous variables. Welch two sample t test was performed for normally distributed data; Age, BMI, FEV1, FVC, FEV1/FVC and FEF 25-75. Mann-Whitney U test was performed for skewed data; smoking pack years. BMI, body mass index; FEV1, forced expiratory volume in one sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; NA, non-applicable; SD, standard deviation.

4.4.2 MiRNA chosen for validation by RT-qPCR

The top 35 targets identified from the differential expression analysis were chosen for validation by RT-qPCR (Table 4.5). Targets were chosen preferentially based on the highest log2FC (thus possibly the most biologically relevant), and also on availability of miRNA assays available for RT-qPCR.

Table 4.5 List of miRNA targets validated by RT-qPCR

Assay	Log2FC	logCPM	P Value	FDR
Upregulated in COPD based on sequencing results				
hsa-miR-223-3p	1.74	10.87	<0.00001	<0.001
hsa-miR-223-5p	1.72	4.34	<0.00001	<0.001
hsa-miR-296-3p	1.05	3.69	<0.0001	0.01
hsa-miR-20b-5p	1.01	4.79	<0.01	0.048
hsa-miR-27b-3p	0.91	5.73	<0.001	0.005
hsa-miR-31-5p	0.89	9.64	<0.001	0.01
hsa-miR-2110	0.83	8.15	<0.0001	0.002
hsa-miR-769-3p	0.79	4.54	<0.01	0.04
hsa-miR-185-5p	0.78	8.04	<0.001	0.005
hsa-miR-146a-5p	0.72	12.38	<0.01	0.03
hsa-miR-191-5p	0.69	13.50	<0.0001	0.004
hsa-miR-25-5p	0.66	4.76	<0.01	0.04
hsa-miR-345-5p	0.65	7.36	<0.01	0.02
hsa-miR-200b-5p	0.57	8.87	<0.00001	0.004
hsa-miR-182-5p	0.53	11.03	<0.000001	0.001
hsa-miR-625-3p	0.55	7.28	<0.01	0.04
hsa-miR-589-5p	0.66	4.48	<0.05	0.048
Downregulated in COPD based on sequencing results				
hsa-miR-138-5p	-1.02	6.45	<0.0001	0.002
hsa-miR-338-3p	-1.18	10.65	<0.0001	0.002
hsa-miR-204-5p	-1.36	6.97	<0.001	0.009
hsa-miR-181a-5p	-0.70	11.11	<0.001	0.005
hsa-miR-20a-5p	-0.61	9.16	<0.001	0.007
hsa-miR-181d-5p	-0.72	4.85	<0.01	0.03
hsa-miR-301a-3p	-0.73	5.70	<0.001	0.01
hsa-miR-17-3p	-0.76	4.24	<0.01	0.046
hsa-miR-181c-5p	-0.80	4.68	<0.01	0.02
hsa-miR-30b-5p	-0.81	11.31	<0.001	0.009
hsa-miR-934	-0.88	5.04	<0.001	0.008
hsa-miR-30d-3p	-0.96	6.86	<0.001	0.006
hsa-miR-374a-5p	-0.56	8.63	<0.001	0.008
hsa-miR-452-5p	-0.52	4.99	<0.01	0.04
hsa-miR-138-1-3p	-0.98	4.54	<0.01	0.048
hsa-miR-374b-5p	-0.58	6.65	<0.01	0.048
hsa-miR-30e-5p	-0.35	12.08	<0.05	0.048

FC, fold change; CPM, counts per million; FDR, false discovery rate; miR, microRNA

In addition, the top six miRNA that were most stably expressed across both COPD and healthy ex-smokers were chosen as normalisers (Table 4.6). These were identified using the NormFinder® software (379).

Table 4.6 miRNA identified from NormFinder analysis of miRNA sequencing data as those most stably expressed across COPD and healthy ex-smoker samples

Assay	Stability	CPM average
hsa-miR-24-3p	0.06	1636
hsa-miR-23b-3p	0.14	1732
hsa-miR-27b-3p	0.16	3185
hsa-miR-93-5p	0.17	1496
hsa-miR-221-3p	0.18	1209
hsa-let-7g-5p	0.18	1736

Stability is the measure of stability across the dataset, low values (<0.2) indicate good stability. CPM (counts per million) average is the abundance of the miRNA across the dataset, CPM > 10 required for detection by qPCR.

4.4.3 RT-qPCR data quality control

RNA spike-ins were used for quality control of RNA isolation (data shown previously section 2.7.2) and cDNA synthesis. The cDNA synthesis control (UniSp6) was added in the reverse transcription step giving the opportunity to evaluate the reaction. In addition to this, a DNA spike-in (UniSp3) was added to indicate any inhibitions at the qPCR level. Figure 4.5 shows the steady level (Cq between 19.12 – 19.66 for UniSp3 and 18.5 – 20.19 for UniSp6) of these assays, indicating the reverse transcription and qPCR were successful. The steady level of the RNA spike-ins, which is comparable to the blank purification, also shows that none of the samples contained inhibitors.

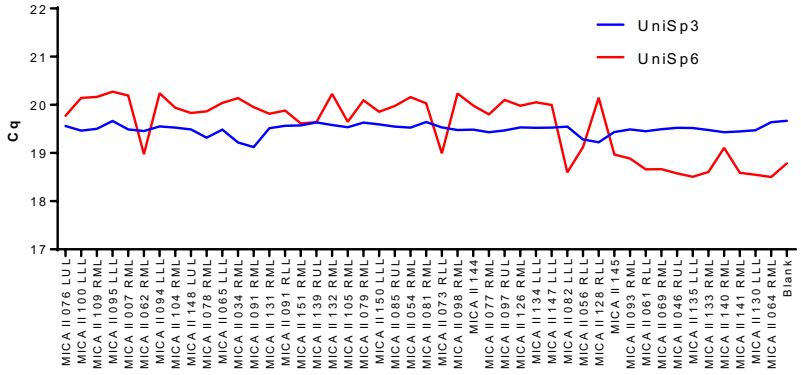


Figure 4.5 Raw Cq values for spike-in assays used to assess quality of cDNA synthesis (UniSp3) and reverse transcription reaction (UniSp6)

4.4.3.1 Number of detected miRNAs

The number of miRNA detected in each sample and the Cq value of the global mean for each sample is shown in Figure 4.6. Only 10 out of the 41 miRNA tested (35 target miRNA and six normaliser miRNA) were detected in all 46 samples. The mean number of miRNA detected per sample was 30 (range 10-38, SD±5). The average global mean Cq value across the dataset was 29.17 (range 26.9-34.6, SD±1.6). Two samples, MICA_II_091_RLL and MICA_II_145, showed much higher average global mean Cq values, at 34.6 and 33.8 respectively. In addition these samples had a lower number of miRNA detected in their samples, at 10 and 21 out of a total of 41 miRNA respectively. Given these findings, these two samples were excluded from the differential expression analysis.

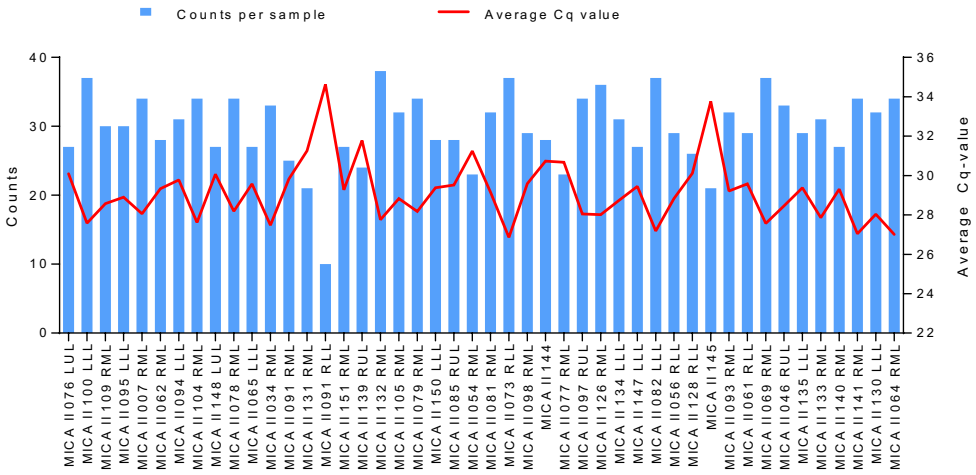


Figure 4.6 Number and expression level of miRNA detected by RT-qPCR for each sample
Number of miRNAs detected across samples shown as blue bars and the average Cq value for the commonly expressed miRNAs shown in red line. N=46; COPD, n=26.

4.4.4 Differential miRNA expression analysis of RT-qPCR data

After outliers (MICA_II_091_RLL and MICA_II_145) were removed, differential miRNA expression analysis was performed on a final cohort of 44 samples, 24 COPD subjects and 20 healthy ex-smokers. This did not alter the overall subject characteristics, and groups were still matched for age, sex, gender, BMI and smoking pack years.

When comparing COPD subjects to healthy ex-smokers, fourteen miRNAs were found to be differentially expressed using a cut-off P-value <0.05. Eight of these passed an FDR correction at a significance level of 0.05 (Table 4.7). Five were up-regulated in COPD (miR-2110, miR-223-3p, miR-625-3p, miR-182-5p and miR-200b-5p) and three were down-regulated in COPD (miR- 204-5p,

miR-138-5p and miR-338-3p). The log2FC of these differentially expressed miRNA are summarised in Table 4.7 and highlighted in red in Figure 4.7. The spread of the normalised expression data across the samples for each of the differentially expressed miRNA is shown in Figure 4.8.

Table 4.7 Significantly differentially expressed miRNA measured by RT-qPCR between COPD subjects and healthy ex-smokers

miRNA	COPD SD	Healthy_ES SD	Log2FC	P value	FDR
Up-regulated in COPD					
hsa-miR-2110	1.21	0.71	2.12	0.001	0.016
hsa-miR-223-3p	1.47	1.37	2.97	0.001	0.016
hsa-miR-625-3p†	0.91	0.76	1.85	0.006	0.041
hsa-miR-182-5p	0.70	0.66	1.52	0.006	0.041
hsa-miR-200b-5p	0.79	0.72	1.52	0.009	0.047
Down-regulated in COPD					
hsa-miR-204-5p	1.32	1.23	-2.37	0.003	0.037
hsa-miR-138-5p	0.90	0.77	-1.66	0.005	0.041
hsa-miR-338-3p	1.15	0.78	-1.72	0.009	0.047

Shapiro-Wilk test for normality was performed and showed data were normally distributed. Unpaired Welch's t test was performed and then adjusted using Benjamini Hochberg to generate an FDR value. COPD, Chronic obstructive pulmonary disease, FC: Fold change. FDR, false discovery rate; Healthy_ES, healthy ex-smoker; miRNA, microRNA, SD: standard deviation.
†missing data points; COPD, n=18; Healthy-ES, n=12

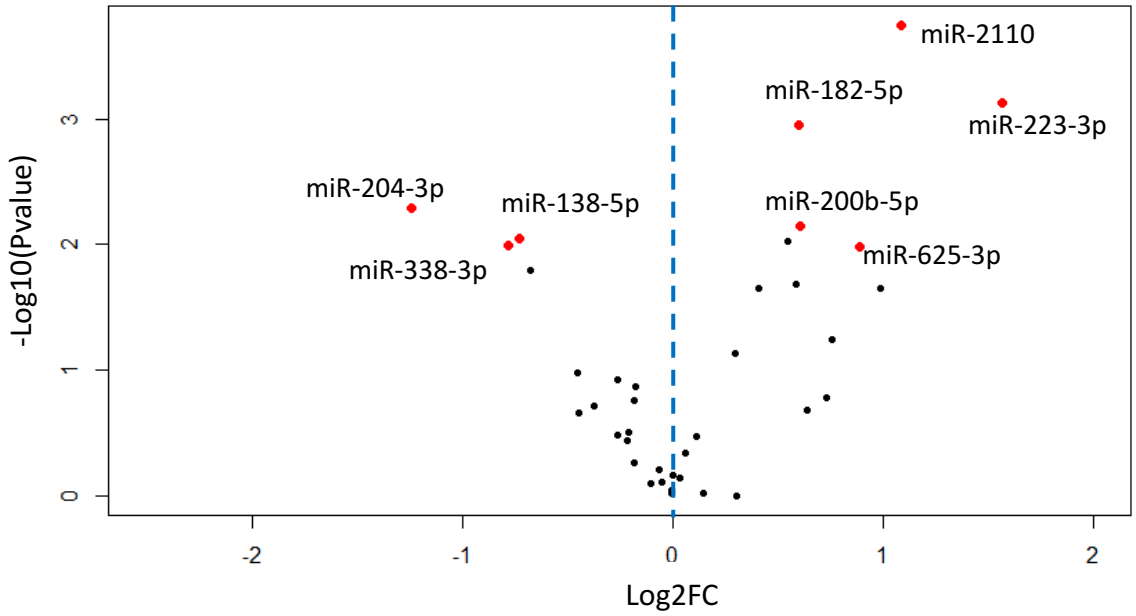


Figure 4.7 Volcano plot showing relationship between P values and expression data
Red dots show miRNA with P values <0.05 after FDR correction for multiple testing. Blue dotted line represents zero Log2FC, points to the right are up-regulated in COPD, and points to the left are down-regulated in COPD. FC, fold change; miRNA, microRNA.

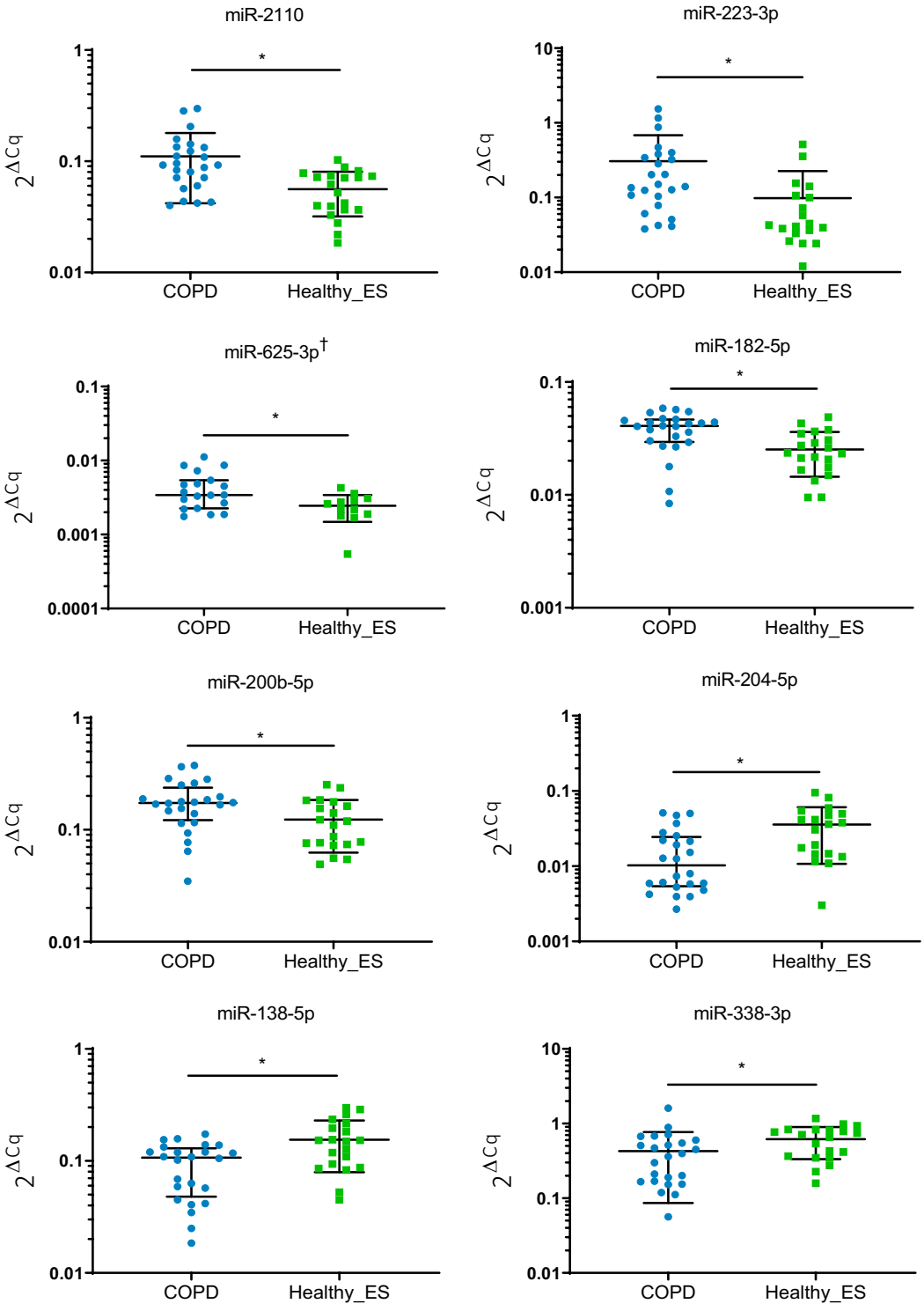


Figure 4.8 Normalised expression levels for significantly differentially expressed miRNA

$2^{\Delta Cq}$, normalised expression levels. Mean and SD shown.
Shapiro-Wilk test for normality demonstrated normally distributed data for all continuous variables. Unpaired Welch's t test was performed and then adjusted using Benjamini-Hochberg to generate an FDR value, *p<0.05. N=44; COPD, n=24.†miR-635-3p = missing data points; COPD, n=18; Healthy-ES, n=12

Of note, in the RT-qPCR validation experiment, there were several missing data points for miR-625-3p (n=31; COPD, n=18; healthy ex-smokers, n=12), as this miRNA was not detected in these samples. There were no obvious clinical characteristics which delineated these samples and missing values were found in both comparator groups. Given this missing data, this may make the results for this miRNA less robust.

When comparing the RT-qPCR validation results with the miRNA sequencing results, several of the top targets identified from the sequencing analysis were validated as significantly differentially expressed by RT-qPCR. The up-regulated miRNAs miR-223-3p, miR-200b-5p and all of the down-regulated miRNAs (miR-204-5p, miR-138-5p and miR-338-3p) were identified as the top differentially expressed miRNA between COPD and healthy controls in the sequencing analysis (Table 4.3). This gives greater certainty to the results, given these differentially expressed miRNA have been identified by multiple methods.

4.5 Discussion

In this chapter I have described the mapping and alignment results with comparison between COPD and healthy ex-smokers, the results of my differential expression analysis of the BALF EV miRNA, and the RT-qPCR validation study in a slightly larger cohort of 44 subjects.

4.5.1 RNA mapping and alignment

The main mapping quality parameter is the percentage of mapped reads, which is a global indicator of the overall sequencing accuracy and of the presence of contaminating DNA (365). On average ~2.8 million reads were obtained for each sample and the average genome mapping rate was 52.7%. It is difficult to find exact estimates of what mapping rates are expected in RNA sequencing experiments, especially when trying to compare different sources of RNA and different mapping tools. Some literature suggests as high as 70-90% (386), however this predominantly includes studies using abundant RNA sources (e.g. tumour) rather than limited sources such as EVs. Reassuringly, after excluding the unmapped reads, the mean number of miRNA mapped reads was 40.5%, (SD±14.3%) which is within the 10-60% range quoted in the literature (418).

Of the two previous studies investigating BALF derived EV RNA, only one used sequencing to profile RNA expression in lung transplant patients with and without acute cellular rejection (358). The library sizes generated were much smaller compared with my sequencing results (166,326

versus ~2.8 million), suggesting a much lower depth of sequencing, and they did not report the mapping rate. The other study investigated BALF derived EV miRNA in asthmatics compared with healthy controls using miRNA microarray (262). Next generation sequencing has now superseded this technology as it provides a more in-depth, unbiased, discovery-based analysis of the transcriptome, generating both known and novel transcript data with increased specificity and sensitivity compared to a microarray platform (419).

This rigorous quality control processes and detailed exploratory analyses (outlined in sections 2.10.1 and 2.10.2) were key in ensuring the quality and reliability of my data. Reporting these quality control metrics will be critical in any forthcoming publication (365). Although advanced RNA sequencing techniques can yield novel transcript discovery, it is a rapidly evolving field and the bioinformatics methods for data analysis are highly diverse and newly emerging.

4.5.2 Extracellular vesicle miRNA packaging

Importantly, my results show a higher proportion of miRNA in COPD BALF EVs than healthy ex-smokers (section 4.3.2). To my knowledge, only one other study has previously shown altered proportions of miRNAs in EVs in disease. Francisco-Garcia et al. showed deficient loading of miRNAs in BALF EVs of severe asthmatics compared with healthy controls. In addition, pathway analysis suggested that these significantly down-regulated miRNAs in severe asthmatics converge on pathways known to be important in asthma pathogenesis (420).

To explore this in more detail it is important to understand the potential sorting mechanisms for miRNA in EVs. Based on current research, there are four potential modes for sorting miRNA into EVs, although the underlying mechanisms remain largely unclear. Firstly, the "*neutral sphingomyelinase (nSMase)2-dependent pathway*": nSMase2 is the first molecule reported to be related to miRNA secretion into EVs. Kosaka et al. found that expression of nSMase2 was proportional to the number of miRNAs packaged in EVs (421). Previous work has also shown the role of nSMase in ceramide generation, aberrant apoptosis and lung injury in response to cigarette smoke exposure (422, 423). Moreover, as previously highlighted (Table 1.2 in section 1.7), Serban et al. showed endothelial microparticles (EMPs) required acid sphingomyelinase (aSMase) for release in response to cigarette smoke exposure, and aSMase enzyme activity was significantly up-regulated in plasma of patients with COPD (252). Together, these studies suggest that SMase activity may be regulated by smoke exposure, and sMase may have a role EVs release and regulation of EV miRNA cargo in COPD.

Secondly, the "*heterogeneous nuclear ribonucleoproteins (hnRNPs)-dependant pathway*": hnRNPs are capable of recognising the GGAG motif in the 3' region of miRNA sequences and cause specific miRNAs to be packed into EVs (424). Furthermore, hnRNPs are more widely involved in telomere stability, cell senescence and cell cycle regulation (425). Overexpression of hnRNP A2/B1 in plasma and primary human bronchial epithelium has a high sensitivity for the presence of Non-small cell lung cancer (NSCLC), and is also present in high-risk smokers' years before they develop lung cancer (426). Thus hnRNPs role in COPD pathogenesis may be beyond just involvement in EV miRNA packaging.

Thirdly, the "*3'-end of the miRNA sequence-dependent pathway*": Koppers-Lalic et al. found that 3'-end adenylated miRNAs isoforms are enriched in cells, whereas 3'-end uridylated miRNA isoforms are over-represented in EVs, suggesting that post-transcriptional modifications (e.g. 3'-end adenylation and uridylation) contribute in part to miRNA packaging in EVs (427). Recently, Zhang et al. showed that 3'-end uridylation mediated the packaging of miR-223 and miR-142 into macrophage-derived microvesicles (MVs) released in the lungs in response to lipopolysaccharide (LPS) and *Klebsiella pneumoniae* (428). Interestingly, miR-223 (along with miR-142) was found to be the selectively enriched miRNA in the macrophage MVs, in-keeping with the findings of this study, which shows miR-223-3p up-regulated in lung-derived EVs in COPD. These studies suggest that 3'-end uridylation mechanism may play a pivotal role in EV miRNA packaging in lungs in response to inflammatory stimuli.

Finally, the "*miRNA induced silencing complex (miRISC) pathway*": as discussed in section 1.8.1, mature miRNA bind with assembly proteins (RISC) to form a complex, the main components of which are: miRNA, target mRNA, GW182 and argonaute 2 (AGO2). Recent studies suggest a correlation between AGO2 and EV miRNA packaging, with the presence of AGO2 increasing abundance of miRNAs in EVs (429). Although, so far, no studies have investigated AGO2 role in EV packaging within the lungs or in inflammatory disease.

In summary, cells have the ability to selectively sort miRNA into EVs for secretion to nearby or distant targets. Broadly these mechanisms include RNA-binding proteins such as hnRNPA2B1, membranous proteins involved in EV biogenesis such as nSMase2, and specific miRNA-binding motifs capable of exerting selectivity over the miRNAs shuttled into EVs. Current EV miRNA literature focuses on the dysregulated EV-miRNA content, however little is known about the role of disease pathogenesis in regulating the EV miRNA selective sorting process. Therefore understanding the sequences and/or proteins responsible for selective sorting of miRNA in COPD

lung-derived EVs may reveal novel mechanisms in the disease pathogenesis, and provide targets for manipulating EV content that could have beneficial disease modifying effects.

4.5.3 Differential expression of EV miRNA in COPD subjects compared with healthy ex-smokers

In addition to the overall proportion of miRNA being higher in COPD BALF EVs than healthy ex-smokers, this is the first study showing differential miRNA expression in lung-derived EVs in COPD. Specifically, five miRNA were found up-regulated in COPD (miR-223-3p, miR-182-5p, miR-2110, miR-200b-5p and miR-625-3p) and three were down-regulated (miR-204-5p, miR-338-3p and miR-138-3p) when compared to healthy ex-smoker controls.

Up-regulation of non-EV miR-223 has been reported in COPD miRNA studies (reviewed in section 1.8.2) both in COPD lung tissue compared with smokers (319), and in human endothelial cells in culture (339). MiR-223 has a crucial role in innate immunity, myeloid cell differentiation, and cell homeostasis and has several gene targets that are involved in pathways implicated in the pathogenesis of COPD. MiR-223 is transcribed from an independent promoter located on the X chromosome and is mainly expressed by haemopoeitic cells (430). Under resting conditions, low levels of miR-223 expression are primarily controlled by binding of nuclear factor I A-type (NFI-A) to the miR-223 promoter (430). During granulocytic differentiation, NFI-A is released from the miR-223 promoter and replaced by CCAAT enhancer protein α (C/EBPα), resulting in up-regulation of miR-223 expression (431). Interestingly, one of the target genes of miR-223 is NFI-A, implicating that up-regulation of miR-223 dampens the expression of NFI-A, resulting in a positive feedback loop. In line with the role of miR-223 in myeloid differentiation, overexpression of miR-223 induces the monocytic and granulocytic differentiation marker CD11b, while inhibition of miR-223 is shown to reduce the expression of CD11b in promyelocytic leukaemia cells (430). In addition to the importance in myeloid differentiation, miR-223 is involved in erythropoiesis by dampening the gene expression and protein translation of LIM-only protein 2, a positive regulator of erythropoiesis (432).

MicroRNA profiling in human blood demonstrated that miR-223 is expressed in hematopoietic stem cells, granulocytes, dendritic cells and monocytes, while lower levels of miR-223 were also found in naïve and memory T cells (430, 433, 434). In induced sputum high expression of miR-223 was measured in monocytes, macrophages and neutrophils (435). Furthermore, *in situ* hybridisation in human bronchial biopsies showed that miR-223 expression was mainly expressed in neutrophils localized in the lamina propria (435).

With regards to smoking, lower miR-223 expression levels have been observed in bronchial brushings from current smokers compared to never smokers (316). Whereas, Ezzie et al. found higher levels of miR-223 in the lung tissue of COPD patients compared with smokers (319). Furthermore, higher levels of miR-223 were also measured in the BALF cell pellet obtained from COPD patients compared to non-COPD controls (342). In serum, miR-223 expression was also higher in women with COPD due to biomass smoke, than healthy controls exposed to biomass smoke (345). Taken together, it is clear that miR-223 expression is differently expressed in response to obstructive lung disease, however no data are available that link miR-223 expression to disease stage, inflammatory phenotype or presence of emphysema. Additionally, smoke (tobacco or biomass) can alter the expression of miR-223, which together with differences in examined samples and patient groups, further adds to the complexity of the observed findings.

Dysregulation of miR-223 has been implicated in the pathogenesis of a number of inflammatory diseases, including acute lung injury, rheumatoid arthritis, inflammatory bowel disease, and type II diabetes (436-438). Furthermore, studies have shown miR-223 as a regulator of macrophage function. Chen et al. demonstrated that expression of miR-223 dampens macrophage inflammatory responses to toll-like receptor (TLR) ligand stimulation; whereby LPS stimulated macrophages showed decreased miR-223 expression, resulting in increased levels of signal transducer and activator of transcription (STAT) 3 (a direct target of miR-223). This in turn, led to an increased production of pro-inflammatory cytokines interleukin (IL)-6 and IL-1β (439). In *Mycobacterium tuberculosis* infected patients, increased miR-223 levels in monocyte-derived macrophages (MDMs) correlate with impaired activation and cytokine production when compared to healthy individuals (440). Given this previous work, future investigation into EV-derived miR-223 regulation of macrophage function could be fundamental in determining the mechanisms underlying the defective macrophage phenotype found in COPD (73).

In addition to intrinsic cellular function of miR-223, studies have also reported EV transfer of miR-223 as a mechanism for intercellular communication (437, 441, 442). Neudecker et al. used a murine acute lung injury model to show in response to pulmonary injury neutrophils secrete EVs containing miR-223, which are shuttled to alveolar epithelial cells leading to reduced cellular inflammation and tissue injury via repression of poly-(ADP-ribose) polymerase (PARP)-1.(437) Therefore, therapeutic manipulation of miR-223 to dampen inflammatory targets could be a potential treatment to control excessive innate immune responses during mucosal inflammation.

MicroRNA-182 is known as an oncogenic miRNA and is implicated in the progression of several cancers owing to its role in promoting cell proliferation and invasion (443). Specifically, miR-182

has been found up-regulated in sputum of patients with lung adenocarcinoma (444) and in serum of patients with early stage NSCLC (445). In addition, recent work has identified miR-182-3p as a regulator of smooth muscle cell proliferation and vascular remodelling, with implications for pulmonary arterial hypertension (446). Moreover, miR-182-5p is overexpressed in lung tissue in pulmonary fibrosis, and modulation of miR-182-5p via mothers against decapentaplegic homolog (*SMAD)7* up-regulation can inhibit the expression of pro-fibrotic proteins such as fibronectin, α-smooth muscle actin and p-SMAD2/p-SMAD3 (447). Finally, miR-182 expression is regulated by smoke exposure, and was up-regulated in the BALF of mice after 4 weeks of cigarette smoke exposure (327). Taken together, miR-182 may be important in COPD pathogenesis as a regulator of smooth muscle cell proliferation, vascular remodelling (128) and fibrosis (448).

MicroRNA-2110 was first reported as one of a group of neurite-inducing miRNAs (449). MiR-2110 was also identified as a tumour-suppressor by targeting Tsukushi (*TSKU)*, inducing cell differentiation and reducing cell survival in neuroblastoma cell lines (450). In patients, low tumour miR-2110 levels were significantly correlated with high tumour *TSKU* mRNA levels, and both low miR-2110 and high *TSKU* mRNA levels were significantly correlated with poor patient survival (450). In addition, miR-2110 has been identified as one of five miRNAs up-regulated in serum exosomes in patients with active *Mycobacterium tuberculosis* infection (451). Little is known about miR-2110 in the context of inflammatory disease, and therefore further work investigating mRNA targets within the airways of COPD patients may establish a role in inflammatory lung disease.

MicroRNA-200b-5p belongs to the miR-200 family, which controls epithelial-mesenchymal transition (EMT) and metastasis in tumour cells (452, 453). The primary cause of organ fibrosis is the production of excessive extracellular collagen by activated fibroblasts (myofibroblasts). A significant proportion of fibroblasts are derived via EMT of resident epithelial cells within the diseased organ itself and this process has been observed in the alveolar epithelial cells (454). Furthermore as discussed in section 1.2, as part of remodelling, peribronchiolar fibrosis is observed in the small airways of patients with COPD contributing to airway obstruction. EMT appears to be involved in the formation of peribronchiolar fibrosis in COPD (115) and may be a precursor to the development of lung cancer in these patients (455). In addition, miR-200b promotes angiostatic effects by silencing v-ets erythroblastosis virus E26 oncogene homolog 1 (ETS1), a transcription factor for controlling angiogenic genes like vascular endothelial growth factor receptor 2 (*VEGFR2*) and matrix metalloproteinase 1 (*MMP1*) expression in endothelial cells (456). Both small airway fibrosis and vascular remodelling have important roles in COPD

progression and therefore manipulation of these pathways may lead to novel targets for COPD treatment.

Finally, miR-625-3p was significantly up-regulated miRNA in lung-derived EVs COPD. MiR-625-3p, is overexpressed in colorectal malignancies where it promotes migration, invasion and apoptosis resistance (457, 458). MiR-625-3p has also been detected in the circulation of malignant pleural mesothelioma patients (459). Additionally, miR-625-3p may have an essential role in the downstream cascade of T cell receptor signals to promote CD8+ T cell proliferation. Inhibition of miR-625-3p expression by the mammalian target of rapamycin (mTOR) inhibitor rapamycin, can influence CD8+ T cell proliferation and functions that mediate the anti-viral immunity and graft-vs-host-disease (GVHD) in stem cell transplant patients (460). MiR-625-5p is also shown to be significantly down-regulated in paediatric asthma (461) and suppresses inflammatory responses by targeting protein kinase B2 (AKT2) and inhibiting the nuclear factor kappa B (NF-κB) signalling pathway in human bronchial epithelial cells (462). Thus, based on this previous work, up-regulation of miR-625 in COPD lung-derived EVs may suggest a role in regulation of CD8+ T cell proliferation within the lungs of COPD and regulation of several inflammatory pathways. However, as noted in the results (section 4.5.4), there were several missing data points in the RT-qPCR validation study, which may suggest this data showing up-regulation in COPD lung-derived EVs may be less robust than for the other significantly dysregulated miRNA.

In this study, miR-338-3p was down-regulated in lung-derived EVs of patients with COPD. MiR-338-3p has previously been reported to be down-regulated in COPD plasma compared with asthmatics and healthy controls (341), and down-regulated in plasma of allergic rhinitis patients, where it was shown to target selected inflammatory genes mitogen-activated protein kinase (*MAPK*) 8 and inhibitor of nuclear factor kappa-B kinase subunit beta (*IKKβ*) (463). In contrast, Lacedonia et al. showed miRNA-338 expression in the sputum was higher in both patients with asthma and COPD compared to controls, however asthmatics showed a significantly higher miR-338 expression compared to COPD patients (464). In addition, studies have shown that miR-338 may be important in cellular apoptosis, differentiation and tumour growth (465-467), via targeting of metastasis associated in colon cancer 1 (*MACC1*) gene (468). Given the aforementioned literature, down-regulation of miR-338-3p in COPD lung-derived EVs may have an important role in regulating key inflammatory pathways (MAPK and NF-κB signalling) and cellular processes important in tumour development.

MicroRNA-204-5p was found down-regulated in COPD lung-derived EVs compared with healthy ex-smokers. Recent work has identified down-regulation of miR-204-5p and its regulatory

network is important in NSCLC (469), with low expression in NSCLC tumours associated with advanced progression, poor prognosis and severe metastatic potential (470, 471). Furthermore, miR-204-5p has reported as a novel diagnostic biomarker in patients with frontotemporal dementia (472) and Parkinson's disease (473). Importantly, as with miR-200b-5p, research suggests a role for miR-204-5p in direct regulation of EMT through its targeting of *SMAD4*, a mediator of transforming growth factor (TGF)-β signalling (474). Wang et al. demonstrated that miR-204-5p overexpression enhanced the repression of TGF-β2–induced EMT in the presence of *SMAD4* small interfering RNA (474). Therefore, a reduction in miR-204-5p, as seen in lung-derived EVs in COPD patients in this study, is likely to lead to an increase in EMT. These findings suggest possible synergistic activity with miR-200b-5p and together these dysregulated miRNA may contribute to the small airway fibrosis seen in COPD.

Finally miR-138-5p was found down-regulated in COPD lung-EVs compared with healthy ex-smokers. miR-138-5p is known to inhibit tumour growth and activate the immune system by down-regulating programmed cell death protein (PD)-1/PD-ligand 1, and thus is a promising therapeutic target for NSCLC (475). Little is known about miR-138-5p in the context of inflammatory disease or immune cell dysfunction, and thus additional work investigating mRNA targets within the airways of COPD patients may establish a role in inflammatory lung disease.

4.5.4 Strengths and limitations

Limitations of these results include the relatively small sample size (n=44 for the validation study), although this is in keeping with other studies using human BALF samples (232, 262, 357-359). This is in part due to the availability of samples, as not only can recruitment to a bronchoscopy study be challenging, but also the resultant sample quantities (i.e. of BALF) can be influenced by the underlying disease processes (discussed in Chapter 3). In addition, the validation study was performed in the same cohort as the microRNA sequencing study, where a more robust experimental design would have conducted these experiments in an entirely separate cohort. Again due to sample availability this was not possible, however I was able to increase the sample size (with an additional 6 COPD subjects and 5 healthy ex-smokers) at this stage.

As previously discussed the COPD patients included in this study had relatively mild disease with a mean FEV1% predicted of 78.9% (SD±14.4). This is in contrast to other EV miRNA studies in COPD (247, 250, 260), which included a broader range and severity of COPD patients (mean FEV1 63.4%, SD±29.54) and current smokers. Although their findings may be applicable to a wider COPD cohort, they are less translatable mechanistically given their broader range of included subject

phenotypes and the inclusion of current smokers, which may attribute effects to active smoking rather than disease alone. In addition, these studies used blood and sputum samples, rather than BALF as a source for EVs. A recent study has shown that major differences in the COPD lung tissue transcriptome were poorly mirrored in sputum and non-representative to those in blood (476). Therefore, a major strength of this work is the EV miRNA were derived from BALF and therefore should relate directly to the altered lung transcriptome in COPD and provide novel mechanistic insights into disease pathogenesis.

4.5.5 Summary

In conclusion, this chapter highlights two novel findings in the field of EV miRNA research in COPD. Firstly, there is an increased proportion of miRNA packaged in COPD lung-derived EVs compared with healthy ex-smokers. Further work needs to be done to establish the mechanisms underlying the alteration in EV RNA content in COPD, and whether these changes have any pathological consequences. Secondly, there are differentially expressed miRNA within the COPD lung-derived EVs compared with healthy ex-smokers. Two of these miRNA (miR-223-3p and miR-338-3p) have been found dysregulated in other COPD studies (319, 339, 341), providing further evidence as to their importance and relevance in this disease. To investigate the effects of these differentially expressed miRNA on target gene expression and their importance to COPD pathogenesis, in-depth pathway analysis was performed, and the results are presented in the following chapter.

Chapter 5 Identification, visualisation and analysis of microRNA-target interaction networks

5.1 Introduction

Identifying the differentially expressed lung-derived extracellular vesicle (EV) microRNA (miRNA) in Chronic Obstructive Pulmonary Disease (COPD) compared with healthy ex-smokers was the first aim of this thesis. However, understanding the impact of these differentially expressed miRNA on gene expression in COPD is key to understanding whether these differences have biological significance and relevance to disease pathology.

Chapter 5 explores the interactions between the differentially expressed miRNA and their differentially expressed target genes in COPD. Firstly, it presents the validated and predicted gene targets of the differentially expressed miRNA. Next, it shows the results of the differential expression analysis for the epithelial brushing transcriptome data comparing COPD with healthy ex-smokers, using the same cohort as the EV miRNA samples. Finally, this chapter explores how these differentially expressed genes integrate with the differentially expressed miRNA using several bioinformatic techniques (outlined in methods section 2.13) to provide further insights into the impact of this regulatory network on the biology of COPD.

5.2 Identifying miRNA target genes

5.2.1 miRNA target identification *in silico* using multiMiR

In silico analysis using the multiMiR package identified 21,343 gene targets of the differentially expressed miRNA in COPD (results summarised in Table 5.1). Of these 7,615 (35.6%) were validated in experimental models. Several of the gene targets identified were co-regulated by the miRNA, suggesting possible synergistic action, which may result in a greater impact on transcript expression.

Table 5.1 Summary results of *in silico* target prediction using multiMiR

miRNA	Number of Validated targets	Number of Predicted targets	Total number of targets
Up-regulated miRNA in COPD	4,627	7487	12,114
miR-2110	534	1538	2072
miR-223-3p	149	1335	1484
miR-625-3p	198	468	666
miR-182-5p	3634	3244	6878
miR-200b-5p	112	902	1014
Down-regulated miRNA in COPD	2,988	6,241	9,229
miR-204-5p	945	3172	4117
miR-138-5p	1053	1557	2610
miR-338-3p	990	1512	2502
Total number of targets	7,615	13,728	21,343

COPD, Chronic Obstructive Pulmonary Disease; miRNA, microRNA

These results were used to identify the putative miRNA-mRNA interactions (section 5.2.3) in combination with the results from the next section (section 5.2.2), which identifies the differentially expressed genes in the epithelial brushings in COPD compared with healthy ex-smokers.

5.2.2 Differential gene expression analysis of epithelial brushings in COPD and healthy ex-smokers

A key strength to this study is that in addition to bronchoalveolar lavage fluid (BALF) EV miRNA isolation, epithelial brushings were also taken from the same patients and from the same lung lobe location as the EVs were recovered from for transcriptome analysis. By identifying the differentially expressed genes in the epithelial cell compartment between COPD and healthy ex-smokers, it is possible to identify those genes that may be under epigenetic control by the differentially expressed EV miRNA.

Raw mRNA counts generated using the NovaSeq 6000 (Illumina®) were filtered, analysed for quality and condensed into a filtered dataset (performed by AstraZeneca) comprising of 44 epithelial brushing samples and 27,229 mRNA. Average reads per sample were 52.6 million.

Differential gene expression were assessed with DESeq2 in the 44 epithelial brushing samples (24 COPD and 20 healthy ex-smokers), and this identified 192 differentially expressed genes between COPD and healthy ex-smokers with a false discovery rate (FDR) of <0.05 (Figure 5.1).

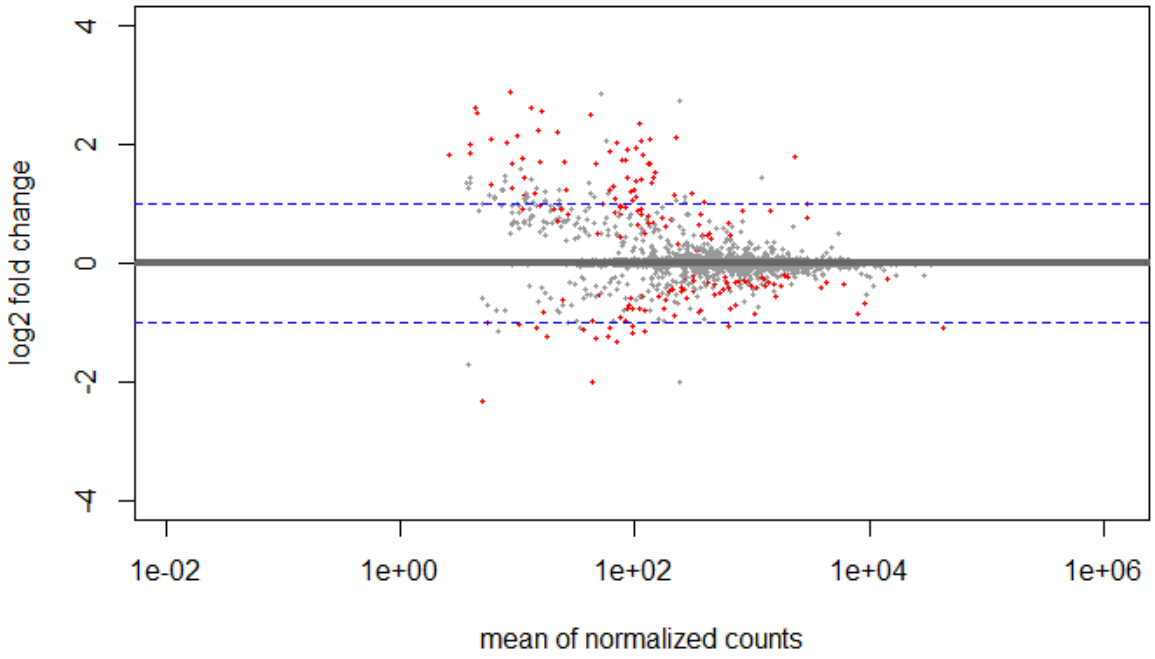


Figure 5.1 MA plot showing differentially expressed mRNA in epithelial brushings between COPD subjects and healthy ex-smokers
Red dots represent mRNA with an FDR <0.05. Blue lines represent a twofold change in expression

Of the 196 differentially expressed genes, there were 57 genes with a log2 fold change (FC) of greater than one (most up-regulated in COPD) and 17 genes with a log2FC less than one (most down-regulated in COPD). Table 5.2 summarises the top 10 up and down-regulated genes in the COPD epithelial brushings, with the entire list of differentially expressed genes summarised in Appendix A, Supplementary Table 1.

Table 5.2 Top ten differentially expressed genes in epithelial brushings in COPD subjects

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
Top up-regulated genes in COPD				
ENSG00000090512	*FETUB*	2.87	4.05E-05	0.013
ENSG00000179593	*ALOX15B*	2.61	1.32E-05	0.0075
ENSG00000273331	*TM4SF19-TCTEX1D2*	2.60	1.09E-05	0.0064
ENSG00000287059	N/A	2.55	6.54E-10	2.94E-06
ENSG00000155918	*RAET1L*	2.52	0.00014	0.028
ENSG00000262406	*MMP12*	2.49	5.40E-05	0.015
ENSG00000115590	*IL1R2*	2.34	2.50E-07	0.00045
ENSG00000287771	N/A lncRNA	2.24	3.69E-05	0.012
ENSG00000111700	*SLCO1B3*	2.20	3.13E-06	0.0028
ENSG00000255833	*TIFAB*	2.14	4.56E-05	0.014
Top down-regulated genes in COPD				
ENSG00000198787	19587 (pseudogene)	-2.34	3.10E-05	0.012
ENSG00000113389	*NPR3*	-2.01	9.96E-08	0.0003
ENSG00000038295	*TLL1*	-1.35	3.42E-05	0.012
ENSG00000174059	*CD34*	-1.27	9.15E-05	0.021
ENSG00000099994	*SUSD2*	-1.25	8.03E-06	0.0056
ENSG00000214870	LOC441204	-1.24	0.000114	0.025
ENSG00000126562	*WNK4*	-1.19	1.17E-06	0.0013
ENSG00000078596	*ITM2A*	-1.14	2.59E-05	0.011
ENSG00000125144	*MT1G*	-1.12	3.66E-05	0.012
ENSG00000283413	N/A lncRNA	-1.11	0.000184	0.033

HGNC, HUGO Gene Nomenclature Committee; FC, fold change; FDR, False discovery rate.

These 196 differentially expressed genes were used in combination with the *in silico* identified targets of the differential expressed miRNA (section 5.1) to identify the putative miRNA-mRNA interactions in the next section.

5.2.3 Identifying putative miRNA-mRNA interactions using multiMiR

In addition to retrieving all validated and predicted target genes of a given miRNA (section 5.1), the multiMiR package has functionality to retrieve interactions between differentially expressed miRNAs and a specific set of differentially expressed genes (methods summarised in section 2.12.4 and an overview in Figure 5.2 – Figure 2.26 re-presented for ease of reference). Given that miRNA lead to down-regulation of their gene targets (an inverse correlation), only down-regulated mRNA in COPD from the epithelial biopsy differential expression analysis were selected as possible targets for up-regulated EV miRNA in COPD and vice versa.

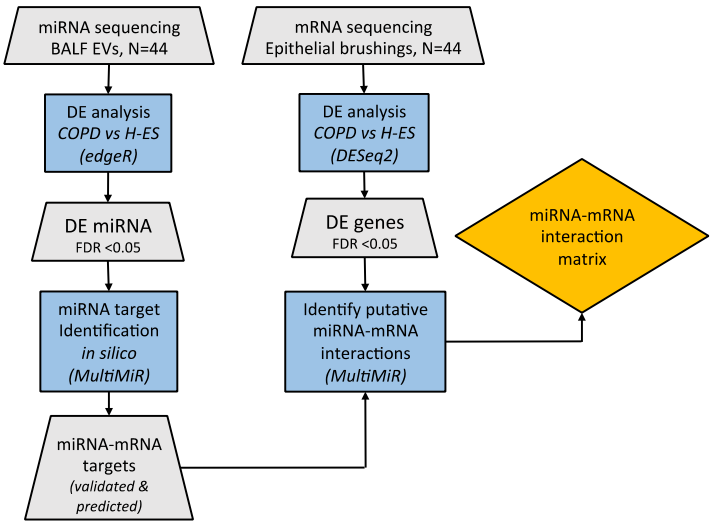


Figure 5.2 An overview of the methods used to identify putative miRNA-mRNA interactions
Blue rectangle indicates a process with the software/tool specified in parenthesis, grey trapezium indicate a data output, yellow diamond is an important checkpoint. BALF, bronchoalveolar lavage fluid; COPD, Chronic obstructive pulmonary disease; DE, differentially expressed; EVs, extracellular vesicles; FDR, false discovery rate; H-ES, healthy ex-smoker; miRNA, microRNA; mRNA, messenger RNA

5.2.3.1 Identification of the up-regulated EV miRNA gene targets in epithelial brushings

The multiMiR package identified 54 gene targets of the five EV miRNA up-regulated in COPD when considering both validated and the top 20% predicted targets. Of these, 16 had been experimentally validated and 38 were predicted targets, summarised in Table 5.3. All of the validated targets had been identified using high-throughput methods, such as high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation (HITS-CLIP). Seven of the target genes (Pleomorphic adenoma gene 1 (*PLAG1),* Kruppel Like Factor 10 (*KLF10*), GLI Family Zinc Finger 3 (*GLI3*), Inhibitor Of Growth Family Member 1 (*ING1*), FXYD Domain Containing Ion Transport Regulator 6 (*FXYD6),* Cysteine Rich Transmembrane BMP Regulator 1 (*CRIM1*), F-Box And Leucine Rich Repeat Protein 14 (*FBXL14*)), were co-regulated by different up-regulated miRNA in COPD suggesting synergistic action.

Table 5.3 Gene targets of the up-regulated EV miRNA in COPD, identified as down-regulated within the epithelial brushing transcriptome in COPD

miRNA	Gene Target*	ENSEMBL ID	Method of validation	Source database
Validated targets				
hsa-miR-200b-5p	***PLAG1***	ENSG00000181690	HITS-CLIP	mirtarbase (477)
hsa-miR-200b-5p	***KLF10***	ENSG00000155090	HITS-CLIP	mirtarbase
hsa-miR-182-5p	*THSD7A*	ENSG00000005108	Degradome sequencing	tarbase (478)
hsa-miR-182-5p	*JADE1*	ENSG00000077684	Degradome sequencing	tarbase
hsa-miR-182-5p	***GLI3***	ENSG00000106571	Degradome sequencing	tarbase
hsa-miR-182-5p	*SLC19A2*	ENSG00000117479	Degradome sequencing	tarbase
hsa-miR-182-5p	*C3*	ENSG00000125730	Degradome sequencing	tarbase
hsa-miR-182-5p	*CILP*	ENSG00000138615	Degradome sequencing	tarbase
hsa-miR-182-5p	*HOMER1*	ENSG00000152413	Degradome sequencing	tarbase
hsa-miR-182-5p	***ING1***	ENSG00000153487	Degradome sequencing	tarbase
hsa-miR-182-5p	***KLF10***	ENSG00000155090	Degradome sequencing	tarbase
hsa-miR-182-5p	*SRGAP2C*	ENSG00000171943	Degradome sequencing	tarbase
hsa-miR-182-5p	*HEG1*	ENSG00000173706	Degradome sequencing	tarbase
hsa-miR-182-5p	*DPY19L3*	ENSG00000178904	Degradome sequencing	tarbase
hsa-miR-182-5p	***PLAG1***	ENSG00000181690	Degradome sequencing	tarbase
hsa-miR-182-5p	*ZDBF2*	ENSG00000204186	Degradome sequencing	tarbase
Predicted targets				
hsa-miR-182-5p	*SCUBE3*	ENSG00000146197	NA	diana_microt (479)
hsa-miR-182-5p	*PIEZO2*	ENSG00000154864	NA	diana_microt
hsa-miR-182-5p	*NPR3*	ENSG00000113389	NA	diana_microt
hsa-miR-182-5p	*PCDH17*	ENSG00000118946	NA	diana_microt
hsa-miR-182-5p	*SIK2*	ENSG00000170145	NA	diana_microt
hsa-miR-182-5p	*CDKN1C*	ENSG00000129757	NA	diana_microt
hsa-miR-182-5p	*PER2*	ENSG00000132326	NA	diana_microt
hsa-miR-182-5p	*ATOH8*	ENSG00000168874	NA	diana_microt
hsa-miR-182-5p	*FAM69B*	ENSG00000165716	NA	elmmo (480)
hsa-miR-182-5p	*SCNN1G*	ENSG00000166828	NA	elmmo
hsa-miR-182-5p	*NEURL1B*	ENSG00000214357	NA	elmmo
hsa-miR-182-5p	*FLNB*	ENSG00000136068	NA	elmmo
hsa-miR-182-5p	*LRIG1*	ENSG00000144749	NA	miranda (481)
hsa-miR-182-5p	*FZD8*	ENSG00000177283	NA	miranda
hsa-miR-182-5p	*ARHGAP6*	ENSG00000047648	NA	mirdb (482)
hsa-miR-182-5p	*TLL1*	ENSG00000038295	NA	mirdb

miRNA	Gene Target*	ENSEMBL ID	Method of validation	Source database
hsa-miR-182-5p	***FXYD6***	ENSG00000137726	NA	pita (483)
hsa-miR-182-5p	***CRIM1***	ENSG00000150938	NA	pita
hsa-miR-182-5p	*TOB1*	ENSG00000141232	NA	targetscan (303)
hsa-miR-200b-5p	*PRDM11*	ENSG00000019485	NA	diana_microt
hsa-miR-200b-5p	*CD34*	ENSG00000174059	NA	miranda
hsa-miR-200b-5p	***CRIM1***	ENSG00000150938	NA	miranda
hsa-miR-2110	***FBXL14***	ENSG00000171823	NA	diana_microt
hsa-miR-2110	*ATP6V1B1*	ENSG00000116039	NA	diana_microt
hsa-miR-2110	*SNTB1*	ENSG00000172164	NA	diana_microt
hsa-miR-2110	***ING1***	ENSG00000153487	NA	diana_microt
hsa-miR-2110	***FXYD6***	ENSG00000137726	NA	diana_microt
hsa-miR-2110	*TSPAN11*	ENSG00000110900	NA	diana_microt
hsa-miR-2110	*INSR*	ENSG00000171105	NA	diana_microt
hsa-miR-2110	*FZD8*	ENSG00000177283	NA	diana_microt
hsa-miR-2110	*SUSD2*	ENSG00000099994	NA	mirdb
hsa-miR-223-3p	*FAM46A*	ENSG00000112773	NA	diana_microt
hsa-miR-223-3p	*CRIM1*	ENSG00000150938	NA	diana_microt
hsa-miR-223-3p	***FBXL14***	ENSG00000171823	NA	diana_microt
hsa-miR-223-3p	***GLI3***	ENSG00000106571	NA	diana_microt
hsa-miR-223-3p	*ADAMTS15*	ENSG00000166106	NA	diana_microt
hsa-miR-223-3p	*ACTR3B*	ENSG00000133627	NA	pita
hsa-miR-625-3p	*TBC1D3C*	ENSG00000278299	NA	diana_microt

*Gene target in bold if co-regulated by more than one EV miRNA. miRNA target databases referenced in table. HITS-CLIP, high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation; NA, non-applicable as a predicted target.

Only two of the five up-regulated EV miRNA (miR-182-5p and miR-200b-5p) had targets identified within the significantly down-regulated genes in epithelial brushings that had been validated in experimental models. Overall miR-182-5p was found to target the most genes with 33, compared with miR-2110 with nine gene targets, mir-200b-5p with five targets, miR-223-3p with six targets, and miR-625-5p with just one identified gene target.

5.2.3.2 Identification of the down-regulated EV miRNA gene targets in epithelial brushings.

The multiMiR package identified 53 gene targets in the epithelial brushing transcriptome of the three EV miRNA down-regulated in COPD when considering both validated and top 20% of predicted targets. Of these, 22 had been experimentally validated and 31 were predicted targets, summarised in Table 5.4. In this instance, some of the validated targets were confirmed by low

throughput techniques such as real-time quantitative polymerase chain reaction (RT-qPCR) (*c.f.* section 5.2.3.1), which are considered to have greater specificity than high-throughput platforms. Six of the target genes (Solute Carrier Family 45 Member 3 (*SLC45A3*), CD44 molecule (*CD44*), Aldo-Keto Reductase Family 1 Member C2 (*AKR1C2*), Brain-derived neurotrophic factor (*BDNF*), Cyclic AMP-Responsive Element-binding Protein 5 (*CREB5*), Solute Carrier Family 39 Member 14 (*SLC39A14*), BTB Domain containing 7 (*BTBD7*)) were co-regulated by different down-regulated EV miRNA in COPD suggesting synergistic action.

Table 5.4 Gene targets of the down-regulated EV miRNA in COPD, identified as up-regulated in the epithelial brushing transcriptome in COPD

miRNA	Gene Target*	ENSEMBL ID	Method of validation	Source database
Validated targets				
hsa-miR-138-5p	***SLC45A3***	ENSG00000158715	WB	mirtarbase (477)
hsa-miR-138-5p	***CD44***	ENSG00000026508	Degradome sequencing	tarbase (478)
hsa-miR-138-5p	*CDC45*	ENSG00000093009	Degradome sequencing	tarbase
hsa-miR-138-5p	*FGFBP1*	ENSG00000137440	Degradome sequencing	tarbase
hsa-miR-138-5p	*CYP1B1*	ENSG00000138061	Degradome sequencing	tarbase
hsa-miR-138-5p	*PHLDA1*	ENSG00000139289	Degradome sequencing	tarbase
hsa-miR-138-5p	*MKI67*	ENSG00000148773	Degradome sequencing	tarbase
hsa-miR-138-5p	***AKR1C2***	ENSG00000151632	Degradome sequencing	tarbase
hsa-miR-138-5p	*CABYR*	ENSG00000154040	Degradome sequencing	tarbase
hsa-miR-138-5p	*PTTG1*	ENSG00000164611	Degradome sequencing	tarbase
hsa-miR-138-5p	*AKR1B10*	ENSG00000198074	Degradome sequencing	tarbase
hsa-miR-204-5p	*ARNTL2*	ENSG00000029153	PAR-CLIP	mirtarbase
hsa-miR-204-5p	***BDNF***	ENSG00000176697	IF/RT-PCR/WB	mirtarbase
hsa-miR-204-5p	***CD44***	ENSG00000026508	HITS-CLIP	mirtarbase
hsa-miR-204-5p	***CREB5***	ENSG00000146592	IF/Microarray/RT-PCR/WB	mirtarbase
hsa-miR-204-5p	*KCNK6*	ENSG00000099337	HITS-CLIP	mirtarbase
hsa-miR-204-5p	*MDFI*	ENSG00000112559	HITS-CLIP	mirtarbase
hsa-miR-204-5p	*PLAT*	ENSG00000104368	Microarray	mirtarbase
hsa-miR-338-3p	***SLC39A14***	ENSG00000104635	Degradome sequencing	tarbase
hsa-miR-338-3p	*FAM49B*	ENSG00000153310	Degradome sequencing	tarbase
hsa-miR-338-3p	*TUBA1C*	ENSG00000167553	Degradome sequencing	tarbase
Predicted targets				
hsa-miR-138-5p	***BTBD7***	ENSG00000011114	NA	diana_microt (479)
hsa-miR-138-5p	*C12orf36*	ENSG00000180861	NA	diana_microt

miRNA	Gene Target*	ENSEMBL ID	Method of validation	Source database
hsa-miR-138-5p	*SCN2A*	ENSG00000136531	NA	elmmo (481)
hsa-miR-138-5p	***CREB5***	ENSG00000146592	NA	elmmo
hsa-miR-138-5p	***BDNF***	ENSG00000176697	NA	pita (483)
hsa-miR-138-5p	*C15orf37*	ENSG00000259642	NA	pita
hsa-miR-138-5p	*CYP3A5*	ENSG00000106258	NA	targetscan
hsa-miR-204-5p	***BTBD7***	ENSG00000011114	NA	elmmo
hsa-miR-204-5p	*CA12*	ENSG00000074410	NA	elmmo
hsa-miR-204-5p	*CHST15*	ENSG00000182022	NA	elmmo
hsa-miR-204-5p	*CLEC5A*	ENSG00000258227	NA	elmmo
hsa-miR-204-5p	*CLIP4*	ENSG00000115295	NA	elmmo
hsa-miR-204-5p	*CYP1B1*	ENSG00000138061	NA	elmmo
hsa-miR-204-5p	*DSG3*	ENSG00000134757	NA	diana_microt
hsa-miR-204-5p	*MKI67*	ENSG00000148773	NA	elmmo
hsa-miR-204-5p	*RHBDL2*	ENSG00000158315	NA	elmmo
hsa-miR-204-5p	*SCN2A*	ENSG00000136531	NA	diana_microt
hsa-miR-204-5p	***SLC39A14***	ENSG00000104635	NA	elmmo
hsa-miR-204-5p	***SLC45A3***	ENSG00000158715	NA	elmmo
hsa-miR-204-5p	*SLC7A11*	ENSG00000151012	NA	elmmo
hsa-miR-338-3p	*KIAA1199*	ENSG00000103888	NA	diana_microt
hsa-miR-338-3p	*ADAMDEC1*	ENSG00000134028	NA	diana_microt
hsa-miR-338-3p	*RASAL1*	ENSG00000111344	NA	diana_microt
hsa-miR-338-3p	***BTBD7***	ENSG00000011114	NA	diana_microt
hsa-miR-338-3p	*SYT8*	ENSG00000149043	NA	diana_microt
hsa-miR-338-3p	*CCL22*	ENSG00000102962	NA	diana_microt
hsa-miR-338-3p	*COL17A1*	ENSG00000065618	NA	microcosm(484)
hsa-miR-338-3p	***AKR1C2***	ENSG00000151632	NA	microcosm
hsa-miR-338-3p	*LILRB2*	ENSG00000131042	NA	microcosm

*Gene target in bold if co-regulated by more than one EV miRNA. miRNA target databases referenced in table. HITS-CLIP, high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation; IF, Immunofluorescence; NA, non-applicable as a predicted target; PAR-CLIP, photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation; RT-qPCR, Real time quantitative polymerase chain reaction; WB, Western blotting;

Compared with the up-regulated EV miRNA, all three down-regulated EV miRNA were found to have validated targets within the significantly up-regulated gene targets identified in COPD epithelial brushings. Overall, miR-138-5p had the greatest number of targets with 21, followed by miR-204-5p with 20 and miR-338-3p with 12 gene targets identified in the up-regulated genes in epithelial brushings transcriptome in COPD.

Together these 107 putative miRNA-mRNA interactions (54 gene targets of the five up-regulated EV miRNA and 53 gene targets of the three down-regulated EV miRNA) formed the "***miRNA-mRNA interaction matrix"*** that was used for the miRNA-mRNA interaction analysis in the next section.

5.2.4 miRNA-mRNA interaction analysis

There is no gold standard method for integrative analysis of miRNA and mRNA expression data. Therefore, a number of different approaches were used to explore and understand the interactions between the differentially expressed lung EV miRNA and mRNA from epithelial brushings in COPD (summarised in Figure 5.3) (394-396, 399-401, 403, 404). The results of each of these approaches will be covered in this section, with the ultimate aim of determining those interactions, which require further investigation and potential validation in disease specific biological models.

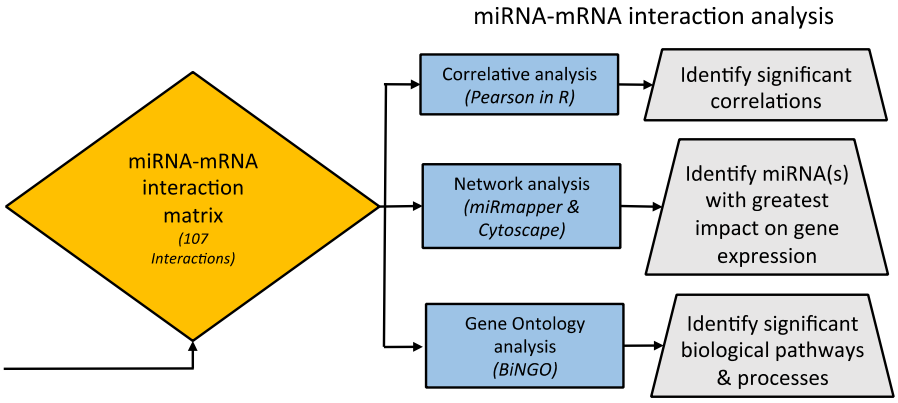


Figure 5.3 Summary of miRNA-mRNA interaction analyses performed in this thesis.
Blue rectangle indicates a process with the software/tool specified in parenthesis, grey trapezium indicate a data output, yellow diamond is an important checkpoint.

5.2.4.1 Pairwise correlations between dysregulated miRNA and mRNA

Pairwise correlations were used to analyse the relationship between the eight differentially expressed EV miRNA and 196 differentially expressed genes from the epithelial brushings.

In this analysis, a valid miRNA-mRNA interaction was reported based on the following conditions:

I. only significantly (i.e. FDR <0.05) differentially expressed miRNA and mRNA in COPD were considered for analysis.
II. miRNA normalised expression significantly correlated with normalised gene expression (i.e. FDR <0.05).

III. directions of effect sizes of miRNA and gene on COPD were consistent with that from the pairwise correlation analysis.

A total of 302 pairs were identified were identified by Pearson correlation test with P <0.05, of those 141 had an FDR of <0.05. Of these, 85 showed negative correlations implying miRNA regulation of mRNA (summarised in Figure 5.4). Comparison of these 85 miRNA-mRNA pairs with the results from the *in silico* target prediction (section 5.2.1) revealed 13 known target pairs, defined as "*direct*" interactions (Table 5.5) and 72 unknown pairs defined as "*indirect*" interactions (Appendix A, Supplementary Table 2).

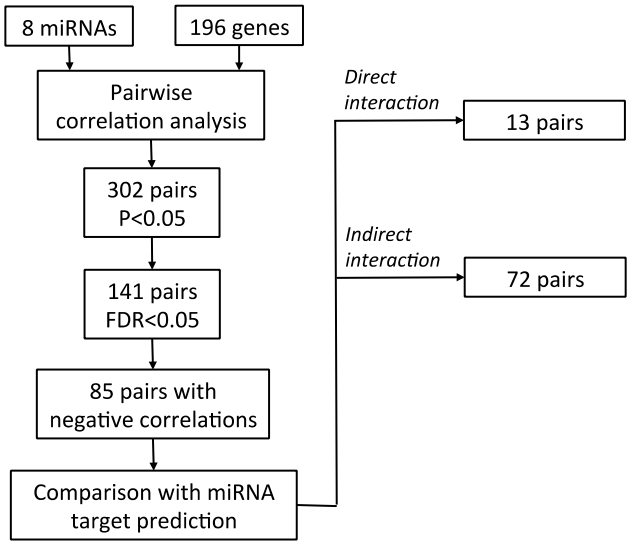


Figure 5.4 Summary of pairwise correlation analyses to identify mRNA targets of miRNA
Direct interaction = known validated or predicted gene target of miRNA; Indirect interaction = unknown gene target of miRNA. miRNA, microRNA; FDR, false discovery rate

Table 5.5 "*Direct*" miRNA-mRNA interactions from pairwise correlation analyses

miRNA	HGNC SYMBOL	Log2FC	Validated/ Predicted	r	P value	FDR
Up-regulated miRNA						
hsa-miR-182-5p	*JADE1*	-0.27	Validated	-0.52	0.0003	0.002
hsa-miR-182-5p	*PIEZO2*	-1.06	Predicted	-0.42	0.005	0.02
hsa-miR-223-3p	*ACTR3B*	-0.26	Predicted	-0.39	0.008	0.02
hsa-miR-182-5p	*ZDBF2*	-0.29	Validated	-0.39	0.008	0.02
hsa-miR-182-5p	*PLAG1*	-0.46	Validated	-0.38	0.01	0.03
hsa-miR-182-5p	*ARHGAP6*	-0.47	Predicted	-0.38	0.009	0.03
Down-regulated miRNA						
hsa-miR-138-5p	*AKR1C2*	0.88	Validated	-0.46	0.002	0.009
hsa-miR-338-3p	*AKR1C2*	0.88	Predicted	-0.44	0.003	0.01
hsa-miR-338-3p	*CEMIP*	1.42	Predicted	-0.4	0.007	0.02
hsa-miR-338-3p	*COL17A1*	1.21	Predicted	-0.4	0.008	0.03
hsa-miR-338-3p	*RASAL1*	0.94	Predicted	-0.37	0.01	0.04
hsa-miR-138-5p	*CYP1B1*	1.83	Validated*	-0.37	0.01	0.04
hsa-miR-338-3p	*ADAMDEC1*	0.009	Predicted	-0.36	0.02	0.046

*Also a predicted target for miR-204-5p. r – generated using Pearson's correlation coefficient. FDR, false discovery rate generated using the Benjamini-Hochberg method, with significance <0.05

Only two of the up-regulated miRNA (miR-182-5p and mIR-223-3p) were found to have "*direct*"/known gene target interactions based on the pairwise correlative analysis. Of these direct interactions, miR-182-5p showed a moderate correlation with both Jade family PHD finger 1 (*JADE1*) (r = -0.52, FDR = 0.002) and Piezo Type Mechanosensitive Ion Channel Component 2 (*PIEZO2*) (r = -0.42, FDR = 0.02), whereas the rest of the up-regulated miRNA demonstrated only weak correlations with their direct targets. Of the down-regulated mRNA, miR-138-5p and miR-338-3p showed a moderate correlation with *AKR1C2* (r = -0.46, FDR = 0.009; r = -0.44, FDR = 0.01 respectively), suggesting synergistic action. In addition, miR-338-3p showed a moderate correlation with Cell migration inducing hyaluronidase (*CEMIP*) (r = -0.4, FDR = 0.02), whereas the rest of the down-regulated miRNA-mRNA pairs demonstrated only weak correlations. The "*indirect*" interactions (summarised in Appendix A, Supplementary Table 2) showed several moderate correlations, however these may be considered to be less reliable as they were not found in the target prediction databases.

5.2.4.2 Identifying the dominant miRNAs of the miRNA-mRNA interactions using miRMapper

The most relevant miRNAs in a cellular context are not necessarily those with the greatest change in expression levels between healthy and diseased tissue. Differentially expressed miRNAs that modulate a large number of mRNA transcripts ultimately have a greater influence in determining phenotypic outcomes and are sometimes considered more important than miRNA that modulate just a few mRNA transcripts.

Using the 107 miRNA-mRNA interactions (54 between up-regulated miRNA and down-regulated mRNA and 53 between down-regulated miRNA and up-regulated mRNA) generated by the analysis in section 5.2.3; and the 196 differentially expressed mRNA identified in section 5.2.2, the mirMapper package was used to identify the dominant miRNA in the miRNA-mRNAs interaction network.

5.2.4.2.1 Adjacency matrix to determine gene targets with the greatest centrality

Firstly, an adjacency matrix was produced, which organises the data for downstream analysis and defines the gene targets with the greatest degree centrality (the number of edges and so in this example regulatory miRNAs) in the miRNA-mRNA interaction network (Table 5.6).

Table 5.6 Adjacency matrix for the gene targets with the greatest degree centrality

Gene	miR-2110	miR-223-3p	miR-182-5p	miR-625-3p	miR-200b-5p	miR-138-5p	miR-204-5p	miR-338-3p	sums
CRIM1	0	1	1	0	1	0	0	0	3
BTBD7	0	0	0	0	0	1	1	1	3
GLI3	0	1	1	0	0	0	0	0	2
ING1	1	0	1	0	0	0	0	0	2
KLF10	0	0	1	0	1	0	0	0	2
PLAG1	0	0	1	0	1	0	0	0	2
FXYD6	1	0	1	0	0	0	0	0	2
FBXL14	1	1	0	0	0	0	0	0	2
FZD8	1	0	1	0	0	0	0	0	2
SLC45A3	0	0	0	0	0	1	1	0	2

0 = not a gene target of the specified miRNA; 1 = known gene target of the specified miRNA

In the epithelial brushings, the genes *CRIM1* and *BTBD7* have the greatest degree centrality in the miRNA-mRNA interaction network, with three regulatory miRNAs in each case (miR-223-3p, miR-182-5p and miR-200b-5p target CRIM1 and miR138-5p, miR-204-5p and miR-338-3p target BTBD7). Interestingly, both *CRIM1* and *BTBD7* had relatively small differential expression in the

epithelial brushings in COPD (Appendix A, Supplementary Table 1 log2FC = -0.39, FDR = 0.01 and log2FC = 0.00044, FDR 0.006 respectively). In contrast, both the most down-regulated gene transcript in COPD (atrial Natriuretic Peptide Receptor 3 *(NPR3)*, log2FC = -2.34, FDR = 0.01) and the most up-regulated gene transcript in COPD (Aldo-Keto Reductase Family 1 Member B10 *(AKR1B10*), log2FC =2.03, FDR < 0.00001) identified as a miRNA target had a degree centrality of 1, with just one miRNA target each (miR-182-5p and miR-138-5p respectively).

5.2.4.2.2 miRNA impact on gene expression in epithelial brushings

The adjacency matrix is also used as input to define the degree centrality of the miRNA itself, shown in Table 5.7 and in Figure 5.5.

Table 5.7 miRNA impact on the differential gene expression

miRNA	Predicted genes identified	Percentage of gene targets	Percentage of differentially expressed genes
hsa-miR-182-5p	33	39.3	16.8
hsa-miR-204-5p	20	23.8	10.2
hsa-miR-138-5p	18	21.4	9.2
hsa-miR-338-3p	12	14.3	6.1
hsa-miR-2110	9	10.7	4.6
hsa-miR-223-3p	6	7.1	3.1
hsa-miR-200b-5p	5	6.0	2.6
hsa-miR-625-3p	1	1.2	0.5

Although miR-223-3p was shown to be the most up-regulated miRNA in lung-derived EVs in COPD, (log2FC = 2.97, FDR = 0.02, Table 4.7), it appears to have only a small effect on the gene targets identified, impacting only 3.1% of all the differentially expressed genes and regulating only 7.1% of all genes targeted by the miRNA. However miR-182-5p, with a linear expression nearly three times smaller than miR-223-3p (log2FC = 1.52, FDR = 0.04, Table 4.7), has the greatest number of targets, impacting on 16.8% of all the differentially expressed genes and regulating 39.3% of all the genes targeted by a miRNA in the dataset (Table 5.7 & Figure 5.5).

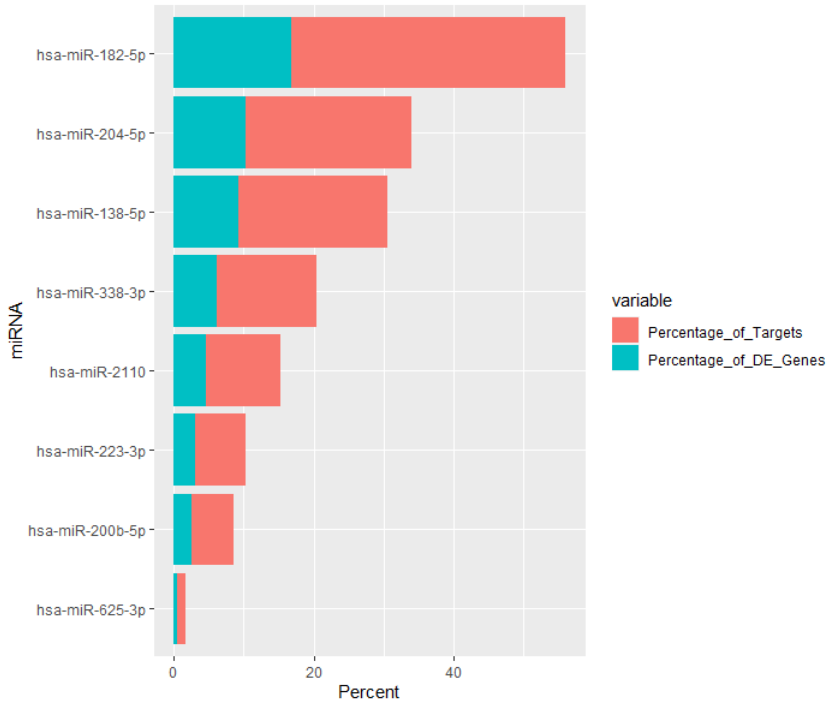


Figure 5.5 Predicted miRNA impact on genes.
Data are presented in the order of the greatest number of impacted genes to the lowest, with percentage of total targets affected by the miRNA in red and the percentage of total differentially expressed genes affected by the miRNA in blue.

5.2.4.2.3 Identification of synergistic miRNA action on differential gene expression

The mirMapper package also identifies those miRNA that are working synergistically (Figure 5.6), which is important as it is normally the action of more than one miRNA on a gene target to cause a significant impact on transcript levels (485). As shown in the dendrogram (Figure 5.6), miR-204-5p and miR-138-5p are represented in a single *clade* (branch of a dendrogram), with the shortest branch points indicating the greatest similarity based on the miRNAs' Jaccard similarity index (a statistical measure of similarity for two sets of data (486). These two down-regulated miRNA belong to two distinct miRNA clusters (487) and therefore it is not possible to infer synergistic action from their sequence analysis alone. The two up-regulated miRNA, miR-223-3p and miR-200b-5p are also present in a single *clade*, but have a longer branch point indicating less similarity than miR-204-5p and miR-138-5p. In addition, miR-625-3p is represented as a single *leaf* (terminal end of a clade) of a *clade* demonstrating no similarity to the other differentially expressed miRNA.

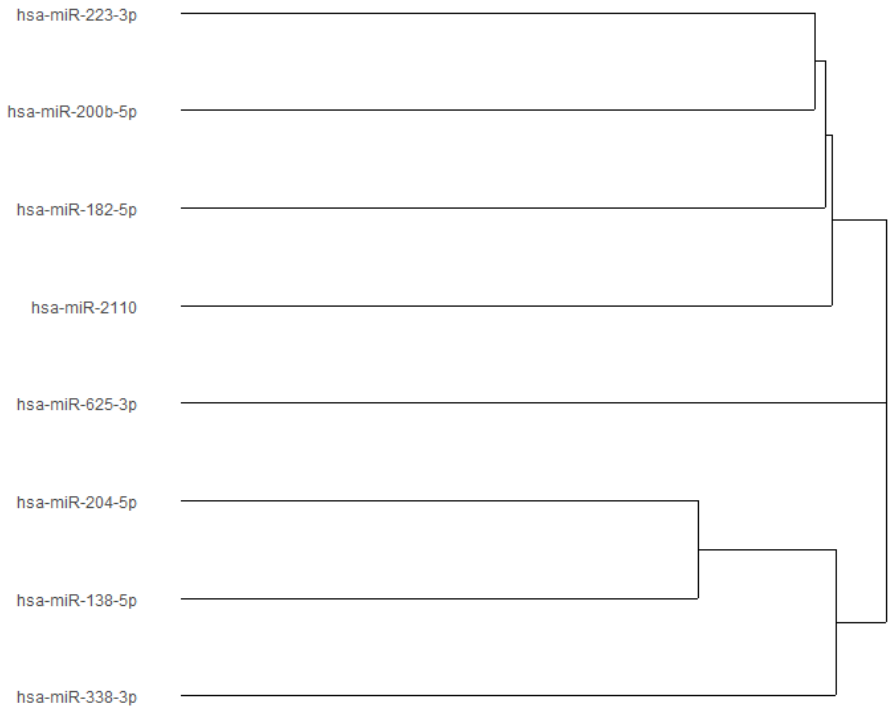


Figure 5.6 Dendrogram showing clustering of differentially expressed EV miRNA based on the similarity of the miRNAs' Jaccard index values to each other

5.3 miRNA-mRNA interaction network topology

The 107 miRNA-mRNA interactions were mapped and visualised in a network using a programme Cytoscape. The network of the up-regulated miRNA and their down-regulated gene targets are shown in Figure 5.7 and the network of the down-regulated miRNA and their up-regulated gene targets are shown in Figure 5.8. These networks display a summary of the findings from the analysis so far in this chapter, with the following attributes:

I. The circular nodes correspond to the differentially expressed genes identified in section 5.2.2. The size of the node is proportional to the differential expression (log2FC) in COPD, i.e. for the up-regulated miRNA network, the larger the node, the more down-regulated the gene was in COPD epithelial brushings.

II. The edges represent an interaction between a miRNA and mRNA. If the edge is a solid line, this represents a validated mRNA target, whereas a dotted line represents a predicted mRNA target (both of these would be considered "*direct*" interactions from the correlation analysis in section 5.2.4.1).

III. The nodes coloured in orange are those identified as significant (FDR<0.05) from the correlative analysis (section 5.2.4.1).

IV. The nodes coloured purple are those mRNA regulated by more than one miRNA and therefore have a degree centrality >1 (identified in section 5.2.3 & 5.2.4.2.1).

V. The nodes coloured yellow satisfy both iii. and iv. and therefore may be of greater significance.

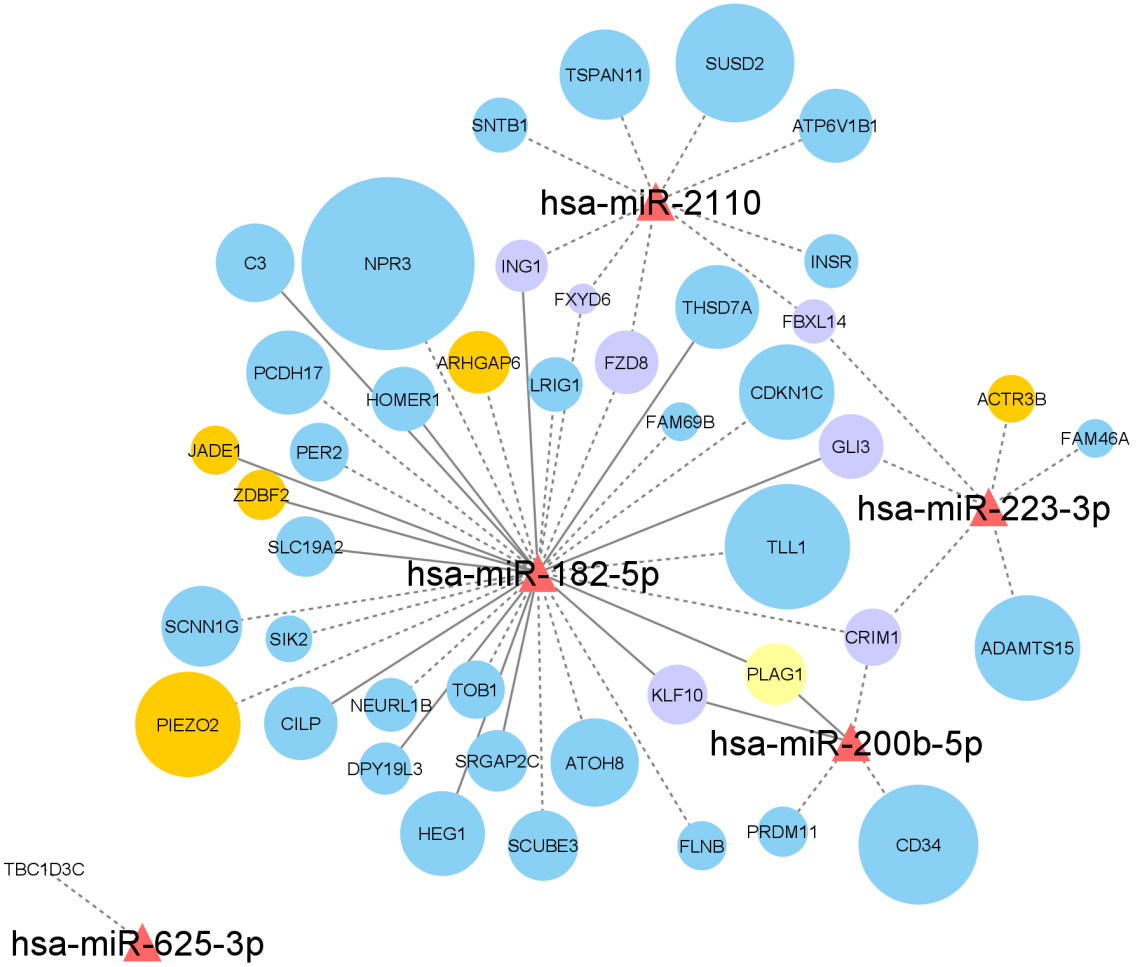


Figure 5.7 Up-regulated EV miRNA- mRNA interaction network in epithelial brushings in COPD
Red triangular nodes correspond to up-regulated EV miRNA in COPD, circular nodes correspond to down-regulated genes in COPD, these are coloured according to significance, with purple nodes showing a degree centrality >1, orange nodes identified as significant interactions from the Pearson correlation analysis, yellow nodes showing both attributes and blue nodes showing neither. The size of the circular node is proportional to the differential expression in COPD (i.e. the bigger the circle, the more down regulated it is in COPD). The edges are solid if the genes are validated targets of the miRNA and dotted if predicted targets

The miRNA with the highest node degree (number of edge incidents to the node) and the highest betweenness centrality (indicator of a node's centrality in a network) is miR-182-5p (Figure 5.7). This reinforces the results presented in section 5.2.4.2.2, which demonstrated miR-182-5p has the greatest number of targets, impacting on 16.8% of all the differentially expressed genes. Whereas, miR-625-3p has only one gene target, TBC1 Domain Family Member 3C (*TBC1D3C*), and is separate from the rest of the network, suggesting this EV

miRNA has the least regulatory effect on the differentially expressed genes and the network as a whole. Interestingly, the genes with the greatest differential expression levels (largest circles) such as *NPR3*, Tolloid-like 1 (*TLL1*), Sushi domain containing 2 (*SUSD2*), *PIEZO2* and CD34 molecule (*CD34*) are not co-targeted by any of the up-regulated miRNA. In addition, with the exception of *PIEZO2* (r = -0.42, FDR = 0.02), they were not significant in the pairwise correlation analysis (section 5.2.4.1).

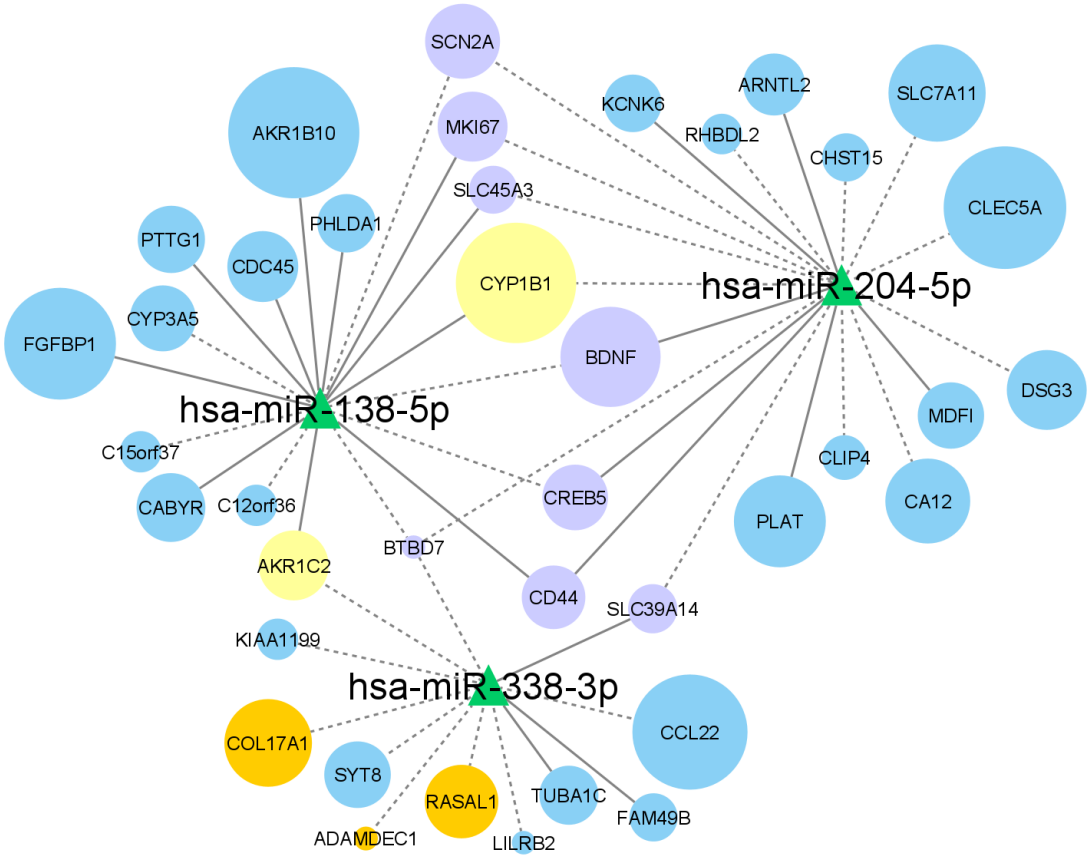


Figure 5.8 Down regulated EV miRNA-mRNA interaction network in epithelial brushings in COPD
Green triangular nodes correspond to down-regulated EV miRNA in COPD, circular nodes correspond to up-regulated genes in COPD. These are coloured according to significance, with purple nodes showing a degree centrality >1, orange nodes identified as significant interactions from the Pearson correlation analysis, yellow nodes showing both attributes and blue nodes showing neither. The size of the circular node is proportional to the differential expression in COPD (i.e. the bigger the circle, the more up regulated it is in COPD). The edges are solid if the genes are validated targets of the miRNA and dotted if predicted targets.

In the down-regulated EV miRNA-mRNA interaction network (Figure 5.8), miR-204-5p showed the highest node degree (20 gene targets) and the highest betweenness centrality, indicating it may be the most important miRNA in the network. All three down-regulated EV miRNA demonstrate possible synergistic activity with several co-targeted genes. In particular, Cytochrome P450 1B1 (*CYP1B1*) (one of the most up-regulated genes in the network, log2FC = 1.83, FDR = 0.02) is targeted by both miR-138-5p and miR-204-5p, and *CYP1B1* gene expression negatively correlated

with expression of miR-138-5p (r = -0.37, FDR = 0.04, Table 5.5). In addition, *AKR1C2* (log2FC = 0.9, FDR = 0.01) is targeted by both miR-138-5p and miR-338-3p, and *AKR1C2* gene expression negatively correlated with both expression of miR-138-5p and miR-338-3p (r = -0.46, FDR = 0.009 and r = -0.44, FDR = 0.01 respectively). This synergism suggests coordinated activity by the down-regulated EV miRNA and may lead to a greater impact on post-transcriptional gene expression and further downstream pathway effects.

5.3.1 Cluster analysis of networks

Usually in biological networks, the nodes that work together (i.e. play a role in the same pathway), are highly interconnected. These interconnected nodes or clusters may represent an important cellular module. These nodes can be identified by visual exploration of the network. The tool ClusterViz within Cytoscape was used to analyse the miRNA-mRNA network using the EAGLE clustering algorithm to identify key clusters/modules within the network. This approach identified five major sub-networks (Figure 5.9).

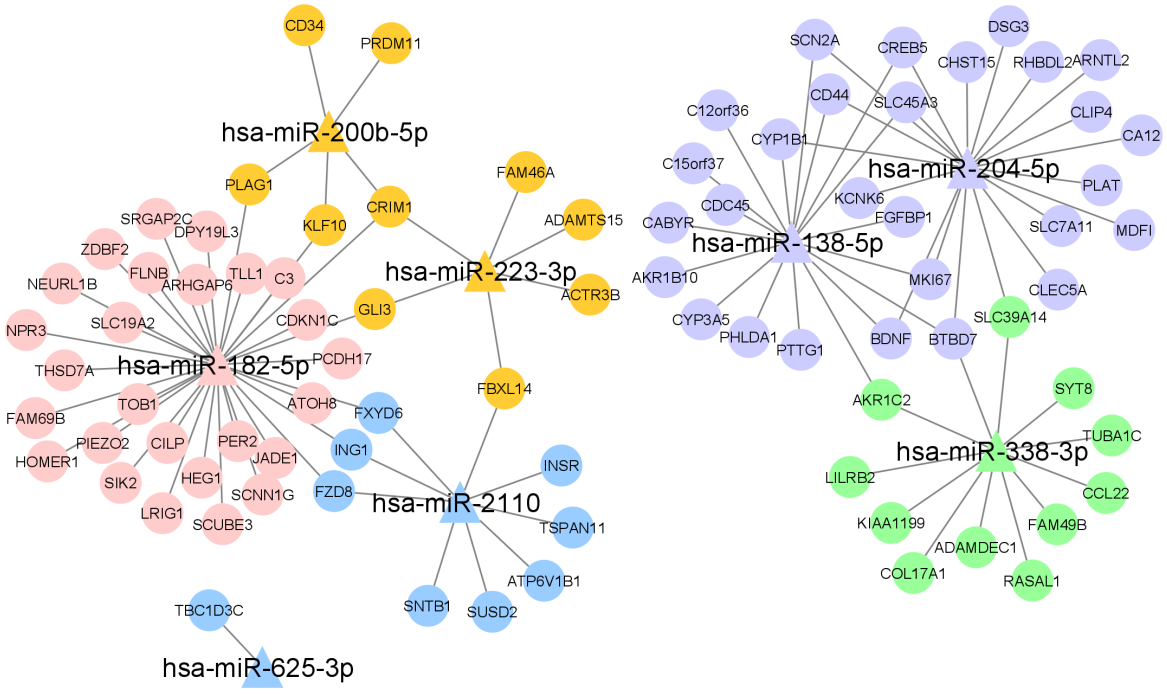


Figure 5.9 Cluster analysis of miRNA-mRNA interaction network
Each colour represents a cluster: Cluster 1 (purple), Cluster 2 (Pink), Cluster 3 (Green), Cluster 4 (Orange), Cluster 5 (Blue). Triangular nodes represent differentially expressed EV miRNA. Circular nodes represent differentially expressed gene targets in epithelial brushings.

The attributes of these five clusters are summarised below Table 5.8. The largest cluster (Cluster 1 - purple) involving two down-regulated EV miRNA (miR-204-5p and miR-138-5p) encompassed 30 nodes, with an in-degree (number of edges within the cluster) centrality of 36, and the highest

modularity of 12. This supported the findings of the miRMapper analysis (section 5.2.4.3.2), which used a different method (similarity of miRNAs' Jaccard's index values) to determine miRNA synergism, and showed that miR-138-5p and miR-204-5p were the most similar of the differentially expressed miRNA (Figure 5.5). Cluster 2 (pink) involving only miR-182-5p, contained 27 nodes and the greatest out-degree (number of edges outside the cluster) centrality (Table 5.8). The smallest cluster (Cluster 5 – blue) involved two up-regulated miRNA (miR-2110 and miR-625-3p) that did not interconnect and had the lowest modularity of 2.25. As mentioned previously, miR-625-3p is likely to be of least importance to the network as a whole with only one interaction (edge) and no shared targets. Whereas, miR-204-5p, miR-138-5p and miR-182-5p are the most central miRNAs to the network and therefore likely to have more biological significance.

Table 5.8 Attributes of the five clusters identified from the miRNA-mRNA network cluster analysis

	miRNA	Number of nodes	Number of edges inside the cluster (In-degree)	Modularity	Number of edges outside the cluster (Out-Degree)
Cluster 1 (Purple)	miR-204-5p, miR-138-5p	30	36	12	3
Cluster 2 (Pink)	miR-182-5p	27	26	3.71	7
Cluster 3 (Green)	miR-338-3p	12	11	3.67	3
Cluster 4 (Orange)	miR-200b-5p, miR-223-3p	12	11	2.2	5
Cluster 5 (Blue)	miR-2110, miR-625-3p	11	9	2.25	4

5.4 Gene Ontology Enrichment Analysis

To identify the key biological processes, molecular functions and pathways, which the identified miRNA-mRNA interactions are involved in, Gene Ontology (GO) enrichment analysis was performed using the Biological Networks Gene Ontology tool (BiNGO) within Cytoscape. BiNGO identifies the GO categories that are statistically overrepresented in a set of genes or a sub-graph of a biological network. BiNGO maps the predominant functional themes of a given gene set on the GO hierarchy and outputs this mapping as a Cytoscape graph. From this analysis, the intention is to prioritise one or a few gene targets to perform functional studies for their interaction validation.

5.4.1 Gene ontology enrichment analysis for GO: biological process

In this analysis the BiNGO output was generated using the hypergeometric statistical test followed by an FDR correction (<0.05) for overrepresentation of the target genes against the GO_biological_process dataset. This output was used to generate an enrichment map using a p value < 0.01 (Figure 5.10).

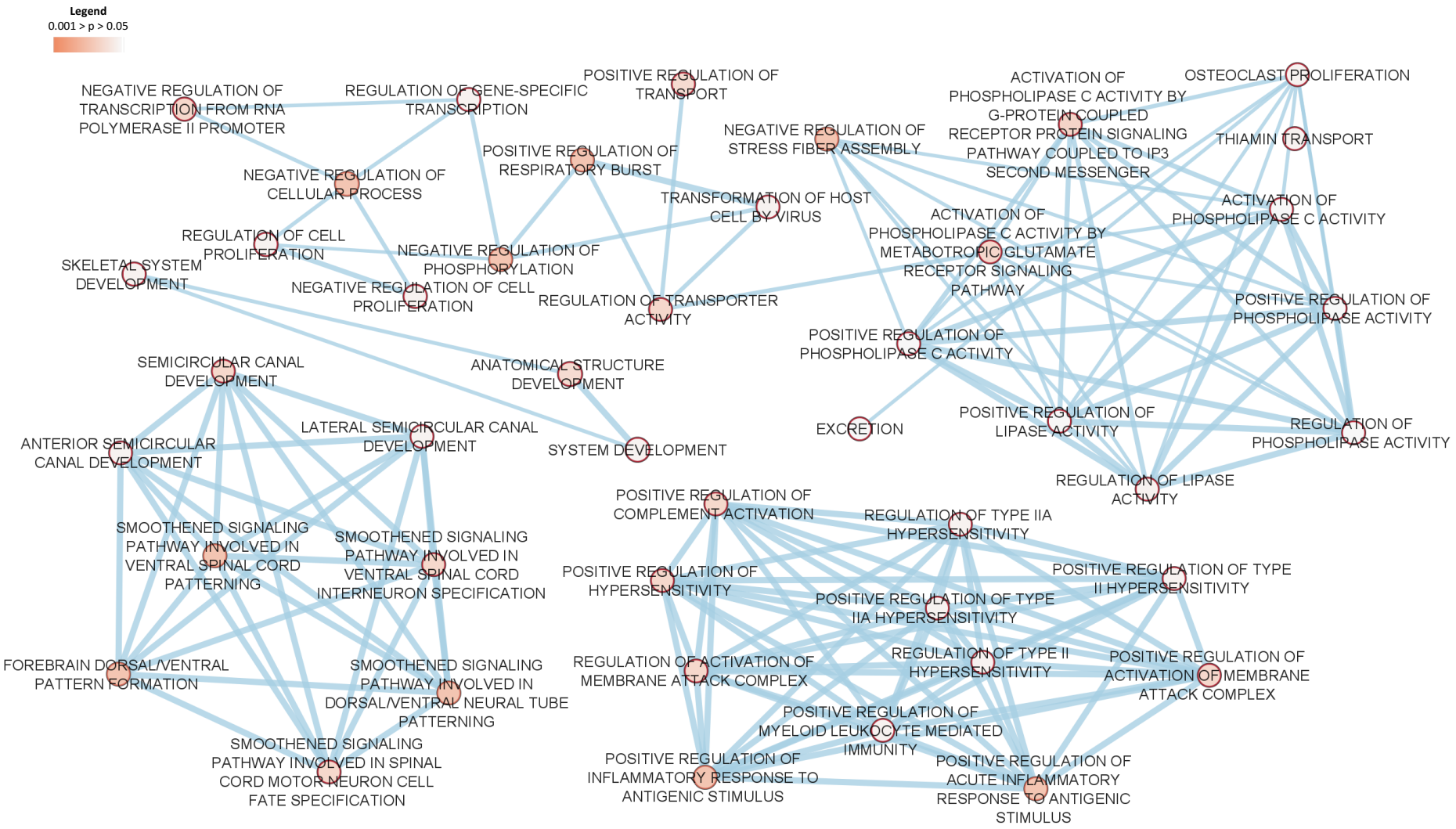


Figure 5.10 Enrichment Map from BiNGO output with most significant GO terms for biological process and their interactions.
Circles represent GO: biological process, with a darker red colour corresponding to increasing significance. Edges in blue represent an interaction between biological processes.

Although this analysis identified several significant (p<0.01) GO: biological processes for the miRNA-mRNA network, none passed the FDR <0.05 threshold. This is likely due to the interaction network being small and only involving 107 target genes. The most significant GO terms for GO: biological processes are summarised in Table 5.9. These include several GO terms relating to regulation of phospholipase activity, which mapped to the following target genes *NPR3*, Homer scaffold protein 1 (*HOMER1*) and Rho GTPase activating protein 6 (*ARHGAP6*).

Table 5.9 Most significant GO terms generated from BiNGO output for GO biological process

GO ID	Description	# Genes	Mapped Genes	P value	FDR
GO:7588	Excretion	3	*SCNN1G*|*NPR3*|*ATP6V1B1*	0.0002	0.08
GO:7202	Activation of phospholipase c activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.0003	0.08
GO:10863	Positive regulation of phospholipase c activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.0003	0.08
GO:60193	Positive regulation of phospholipase activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.0004	0.08
GO:10517	Regulation of phospholipase activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.0005	0.08
GO:60193	Positive regulation of lipase activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.0005	0.08
GO:1501	Skeletal system development	5	*NPR3*|*ATP6V1B1*|*KLF10*| *TLL1*|*GLI3*	0.0007	0.09
GO:60191	Regulation of lipase activity	3	*NPR3*|*HOMER1*|*ARHGAP6*	0.001	0.09
GO:32583	Regulation of gene-specific transcription	4	*GLI3*|*CDKN1C*|*PER2*|*INSR*	0.001	0.09
GO:48731	System development	13	*NPR3*|*ATP6V1B1*|*HOMER1*| *KLF10*|*TLL1*|*GLI3*|*CDKN1C*| *INSR*|*PLAG1*|*CRIM1*|*FZD8*| *ATOH8*|*FLNB*	0.001	0.09

FDR, False discovery rate; GO, Gene ontology

5.4.2 Gene ontology enrichment analysis for GO: molecular function

In this analysis, the BiNGO output was generated using the hypergeometric statistical test followed by an FDR correction (<0.05) for overrepresentation of the target genes against the GO_molecular function dataset. This output was used to generate an enrichment map using a p value < 0.01 (Figure 5.11).

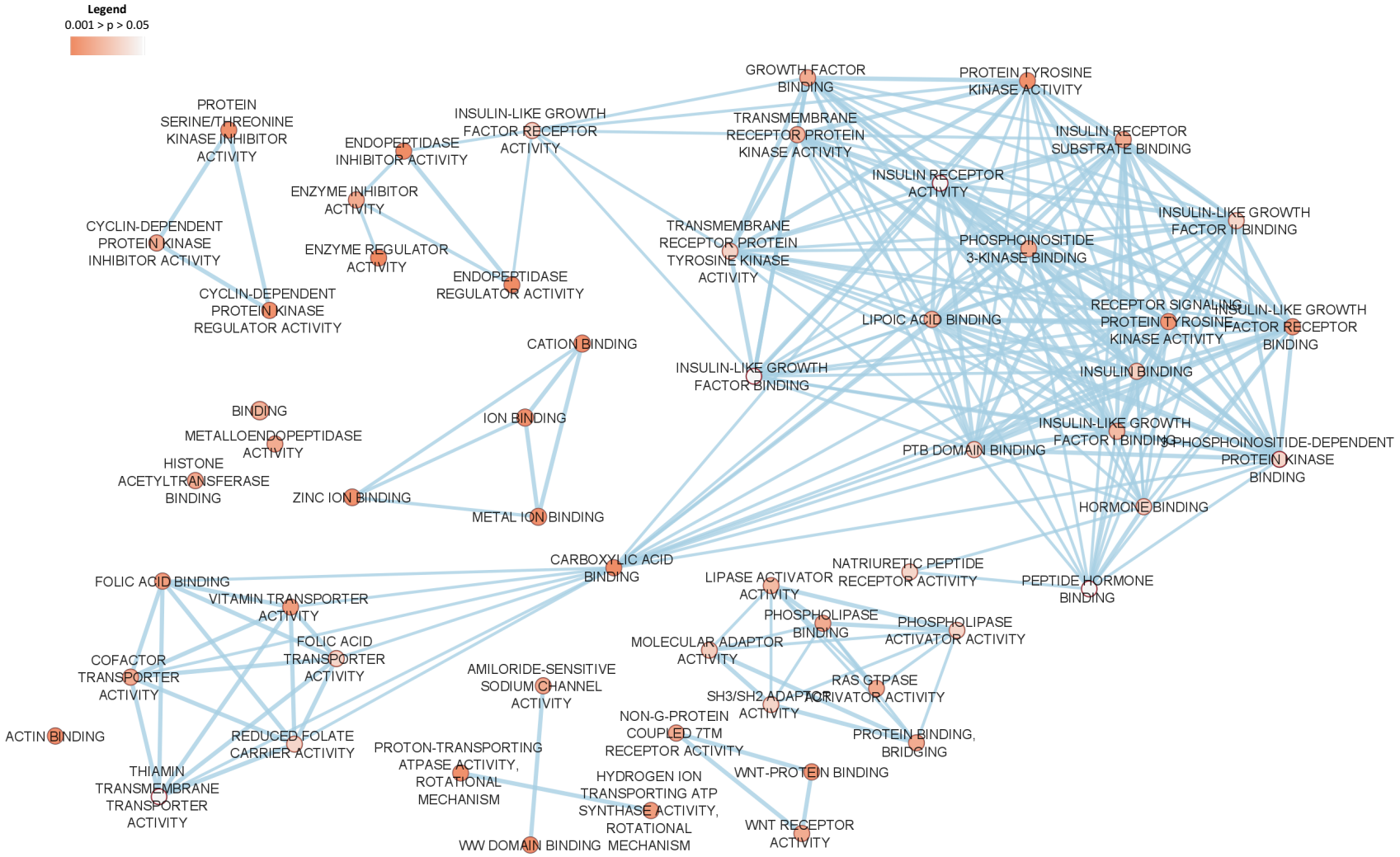


Figure 5.11 Enrichment Map from BiNGO output with most significant GO terms for molecular function and their interactions.
Circles represent GO: molecular functions, with a darker red colour corresponding to increasing significance. Edges in blue represent an interaction between molecular functions.

Again, although this analysis identified several significant (p<0.01) GO: molecular function terms for this miRNA-mRNA network, however none passed the FDR <0.05 threshold. The most significant GO terms for GO molecular function are summarised in Table 5.10. These include several GO terms relating to regulation of hormone activity, with the following mapped genes Insulin receptor (*INSR*), *CRIM1* and *NPR3*.

Table 5.10 Most significant GO terms generated from BiNGO output for GO molecular function

GO ID	Description	# Genes	Mapped Genes	P value	FDR
GO:5200	Insulin-like growth factor binding	2	*INSR*|*CRIM1*	0.001	0.1
GO:17046	Peptide hormone binding	2	*INSR*|*NPR3*	0.002	0.1
GO:5009	Insulin receptor activity	1	*INSR*	0.002	0.1
GO:15234	Thiamine transmembrane transporter activity	1	*SLC19A2*	0.002	0.1
GO:43423	3-phosphoinositide-dependent protein kinase binding	1	*INSR*	0.005	0.1
GO:5070	Sh3/sh2 adaptor activity	2	*ARHGAP6*|*TOB1*	0.006	0.1
GO:42562	Hormone binding	2	*INSR*|*NPR3*	0.007	0.1
GO:51425	Ptb domain binding	1	*INSR*	0.007	0.1
GO:16941	Natriuretic peptide receptor activity	1	*NPR3*	0.007	0.1
GO:16004	Phospholipase activator activity	1	*ARHGAP6*	0.009	0.1

FDR, False discovery rate; GO, Gene ontology

5.5 Discussion

Chapter 5 presents the findings of EV miRNA gene target identification, differential expression analysis of the epithelial brushing transcriptome in paired samples, and the interactions of the differentially expressed lung EV miRNA with the differentially expressed genes in these epithelial brushings. The chapter explores these miRNA-mRNA interactions further to identify key pathways and biological processes under miRNA regulation that may be pertinent to COPD pathology.

5.5.1 Identifying differentially expressed EV miRNA target genes within epithelial brushings

The results of the *in silico* analysis identified over 21,000 target genes of the eight differentially expressed lung-derived EV miRNA. The majority of these genes were predicted targets (64.3%) rather than validated in experimental models. This distinction is important to note as miRNA target prediction methods mainly focus on programming alignment to identify complementary elements in the 3'-untranslated region (UTR) within the seed sequence of the miRNA and the phylogenetic conservation of the complementary sequences in the 3'-UTRs of orthologous genes. Although, evidence suggests that perfect seed pairing may not necessarily be a reliable predictor for miRNA interactions (488). Therefore experimental validation is important to endorse the physiological functions and clinical relevance of specific miRNA on their predicted gene targets. However in this study, using both predicted and validated targets for analysis may reveal both novel and known miRNA-mRNA interactions that could be important in COPD pathogenesis.

A major strength of the overall MICA II study is the paired sampling approach, which encompassed epithelial brushings, alveolar macrophages and epithelial biopsies from the same patients as the lung EVs were isolated from. Epithelial brushings were chosen as a focus of miRNA regulation as they demonstrated the greatest differential gene expression between COPD and healthy ex-smokers, compared with the alveolar macrophage and epithelial biopsy samples (489). Furthermore, they are a key regulatory cell in COPD, not only forming a physical barrier to harmful pathogens and toxins, but also a coordinator of the innate immune response (98-100). In addition, there is clear evidence of epithelial cell compromise in COPD (94-97, 302).

The differential expression analysis of the epithelial brushing transcriptome revealed 196 differentially expressed genes (FDR <0.05). Nearly 40% of these had a log2FC >1/-1 demonstrating significant up/down regulation in COPD. Although not all of these genes were found to be regulated by the differentially expressed miRNA, these dysregulated genes are discussed in more

detail in the next section, as they may be key to mechanisms and pathways important to the pathophysiology in COPD.

5.5.1.1 Up-regulated genes in COPD epithelial brushings

Several of the up-regulated genes in the epithelial brushings have been identified previously as dysregulated in COPD. Fetuin B (*FETUB*) was the most up-regulated gene in the COPD epithelial brushings compared with healthy ex-smokers (Table 5.2, log2FC = 2.87, FDR = 0.01). FETUB is a liver-derived plasma protein (490) and has been reported to influence glucose metabolism (491). Raised plasma FETUB in COPD compared with controls was identified by isobaric tags for relative and absolute quantitation (iTRAQ)-based proteomics and validated in a second cohort by enzyme-linked immunosorbent assay (ELISA) (492). This latter study also showed that plasma FETUB predicted the occurrence of acute exacerbations and levels negatively correlated with disease severity (FEV1 % predicted) (492). Furthermore, the EvA study identified *FETUB* and retinoic acid early transcript-1 (*RAETL1*) (another of the up-regulated genes in this study) as two of the top differentially expressed genes in epithelial brushings of patients with COPD and high blood eosinophil counts (>200 eosinophils/μL) (493). This suggests that *FETUB* and *RAETL1* may play a role in eosinophilic COPD (to be explored further in Chapter 6).

Additionally, other up-regulated genes in the COPD epithelial brushings included arachidonate 15-lipoxygenase type B (*ALOX15B*) (Table 5.2, log2FC = 2.61, FDR = 0.008), which is a member of the lipoxygenase family and is involved in resolution of inflammation (494). *ALOX15B* is implicated in conditions that complicate inflammatory lung disease such as nasal polyps and chronic rhinosinusitis (495), and the mechanism by which roflumilast (phosphodiesterase - 4 (PDE4) inhibitor) works in COPD (496). The matrix metalloproteinase 12 (*MMP12*) gene was also significantly up-regulated in COPD epithelial brushings (Table 5.2, log2FC = 2.49, FDR = 0.02). *MMP12* is considered to play a key role in the development of emphysema and small airway remodelling (497). Interestingly there is also published data suggested that selective inhibition of *MMP12* might be a viable therapy for COPD (497), raising the possibility of miRNA post-transcriptional regulation of *MMP12* as a possible future therapy. Interleukin 1 receptor type 2 (*IL1R2*) was also found up-regulated in COPD epithelial brushings compared with healthy ex-smokers (Table 5.2, log2FC = 2.34, FDR = 0.0005). Baines et al. demonstrated raised IL1R2 in sputum of patients with COPD and that sputum IL1R2 may predict future exacerbations (498). Importantly, the *IL1R2* gene encodes for one of the IL-1 cytokine receptors that binds IL-1β, a potent activator of alveolar macrophages in COPD leading to MMP-9 production (108), and thus together these findings further illustrate the importance of *IL1R2* dysregulation in COPD.

5.5.1.2 Down-regulated genes in COPD epithelial brushings

Atrial natriuretic peptide receptor 3 (*NPR3)* was the most down-regulated known transcript in COPD epithelial brushings compared with healthy ex-smokers (Table 5.2, log2FC = -2.01, FDR = 0.0003), and has been implicated in COPD previously (499). *NPR3* encodes for one of the three natriuretic peptide receptors, which have an essential role in the regulation of blood pressure, intravascular volume and electrolyte homeostasis (500). Specifically, the NPR3 receptor is primarily responsible for clearing circulating and extracellular natriuretic peptides (501). However, several studies indicate that NPR3 receptor activation may also trigger several intracellular signalling pathways, including cyclic adenosine monophosphate (cAMP) inhibition (501), a potent regulator of innate and adaptive immune cell functions, and phospholipase C signalling (502) (to be discussed in more detail in section 5.5.3). NPR3 can also activate mitogen-activated protein kinase (MAPK)/ phosphoinositide 3-kinase (PI3K) signalling pathways, which are involved in cell proliferation in vascular smooth muscle cells (503). Furthermore, in the context of the lung, *NPR3* gene expression is selectively down-regulated in the lungs of rats and mice in response to hypoxia and may enhance the vasodilator effects of atrial natriuretic peptide (ANP) in the lung, thus modulating hypoxic pulmonary vasoconstriction/hypertension (504). Most recently, Kachroo et al. showed *NPR3* was differentially methylated in smoke exposed fetal lung samples (505) and *NPR3* presence (along with *TLL1),* another significantly down-regulated gene in this study) correlated with diseased parenchyma in lung biopsies (506). Furthermore, in a COPD mouse model, *NPR3* regulation was implicated in treatment responsiveness to a glucagon-like peptide (GLP)-1 agonist (499).

In contrast to this study (where subjects were all ex-smokers and *TLL1* was down-regulated), *TLL1* expression was found significantly up-regulated in bronchial brushes compared with nasal epithelium after exposure to smoke, implying that *TLL1* gene expression may vary depending on the temporal nature of smoke exposure (507). In addition, *TLL1* was found significantly dysregulated in subtypes of asthma (508), and importantly encodes a metalloprotease that is involved in degradation of the extracellular matrix suggesting a role in airway remodelling (509).

CD34 is a protein coding gene for haemapoietic stem cells, these are found be reduced in peripheral blood of smokers (510) and patients with COPD (511), and are important for vascular endothelial repair (512). *CD34* transcript was down-regulated in COPD epithelial brushings compared with healthy ex-smokers (Table 5.2, log2FC = -1.27, FDR = 0.02), which may suggest a lack of recruitment of these stem cells to the lungs in COPD, impairing tissue repair and regeneration. In addition, other down-regulated genes, WNK lysine deficient protein kinase 4

(*WNK4)* and *SUSD2* have been identified as altered in co-methylation analysis in lung tissue in COPD (505). Integral membrane protein 2A (*ITM2A)* has previously been identified as down-regulated in COPD lung tissue (513), whereas metallothionein 1G (*MT1G)* was up-regulated in the same study in contrast to the results of this study (Table 5.2, log2FC = -1.12, FDR = 0.012). It is worth noting that *MT1G* codes for the protein metallothionein 1 which is transcriptionally regulated by glucocorticoids, with increased metallothionein expression reflecting steroid resistance (514).

A number of the transcripts identified from the differential expression analysis were new transcripts with very little literature base, such as ENSG00000287059, ENSG00000287771, ENSG00000198787, ENSG00000214870, ENSG00000283413.

Overall the differentially expressed genes in the epithelial brushings in COPD revealed several possible (and some novel) mechanisms underlying COPD pathophysiology. Encouragingly, many of these dysregulated genes were already found in the COPD literature (e.g. *ALOX15B*, *MMP12* and *IL1R2*), which increases the likelihood of their significance to underlying biology. Several of the genes were shown to be epigenetically modified in COPD, (*NPR3*, *WNK4*, *SUSD2* modified by DNA methylation rather than miRNA (505)) suggesting that there may be dual epigenetic control of some genes related to COPD pathogenesis. Furthermore, certain genes were found to have implications for treatment in COPD, such as steroid resistance (e.g. *MTIG*) or an association with eosinophilic disease (e.g. *FETUB*). Finally the dysregulation of *CD34* raised important insights into the role of stem cells in vascular remodelling in COPD and whether this is a key modifiable factor that is currently under-explored in this disease.

5.5.2 Identifying putative miRNA-mRNA interactions

The multiMiR package identified 107 putative interactions between the eight differentially expressed EV miRNA and the 196 differentially expressed genes in the COPD epithelial brushings.

Further analysis suggested 13 of these interactions had a significant negative correlation when comparing their normalised expression profiles, increasing the likelihood of a meaningful interaction within this sample. The strongest correlation in the up-regulated EV miRNA was *JADE1* expression with miR-182-5p (Table 5.5, r = -0.52, p = 0.0003). *JADE1* is a negative regulator of Wingless/Integrase-1 (WNT) signalling, where previous studies have suggested the shift from canonical to non-canonical WNT signalling in the COPD alveolar epithelium promotes emphysema through abnormal alveolar repair (515). *JADE1* has previously been shown to be differentially methylated in COPD and was one of the biomarkers shown to predict future risk of COPD (516).

Of the down-regulated EV miRNA gene targets, *AKR1C2* showed significant negative correlations with both miR-138-5p and miR-338-3p. This synergistic activity increases the likely biological impact of these interactions. The AKR1C family, including *AKR1C2,* code for enzymes implicated in steroid metabolism and their expression levels are localised in the normal tissues of the lung, liver, prostate, testis and mammary glands (517). Their regulation is associated with several cancers, however their exact mechanism in promoting tumourigenenis is not known (518). Furthermore, *AKR1C2* has been found up-regulated in the airways of COPD and healthy smokers (519), in keeping with this study (Table 5.3, log2FC = 0.88, FDR = 0.01).

The gene *CEMIP* was up-regulated in COPD epithelial brushings (Table 5.2, log2FC = 1.42, FDR = 0.002) and significantly correlated with miR-338-3p (Table 5.5, r = -0.4, FDR = 0.02). *CEMIP* has been implicated in pulmonary fibrosis (520), inflammation (521) and tumorigenesis (522) and therefore may have a significant role in COPD pathogenesis. Collagen type XVII alpha 1 chain (*COL17A1*), a protein-coding gene for XVII collagen, was also found up-regulated in COPD epithelial brushings and negatively correlated with miR-338-3p expression (Table 5.5, r = -0.39, FDR = 0.03). Up-regulation of *COL17A1* has been shown previously in COPD (523) and given its role in collagen formation, dysregulation of *COL17A1* has implications for extracellular matrix remodelling and emphysema pathogenesis (524).

5.5.2.1 Identifying the most dominant miRNA-mRNA interactions

Overall, several of these interactions have biological importance to COPD pathogenesis. However, using this correlative approach to analyse the relationships between miRNA and mRNA can miss/under-represent targets of biological importance that are regulated by multiple miRNA (525). The mirMapper analysis focused on identifying the most dominant miRNA-mRNA interactions. This analysis identified *CRIM 1* and *BTBD7* as the two most targeted genes with three miRNA targets each (Table 5.6). The gene *CRIM1* was found to be down-regulated in COPD epithelial brushings and is targeted by three up-regulated EV miRNA (miR-223-3p, miR-182-5p and miR-200b-5p). *CRIM1* has been implicated in inhibiting the invasion and metastasis of lung adenocarcinoma cells via regulation of miR-182 (526), and may play a role in capillary formation and maintenance during angiogenesis (527). Furthermore, *BTBD7* was found to be up-regulated in COPD epithelial brushings and is regulated by all three down-regulated miRNA. Altered methylation of *BTBD7* has been described in asthma (528) and implicated in lung cancer progression (529). Yet, although these findings suggest both *CRIM1* and *BTBD7* have the greatest degree of centrality to the miRNA-mRNA interaction network, their relatively small differential expression in the COPD epithelial brushings (log2FC = -0.39 FDR = 0.02; log2FC = 0.0004, FDR =

0.006 respectively) are unlikely to translate into a large impact on disease specific pathways. In contrast, *BDNF* and sodium voltage-gated channel α subunit 2 (*SCN2A*) had much higher differential expression in COPD epithelial brushings (log2FC = 1.44 and 0.97 respectively), and were also co-regulated by miRNA (miR-138-5p and miR-204-5p) suggesting a greater potential impact on downstream pathways. Levels of serum BDNF have been shown in a number of studies to correlate with disease severity in COPD (530-532) and its role as a mediator of neuronal plasticity, has been shown to be key in acute and chronic inflammatory conditions of the airways (533, 534). Furthermore, mutations in the *SCNA* family have a strong association with pulmonary emphysema (535) and therefore modulation of *SCN2A* in the airway epithelium may have important consequences for COPD pathogenesis.

In addition to focusing on gene targets, the miRmapper analysis also highlighted the miRNA with the greatest degree centrality in the miRNA-mRNA interaction network and any synergism within the differentially expressed EV miRNA. The up-regulated EV miRNA, miR-182-5p, had the highest degree centrality, targeting 39.3% of all differentially expressed targets (Table 5.7) and therefore is presumed to have a greater impact on biological function, even though it is not the most differentially expressed miRNA. However, more recent studies suggest rather than a single miRNA leading to meaningful biological effect, that cooperative regulation of miRNA on a group of genes is far more likely (536). Therefore, two cluster-based methods (Jaccard similarity index and EAGLE algorithm) were used to interrogate the miRNA-miRNA interaction network, and these methods both identified miR-204-3p and miR-138-5p with the greatest synergistic activity based on shared targets (Figure 5.5 and Figure 5.8). Interestingly, two of these gene targets have already been discussed as potentially important to COPD pathogenesis (*SCN2A* and *BDNF*). In addition, the up-regulated *CYP1B1* gene in epithelial brushings in COPD (Table 5.5. log2FC = 1.83, FDR = 0.01) is also co-targeted by miR-204-3p and miR-138-5p, and was found to be negatively correlated with miR-138-3p expression (Table 5.5, r = -0.37, FDR = 0.04). The gene *CYP1B1* encodes a member of the cytochrome P450 superfamily of enzymes, which catalyse many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids (537). Specifically, *CYP1B1* has been shown to metabolically activate polycyclic aromatic hydrocarbons (PAHs), such as those found in tobacco smoke, to ultimately generate carcinogenic compounds (538). Moreover, Smerdova et al. demonstrated pro-inflammatory cytokines, such as tumour necrosis factor (TNF)-α augmented the effects of PAHs through *CYP1B1* (539). Finally, in keeping with this study, up-regulation of *CYPB1* expression has been previously shown in type II alveolar epithelial cells from patients with COPD (540). Taken together these findings suggest an important role for *CYPB1* regulation in the development of lung cancer in pro-inflammatory lung conditions.

5.5.2.2 Visualising the miRNA-mRNA interactions in a network

To summarise the findings of the miRNA-mRNA interaction analysis, an open source network visualisation software (Cytoscape) was used (398). Although node attributes, such differential expression, are often incorporated into the network structure, to my knowledge this is the first miRNA-mRNA network representation to include results of the analytical approaches used in this thesis. In order to determine the important miRNA-mRNA interactions, results of the *in silico* analysis identifying predicted and validated gene targets, correlation analysis and miRmapper analysis were all displayed in the network to give a comprehensive overview of the analytical methods and results. This enabled specific miRNA-mRNA interactions with likely greater biological significance to be highlighted and these interactions have been discussed above. However, although network visualisation gives insights into the individual miRNA-mRNA interactions and clusters, GO enrichment analysis highlights biological pathways where these interactions are prominent and thereby identifies processes that may by key to COPD pathogenesis.

5.5.3 GO enrichment analysis to identify key pathways regulated by miRNA-mRNA network

Gene ontology enrichment analysis identified several significant biological processes, however none passed the FDR of <0.05. This is likely due to the small sample size, but despite this conclusions can be drawn from those processes highlighted in this analysis. Strikingly, processes involving regulation of phospholipase (PL) were repeatedly found to be significant in this analysis and mapped to three significantly differently expressed genes *NPR3*, *HOMER1* and *ARHGAP6*. PLs are a ubiquitous group of enzymes that have a diverse set of functions, and include the signalling effector PLC family which regulate various cellular activities through activated protein kinase C and calcium metabolism (541). In this study, activation and regulation of PLC activity were highlighted in the top three GO: biological processes. Studies have shown that PLC activation is important for influenza virus infection of lung epithelial cells (542), and is involved in up-regulation of pro-inflammatory cytokines (via NFκB-MAPK signalling), as well as the generation of intracellular reactive oxygen species (ROS) (543). Furthermore, accumulated evidence has suggests that the PLC signalling inhibitor U73122 attenuates both acute and chronic inflammation (544). Little is known about PLC function and expression in COPD and therefore these findings highlight possible novel pathways for exploration in this disease.

Of the genes mapped to these biological processes, *NPR3* was identified as the most down-regulated in epithelial brushings (see above for activity related to COPD), whereas *ARHGAP6*

significantly correlated with miR-182-5p expression (Table 5.5 r = -0.38, FDR = 0.03). The latter encodes a member of the rhoGAp family of proteins, which play a role in the regulation of actin polymerization at the plasma membrane during several cellular processes and have also been implicated in COPD pathogenesis (25). *HOMER1* encodes proteins that serve as adaptors linking receptors that activate the phosphoionositide 3-kinase (PI3K) - mammalian target of rapamycin (mTOR) pathway (545), which plays a key role in cellular senescence and autophagy. Therefore, dysregulation of *HOMER1* is may be implicated in the accelerated ageing process implicated in COPD progression (see section 1.1.3.2).

The GO enrichment analysis for molecular function again identified several significant molecular functions, however none passed the FDR of 5% for the same reasons outlined above. These results identified a number of hormone regulatory pathways such as insulin-like growth factor (IGF) binding with key mapped genes *INSR*, *NPR3* and *ARHGAP6*. IGF-1 signaling modulates the development and differentiation of many types of lung cells, including airway basal cells, club cells, alveolar epithelial cells, and fibroblasts. Although, the components of the IGF-1 signalling pathway are potentiated as biomarkers as they are dysregulated locally or systemically in COPD, much of data may be inconsistent or even paradoxical among different studies (546). Therefore, further work elucidating the exact mechanisms of this dysregulation in COPD is required. The significance of this dyrsegulation of insulin receptor activity may also have a bearing on the systemic nature of COPD and provide insights into the pathological mechanisms linking COPD with metabolic multimorbidity such as diabetes and obesity (547).

5.5.4 Strengths and limitations

A major strength of this work is the paired dataset between the EV miRNA and epithelial brushings from the same lung lobe, in the same subject. This enables a more accurate description of the miRNA-mRNA interactions, rather than relying on *in silico* analysis alone or even experimental models, which are an over-simplification of the disease process. Next the combinatorial analytical approach of these miRNA-mRNA interactions, with correlations, synergistic analysis, network visualisation and clustering enables emphasis of specific interactions, which may have more biological relevance. This enables a more focused insight into the interaction network and highlights targets that could be taken forward for validation in experimental or therapeutic models.

As mentioned above, there is a limitation to mining miRNA target prediction databases which mainly focus aligning complementary features in the 3'- UTR within the miRNA and genes (488). In

addition, the *in silico* analysis was limited to the top 20% of predicted genes, which may have missed some additional novel findings. However, the search identified over 21,000 target genes and subsequent analysis has focused on narrowing these to a few for further investigation, which ultimately will provide more certainty into the importance of these miRNA-mRNA interactions.

As seen with the GO enrichment analysis, the interaction miRNA-mRNA network was relatively small and therefore did not highlight biological processes or molecular functions that passed the FDR of 5%. The sample size was restricted by the availability of large volumes of BALF for EV isolation, which can be hard to achieve particularly in patients with COPD. Regardless, this study is the largest looking at lung derived EV miRNA and their impact on target genes in the epithelium.

One of the questions this study raises, is whether the airway epithelium is the most applicable target cell type to study. Lung EVs are known to target several cell types and an alternative approach could include analysis of the alveolar macrophage and bronchial biopsy transcriptomes. The epithelial brushing mRNA data was chosen initially as it had the greatest number of samples available which matched the EV miRNA lung samples and the greatest number of differentially expressed genes. However, given alveolar macrophages are key inflammatory cell mediators in COPD, the macrophage transcriptome may be the next dataset to explore.

5.5.5 Summary

In conclusion, these analyses describe the possible interactions and regulation of lung-derived EV miRNA with differentially expressed genes identified in the epithelial brushings of patients with COPD. Using combinatorial bioinformatic methods, these results identified a novel miRNA-mRNA interaction network and highlighted key signalling pathways, which may be helpful in understanding the pathophysiological changes of COPD at a transcriptome level.

Chapter 6 Diagnostic use of extracellular vesicle miRNA and the relationship with COPD inflammatory endotypes

6.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is an umbrella diagnosis currently defined by the presence of airflow limitation (measured by spirometry) and characteristic symptoms alone. Although spirometry is the most reproducible and objective measurement of airflow limitation, there are concerns that spirometry is an effort dependent test and can be insensitive to early disease (141). Therefore exploring additional diagnostic strategies may improve early diagnosis and give potential insights into the underlying biology of the disease.

Despite the clinical heterogeneity of COPD, it has proved difficult to identify distinct endotypes of disease. Endotypes are important as they describe a distinct pathophysiological mechanism at a cellular or molecular level, with the aim of identifying specific treatment targets for a greater impact on disease control. Different inflammatory patterns have been described in COPD and are referred to as "*inflammatory endotypes*" (described in section 1.5), however the true molecular mechanisms underlying these remain uncertain. Most patients have increased neutrophils and macrophages in their lungs reflecting the inflammatory nature of the disease (548). Some patients also have increased eosinophils, which are associated with more frequent exacerbations (549) and importantly, predict good response to corticosteroid treatment (549, 550). However, studies have failed to show the same improvement with other treatments targeting eosinophilia (e.g. anti-interleukin (IL)-5 antibodies) (178) therefore further research is needed to understand the mechanisms behind these inflammatory endotypes in COPD.

Chapter 6 describes the clinical characteristics of the subjects included in the bronchoalveolar lavage fluid (BALF) extracellular vesicle (EV) microRNA (miRNA) validation analysis (section 4.5) and their relationship with lung-derived EV miRNA expression. Furthermore, it explores the predictive ability of these differentially expressed miRNA to differentiate between health and COPD and whether they can differentiate between different inflammatory endotypes in COPD. Finally, this chapter explores whether the differential EV miRNA expression is translated from the lung into the peripheral blood, which would lead to a greater potential for its clinical utility as a possible biomarker of disease.

6.2 Characteristics of the subject included in the analysis exploring the predictive ability of EV miRNA to differentiate COPD from health and associate with inflammatory endotypes

For the analyses covered sections 6.3, 6.4 and 6.5, the larger cohort from the differential miRNA expression analysis performed in the validation study (section 4.5) was used as this cohort contained the greatest number of subjects (N=44), with 24 COPD subjects and 20 healthy ex-smokers included. The clinical characteristics are summarised in Table 6.1.

Table 6.1 Characteristics of subjects included in the analysis exploring the diagnostic use of BALF EV miRNA and associations with inflammatory endotypes, N=44

Subject/sample characteristics	COPD (n=24)	Healthy ex-smoker (n=20)	P value
Age, mean ±SD	70.1 ±6.9	68 ±7.3	0.34
Male, n (%)	20 (83)	11 (55)	0.06
Smoking pack years, mean ±SD	47 ±29.2	27.8 ±13	0.06
BMI, mean ±SD	29.6 ±4	28.4 ±4	0.3
FEV1 (% predicted), mean ±SD	77.5 ±14.8	101.8 ±14.6	**<0.00001**
FVC (% predicted), mean ±SD	102.8 ±16	100.6 ±16.4	0.65
FEV1/FVC%, mean ±SD	57.7 ±8.3	78.2 ±4.2	**<0.00001**
FEF 25-75 (% predicted), mean ±SD	41.2 ±16.7	106 ±25.4	**<0.000001**
DLCO (% predicted), mean ±SD	75.1 ±13.3	88.39 ±4.4	**0.004**
COPD status, GOLD stage, n (%)			0.41
Mild	10 (42)	NA	
Moderate	14 (58)	NA	
Baseline & historic blood counts			
Total blood leucocytes (10^9/L), mean ±SD	7.4 ±1.4	6.7 ±1.4	0.09
Absolute neutrophil count (10^9/L), mean ±SD	4.5 ±1.2	3.9 ±1.1	0.12
Absolute eosinophil count (10^9/L), median (IQR)	0.2 (0.18)	0.1 (0.1)	**0.01**
Historic eosinophils (10^9/L), median (IQR)	0.35 (0.2)	0.1 (0.1)	**<0.0001**
HRCT measurements			
E/I MLD, mean ±SD	0.85 ±0.05	0.8 ±0.05	**0.003**
$\%LAA_{<-950,}$ mean ±SD	10.9 ±5.1	6.6 ±4.5	**0.005**
BALF differential cell count			
Neutrophil %, median (IQR)	3.6 (8.4)	0.8 (1.2)	**0.02**
Macrophage %, median (IQR)	63.7 (53.2)	70 (28.4)	0.4
Eosinophil %, median (IQR)	1 (2.95)	0.4 (0.6)	**0.04**
Lymphocyte %, median (IQR)	0 (0.55)	0 (1.85)	0.08

Fisher's exact test was performed for gender given small sample size. Chi-squared test used for COPD status. Shapiro-Wilk test for normality was performed for all continuous variables. Welch two sample t test was performed for normally distributed data; Age, BMI, FEV1, FVC, FEV1/FVC and FEF 25-75, TLCO, RV/TLC SR, total blood leucocytes, absolute neutrophil count, E/I MLD and %LAA<-950. Mann-Whitney U test was performed for skewed data; smoking pack years, eosinophil blood counts and BALF differential cell count. BMI, body mass index; FEV1, forced expiratory volume in 1 sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; DLCO, diffusion capacity of the lung for carbon monoxide; E/I MLD, ratio of mean lung attenuation on expiratory and inspiratory scans; HRCT, high resolution computer tomography; %LAA<-950, percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units; IQR, interquartile range; NA, non-applicable; SD, standard deviation.

The subjects were matched for age, smoking pack years and body mass index (BMI). There were more males in both groups, and although it appeared as though the COPD subjects had a greater proportion of males, this wasn't statistically significant ($p = 0.06$). As expected post-bronchodilator forced expiratory volume in one second (FEV1) % predicted, FEV1/forced vital capacity (FVC) ratio, and forced expiratory flow rate (FEF) 25-75% predicted were all significantly reduced in the COPD group. The COPD subjects varied from mild to moderate disease, with a mean FEV1% predicted of 77.5% (SD±14.8).

Similar to the smaller cohort (n=35 – see section 3.2), COPD patients had higher levels of historic eosinophil counts ($p <0.0001$). However, in addition, in this larger cohort they also demonstrated higher levels of blood eosinophils at their baseline test ($p = 0.01$).

As expected, COPD subjects demonstrated more evidence of small airways and emphysema, with a higher ratio mean lung attenuation on expiratory and inspiratory scans (E/I MLD) ($p = 0.02$) and lower diffusion capacity of the lung for carbon monoxide (DLCO) % predicted ($p = 0.004$) respectively compared with healthy ex-smokers. Furthermore, in contrast to the smaller cohort (n=35 – see section 3.2), COPD patients also had higher percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units (%LAA<-950), suggestive of underlying emphysema ($p = 0.005$).

In addition, COPD subjects also demonstrated higher proportions of neutrophils and eosinophils in the BALF, which is in keeping with the findings from the smaller cohort (n=35, see section 3.2). Again, similar to the smaller cohort macrophages were the predominant cell type in the BALF (median proportion 68% across the whole cohort), and there was no difference in macrophage proportions between COPD and healthy ex-smokers ($p = 0.4$). Relative numbers of lymphocytes in BALF were low and there were no significant differences between groups ($p = 0.08$). The BALF differential cell counts in relation to COPD inflammatory endotypes will be explored further in section 6.5.

6.3 Lung-derived EV miRNA in relation to COPD clinical characteristics

Spearman correlation coefficients were generated for the lung-derived EV miRNA normalised expression data and the clinical phenotypic characteristics of COPD (Table 6.2). Although significant correlations between clinical variables and EV miRNA expression data were shown when analysing the cohort as a whole (n=44), most of these became non-significant when analysing just the COPD subjects alone (n=24) (Table 6.2). These results suggest significant

correlations are due to the presence of disease, rather than disease specific characteristics. Of note, significant correlations were identified for miR-2110 and miR-200b-5p expression with DLCO % predicted (r = -0.43, p = 0.04 and r = -0.6, p = 0.003 respectively) in COPD patients alone, suggesting these associations may be due to a specific disease pathology (e.g. emphysema in the context of DLCO %).

Table 6.2 Correlations of up-regulated lung-derived EV miRNA expression with COPD phenotypic disease characteristics

	FEV1	FVC	FEV1/ FVC	FEF 25-75	DLCO	E/L MLD	%LAA $_{<-950}$	Historic Eosinophils (10^9/L)
Whole cohort, N=44								
miR-2110	**-0.4****	-0.07	**-0.46****	**-0.47****	**-0.37***	0.19	0.26	**0.43****
miR-223-3p	-0.26	0.05	**-0.42****	**-0.44****	**-0.38***	**0.34***	0.14	**0.43****
miR-182-5p	**-0.3***	0.05	**-0.43****	-0.38*	**-0.4***	0.18	0.24	**0.32***
miR-625-3p[r]	0.005	-0.06	0.007	-0.0005	0.11	-0.05	0.03	0.08
miR-200b-5p	-0.16	0.07	-0.24	-0.23	**-0.35***	0.08	0.14	0.28
COPD subjects alone, N=24								
miR-2110	-0.19	-0.04	-0.12	-0.18	**-0.43***	-0.04	0.04	0.01
miR-223-3p	0.02	0.17	-0.07	-0.03	-0.22	-0.03	-0.37	0.02
miR-182-5p	0.04	0.09	-0.07	-0.06	-0.3	-0.25	-0.04	0.03
miR-625-3p[r]	0.18	0.01	0.19	0.17	-0.21	-0.07	-0.12	0.1
miR-200b-5p	0.22	0.25	0.07	0.05	**-0.6****	-0.05	0.01	-0.04

[r]missing data for 13 COPD subjects, N=11. Spearman's correlation coefficient.*p<0.05, **p<0.005. FEV1, FVC, FEF 25-75 and DLCO are all measured as percent predicted. Historic eosinophil refers to highest ever recorded eosinophil count. FEV1, forced expiratory volume in 1 sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; DLCO, diffusion capacity of the lung for carbon monoxide; E/I MLD, ratio of mean lung attenuation on expiratory and inspiratory scans; %LAA<-950, percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units

Likewise, in the down-regulated lung-derived EV miRNA correlations with clinical characteristics of COPD (Table 6.3), all of the miRNA significantly correlated with FEV1% predicted, (a marker of disease severity), FEF 25-75% predicted, (a marker of small airways disease), and historic eosinophils when analysing the cohort as a whole (n=44). However, when examining the associations in COPD subjects alone (n=24), only miR-338-3p expression correlated significantly with FEF 25-75% suggesting possible association with small airways disease (r = 0.44, p = 0.03) and DLCO % suggesting a possible link with underlying emphysema (r = 0.48, p = 0.03).

Table 6.3 Correlations of down-regulated BAL EV miRNA expression with COPD phenotypic disease characteristics

	FEV1	FVC	FEV1/ FVC	FEF 25-75	DLCO	E/L MLD	%LAA$_{<-950}$	Historic Eosinophils (10^9/L)
Whole cohort, N=44								
miR-204-5p	**0.35***	-0.13	**0.52****	**0.47****	0.2	**-0.31***	-0.2	**-0.33***
miR-138-5p	**0.32***	-0.07	**0.43****	**0.42****	0.28	-0.3	-0.22	**-0.35***
miR-338-3p	**0.34***	-0.06	**0.41****	**0.4****	**0.4***	-0.26	-0.24	**-0.35***
COPD subjects alone, N=24								
miR-204-5p	0.28	-0.08	0.36	0.33	0.13	-0.2	-0.05	-0.03
miR-138-5p	0.2	-0.05	0.2	0.26	0.3	-0.2	-0.07	0.008
miR-338-3p	0.29	-0.21	0.43*	**0.44***	**0.48***	-0.25	-0.28	0.004

Spearman's correlation coefficient.*p<0.05, **p<0.005. FEV1, FVC, FEF 25-75 and DLCO are all measured as percent predicted. Historic eosinophil refers to highest ever recorded eosinophil count. FEV1, forced expiratory volume in 1 sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; DLCO, diffusion capacity of the lung for carbon monoxide; E/l MLD, ratio of mean lung attenuation on expiratory and inspiratory scans; %LAA<-950, percent of lung voxels on the inspiratory scan with attenuation values below -950 Hounsfield Units

6.4 Predictive ability of miRNA to differentiate between COPD and health

Given the lung-derived EV miRNA were associated with many of the clinical phenotypic characteristics across the cohort as a whole, the predictive ability of the up-regulated miRNA to differentiate between health and COPD was assessed. Importantly, these up-regulated miRNA were identified from comparison of a relatively mild COPD cohort, with a mean post-bronchodilator FEV1 % predicted of 77.5% (SD±14.8) with a healthy ex-smoker age-matched population.

Receiver operative characteristic (ROC) curves were generated using the miRNA normalised expression data in SPSS® and results are summarised in Table 6.4.

Table 6.4 ROC analysis for predictive ability of up-regulated miRNA to differentiate between COPD and healthy ex-smokers

miRNA	Log2FC	AUC (95% CI)	Standard Error[a]	P value
miR-2110	2.12	0.81 (0.68 -0.93)	0.06	<0.0001
miR-223-3p	2.97	0.79 (0.65 – 0.93)	0.07	0.001
miR-182-5p	1.52	0.78 (0.64 – 0.92)	0.07	0.001
miR-625-3p[r]	1.85	0.76 (0.59 – 0.93)	0.09	0.02
miR-200b-5p	1.52	0.71 (0.56 – 0.87)	0.08	0.02

[r] data missing for 13 COPD subjects, N=11. [a.] under the nonparametric assumption. AUC, area under receiver operator characteristic curve; CI, confidence interval; miRNA, microRNA.

The ROC curve analysis showed that in isolation miR-2110, miR-223-3p and miR-182-5p have moderate predictive ability to differentiate between COPD and healthy ex-smokers, with an area under the curve (AUC) either >0.8 for miR-2110 or approaching 0.8 for the latter two. Although miR-625-3p performed nearly as well, this was excluded from further analysis based on the number of missing data points (n=11 for this analysis only). MiR-200b-5p performed least well with an AUC 0.71.

Different combinations of the miRNA were tested using the ROC curve analysis to determine the best combination of miRNA for predicting disease presence (Table 6.5).

Table 6.5 ROC analysis to determine optimal combination of EV miRNA in differentiating between COPD and healthy ex-smokers

miRNA	AUC (95% CI)	Standard Error[a]	P value
miR-2110, miR-223-3p, miR-182-5p	**0.91 (0.8 -0.98)**	**0.05**	**<0.0001**
miR-2110, miR-223-3p, miR-182-5p, miR-200b-5p	0.85 (0.73 – 0.97)	0.06	<0.0001
miR-2110, miR-223-3p	0.84 (0.72 – 0.96)	0.06	<0.0001
miR-2110, miR-182-5p	0.84 (0.73 – 0.96)	0.06	<0.0001
miR-223-3p, miR-182-5p	0.83 (0.72 – 0.96)	0.06	<0.0001

[a.] under the nonparametric assumption. AUC, area under receiver operator characteristic curve; CI, confidence interval; miRNA, microRNA.

These analyses showed that using the combination of miR-2110, miR-223-3p and miR-182-5p improved the predictive ability to discriminate between COPD and healthy ex-smokers, with an AUC 0.91 (Figure 6.1).

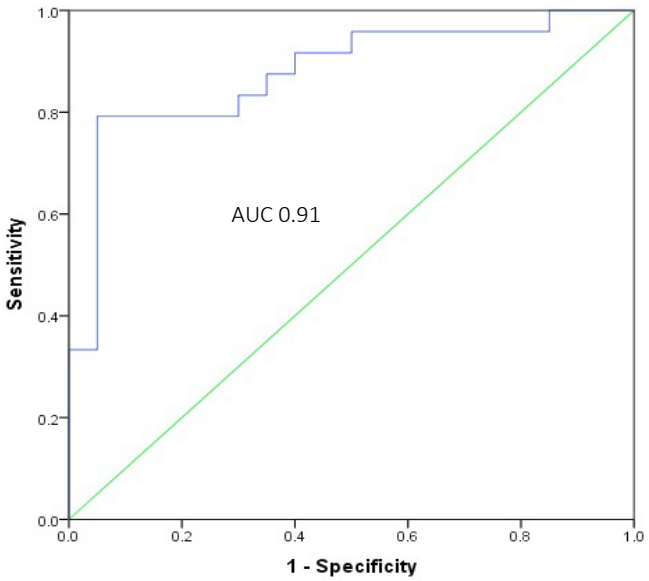


Figure 6.1 ROC curve for miR-2110, miR-223-3p and miR-182-5p for the predicting the presence of COPD in the cohort

AUC = area under the curve.

6.5 EV miRNA in relation to COPD inflammatory endotypes

6.5.1 Describing inflammatory endotypes in COPD subjects

BALF differential cell counts were analysed in all 24 COPD subjects and 20 healthy ex-smokers. As previously described each subject had two lobes sampled and for this analysis the cell counts which corresponded to the lobe from which the EVs were recovered were used. There were significantly increased neutrophils and eosinophils in the BALF of COPD subjects compared with the healthy ex-smokers. However, there was no difference in macrophage proportions between COPD subjects and healthy ex-smokers (Figure 6.2). Macrophages were the predominant cell type in the airways (median proportion 68%). However the 95% confidence intervals for the proportion of macrophages found in COPD BALF samples were wide (Figure 6.2 C), reflecting heterogeneity across disease and the limits of a proportional analysis.

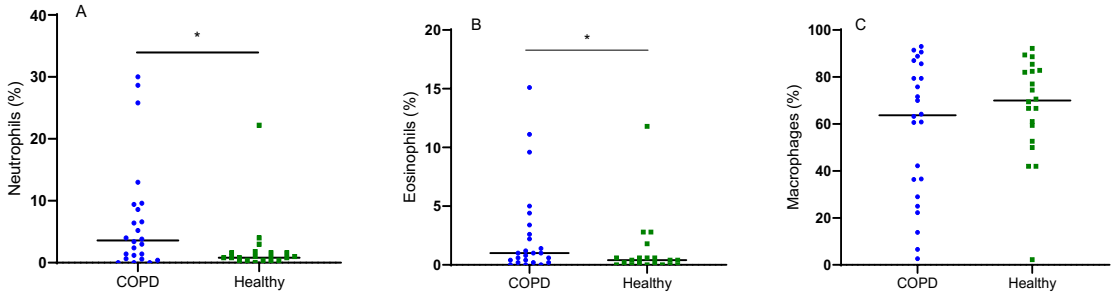


Figure 6.2 BALF expression of immune cells in COPD subjects and healthy ex-smokers.
(A) Neutrophils, (B) Eosinophils, (C) Macrophages. Data represents median with 95% confidence interval. Each dot represents BALF concentration of individual value in a specific patient. N=44; COPD, n=24. **p<0.01, *p<0.05 using Mann-Whitney U test.

6.5.2 Relationship between EV miRNA expression and levels of inflammatory cells in BAL

The relationship between BAL EV miRNA expression and levels of inflammatory cells was assessed using Spearman's correlations in the COPD subjects alone (N=24), (Table 6.6).

Table 6.6 Correlations between EV miRNA expression and immune cells proportions within BALF

BALF EV miRNA	Neutrophils %	Eosinophils %
Up-regulated in COPD		
miR-2110	**0.47***	0.2
miR-223-3p	0.35	**0.47***
miR-182-5p	**0.46***	0.1
miR-625-3p[r]	0.12	0.23
miR-200b-5p	0.33	0.02
Down-regulated in COPD		
miR-204-5p	**-0.49***	**-0.57****
miR-138-5p	-0.11	-0.22
miR-338-3p	-0.22	**-0.42***

[r]missing data for 13 COPD subjects, N=11. Spearman's correlation coefficient. N = 24 *p<0.05, **p<0.005. BAL; bronchoalveolar lavage fluid; EVs, extracellular vesicles; miR, microRNA.

There were significant positive correlations between levels of neutrophils and two of the up-regulated miRNA in COPD (miR-2110 and miR-182-5p). Whereas, miR-223-3p significantly correlated with eosinophil expression. Conversely, in the down-regulated miRNA, miR-204-5p showed significant negative correlations with both neutrophils and eosinophil expression, whereas miR-338-3p only significantly correlated with eosinophils.

The significant correlation with miR-223-3p, miR-204-5p and miR-338-3p with BALF eosinophil levels, prompted further analysis with blood eosinophil levels given the clinical utility of highest-ever historic blood eosinophil count in defining eosinophilic disease in COPD. However, there was no correlation between BALF eosinophil levels and highest-ever historic blood eosinophil count ($r = 0.1$, $p = 0.65$). Furthermore, there was no association between highest-ever historic eosinophil count and miR-223-3p ($r = 0.02$, $p = 0.9$), miR-204-5p ($r = -0.03$, $p = 0.9$) and miR-338-3p ($r = 0.004$, $p = 0.9$) expression levels in the lung-derived EVs.

6.5.3 Using EV miRNA to predict inflammatory endotypes in COPD

Given the significant correlations between levels of inflammatory cells and specific EV miRNA in COPD subjects, this raises the possibility of EV miRNA ability to predict specific inflammatory endotypes in COPD.

The above analyses were used to categorise the COPD subjects into inflammatory endotypes based on the cut-offs defined in the American Thoracic Society Clinical Practice Guidelines for BAL analysis (551), which determined neutrophilia as >3% and eosinophilia as >1%.

Using these criteria, the 24 COPD subjects were classified into four inflammatory endotypes based on the inflammatory cut-offs defined below (summarised in Table 6.7).

Table 6.7 Definitions of inflammatory endotypes in COPD using %neutrophil and eosinophil pre-defined cut-offs

Inflammatory endotype	% Neutrophils in BALF	% Eosinophils in BALF
Eosinophilic	NA	>1%
Neutrophilic	>3%	NA
Mixed granulocytic	>3%	>1%
Paucigranulocytic	≤3%	≤1%

BALF, bronchoalveolar lavage fluid; NA, non-applicable

The Venn diagram below describes the distribution of inflammatory endotypes within this cohort, with 10 COPD subjects defined as eosinophilic (41.7%), 13 defined as neutrophilic (54.2%), 6 defined as mixed granulocytic (25%) and 7 defined as paucigranulocytic (29.2%).

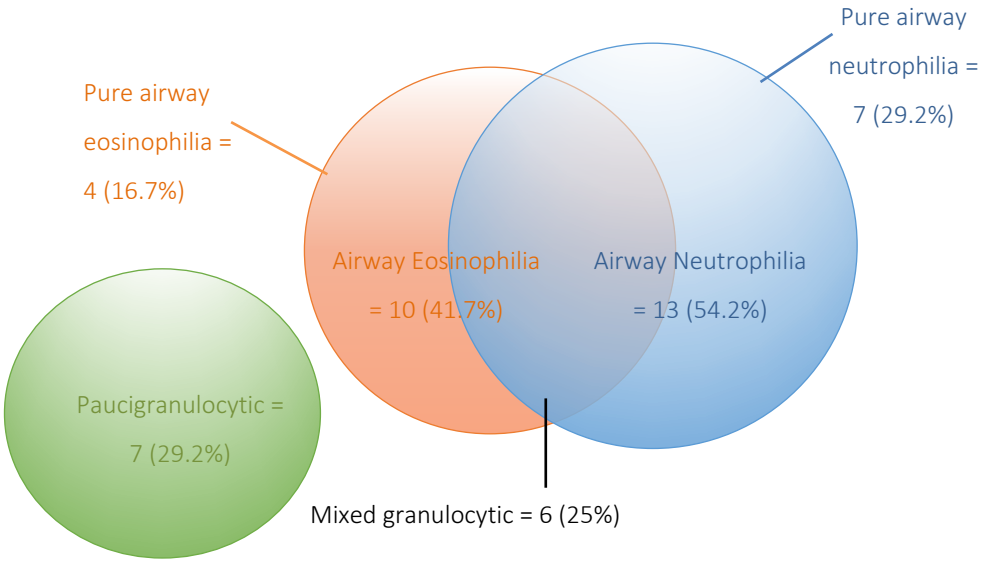


Figure 6.2 Venn diagram to describe the inflammatory endotypes in the COPD subjects based on pre-defined cut-offs

A series of ROC analyses were performed to determine the predictive ability of miRNA to determine inflammatory endotypes. Firstly, the eosinophilic subjects with COPD (n=10) were compared against the non-eosinophilic COPD subjects (n=14; pure airway neutrophilia, n = 7 and paucigranulocytic, n = 7), to determine the predictive ability of the dysregulated miRNA in distinguishing between eosinophilic and non-eosinophilic disease (Table 6.8).

Table 6.8 ROC analyses to determine predictive ability of miRNA to differentiate between eosinophilic and non-eosinophilic subtypes in COPD

miRNA	AUC (95% CI)	Standard Error[a]	P value
miR-2110	0.51 (0.2 – 0.8)	0.14	0.9
miR-223-3p	**0.78** (0.6-1)	0.14	**0.04**
miR-182-5p	0.57 (0.3 – 0.8)	0.14	0.6
miR-625-3p[r]	0.30 (0.04 – 0.6)	0.14	0.14
miR-200b-5p	0.51 (0.2 – 0.8)	0.14	0.9
miR-204-5p	0.74 (0.5 – 0.9)	0.05	0.05
miR-138-3p	0.60 (0.4 – 0.8)	0.41	0.4
miR-338-3p	**0.74** (0.5 – 0.9)	0.05	**0.046**

[a] Standard error under the nonparametric assumption
[r] data missing for 13 subjects

MiR-223-3p and miR-338-3p showed fair predictive ability to distinguish between eosinophilic and non-eosinophilic disease with AUC >0.7 ($p < 0.05$). When combining these measures the AUC improved to 0.83 (95% CI 0.7 – 0.9, $p = 0.007$) and therefore using a combination of miRNA markers may be a more accurate approach.

Given the combination of miR-223-3p and miR-338-3p showed good predictive ability in distinguishing eosinophilia when also including the subjects with a mixed granulocytic picture; further analysis was performed to see whether these miRNA were even more specific at distinguishing eosinophilic disease when considering just pure eosinophilic disease (n=4).

Table 6.9 ROC analysis to differentiate between pure eosinophilic COPD and paucigranulocytic or neutrophilic COPD

miRNA	AUC (95% CI)	Standard Error[a]	P value
miR-2110	0.75 (0.3 – 1)	0.21	0.25
miR-223-3p	**0.94 (0.8 – 1)**	**0.09**	**0.04**
miR-182-5p	0.81 (0.5 – 1)	0.18	0.15
miR-625-3p[Γ]	0.75 (0.4 – 1)	0.19	0.25
miR-200b-5p	0.88 (0.6 – 1)	0.14	0.08
miR-204-5p	0.86 (0.6 -1)	0.13	0.06
miR-138-3p	0.68 (0.3 – 1)	0.18	0.35
miR-338-3p	**0.85 (0.6 – 1)**	**0.08**	**0.03**
miR-223-3p, miR-338-3p	**0.81 (0.6-1)**	**0.09**	**0.04**

[a] Standard error under the nonparametric assumption
[Γ] data missing for 13 subjects

MiR-223-3p showed excellent predictive ability of differentiating pure airway eosinophilia from paucigranulocytic and pure airway neutrophilic disease with an AUC 0.94 ($p = 0.04$). MiR-338 did not perform as well, but was significant in this analysis at distinguishing pure airway eosinophilia with an AUC 0.85 ($p = 0.03$). The combination of the two miRNA did not improve the specificity with an AUC 0.81 (see Table 6.9).

Next, the neutrophilic subjects with COPD (n = 7) were compared against the non-neutrophilic COPD subjects (n=11; pure airway eosinophilia, n= 4 and paucigranulocytic, n = 7)) to determine the predictive ability of the miRNA in distinguishing between neutrophilic and non-neutrophilic disease (Table 6.10).

Table 6.10 ROC analyses to determine predictive ability of miRNA to differentiate between neutrophilic and non-neutrophilic subtypes in COPD

miRNA	AUC (95% CI)	Standard Error[a]	P value
miR-2110	0.69 (0.5 – 0.9)	0.12	0.2
miR-223-3p	0.67 (0.4 – 0.9)	0.12	0.2
miR-182-5p	0.58 (0.3 – 0.9)	0.14	0.6
miR-625-3p[r]	0.34 (0.09 – 0.6)	0.13	0.2
miR-200b-5p	0.52 (0.2 – 0.8)	0.14	0.9
miR-204-5p	0.60 (0.4 – 0.8)	0.12	0.4
miR-138-3p	0.4 (0.2 – 0.6)	0.12	0.4
miR-338-3p	0.44 (0.2 – 0.7)	0.12	0.6

[a.] Standard error under the nonparametric assumption
[r] data missing for 13 subjects

However, none of the miRNA showed significant predictive ability for distinguishing between neutrophilic and non-neutrophilic disease.

6.6 Differences in serum extracellular vesicle miRNA expression

Given the promise of the ROC analyses showing specific BAL EV miRNA (miR-2110, miR-223-3p and miR-182-5p) had good predictive ability in differentiating between COPD and healthy ex-smokers, serum EV miRNA were analysed to assess whether the changes were also present in blood. Peripheral blood was chosen as if EV miRNA were to be used as an additional diagnostic marker for COPD, it would need to be identified via a non-invasive method (i.e. blood test rather than bronchoscopy) to have clinical utility.

6.6.1 Subject characteristics

Fourteen COPD subjects and ten healthy ex-smokers were included in the serum EV miRNA quantification by RT-qPCR based on sample availability. Not all of the subjects included in this analysis were part of the larger cohort (N=44). To increase sample numbers, an additional seven serum samples were used from an additional seven COPD subjects from the overall MICA II cohort. The characteristics of this smaller cohort (N=24) is summarised in Table 6.11.

Table 6.11 Characteristics of subjects included in serum EV miRNA target validation by RT-qPCR, N=24

Subject/sample characteristics	COPD (n=14)	Healthy ex-smoker (n=10)	P value
Age, mean ±SD	69.1 ±6.6	67.7 ±9.3	0.68
Male, n (%)	5 (36%)	4 (40%)	1
Smoking pack years, mean ±SD	63.8 ±50.3	25.5 ±15.6	0.003
BMI, mean ±SD	28.2 ±3.6	27.5 ±4.5	0.66
FEV1 (% predicted), mean ±SD	62.9 ±13.2	101.8 ±11.9	<0.00001
FVC (% predicted), mean ±SD	93.3 ±18.2	101.6 ±12.5	0.17
FEV1/FVC%, mean ±SD	52.9 ±9.6	77.5 ±3.7	<0.0001
FEF 25-75 (% predicted), mean ±SD	28.1 ±12.9	103.6 ±21.8	<0.000001
COPD status, GOLD stage, n (%)			0.003
Mild	1 (7)	NA	
Moderate	11 (79)	NA	
Severe	2 (14)	NA	

Fisher's exact test for Gender given small sample size. Chi-squared test used for COPD status. Shapiro-Wilk test for normality was performed for all continuous variables. Welch two sample t test was performed for normally distributed data; Age, BMI, FEV1, FVC and FEF 25-75. Mann-Whitney U test was performed for skewed data; smoking pack years and FEV1/FVC. BMI, body mass index; FEV1, forced expiratory volume in one sec, FVC, forced vital capacity; FEF, Forced expiratory flow rate; NA, non-applicable; SD, standard deviation.

Despite the smaller cohort, the subjects were still matched for age, gender, and BMI. However there were a higher number of smoking pack years in the COPD subjects (mean 63.8 years, ±SD50.3) than the healthy ex-smokers (mean 25.5 years, ±SD15.5). This is likely to be due to including COPD subjects that had more severe disease (mean FEV1 % predicted 62.9, ±SD13.2), and therefore were not eligible for a research bronchoscopy. As expected FEV1% predicted, FEV1/FVC and FEF 25-75% predicted was significantly reduced in the COPD group. Due to the additional seven patients to this cohort, the COPD subjects now included a full spectrum of disease, with mild (7%), moderate (79%) and severe (14%) GOLD stage represented. Not all of these patients underwent gas transfer measurement (DLCO) and HRCT assessment and therefore these parameters were not included in any analysis.

It is important to note that the serum EVs were isolated using the ExoMiR™ filtration kit, a different method to the BALF EV isolation (see section 2.5.5). This method was chosen due to the previous experience of the laboratory group and the abundance of EV RNA recovered from serum using this approach. This may have a significant bearing on the results obtained from this analysis, especially when comparing them to the BALF EV miRNA results.

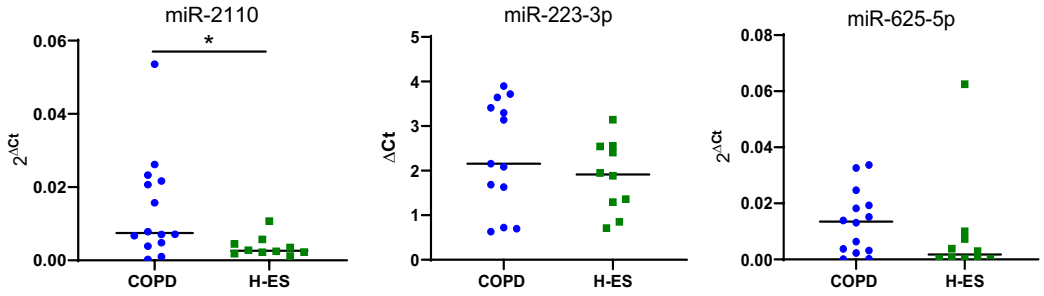
6.6.2 Differential miRNA expression analysis from serum EV miRNA

The five up-regulated and three down-regulated miRNA identified from the differential expression analysis of the BAL EV miRNA were chosen for analysis in serum EVs by RT-qPCR. In addition, two miRNA (miR-16-5p and miR-24-3p) known to be most stably expressed in serum EVs (552) were chosen as normalisers.

The normaliser miRNA showed stable expression across the data set and between groups, with a mean Ct 24.6 (±SD 1.6) in healthy ex-smokers and a mean Ct 24.5 (±SD 1.2) in COPD for miR-16-5p and a mean Ct 27.1 (±SD 1.0) in healthy ex-smokers and a mean Ct 26.9 (±SD 1.2) in COPD for miR-24-3p. These data were combined as a geomean to normalise the results of the other miRNA.

Only four of the eight miRNA were detectable in serum EVs (miR-2110, miR-223-3p, miR-625-3p and miR-338-3p), with miR-223-3p showing the highest level of expression in serum EV with a mean Ct 24.6 (±SD 2.2) in all samples (compared with mean Ct >30 for the other three miRNA). However, only miR-2110 was found to be differentially expressed in COPD compared with healthy ex-smokers in the serum EV samples (Figure 6.4). This significance may in part be driven by the sample from MICA_II_052 which had the highest normalised expression level of 0.054. This subject had the most severe disease with a post-bronchodilator FEV1 % predicted of 46%.

Upregulated miRNA in COPD BAL EVs detected in serum EVs



Downregulated miRNA in COPD BAL EVs detected in serum EVs

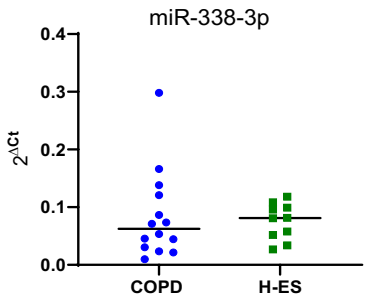


Figure 6.3 Normalised expression levels for miRNA in serum EVs

$2^{\Delta Ct}$, normalised expression levels. Median and 95% confidence intervals shown. Unpaired, Welch's t test, *p <-0.05. N=24; COPD, n=14

This prompted a further analysis to examine whether miR-2110 expression in serum EVs correlated with disease severity (post-bronchodilator FEV1 % predicted). There was a significant association between miR-2110 expression and post-bronchodilator FEV1 % predicted (Figure 6.4), with a higher miR-2110 expression in the more severe COPD subjects (r = -0.51, p = 0.03). There was no association with other markers of disease severity such as FEF 25-75% predicted (r = -0.33, p = 0.25).

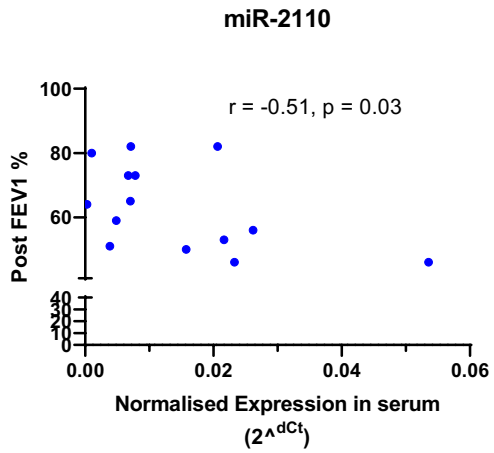


Figure 6.4 Relationship between Post FEV1% predicted and miR-2110 expression in serum.
Pearson correlation performed given FEV1% predicted normally distributed as determined by Shapiro-Wilk test with a one-tail hypothesis. N=14.

6.7 Discussion

Chapter 6 presents the clinical characteristics of the cohort used for BALF EV miRNA validation study and their relationship with the differentially expressed miRNA. It presents the findings of the ROC curve analysis examining the predictive ability of the EV miRNA to discriminate between COPD and healthy ex-smokers and explores the potential of these miRNA to differentiate between different inflammatory endotypes of COPD, which may have important implications for treatment strategies. Finally, this chapter explores the potential of these dysregulated EV miRNA to be used as biomarkers for early disease in the serum of patients with COPD.

6.7.1 Predictive ability of EV miRNA to differentiate between health and disease

Results from the ROC curve analysis demonstrated the combination of miR-2110, miR-223-3p and miR-182-5p had excellent predictive ability (AUC 0.91) in discriminating between COPD and healthy ex-smokers. Importantly this was shown in a relatively mild COPD cohort (mean FEV1 % predicted 77.5%; SD±14.8). Currently the diagnosis of COPD depends on the use of spirometry to define lung function impairment, however lung function decline can occur before the disease is

clinically apparent and therefore decline in FEV1 may not detect early stage disease. Therefore having a more sophisticated biomarker of disease that is able to detect pre-clinical disease could have significant implications for treatment initiation and long term prognosis in COPD.

MicroRNA are posed as ideal biomarker candidates as they are easily measurable in liquid biopsies (e.g. blood, urine, sputum and BALF) and have demonstrated high sensitivity for differentiating stages of disease and even treatment responsiveness (553). Specifically urinary exosomal miRNA have been shown to detect early renal fibrosis in lupus nephritis (554) and a nine-miRNA multimarker panel for breast carcinoma has been shown to significantly improve reliability of breast cancer diagnosis (555). Furthermore, the technologies for detection of these small non-coding RNAs are advancing at speed with the development of newer assays requiring less time and lower costs in comparison to producing new antibodies for protein biomarkers.

However in this study, although, a differential EV miRNA signature was found in the lungs of patients with COPD, this did not translate into serum. Only four of the differentially expressed miRNA were found in the serum EVs and of these, only one (miR-2110) showed differential expression in COPD compared with healthy ex-smokers. Interestingly, miR-2110 was also shown to correlate with disease severity and therefore highlights its utility as a possible marker of early disease. EV miRNA content is known to vary depending on sample type and even plasma and serum EVs have shown differences in miRNA expression (556). Therefore, it is perhaps unsurprising that the signal detected in the lungs was not translated into peripheral blood. Furthermore a different method of EV isolation was used in the serum (ExoMiR™ kit, section 2.5.5) compared with BALF (ultrafiltration and SEC, sections 2.5.1 and 2.5.2). Thus direct comparison of the two EV populations may not be possible.

One of the challenges of working in the EV field is the lack of cell/tissue of origin specificity exosomes display. As discussed in section 1.7, exosomal surface markers correspond to proteins from the endosomal compartment (CD9, CD63 and CD81), endocytic pathway (Tsg101 and alix), and cytosol (actin and tubulin) (207, 227), which are generic cellular markers. However, unlike microvesicles, which express the same surface antigens found on the parent cells, there is no consensus on markers that distinguish the origin of exosomes once they have left the cell (557). Ideally when evaluating the lung EV miRNA signature in serum, one would be able to stratify the EV serum sample for the lung “specific” EVs and hope to find a more specific disease signature. More work needs to be done to characterise markers of exosomal origin to allow more specific characterisation, particularly in circulatory EV populations which may originate from multiple organs and tissue types.

Although this thesis highlights novel differences in BALF-derived EV miRNA between health and COPD, for this work to be translated into a useful clinical biomarker, the EV miRNA signature needs to be measureable in easily accessible biofluids. Given the above mentioned challenges of measuring a lung EV specific signature in the peripheral blood, next steps should perhaps focus on lung-derived biofluids such as sputum and exhaled breath. Sputum microparticles have already been identified in patients with COPD (247) and differential miRNA expression has been identified in both induced sputum (309, 327) and exhaled breath (337). Therefore, methodology focused on rapid EV isolation and miRNA quantification in these samples may provide a useful platform to translate this novel, COPD specific EV miRNA signature into clinical practice. Furtherwork also needs to be done to establish the stability of the EV miRNA signature over time and whether miRNA expression correlates with disease severity, with studies so far showing high stability in blood in cancer (558). In addition, the study of a lung EV miRNA signature in pre-clinical disease (i.e in smokers with normal spirometry) may have huge potential in broadening understanding of why only some smokers go on to develop COPD.

6.7.2 Relationships between lung EVs and inflammatory endotypes in COPD

The heterogeneous clinical manifestations of COPD and differences in response to therapy suggest there may be different endotypes of disease that in future may be treated with more precision than the current broad-spectrum therapies. Endotypes imply that the underlying molecular mechanisms that drive the clinical manifestations of the disease are known but this is rarely the case in COPD. Therefore exploring the relationship with lung EV miRNA expression and inflammatory cells within the lung may shed new light on these poorly defined pathways.

Correlative analysis showed there were significant positive correlations between neutrophil expression and miR-2110 and miR-182-5p, and eosinophil expression and miR-223-3p. In the down-regulated miRNA, miR-204-5p showed significant negative correlations with both neutrophil and eosinophil expression, whereas miR-338-3p only significantly correlated with eosinophils. These associations raise questions about the origin of these lung-derived EVs and their possible target cells. For example, a positive correlation may suggest that a specific cell type (e.g. neutrophil) may be the dominant source of a particular EV miRNA (e.g. miR-182-5p) or the primary recipient.

MicroRNA-182-5p is already known to regulate neutrophils, with Li et al. showing miR-182-5p enhances neutrophil migration into the vascular endothelium as a mechanism for coronary artery scar formation in Kawasaki patients (559). In addition, miR-182 has been shown to regulate

granulopoiesis via inhibition of C/EBPα (a master regulator of granulopoiesis) suggesting a role in neutrophil generation (560). To date this is the first study linking miR-2110 to neutrophil accumulation in the airways, where previous work has focused solely on its role in tumorigenesis (450). Interestingly, previous work demonstrated gene ontology (GO) terms enriched among genes down-regulated by miR-204 (e.g. transforming growth factor β2 (*TGFB2*) and complement C3a receptor 1 (*C3AR1*)), were related to neutrophil chemotaxis (561), which is in keeping with the negative correlation with neutrophil expression seen in this study. Furthermore, studies show that miR-204-5p inhibits inflammation and chemokine generation in renal tubular epithelial cells by modulating IL-6 expression (562). This may suggest an anti-inflammatory role within airway epithelium, where IL-6 is a potent recruiter of neutrophils during acute inflammation (563). Thus, down-regulation of lung-derived EV miR-204-5p may lead to airway neutrophilia via an IL-6 dependant pathway in patients with COPD. This may suggest targeted treatment promoting miR-204-5p expression may prevent excessive airway neutrophilia and lead to a reduction in airway inflammation and tissue destruction. Thus, exploring the mechanisms of miRNA regulation of neutrophil function in COPD may provide key insights into neutrophil dysfunction in this disease and identify alternative targets for treatment.

In this study, eosinophil expression was shown to significantly correlate with miR-204-5p, miR-223-3p and miR-338-3p expression (Table 6.5, r = -0.57, p = 0.003; r = -0.47, p = 0.03; r = -0.42, p = 0.03 respectively). MiR-204-5p has been shown to regulate toll-like receptor (TLR) 4 protein expression via targeting the transcription factor (SRY-related HMG-box) SOX-11 (564). TLR4 is a known activator of eosinophils (565) and therefore miR-204-5p modulation of TLR4 may play a role in driving eosinophilia in COPD. MiR-223-3p expression in bronchial biopsies was previously shown to correlate with eosinophils in asthmatics (566). Overexpression of miR-223-3p significantly reduced granulocyte-macrophage colony stimulating factor (GM-CSF) at baseline and in response to house dust mite, poly-(I:C) and cigarette smoke extract stimulation (567). GM-CSF is known to promote eosinophil recruitment and cell survival (568) and therefore miR-223-3p may be involved in a negative feedback loop, whereby miR-223-3p reduces GM-CSF expression, when eosinophil levels are high. Furthermore, miR-223-3p levels are significantly increased in the serum and nasal epithelium of patients with allergic rhinitis (569), where it is shown to enhance eosinophilic infiltration by targeting the inositol phosphatase *INPP4A* (a molecular checkpoint in control of the phosphoionositide 3-kinase (PI3K) - mammalian target of rapamycin (mTOR) pathway (570)). In another study of allergic rhinitis and asthma, miR-338-3p was found down-regulated in plasma of allergic rhinitis patients and was shown to target selected inflammatory genes mitogen activated protein kinase (*MAPK)-8* and inhibitor of nuclear factor kappa-B kinase

subunit beta (*IKKβ)*, but had no specific relationship with eosinophil expression (463). A more recent study exploring the role of miR-338 specifically in obstructive airways disease demonstrated that the expression level of miRNA-338 in the sputum was higher in all patient groups compared to controls; however, asthmatics showed a significantly higher miR-338 expression compared to COPD patients (464). Although these results are in contrast to the above results, where lung EVs from COPD patients had reduced levels of miR-338-3p, Lacedonia et al. data are from a different compartment (sputum) and did not explore the relationship with eosinophilia (464).

Further analysis exploring the predictive ability of EV miRNA discriminating between different inflammatory endotypes in COPD demonstrated miR-223-3p and miR-338-5p showed good predictive ability at identifying eosinophilic disease when using a pre-defined cut-off of >1% eosinophils in BALF. Furthermore this predictive ability improved when using these miRNA in combination, or, when discriminating pure airway eosinophilia from neutrophilic or paucigranulocytic disease. This study in combination with previous work suggests that both these miRNA (miR-223-3p and miR-338-5p) may play a role in defining eosinophilic airways disease in COPD, however the underlying mechanisms are yet to be elucidated.

There has been considerable interest in the role of blood eosinophil count in predicting treatment responsiveness to corticosteroids in COPD patients, based on the premise that they reflect and correlate with tissue eosinophilic inflammation (549, 571). However in this study, further analysis of historic blood eosinophil expression in COPD patients showed no relationship with lung EV miR-223-3p, miR-204-p and miR-338-3p expression. This is in keeping with more recent work which suggests blood eosinophils do not correlate with lung tissue eosinophilia (411). Therefore the mechanisms underlying the treatment responsiveness seen with corticosteroids in high blood eosinophilic patients' needs to be explored further, as it is unlikely this is mediated by lung tissue eosinophilic inflammation alone.

6.7.3 Strengths and limitations

Importantly, these results showed the excellent predictive ability of specific lung EV miRNA (miR-2110, miR-223-3p and miR-182-5p) to discriminate between COPD and healthy ex-smokers. These findings were shown in a relatively mild COPD cohort, and the importance of early diagnosis as well as the limitations of the current diagnostic method – spirometry, has been discussed in section 1.3. However, to be used as a diagnostic tool, a biomarker must be easily measurable in an easily accessible biofluid. Bronchoscopy is an invasive procedure with risk in those with

significant respiratory failure. It therefore is not an appropriate method of diagnosis and exploring the role of EVs miRNA found in blood is an important one. There is little consensus in the literature with regards to the specificity of circulating EV miRNA in health or disease and extensive literature has shown the importance of standardization in EV isolation and analysis techniques to facilitate comparison of results (291). A major limitation in this study was the use of different EV isolation techniques in serum and BALF, which limited the comparisons between the two EV populations. However, these methods were employed at the time as a result of the current expertise within the research group and have now evolved to reflect these identified limitations. As mentioned above, one of the key challenges is identifying the origin of these circulating EVs and until we have fully identified a method of doing this, it will be hard to justify their use as a biomarker of disease.

A major strength of this study is the phenotypic characterisation of the subjects included. This allows exploration into the association of the differentially expressed lung EV miRNA with different subgroups of disease. A fascinating discovery was the association between certain lung EV miRNA (specifically mir-223-3p and miR-338-5p) and eosinophilia within the lungs of patients with COPD. Although associations do not imply a causal relationship and the subgroups in this analysis are small, these findings are promising for discovery of new inflammatory endotypes in COPD and possible identification of new targets for precision based medicine.

6.7.4 Summary

In conclusion, these findings suggest specific lung-derived EV miRNA are a strong predictor of disease presence in COPD, even in mild disease, and further work should be directed into whether these findings could be translated into other bodily compartments (e.g. blood or exhaled breath) to increase their utility as a diagnostic biomarker. Furthermore, specific lung EV miRNA correlate with expression of inflammatory cells in COPD, and may have a role in defining inflammatory endotypes, which could be important in future treatment stratification.

Chapter 7 Summary discussion and future work

7.1 Overview

COPD morbidity and mortality continues to rise (3), in marked contrast to other chronic diseases, where considerable progress has been made due to targeted treatments (2). COPD is a complex, heterogeneous disease with many distinct phenotypes and this complexity has limited understanding of disease mechanisms and therefore hindered development of effective, novel therapies.

COPD is characterised by persistent airway inflammation with several known underlying mechanisms described such as oxidative stress, protease-anti-protease imbalance and immune cell dysfunction (548). However, therapy targeted towards these pathological mechanisms have had limited success, suggesting that there is more to discover and understand in this disease.

Extracellular vesicles (EVs) are a novel area of biology to explore in the context of COPD. EVs are key intercellular messengers and have been identified as playing an important role in inflammatory regulation. They encompass several components including miRNA, which are important epigenetic modulators of gene expression in recipient cells. Therefore, understanding the impact of EV miRNA in the lungs of COPD patients may uncover new insights and pathological mechanisms which could be pertinent to disease progression, but also manipulated for novel therapies.

Thus, the primary goal of my PhD was to understand the targets and functions of EV miRNA in the lungs of COPD patients with the main hypothesis being:

MicroRNA is differentially expressed in extracellular vesicles in the airways of patients with COPD, and leads to differential gene expression, which drives chronic inflammation in COPD

Many of my findings are novel and support the above hypothesis. The main findings are as follows:-

- EVs can be isolated from the bronchoalveolar lavage fluid (BALF) of patients with COPD and healthy ex-smokers using a combination of ultrafiltration and size exclusion chromatography (SEC). The EV yield and resultant RNA quantity and quality is sufficient to examine the miRNA content using next generation sequencing.

- There is an increased proportion of miRNA packaged in COPD lung-derived EVs compared with healthy ex-smokers.
- Lung-derived EV miRNA are differentially expressed between patients with COPD and healthy ex-smokers, with five upregulated miRNA and three downregulated miRNA in COPD.
- These differentially expressed lung-derived EV miRNA may be involved in epigenetic regulation of differentially expressed genes in the airway epithelium in patients with COPD.
- These targeted miRNA-mRNA interactions form a network which may have a significant impact on key metabolic and inflammatory pathways and provide new insights into the biology of COPD.
- Lung-derived EV miRNA are a strong predictor of disease presence in patients with COPD, even in a mild disease cohort. Although, these findings were not translated into the peripheral blood in a smaller, sub-cohort of patients.
- Specific lung-derived EV miRNAs correlate with expression of inflammatory cells within the airways of patients with COPD, which may provide novel insights into the distinct inflammatory endotypic disease mechanisms, and could help future treatment stratification.

Next, I will discuss my results in consideration of each of my specific aims (outlined in section 1.5 and below) in the context of the known literature:

Overall study aims:

1. To isolate EVs from the BALF of COPD subjects and healthy ex-smokers.
2. To identify differentially expressed miRNA in lung-derived EVs in COPD subjects compared with healthy ex-smokers.
3. To identify the biologically significant targets of these differentially expressed miRNA in the airway epithelium.
4. To investigate the diagnostic use of the lung-derived EV miRNA and explore their relationship with specific COPD inflammatory endotypes.

7.2 Aim 1: Isolate and characterise EVs from bronchoalveolar lavage fluid of COPD subjects and healthy ex-smokers

To my knowledge, this is the first study demonstrating EV isolation from BALF using SEC. All other published studies have used ultracentrifugation as their method of choice (232, 233, 354-360). SEC was chosen on the basis that it resulted in an adequate yield of EV from BALF and ensured separation of the EV fraction from the soluble proteins. This soluble protein separation was particularly important for my downstream RNA analysis, as extra-vesicular RNA can be bound to soluble protein complexes, such as argonaute 2 (AGO2) (362), which would have contaminated my final EV RNA sample.

During method development, I characterised my EV fractions using a CD9 enzyme-linked immunosorbent assay and transmission electron microscopy. These techniques demonstrated expression of CD9 (a known EV surface marker) and the characteristic size (30-150 nm) and cup-shaped morphology of EVs (279). These results confirmed the presence of EVs from BALF using my chosen method of isolation. However, I did not compare the EV characteristics between my healthy and disease populations. In doing so, I may have identified differences in EV surface markers in those derived from the lungs of patients with COPD compared with healthy ex-smokers. These EV surface marker differences may have led to important insights into the biological activity of these EVs or inferred a dominant cell of origin. For example, previous work has demonstrated that EVs isolated from BALF can express MHC class I and II, CD54 and CD63 and the co-stimulatory molecule CD86 (232). Based on these findings, Admyre et al. concluded that these EVs were likely to originate from antigen presenting cells (232). Whereas Kesimer et al. demonstrated that epithelial cells secrete EVs enriched for mucins (234) and thus comparing these two groups of EV surface markers across health and disease may identify differences in cellular origin of EVs. Importantly, understanding the origin of these EVs may reveal the key functions of EV miRNA.

The primary reason for not examining the EV surface markers in more detail was sample availability. In order to achieve the EV RNA quantity and quality required for next generation sequencing, the entire sample was processed for RNA isolation. Access to bronchoscopy samples is limited as this is an invasive procedure for patients to undergo and not without risk. In addition, it was not always possible to recover adequate volumes of BALF (at least 15 mL) to undergo EV isolation with adequate yield. Previous studies have already analysed different subtypes of EVs in samples such as sputum and blood. However, almost all of these studies have focused solely on

microparticles (the larger subtype of EVs) in patients with COPD (247-252), except for one study which studied circulating plasma exosomes (254). Interestingly this latter study found circulating plasma exosome levels were significantly higher in COPD patients and these exosomes were predominantly CD9 positive which is consistent with my findings in lung EVs, which showed CD9 expression rather than CD63 (see section 2.6.1).

EVs were isolated from serum by filtration using the ExoMir™ kit (methods outlined in section 2.5.5). This method was chosen due to the experience of the wider research group, and resulted in an abundant EV population for downstream RNA analysis. However, there are several limitations of using this sample type and method of isolation. Firstly, retrieved vesicles are consistently more abundant in sera than plasma (572) due to platelet-derived EV released *after* blood collection during the process of clot formation (573), which may account for over 50% of EV in serum. Thus, although serum has been shown to exhibit different EV RNA expression in health and disease (574), plasma is recommended as the sample type of choice for investigating circulating EVs (291). However, in preliminary experiments (not shown in the thesis), I compared serum versus plasma EV miRNA signatures using real time quantitative polymerase chain reaction (RT-qPCR) and found no detectable levels of miRNA in the plasma EV samples and thus proceeded to use serum for my subsequent experiments. Secondly, using filtration as a method for EV isolation may not be the optimal method for studying EV miRNA. Filtration may result in contamination from non-EV proteins, which are capable of transporting RNA (e.g. AGO2) (575). In addition, EVs may bind to the filtration membrane resulting in lower yields and the forces applied during filtration may result in EV deformation or rupture (576). This influenced my choice of EV isolation method for the BALF samples where I refined the filtration method using the additional step of SEC to isolate the EVs from soluble proteins.

Given the above factors, it may be difficult to compare the EV miRNA signatures of serum and BALF using different EV isolation methods. In Chapter 6 (section 6.6), I assessed whether the differentially expressed EV miRNA in BALF were also found in the peripheral blood. Of the eight differentially expressed EV miRNA in the lungs of COPD patients, only four were detectable in the serum. Of these, only one miRNA (miR-2110) was found differentially expressed between COPD and health. Expression of miR-2110 in the serum EVs was also shown to significantly correlate with disease severity (FEV1, $r = -0.51$, $p = 0.03$). Given, the limited sample numbers (n=24) and different EV isolation methods, it may be difficult to draw any firm conclusions from these results. Therefore to further investigate the possibility of a lung EV signature presence in the peripheral circulation, the methods of EV isolation should be similar but optimised for the sample type and

involve more in-depth characterisation of the EVs themselves (e.g. proteomic/lipidomic description). This approach may identify cell/tissue specific markers which may give novel insights into EV biology.

Although the EV field is rapidly evolving and expanding, there are still several unknowns with regards to the biology of EVs. A major challenge is the huge and underappreciated vesicle diversity; with a lack of understanding of EV function stemming from our inability to separate complex populations of vesicles into subclasses of particular sizes, compositions and biogenetic pathways (291). Furthermore, most EV research has been conducted *in vitro*, in which the experiment specific culture conditions may affect the biochemical and biophysical features of EVs. However, recent experiments using animal models have tracked EVs to their cells of origin, provide strong evidence in support of an important EV function *in vivo* (577-579). Furthermore, data reproducibility remains a challenge in the field, which is amplified by the EV diversity in both cell culture systems and bodily fluids. Small deviations in isolation protocols may result in collection of very different EV populations. The position papers by the International Society for Extracellular Vesicles (ISEV) and the Extracellular RNA Communication Consortium (ERCC) aim to address these issues and outline developments of effective technologies and strategies that allow better EV isolation, size characterisation, and definition of cargo composition (580-582). In future research, techniques such as single-vesicle analysis (583, 584), will provide greater insights into vesicle type and diversity and will reveal new functional and structural properties of EVs, which could be manipulated for the treatment of human disease.

7.3 Aims 2&3: Identify differentially expressed lung-derived EV miRNA in patients with COPD compared with healthy ex-smokers and their biological significant targets

Using next generation sequencing, I characterised the small RNA content of the EVs from patients with COPD compared with healthy controls. Firstly, my results show a higher proportion of miRNA and smallRNA in COPD BALF EVs than healthy ex-smokers (section 4.2.2). To my knowledge, only one other study has previously shown altered proportions of miRNAs in EVs in disease, with Francisco-Garcia et al. showing deficient loading of miRNAs in the BALF EVs of severe asthmatics (420). During their biogenesis, EVs may selectively capture cell-specific proteins, lipids, RNAs and DNA, which may become part of the EV's membrane or cargo (585). However, the exact mechanism of such selective packaging remains unknown. Specific sequences present in certain miRNA and/or certain proteins may guide the incorporation of specific miRNA into EVs (586-589).

These sequences and/or proteins may be dysregulated in COPD leading to higher levels of miRNA in COPD lung EVs.

Different molecules may be concomitantly incorporated into EVs, as illustrated by the different subclasses of smallRNA identified in this study (section 4.2.2). In general, biological communication systems are characterised by redundancies and interdependence, thus if EVs constitute a system of cell-to-cell communication, it has been suggested that EVs are likely to harbour some redundancy in surface characteristics and cargo (590). Therefore, to fully understand the functionality of EV cargo, technologies such as single-vesicle analysis may help decipher the mechanism of selective encapsulation of EV cargo and identify both active and redundant elements.

Moreover, this is the first study showing differential miRNA expression in lung-derived EVs in COPD. Specifically, five miRNA (miR-2110, miR-223-3p, miR-182-5p, miR-625-3p and miR-200b-5p) were found up-regulated in COPD and three were down-regulated (miR-338-3p, miR-138-5p, miR-204-5p) when compared to healthy controls. In comparison with previous literature (summarised in Table 1.2), this study used a cohort of subjects with COPD and healthy ex-smokers, whereas previous work in the lung used samples from smokers (244) and/or murine models (245). None of the miRNA identified in the study of smoker BALF EVs (let-7e, let-7g and miR-26b) (244) were identified as dysregulated in this study. In addition, the sample size was greater in this study (n=44) in comparison with those previously. Therefore the findings in this study are likely to represent changes as a result of disease, rather than smoke exposure alone, and are more applicable to human disease than a murine model, especially as disease continues after smoking cessation (591).

Several of the differentially expressed lung-derived EV miRNA have been found dysregulated in COPD in other sample types. Importantly, miR-223 has been identified as a key regulator in the innate immune response in airway disease (592), and up-regulation of non-EV miR-223 has been reported in COPD miRNA studies (summarised in Table 1.3). Differential expression of miR-223 has been observed in several lung sample types (e.g. bronchial brushings, lung tissue and BALF) in smokers and patients with COPD (316, 319, 342). In addition, biomass fuel exposure has been shown to increase miR-223 expression in serum from women with COPD (345). Yet no studies so far have linked miR-223 expression to disease stage, inflammatory phenotype or presence of emphysema.

In this study of lung-derived EVs, miR-223-3p was the most up-regulated miRNA in COPD compared with healthy ex-smokers with a log2FC 2.97 (Table 4.7, FDR = 0.016). In addition, miR-223-3p was one of the three miRNA (along with miR-182-5p and miR-2110) which showed the greatest predictive ability to discriminate between COPD and healthy ex-smokers (Table 6.5, AUC 0.91). Interestingly, although previous work in asthma suggested that miR-223 expression in sputum is associated with neutrophilic disease (435), this study shows lung-EV miR-223-3p expression correlates with airway eosinophilia (Table 6.6, r = 0.47, p = 0.03), and moreover strongly predicted the presence of pure airway eosinophilia from other inflammatory endotypes in COPD (Table 6.9, AUC 0.94). In support of this, miR-223-3p expression in bronchial biopsies was previously shown to correlate with eosinophils in asthmatics (566). Overexpression of miR-223-3p significantly increased GM-CSF in response to cigarette smoke extract (566), whereby GM-CSF is known to promote eosinophil recruitment and cell survival (568). Therefore miR-223-3p may be involved in eosinophil recruitment control by increasing GM-CSF expression leading to an increase in eosinophil infiltration and cell survival.

MicroRNA-182 is another candidate which has previously been implicated in COPD pathogenesis, with up-regulation found in BALF of smoke exposed mice (327). This thesis showed up-regulation of miR-182-5p in lung-derived EVs in COPD compared with healthy ex-smokers (Table 4.7, log2FC 1.52, FDR = 0.04). Furthermore, miR-182-5p expression correlated with airway neutrophilia in the lungs of patients with COPD (Table 6.6, r = 0.46, p = 0.03). MiR-182-5p is already known to regulate neutrophil proliferation and migration and therefore EV delivery of miR-182-5p in the airways of COPD patients may contribute to neutrophilia. In addition, miR-182-5p has been implicated in vascular remodelling (446) and small airway fibrosis (447), both which are key pathological processes in COPD.

MiR-182-5p demonstrated the greatest number of putative mRNA targets within the paired epithelial brushing transcriptome (Table 5.7 – 33 predicted genes identified, Figure 5.4 - 16.8% of all differentially expressed genes). In addition, cluster analysis identified miR-182-5p to be a central component to Cluster 2 (shown in pink in Figure 6.9) suggesting miR-182-5p is key to the miR-mRNA interaction network. In addition, lung-EV miR-182-5p expression significantly negatively correlated with several target genes, including jade family PHD finger 1 (*JADE1)* expression (Table 5.5, r = -0.52, FDR = 0.002). *JADE1* is a negative regulator of Wingless/Integrase-1 (WNT) signalling, which has been previously implicated in promoting emphysema through abnormal alveolar repair (515). Thus, lung-derived EV miR-182-5p reduction of *JADE1* expression in the airway epithelium could be one mechanism underlying the development of emphysema in COPD. MiR-182-5p was also found to co-regulate several target genes in the miR-mRNA

interaction network suggesting synergistic activity. For example, *CRIM1*, which was down-regulated in COPD epithelial brushings was targeted by three up-regulated miRNA, miR-223-3p, miR-182-5p and miR-200b-5p (Figure 5.7). *CRIM1* has been implicated in inhibiting the invasion and metastasis of lung adenocarcinoma cells via regulation of miR-182 (526) and may play a role in capillary formation and maintenance during angiogenesis (527). Thus, down-regulation of *CRIM1* expression in COPD airway epithelium may have implications for the early development of lung cancer and the vascular remodelling in COPD.

Of the other up-regulated miRNA, very little is known about their regulation in COPD. MiR-2110 was first reported as one of a group of neurite-inducing miRNAs (449) and may have a role in tumour suppression in neuroblastoma (450). Furthermore, it has been identified as one of five miRNAs up-regulated in serum exosomes in patients with active tuberculosis infection (451). In this thesis, miR-2110 was up-regulated in the lung-derived EVs of patients with COPD compared with healthy ex-smokers. MiR-2110 showed strong predictive ability to differentiate between health and disease (Table 6.4, AUC 0.81), correlated with neutrophilic airway inflammation (Table 6.6, $r = 0.47$, $p = 0.03$) and was also one of two miRNA to correlate with a physiological measure of emphysema (diffusion capacity of the lung for carbon monoxide, DLCO) in COPD patients alone (Table 6.2, $r = -0.43$, $p = 0.02$). Furthermore, in serum, miR-2110 was the only miRNA significantly differentially expressed (Figure 6.3, $p = 0.03$) and correlated with disease severity as measured by FEV1 (Figure 6.4, $r = -0.51$, $p = 0.03$). Taken together, miR-2110 is likely to have a role in COPD progression and given its association with neutrophilia and DLCO (a surrogate for emphysema), it may have a role in potentiating neutrophilic tissue destruction in the lungs of patients with COPD.

Interestingly, although miR-2110 had far fewer putative mRNA targets identified in the airway epithelial transcriptome in COPD (Figure 5.7) than miR-182-5p (likely due to its more recent discovery), a number of these targets had larger difference in fold change, suggesting more regulation. For example, miR-2110 targeted *Sushi Domain Containing 2* (*SUSD2*) gene was found down-regulated in epithelial brushings in COPD (log2FC -1.25, FDR 0.005) and is known to function as a tumour suppressor in lung cancer (593). Cheng et al. demonstrated that knockdown of *SUSD2* promoted cell growth in human alveolar epithelial cells (A549 cells) (593), thus miR-2110 suppression of *SUSD2* has implications for COPD as a possible mechanism for aberrant cellular repair.

It is well known that COPD is a significant risk factor for lung cancer independent of cigarette smoking (594) and activation of specific inflammatory pathways (such as NFκB signalling) are common to both pathologies. MiR-200b belongs to the miR-200 family, which controls epithelial-

mesenchymal transition (EMT) and metastasis in tumour cells (452, 453). EMT has been implicated in the formation of peribronchiolar fibrosis in COPD (115) and may be a precursor of lung cancer in these patients (455). Indeed, miR-200b is significantly over-expressed lung cancer tissue with estimated fold change exceeding 37 times (595). In this thesis, miR-200b was up-regulated in the lung-derived EVs of patients with COPD compared with healthy ex-smokers. In addition miR-200b-5p showed a strong negative correlation with the physiological measure DLCO in COPD patients alone (Table 6.2, r = -0.6, p = 0.001). DLCO is a measure of the total ability of the lungs to transfer carbon monoxide across into the bloodstream (596). When the alveolar-capillary membrane is thickened, as in fibrosis, the distance which the test gas has to travel to reach the blood is increased and the DLCO will be lower. Therefore, these results may suggest that EV delivered miR-200b-5p is increasing EMT, resulting in increased airway fibrosis and reduced gas transfer (DLCO) in patients with COPD. Thus, modulating this signal may offer a therapeutic opportunity to prevent small airway fibrosis and improve gas transfer in patients with COPD.

As discussed in Chapter 4, miR-625-3p has been implicated in several malignancies (457-459) and also has a role in CD8+ T cell proliferation in stem cell transplant patients (460). Furthermore it has been implicated in paediatric asthma (461) and has been shown to suppress inflammation by targeting protein kinase B2 (AKT2) (462). It is worth noting that there were several missing data points for miR-625-3p (n=13 missing points) and therefore the up-regulation in COPD lung-EVs may be a less reliable signal. In keeping with this, miR-625-3p was not as strong predictor of disease presence in COPD (AUC 0.76) and did not correlate with any inflammatory indices within the airways. Moreover, in the miRNA-mRNA interaction network (see Figure 5.7), miR-625-3p had only one mRNA target identified (*TBC1D3C*) within the paired epithelial brushing transcriptome and this demonstrated only a tiny down-regulation in COPD with a log2FC -0.0017 (FDR 0.03). Therefore in this study miR-625-3p regulation is unlikely to have any significant bearing on disease mechanisms.

In this thesis, miR-338-3p was down-regulated in lung-derived EVs of patients with COPD compared with healthy ex-smokers. MiR-338-3p has previously been reported to be down-regulated in COPD plasma compared with asthmatics and healthy controls (341), and down-regulated in the plasma of allergic rhinitis patients, where it was shown to target selected inflammatory genes *MAPK8* and *IKKβ* (463). In contrast, Lacedonia et al. showed miRNA-338 expression in the sputum was higher in both patients with asthma and COPD compared to controls, however asthmatics showed a significantly higher miR-338 expression compared to COPD patients (464). Furthermore, in this thesis, miR-338-3p was the only down-regulated miRNA to correlate with a clinical characteristic of COPD, with a moderate correlation with a marker of

small airways disease (FEF 25-75%) in COPD patients alone (Table 6.3, r = 0.44, p = 0.02). In addition, as with miR-223-3p, miR-338-3p correlated with airway eosinophilia (Table 6.6, r = -0.42, p = 0.03), suggesting a role in regulating T helper-2 (Th2)-driven inflammation. Thus given this association with eosinophilia, these miRNAs may serve as biomarkers for steroid responsive disease in COPD.

MicroRNA-204-5p was the most significantly down-regulated miRNA in the lung-derived EVs in COPD compared with healthy ex-smokers (Table 4.7, log2FC -2.37, FDR 0.037) and had the greatest number of putative mRNA targets identified of the down-regulated miRNA (Table 5.7, 20 predicted genes identified; Figure 5.4, 10.2% of the differentially expressed genes). Importantly, as with miR-200b-5p, research suggests a role for miR-204-5p in direct regulation of EMT through its targeting of *SMAD4*, a mediator of TGF-β signalling (474). Wang et al. demonstrated that miR-204-5p overexpression enhanced the repression of TGF-β2–induced EMT in the presence of *SMAD4* small interfering RNA (474). Therefore a reduction in miR-204-5p, as seen in lung-derived EVs in COPD patients in this study, may lead to an increase in EMT. MiR-204-5p's regulation of EMT suggests synergistic activity with miR-200b-5p (452, 453) and together they may contribute to the small airway fibrosis seen in COPD.

Furthermore, miR-204-5p negatively correlated with both airway neutrophilia and eosinophilia (Table 6.6, r = -0.49, p = 0.01; r = -0.57, p = 0.002 respectively). Previous research shows that miR-204-5p inhibits inflammation and chemokine generation by modulating interleukin (IL)-6 expression (562). This may suggest down-regulation of lung-derived EV miR-204-5p may contribute to airway neutrophilia via an IL-6 dependant pathway in patients with COPD, since IL-6 is a potent recruiter of neutrophils during acute inflammation (563). Thus, enhanced miR-204-5p expression may prevent excessive airway neutrophilia and lead to a reduction in airway inflammation and tissue destruction. Whilst there is no direct evidence of miR-204-5p regulation of eosinophils, miR-204-5p has been shown to regulate TLR4 (564), which in turn activates eosinophils (565) and therefore miR-204-5p modulation of TLR4 may play a role in driving eosinophilia in COPD. Moreover neutrophils and eosinophils are both granulocytes with a common progenitor cell (myeloblast) and therefore regulation of miR-204-5p may occur before differentiation in precursor cell types within the bone marrow.

MiRNA-138-5p was significantly down-regulated in lung-derived EVs in COPD compared with healthy ex-smokers (Table 4.7, log2FC = -1.66, FDR = 0.041). Along with miR-338-3p, miR-138-5p was found to negatively correlate with target gene *Aldo-keto reductase type 1C 2* (*AKR1C2*) (Table 5.5, r = -0.44, FDR 0.01; r = -0.46, FDR 0.009 respectively). The AKR1C family, including *AKR1C2*,

code for enzymes implicated in steroid metabolism and their expression levels are localised in the normal tissues of the lung, liver, prostate, testis and mammary glands (517). *AKR1C2* regulation is associated with several cancers, however their exact mechanism in promoting tumourigenesis is not known (518). In-keeping with this thesis, *AKR1C2* has been found up-regulated in the airways of COPD and healthy smokers (519),

Cluster analysis (Figure 5.9 – cluster 1 shown in purple) and comparison of the miRNA's Jaccard index values to each other (Figure 5.6) revealed similarity of miR-138-5p with miR-204-5p suggesting possible synergistic action. Several genes were co-regulated by miR-138-5p and miR-204-5p (Figure 5.8), including brain-derived neurotrophic factor (*BDNF*) gene and sodium voltage-gated channel alpha subunit 2 (*SCN2A*) gene. Levels of serum BDNF have been shown in a number of studies to correlate with disease severity in COPD (530-532) and its role as a mediator of neuronal plasticity, has been shown to be key in acute and chronic inflammatory conditions of the airways (533, 534). Furthermore, mutations in *SCN1A* have a strong association with pulmonary emphysema (535) and therefore modulation of this gene in the airway epithelium may have important consequences for COPD pathogenesis.

The quantity of cargo material carried within an EV is extremely small (597) and therefore it is assumed that there must be a highly efficient and EV-specific recognition tool in recipient cells (585). Furthermore, it is unlikely that a single miRNA is responsible for phenotypic change, where rather a panel or subset specific for the task would be released by the donor cell. Considering this multi-miRNA regulation, a network approach was employed to understand the synergistic activities of the dysregulated miRNA, and gene ontology (GO) enrichment analysis of the miRNA target genes to explore pathway regulation.

To identify the possible EV miRNA–mRNA interactions a number of analytical approaches were used. A major strength of this work is the paired dataset between the EV miRNA and epithelial brushings from the same lung lobe, in the same subject. This enables a more accurate description of the miRNA-mRNA interactions, rather than relying on *in silico* analysis alone or even experimental models, which are an over-simplification of the disease process. The combinatorial analytical approach of these miRNA-mRNA interactions, with correlations, synergism analysis and clustering enables emphasis of specific interactions, which may have more biological relevance and highlights targets that could be taken forward for validation in experimental or therapeutic models.

The miRNA-mRNA interaction network analysis identified several pathways which may be key to COPD pathogenesis. Specific miRNA-mRNA interactions have been discussed above, however to

examine the overall impact of the miRNA-mRNA interaction network on biological pathways, GO enrichment analysis was performed. Pathways involving regulation of phospholipase (PL) were found to be significantly targeted in this analysis and mapped to three significantly differently expressed genes *NPR3*, *HOMER1* and *ARHGAP6*, all of which are regulated by miR-182-5p. As previously discussed in Chapter 5, PLs are important in cellular homeostasis and are recognised as an important mediator of lung inflammation and infection (598). In addition, GO analysis for molecular function, identified insulin-like growth factor (IGF) binding as dysregulated with key mapped genes *INSR*, *NPR3* and *ARHGAP6*. Components of the IGF-1 signalling pathway are potentiated as biomarkers as they are dysregulated locally or systemically in COPD, however these findings vary among different studies (546). In both GO analyses, *NPR3* was identified as a key target. *NPR3* was the most downregulated mRNA in COPD epithelial brushings compared with healthy ex-smokers (Table 5.2, log2FC = -2.01, FDR = 0.0003) and has been implicated in COPD previously. Most recently, Kachroo et al. showed *NPR3* was differentially methylated in smoke exposed fetal lung samples (505) and its presence correlated with diseased parenchyma in lung biopsies (506). Furthermore, in a COPD mouse model, *NPR3* regulation was implicated in treatment responsiveness to a GLP-1 agonist (499). IGF-1 has also been implicated in right heart failure in COPD via modulation by miR-223-3p (599), and in skeletal muscle wasting (600). This invites the possibility that EV miRNA may have a role in multimorbidity in COPD, a condition that has been identified as a key modifiable factor in COPD (601).

In summary, the eight differentially expressed miRNA found in lung-derived EVs in patients with COPD were found to target 196 differentially expressed genes in epithelial brushings from the same cohort. Several of these miRNA-mRNA interactions are involved in central cellular inflammatory and metabolic pathways and are thus likely to play important roles in COPD pathophysiology. In addition, these novel findings may have therapeutic potential and lead to the development of disease modifying treatments.

7.4 Aim 4 Explore the diagnostic use of the lung-derived EV miRNA and their relationship with specific COPD inflammatory endotypes

The combination of miR-2110, miR-223-3p and miR-182-5p showed excellent predictive ability (AUC 0.91) in discriminating between COPD and healthy ex-smokers. Importantly these findings were shown in a relatively mild COPD cohort (mean FEV1 % predicted 77.5%; SD±14.8), with no established emphysema on high resolution computer tomography. Thus, these findings may have implications for diagnosis and disease mechanisms in early disease.

There is growing interest in the origins of COPD as it is envisaged that preventative efforts and treatment can modify its clinical course. Furthermore, it is recognised that our current diagnostic classifier of FEV1/FVC (4) is a crude tool, which may miss early disease and correlates poorly with symptoms particularly in mild disease (602). COPD pathogenesis may begin *in utero*, where passive smoke exposure is associated with adult COPD risk, independent of active or passive exposure during lifetime (603). In addition, childhood respiratory impairment, either through smoke exposure or infection, leads to an increased risk of reduced adult lung function (604). So far, neither measures of inflammation nor other biomarkers can identify individuals with lung function in the healthy range who will develop COPD (602), however studies designed purposefully to investigate the markers of early disease are ongoing (605). Thus, the potential role of EV miRNA as a biomarker of early disease or as a potential mechanisms for early epigenetic regulation in COPD may be important to investigate further.

In the present study, to pursue the role of these differentially expressed miRNA as a biomarker for disease, I investigated whether the lung-derived EV miRNA signature was also present in the peripheral blood. These results showed only one of the eight miRNA (miR-2110) found in the lung EVs was also significantly dysregulated in serum EVs in COPD patients. As discussed in section 7.2 this lack of translation into the peripheral circulation may be due to the different methodologies used for EV isolation. The role of biomarkers in COPD has been the focus of intensive research, however due to the complexity and heterogeneity of the disease, it has been suggested that future studies should progress from a simplistic approach of comparing patients with COPD with control subjects and focus on more specific patient groups or endotypes (606).

Understanding of COPD endotypes is still limited given the poor understanding of the underlying cellular and molecular mechanisms of COPD and how these mechanisms may vary between patients. However, the variable clinical manifestations of COPD and differences in response to therapy suggest that there may be different endotypes of disease that in future may be treated with more precision than current broad spectrum therapies.

As discussed above, my results suggest that lung-derived EV miR-223-3p and miR-338-5p are associated with eosinophilia within the lung. These results may have implications for treatment stratification for eosinophilic targeted therapy, given that eosinophilic COPD patients have more frequent exacerbations (102, 103) and are more responsive to corticosteroid treatment (104). However, a major issue facing the COPD research community is the lack of consensus on the appropriate thresholds used to define eosinophilic inflammation in this disease. Eosinophil numbers differ during stable disease, exacerbations, and following treatment (607), with blood

eosinophil counts known to fluctuate in individuals during a 24-hour period (608). Therefore, it is unlikely that a single threshold will be recommended for guiding all treatment decisions. Furthermore, evidence varies on whether blood eosinophil levels correlate with eosinophil expression within the airways (169, 171, 411) and the mechanism of eosinophilia in COPD is not yet certain. Eosinophilic promoters such as IL-5, are all increased in patients with eosinophilic COPD (173), however targeted anti-IL-5 treatments (e.g. mepolizumab) have only a minor clinical benefit (178). Therefore exploring novel mechanisms for airway eosinophilia in COPD, possibly through an EV miRNA mechanism, could provide new therapeutic targets.

7.5 Summary of the implications of study findings

7.5.1 EV miRNA as a biomarker in COPD

To my knowledge, this is the first study to identify differentially expressed miRNA in lung-derived EVs in patients with COPD compared with healthy controls. One of the major implications of these findings, is whether these EV miRNA could be used as biomarkers of disease, particularly given these results were shown in a mild disease cohort. EVs have been identified as novel disease biomarkers due to their capacity to reflect parent cells physiological state and microenvironment, as well as being highly stable in circulating bodily fluids, with the ability to package an array of disease associated molecules (199, 223). A number of studies have already demonstrated the use of circulating microvesicles (MVs) as possible biomarker candidates in COPD (250, 260). Although, my findings did not translate into the peripheral blood, future studies investigating EV miRNA as a biomarker in easily sampled biofluids, such as blood or exhaled breath, may identify novel biomarkers for COPD. This could have important implications for early diagnosis and initiation of treatment in a disease with growing prevalence.

7.5.2 Therapeutic potential of EVs in COPD

Molecular engineering techniques have been employed to modify EV cargo for therapeutic use. With particular reference to the findings in this thesis, manipulation of EV miRNAs has been performed, with high efficiency of miRNA delivery to recipient cells (275). EVs provide the ideal transportation method for therapeutic miRNA cargo given they protect from digestion and degradation whilst evading the host immune surveillance system due to their surface markers reflecting host cell origin. Given the implications of some of the differentially expressed miRNA targets in this study, (e.g. miR-200b-5p and miR-204-3p in possible direct synergistic regulation of

EMT), manipulation of these EV miRNA either with mimics or antagomiRs could result in a significant impact on COPD pathological processes, such as small airways fibrosis. Given the lack of disease modifying treatments, this could be a real step-change in the way we manage this condition.

In summary, this thesis demonstrates that dysregulated EV miRNA in the lungs of COPD patients may in part be driving aberrant cellular regulation, specifically in the context of phospholipase C and IGF-1 signalling. Both of these pathways are integral to cellular homeostasis, and therefore altered epigenetic control via miRNA post-transcriptional regulation of gene expression may lead to significant downstream consequences. To further prove this hypothesis, *ex vivo* modelling demonstrating the impact on these miRNA on specific gene expression (e.g. *NPR3*, *HOMER1* and *ARHGAP6*) would provide experimental validation in a disease relevant system. In addition, exploring whether altered gene expression is a consequence of a single miRNA target or multiple miRNA in synergism would be important in determining the role of EVs as targeted messengers of a specific disease miRNA signature. Importantly, examining the functional consequences of changes in gene expression (e.g. loss of epithelial barrier integrity) would be an important additional step in understanding the consequences of this possible epigenetic regulation. Finally, given the high prevalence of comorbidity in COPD, the role of lung derived EV miRNA in driving IGF-1 dysregulation may be a key to understanding the complex interplay between lung inflammation and systemic comorbidity. Future work could focus on the role of lung EV miRNA in driving systemic disease, such as epigenetic control of IGF-1 signalling in potentiating skeletal muscle wasting. The full scope of the future work will be discussed in section 7.7.

7.6 Strengths & Limitations

The strengths and limitations of this thesis have been discussed throughout this chapter. Importantly, the in-depth patient/subject characterisation and paired sampling approach of the MICA II cohort has been crucial in providing a platform for investigation of EV miRNA differences in COPD. Furthermore, in comparison to other patient focused studies in COPD EV miRNA, the sample size was comparable if not larger, leading to increased power of these results. To answer the fundamental unknowns in COPD pathogenesis, detailed characterisation of COPD cohorts is key to exploring the heterogeneity and complexity of this disease.

COPD is inherently a disease characterised by pulmonary inflammation. Thus, focusing on lung-derived samples such as BALF and epithelial brushings to explore the differences and key targets of EV miRNA in COPD is likely to gain novel and relevant insights into COPD biology. However, as

discussed, bronchoscopy is an invasive procedure with limits on sample availability, and in particular reference to biomarker discovery, sampling of more readily available biofluids such as blood or sputum will be key to determining the utility of EV miRNA as biomarkers in future. In this study, the lack of translatability of the EV miRNA differential signal in the peripheral blood limits its value as a diagnostic marker. Thus further work should be done to explore the EV miRNA signature in peripheral blood in COPD, with a focus on cell/tissue specific surface marker identification, which may have increase EV miRNA disease specificity.

7.7 Further work

To fully understand the role of these differentially expressed lung-derived EV miRNA in COPD and prove the overall hypothesis, there is a significant amount of further work to be explored.

7.7.1 *Ex vivo* cell culture models

A next step would be to validate the miRNA-mRNA targets in an established experimental *ex vivo* human bronchial epithelial cell model (609). Using a lipofectamine transfection system, which mimics EV delivery machinery, transfection of primary human bronchial epithelial cells with the differentially expressed miRNA (either as precursor mimics for up-regulated miRNA or as antagomiRs for down-regulated miRNA) could be performed. Measurement of downstream target gene expression with RT-qPCR would validate the predicted targets of the differentially expressed miRNA in a disease relevant cell system and provide an opportunity to further explore synergistic activity of different miRNA combinations. The experimental model can be altered to reflect different disease specific environments, for example EV miRNA function in the context of viral infection. Furthermore, downstream consequences of altered gene expression could be measured, for example using immunofluorescence for tight junction proteins (e.g. Zonula occludens-1) to assess epithelial cell barrier function.

7.7.2 Explore EV miRNA-mRNA interactions in other cell types

As part of the MICA II cohort study, subjects had epithelial biopsies and alveolar macrophages collected processed for RNA sequencing. Therefore, it is possible to examine the mRNA targets of the differentially expressed miRNA in these tissues and cells. Epithelial biopsies contain a mixture of cell types and may give further insights into the likely EV miRNA functions. Whereas alveolar macrophages are the dominant inflammatory cell in the airways (67) and have an altered phenotype in COPD (70-72). Thus exploring the possible interaction of EV miRNA on macrophage

gene expression could identify disease specific pathological mechanisms. Furthermore, by using an established *in vitro* model of naive macrophages derived from monocytes harvested from healthy blood (610), it is possible to transfect macrophages in culture with the dysregulated miRNA (or relevant antagomiRs) and measure phagocytic ability using green fluorescent protein (GFP)-labelled bacteria. In addition, using these established macrophage infection models, measurement of cytokine production and expression of cell surface markers (e.g. MHC class I and II, CD54 and TLRs) in transfected cells following infection with influenza (611, 612) would demonstrate the functional consequences of any miR-mRNA regulation. Together these models will validate the findings of the *in silico* analysis and provide novel mechanistic insights into the pathogenesis underlying COPD.

7.7.3 Interrogate multi-omic readouts to identify downstream effects of EV miRNA

In addition to RNA sequencing, proteomic, metabolomic and lipidomic analysis was performed on the BALF samples in the MICA II cohort subjects. Therefore, it is possible to investigate the potential downstream effects of these miRNA-miRNA interactions in each of these 'omic outputs. Importantly, changes in mRNA expression, do not necessarily result in variation at a protein level and therefore investigating specific mRNA-protein interactions will add further evidence to the importance of these changes in COPD pathogenesis. Lipidomic and metabolomic changes in COPD are somewhat under investigated in comparison with the other 'omic platforms and therefore exploring these changes in relation to EV miRNA may give novel insights into pathways and cellular processes involved in COPD pathogenesis. With reference to this work, I have already started a working collaboration with the AstraZeneca multi-omics bioinformatics team and have preliminary expertise in interpreting multi-omic results using strategies such as multi omics data integrative clustering and gene set analysis (MOGSA).

7.7.4 Study of EV miRNA signature in early COPD

Finally, as discussed in section 7.4, there is renewed focus in the early origins of COPD with the prospect of early intervention leading to more favourable prognostic course. Given the results of this thesis showing the excellent predictive ability of miR-2110, miR-223-3p and miR-182-5p in discriminating between mild COPD and healthy ex-smokers, there is a possible role for EV miRNA as a biomarker of early disease. Therefore investigation of EV miRNA in an early disease COPD cohort, such as the British Lung Foundation Early COPD cohort (605) may give important insights into the role of EV miRNA in the potential mechanisms for the development of COPD in some smokers.

7.7.5 Summary of future work

In summary, the investigation of lung-derived EV miRNA in COPD has led to some novel insights into possible inflammatory and metabolic mechanisms underlying COPD disease progression. Further work pursing the EV miRNA impact on gene expression and downstream protein function both at a molecular and cellular level may lead to novel targets for manipulation into new therapeutics. Moreover, the combinatorial multi-omic analysis approach provides an exciting opportunity to explore in-depth mechanistic processes in a complex disease, which is vital to the evolution of novel disease-modifying treatments in COPD.

Appendix A Supplementary results from microRNA-miRNA interaction

A.1 Differentially expressed genes in epithelial brushings in COPD

Differential gene expression were assessed with DESeq2 in the 44 epithelial brushing samples (24 COPD and 20 healthy ex-smokers) and this identified 192 differentially expressed genes between COPD and healthy ex-smokers with a false discovery rate (FDR) of <0.05 (summarised in Supplementary Table 1).

Supplementary Table 1 Differentially expressed genes in epithelial brushings comparing COPD subjects with healthy ex-smokers.

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000090512	*FETUB*	2.87	0.00004	0.013
ENSG00000179593	*ALOX15B*	2.61	0.00001	0.008
ENSG00000273331	*TM4SF19-DYNLT2B*	2.60	0.00001	0.006
ENSG00000287059	*lncRNA*	2.55	6.54E-10	2.94E-06
ENSG00000155918	*RAET1L*	2.52	0.00014	0.028
ENSG00000262406	*MMP12*	2.49	0.00005	0.015
ENSG00000115590	*IL1R2*	2.34	2.50E-07	4.50E-04
ENSG00000287771	*Lnc-IL1R1-1*	2.24	0.00004	0.012
ENSG00000111700	*SLCO1B3*	2.20	3.13E-06	0.003
ENSG00000255833	*TIFAB*	2.14	0.00005	0.014
ENSG00000198488	*B3GNT6*	2.10	0.00001	0.006
ENSG00000238266	*LINC00707*	2.09	0.00001	0.006
ENSG00000118785	*SPP1*	2.07	0.00004	0.012
ENSG00000231683	*LOC101927136*	2.06	1.08E-06	0.001
ENSG00000198074	*AKR1B10*	2.03	2.36E-09	9.09E-06
ENSG00000187054	*TMPRSS11A*	2.01	0.00025	0.039
ENSG00000283994	*Lnc-KYNU-14*	2.00	0.00032	0.047
ENSG00000180438	*TPRXL*	1.94	9.44E-12	5.10E-08
ENSG00000134827	*TCN1*	1.90	2.98E-06	0.003
ENSG00000258227	*CLEC5A*	1.87	0.00023	0.038

Appendix A

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000145832	*SLC25A48*	1.84	0.00003	0.012
ENSG00000270164	*LINC01480*	1.83	0.00023	0.038
ENSG00000138061	*CYP1B1*	1.83	4.12E-06	0.003
ENSG00000105388	*CEACAM5*	1.79	1.63E-06	0.002
ENSG00000186832	*KRT16*	1.77	0.00021	0.036
ENSG00000102962	*CCL22*	1.73	0.00013	0.027
ENSG00000167680	*SEMA6B*	1.72	0.00006	0.016
ENSG00000180861	*LINC01559*	1.71	1.77E-06	0.002
ENSG00000106178	*CCL24*	1.69	0.00018	0.033
ENSG00000272405	*Lnc-NES-2*	1.67	0.00002	0.010
ENSG00000182885	*ADGRG3*	1.67	0.00017	0.032
ENSG00000137440	*FGFBP1*	1.66	0.00005	0.015
ENSG00000163421	*PROK2*	1.66	0.00005	0.014
ENSG00000080031	*PTPRH*	1.51	0.00001	0.005
ENSG00000176697	*BDNF*	1.44	0.00003	0.012
ENSG00000253339	*Lnc-RDH10-1*	1.43	1.29E-07	3.17E-04
ENSG00000103888	*CEMIP*	1.42	2.37E-06	0.002
ENSG00000151012	*SLC7A11*	1.39	4.21E-06	0.003
ENSG00000167210	*LOXHD1*	1.39	0.00007	0.017
ENSG00000160862	*AZGP1*	1.34	0.00001	0.005
ENSG00000146013	*GFRA3*	1.32	0.00006	0.015
ENSG00000104368	*PLAT*	1.29	5.44E-07	0.001
ENSG00000142224	*IL19*	1.26	0.00033	0.047
ENSG00000104783	*KCNN4*	1.23	1.12E-06	0.001
ENSG00000128591	*FLNC*	1.23	0.00009	0.021
ENSG00000232079	*LINC01697*	1.21	0.00003	0.011
ENSG00000065618	*COL17A1*	1.21	2.48E-07	4.50E-04
ENSG00000100024	*UPB1*	1.17	0.00016	0.031
ENSG00000148926	*ADM*	1.16	0.00018	0.033
ENSG00000074410	*CA12*	1.16	0.00001	0.005
ENSG00000176153	*GPX2*	1.14	0.00000	0.000
ENSG00000080007	*DDX43*	1.13	0.00036	0.049
ENSG00000233013	*FAM157B*	1.11	0.00030	0.044

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000178363	*CALML3*	1.09	0.00030	0.045
ENSG00000134757	*DSG3*	1.06	0.00002	0.009
ENSG00000237523	*LINC00857*	1.02	0.00007	0.017
ENSG00000248323	*LUCAT1*	1.01	0.00000	0.001
ENSG00000130598	*TNNI2*	0.99	0.00003	0.011
ENSG00000232931	*LINC00342*	0.99	0.00002	0.009
ENSG00000187583	*PLEKHN1*	0.97	0.00004	0.012
ENSG00000136531	*SCN2A*	0.97	0.00031	0.046
ENSG00000111344	*RASAL1*	0.94	0.00005	0.015
ENSG00000158292	*GPR153*	0.93	0.00004	0.012
ENSG00000204264	*PSMB8*	0.91	0.00007	0.017
ENSG00000090659	*CD209*	0.91	0.00013	0.027
ENSG00000232977	*LINC00327*	0.90	0.00007	0.017
ENSG00000093009	*CDC45*	0.89	0.00013	0.027
ENSG00000151632	*AKR1C2*	0.88	0.00002	0.010
ENSG00000148773	*MKI67*	0.87	0.00003	0.012
ENSG00000196344	*ADH7*	0.87	0.00010	0.024
ENSG00000154040	*CABYR*	0.85	0.00002	0.009
ENSG00000164611	*PTTG1*	0.82	0.00001	0.008
ENSG00000149043	*SYT8*	0.81	0.00013	0.027
ENSG00000112559	*MDFI*	0.81	0.00019	0.034
ENSG00000106258	*CYP3A5*	0.80	0.00006	0.015
ENSG00000146592	*CREB5*	0.79	0.00035	0.049
ENSG00000065911	*MTHFD2*	0.76	0.00033	0.047
ENSG00000026508	*CD44*	0.76	0.00013	0.027
ENSG00000029153	*ARNTL2*	0.73	0.00001	0.005
ENSG00000260658	*Lnc-CDH8-10*	0.68	0.00036	0.050
ENSG00000167553	*TUBA1C*	0.67	0.00007	0.017
ENSG00000139289	*PHLDA1*	0.66	0.00016	0.031
ENSG00000065833	*ME1*	0.65	0.00025	0.040
ENSG00000184731	*FAM110C*	0.65	0.00005	0.014
ENSG00000099337	*KCNK6*	0.64	0.00014	0.028

Appendix A

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000166401	*SERPINB8*	0.60	0.00000	0.003
ENSG00000121380	*BCL2L14*	0.53	0.00004	0.013
ENSG00000259642	*ST20-AS1*	0.49	0.00022	0.037
ENSG00000112699	*GMDS*	0.49	0.00014	0.028
ENSG00000104635	*SLC39A14*	0.49	0.00023	0.038
ENSG00000182022	*CHST15*	0.47	0.00006	0.017
ENSG00000153310	*FAM49B*	0.47	0.00020	0.035
ENSG00000158715	*SLC45A3*	0.46	0.00011	0.025
ENSG00000213186	*TRIM59*	0.43	0.00018	0.033
ENSG00000115295	*CLIP4*	0.40	0.00001	0.005
ENSG00000158315	*RHBDL2*	0.32	0.00021	0.036
ENSG00000140374	*ETFA*	0.21	0.00023	0.038
ENSG00000105641	*SLC5A5*	0.01	0.00015	0.029
ENSG00000134028	*ADAMDEC1*	0.01	0.00018	0.033
ENSG00000131042	*LILRB2*	0.0011	2.27E-07	4.50E-04
ENSG00000229186	*ADAM1A*	0.0003	8.70E-14	5.87E-10
ENSG00000179886	*TIGD5*	0.0003	0.00024	0.039
ENSG00000273003	*ARL2-SNX15*	0.0002	0.00000	0.001
ENSG00000011114	*BTBD7*	0.00004	0.00001	0.006
ENSG00000234882	*EIF3EP*	-0.0001	2.49E-07	4.50E-04
ENSG00000251992	*SCARNA17*	-0.0003	2.80E-14	2.52E-10
ENSG00000227508	*LINC01624*	-0.0005	1.36E-14	1.83E-10
ENSG00000278599	*TBC1D3E*	-0.0005	3.71E-18	1.00E-13
ENSG00000278299	*TBC1D3C*	-0.0017	0.00023	0.038
ENSG00000285645	*AL133410.3*	-0.01	0.00016	0.031
ENSG00000165716	*DIPK1B*	-0.01	0.00026	0.040
ENSG00000113721	*PDGFRB*	-0.01	0.00025	0.039
ENSG00000250381	*UNC93B4*	-0.01	0.00011	0.024
ENSG00000121075	*TBX4*	-0.01	0.00024	0.039
ENSG00000137726	*FXYD6*	-0.02	0.00005	0.014
ENSG00000189319	*FAM53B*	-0.21	0.00012	0.026
ENSG00000171823	*FBXL14*	-0.21	0.00032	0.046
ENSG00000154930	*ACSS1*	-0.24	0.00029	0.044

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000170145	*SIK2*	-0.24	0.00008	0.019
ENSG00000133627	*ACTR3B*	-0.26	0.00015	0.030
ENSG00000176842	*IRX5*	-0.26	0.00016	0.031
ENSG00000077684	*JADE1*	-0.27	0.00004	0.013
ENSG00000136068	*FLNB*	-0.28	0.00036	0.049
ENSG00000204186	*ZDBF2*	-0.29	0.00023	0.038
ENSG00000019485	*PRDM11*	-0.29	0.00009	0.021
ENSG00000154856	*APCDD1*	-0.31	0.00028	0.043
ENSG00000149929	*HIRIP3*	-0.31	0.00003	0.011
ENSG00000178904	*DPY19L3*	-0.32	0.00005	0.014
ENSG00000153487	*ING1*	-0.32	0.00008	0.019
ENSG00000118960	*HS1BP3*	-0.33	0.00000	0.001
ENSG00000168350	*DEGS2*	-0.34	0.00026	0.040
ENSG00000177508	*IRX3*	-0.34	0.00029	0.044
ENSG00000214357	*NEURL1B*	-0.34	0.00022	0.037
ENSG00000171105	*INSR*	-0.35	0.00003	0.012
ENSG00000144749	*LRIG1*	-0.35	0.00002	0.010
ENSG00000172164	*SNTB1*	-0.36	0.00019	0.033
ENSG00000112773	*TENT5A*	-0.36	0.00006	0.015
ENSG00000143365	*RORC*	-0.38	0.00000	0.001
ENSG00000150938	*CRIM1*	-0.39	0.00007	0.017
ENSG00000141232	*TOB1*	-0.41	0.00033	0.047
ENSG00000185010	*F8*	-0.41	0.00031	0.045
ENSG00000132326	*PER2*	-0.41	0.00005	0.014
ENSG00000137486	*ARRB1*	-0.42	0.00016	0.031
ENSG00000155090	*KLF10*	-0.43	0.00004	0.012
ENSG00000117479	*SLC19A2*	-0.44	0.00002	0.010
ENSG00000171943	*SRGAP2C*	-0.45	0.00001	0.008
ENSG00000181690	*PLAG1*	-0.46	0.00011	0.024
ENSG00000047648	*ARHGAP6*	-0.47	0.00019	0.033
ENSG00000177283	*FZD8*	-0.48	0.00011	0.024
ENSG00000152413	*HOMER1*	-0.49	0.00003	0.012

Appendix A

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000106571	*GLI3*	-0.50	0.00003	0.012
ENSG00000164776	*PHKG1*	-0.54	0.00036	0.049
ENSG00000169515	*CCDC8*	-0.55	0.00006	0.017
ENSG00000184986	*TMEM121*	-0.56	0.00001	0.006
ENSG00000006210	*CX3CL1*	-0.57	0.00002	0.011
ENSG00000146197	*SCUBE3*	-0.57	0.00018	0.033
ENSG00000144712	*CAND2*	-0.58	0.00030	0.045
ENSG00000181458	*TMEM45A*	-0.59	0.00003	0.011
ENSG00000187987	*ZSCAN23*	-0.59	0.00003	0.012
ENSG00000138615	*CILP*	-0.62	0.00035	0.049
ENSG00000116039	*ATP6V1B1*	-0.63	0.00000	0.003
ENSG00000125730	*C3*	-0.69	0.00029	0.044
ENSG00000152931	*PART1*	-0.71	0.00010	0.022
ENSG00000166828	*SCNN1G*	-0.72	0.00003	0.011
ENSG00000073067	*CYP2W1*	-0.76	0.00025	0.040
ENSG00000005108	*THSD7A*	-0.76	0.00021	0.036
ENSG00000197291	*RAMP2-AS1*	-0.77	0.00004	0.012
ENSG00000118946	*PCDH17*	-0.77	0.00031	0.046
ENSG00000173706	*HEG1*	-0.78	0.00004	0.012
ENSG00000079819	*EPB41L2*	-0.80	0.00001	0.006
ENSG00000205835	*GMNC*	-0.80	0.00014	0.028
ENSG00000255471	*Lnc-FZD4-1*	-0.82	0.00004	0.012
ENSG00000168874	*ATOH8*	-0.83	0.00005	0.014
ENSG00000110900	*TSPAN11*	-0.86	0.00012	0.026
ENSG00000161055	*SCGB3A1*	-0.86	0.00012	0.026
ENSG00000197838	*CYP2A13*	-0.89	0.00002	0.011
ENSG00000129757	*CDKN1C*	-0.93	0.00000	0.001
ENSG00000205502	*C2CD4B*	-0.96	0.00004	0.014
ENSG00000007216	*SLC13A2*	-0.98	0.00002	0.010
ENSG00000130988	*RGN*	-1.00	0.00021	0.036
ENSG00000263063	*LOC101929552*	-1.02	0.00017	0.033
ENSG00000129437	*KLK14*	-1.04	0.00020	0.035
ENSG00000154864	*PIEZO2*	-1.06	0.00001	0.006

ENSEMBL ID	HGNC SYMBOL	Log2FC	P value	FDR
ENSG00000166106	*ADAMTS15*	-1.07	0.0000002	0.0001
ENSG00000149021	*SCGB1A1*	-1.09	0.00004	0.012
ENSG00000171476	*HOPX*	-1.10	0.00006	0.015
ENSG00000283413	*Lnc-IL12B-3*	-1.11	0.00018	0.033
ENSG00000125144	*MT1G*	-1.12	0.00004	0.012
ENSG00000078596	*ITM2A*	-1.14	0.00003	0.011
ENSG00000126562	*WNK4*	-1.20	0.000001	0.001
ENSG00000214870	*LOC441204*	-1.24	0.00011	0.025
ENSG00000099994	*SUSD2*	-1.25	0.00001	0.006
ENSG00000174059	*CD34*	-1.27	0.00009	0.021
ENSG00000038295	*TLL1*	-1.35	0.00003	0.012
ENSG00000113389	*NPR3*	-2.01	0.00000	0.000
ENSG00000198787	*OR7E103P*	-2.34	0.00003	0.012

HGNC, HUGO Gene Nomenclature Committee; FC, fold change; FDR, False discovery rate.

A.2 Correlation analysis of miRNA-mRNA interactions

Pearson correlations were generated to explore the relationship between the differentially expressed miRNA and differentially expressed genes in the epithelial brushings in COPD (see section 5.2.4.1). A total of 302 pairs were identified were identified by Pearson correlation test with P <0.05, of those 141 had an FDR <0.05. Of these, 85 showed negative correlations implying miRNA regulation of mRNA (summarised in Figure 5.3). Comparison of these 85 miRNA-mRNA pairs with the results from the *in silico* target prediction (section 5.2.1) revealed 13 known target pairs, defined as direct interactions (summarised in Table 6.5) and 72 unknown pairs defined as indirect interactions (summarised in Appendix A, Supplementary Table 2).

Supplementary Table 2 Indirect miRNA-mRNA interactions from pairwise correlation analyses

miRNA	HGNC SYMBOL	r	P value	FDR
hsa_mir_182_5p	*RAMP2-AS1*	-0.58	0.00004	0.000
hsa_mir_223_3p	*PIEZO2*	-0.57	0.00005	0.000
hsa_mir_2110	*SCGB1A1*	-0.53	0.0002	0.002
hsa_mir_223_3p	*AC008703.1*	-0.52	0.0003	0.002
hsa_mir_138_5p	*SERPINB8*	-0.50	0.0006	0.003
hsa_mir_223_3p	*FLNB*	-0.48	0.001	0.005
hsa_mir_223_3p	*ZSCAN23*	-0.48	0.001	0.006
hsa_mir_204_5p	*AKR1C2*	-0.48	0.001	0.006
hsa_mir_138_5p	*FAM110C*	-0.47	0.001	0.007
hsa_mir_2110	*ZSCAN23*	-0.46	0.002	0.008
hsa_mir_182_5p	*OR4D12P*	-0.45	0.002	0.009
hsa_mir_204_5p	*AL365181.3*	-0.45	0.002	0.009
hsa_mir_338_3p	*AZGP1*	-0.45	0.002	0.009
hsa_mir_338_3p	*CYP1B1*	-0.45	0.002	0.01
hsa_mir_223_3p	*ACSS1*	-0.45	0.002	0.01
hsa_mir_338_3p	*FAM110C*	-0.45	0.002	0.01
hsa_mir_138_5p	*CA12*	-0.44	0.003	0.01
hsa_mir_138_5p	*DSG3*	-0.44	0.003	0.01
hsa_mir_182_5p	*CCDC8*	-0.44	0.003	0.01
hsa_mir_138_5p	*CEMIP*	-0.44	0.003	0.01
hsa_mir_138_5p	*ALOX15B*	-0.44	0.003	0.01
hsa_mir_223_3p	*PHKG1*	-0.44	0.003	0.01
hsa_mir_2110	*GMNC*	-0.43	0.003	0.01

hsa_mir_223_3p	*DEGS2*	-0.43	0.004	0.02
hsa_mir_338_3p	*CHST15*	-0.42	0.004	0.02
hsa_mir_338_3p	*ALOX15B*	-0.42	0.004	0.02
hsa_mir_182_5p	*WNK4*	-0.42	0.004	0.02
hsa_mir_338_3p	*CLIP4*	-0.42	0.005	0.02
hsa_mir_138_5p	*GPX2*	-0.41	0.005	0.02
hsa_mir_2110	*PLAG1*	-0.41	0.005	0.02
hsa_mir_204_5p	*ADAMDEC1*	-0.41	0.005	0.02
hsa_mir_2110	*CYP2A13*	-0.41	0.006	0.02
hsa_mir_223_3p	*SNTB1*	-0.41	0.006	0.02
hsa_mir_200b_5p	*SCGB1A1*	-0.41	0.006	0.02
hsa_mir_2110	*HS1BP3*	-0.41	0.006	0.02
hsa_mir_182_5p	*C2CD4B*	-0.40	0.008	0.03
hsa_mir_204_5p	*CDC45*	-0.40	0.008	0.03
hsa_mir_204_5p	*ALOX15B*	-0.39	0.008	0.03
hsa_mir_338_3p	*SERPINB8*	-0.39	0.009	0.03
hsa_mir_338_3p	*DSG3*	-0.39	0.009	0.03
hsa_mir_138_5p	*ENSG00000287059*	-0.39	0.009	0.03
hsa_mir_204_5p	*SLC25A48*	-0.38	0.01	0.03
hsa_mir_223_3p	*TMEM121*	-0.38	0.01	0.03
hsa_mir_338_3p	*FLNC*	-0.38	0.01	0.03
hsa_mir_2110	***PIEZO2***	-0.38	0.01	0.04
hsa_mir_223_3p	*OR4D12P*	-0.38	0.01	0.04
hsa_mir_138_5p	***COL17A1***	-0.37	0.01	0.04
hsa_mir_2110	*AC004540.1*	-0.37	0.01	0.04
hsa_mir_338_3p	*CA12*	-0.37	0.01	0.04
hsa_mir_138_5p	*CALML3*	-0.37	0.01	0.04
hsa_mir_625_3p	*LINC01480*	-0.37	0.01	0.04
hsa_mir_338_3p	*MDFI*	-0.37	0.01	0.04
hsa_mir_182_5p	*ACTR3B*	-0.36	0.02	0.045
hsa_mir_338_3p	*IL1R2*	-0.36	0.02	0.047
hsa_mir_182_5p	*TSPAN11*	-0.36	0.02	0.047
hsa_mir_182_5p	*GMNC*	-0.36	0.02	0.048

hsa_mir_204_5p	*DDX43*	-0.36	0.02	0.049

Genes in bold also direct targets for alternative miRNA. r – generated using Pearson's correlation coefficient. FDR calculated using the Benjamini-Hochberg method.

List of References

1. Adeloye D, Chua S, Lee C, Basquill C, Papana A, Theodoratou E, et al. Global and regional estimates of COPD prevalence: Systematic review and meta-analysis. Journal of global health. 2015;5(2):020415.

2. World Health Organisation: Global Health Estimates 2016: Disease burden by Cause, Age, Sex, by Country and by Region, 2000-2016. 2018.

3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS medicine. 2006;3(11):e442.

4. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease, 2019 Report. www.goldcopd.org; 2018.

5. Fletcher C, Peto R. The natural history of chronic airflow obstruction. British Medical Journal. 1977;1(6077):1645-8.

6. Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. 2004;23(6):932-46.

7. Sinden NJ, Stockley RA. Systemic inflammation and comorbidity in COPD: a result of 'overspill' of inflammatory mediators from the lungs? Review of the evidence. Thorax. 2010;65(10):930-6.

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

9. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96.

10. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. Eur Respir J. 2006;28(3):523-32.

11. Quach A, Giovannelli J, Cherot-Kornobis N, Ciuchete A, Clement G, Matran R, et al. Prevalence and underdiagnosis of airway obstruction among middle-aged adults in northern France: The ELISABET study 2011-2013. Respiratory medicine. 2015;109(12):1553-61.

12. American Thoracic Society Foundation: The Global Burden of Lung Disease 2014 [Available from: http://foundation.thoracic.org/news/global-burden.php.

13. Halpin DM, Miravitlles M. Chronic obstructive pulmonary disease: the disease and its burden to society. Proc Am Thorac Soc. 2006;3(7):619-23.

14. Stern DA, Morgan WJ, Wright AL, Guerra S, Martinez FD. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. Lancet. 2007;370(9589):758-64.

15. Tashkin DP, Altose MD, Bleecker ER, Connett JE, Kanner RE, Lee WW, et al. The lung health study: airway responsiveness to inhaled methacholine in smokers with mild to moderate airflow limitation. The Lung Health Study Research Group. Am Rev Respir Dis. 1992;145(2 Pt 1):301-10.

16. Lange P, Celli B, Agusti A. Lung-Function Trajectories and Chronic Obstructive Pulmonary Disease. The New England journal of medicine. 2015;373(16):1575.

17. Sveger T. Liver disease in alpha1-antitrypsin deficiency detected by screening of 200,000 infants. The New England journal of medicine. 1976;294(24):1316-21.

18. Rennard SI, Vestbo J. COPD: the dangerous underestimate of 15%. Lancet. 2006;367(9518):1216-9.

19. Hutchison DC. Alpha 1-antitrypsin deficiency in Europe: geographical distribution of Pi types S and Z. Respiratory medicine. 1998;92(3):367-77.

20. Stanley SE, Chen JJ, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, et al. Telomerase mutations in smokers with severe emphysema. J Clin Invest. 2015;125(2):563-70.

21. Cohen BH, Ball WC, Jr., Brashears S, Diamond EL, Kreiss P, Levy DA, et al. Risk factors in chronic obstructive pulmonary disease (COPD). American journal of epidemiology. 1977;105(3):223-32.

22. Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, et al. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. Am J Respir Crit Care Med. 1998;157(6 Pt 1):1770-8.

23. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. American journal of human genetics. 2002;70(5):1229-39.

24. Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. Am J Respir Crit Care Med. 2011;184(7):786-95.

25. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nature genetics. 2010;42(3):200-2.

26. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS genetics. 2009;5(3):e1000421.

27. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nature genetics. 2017;49(3):416-25.

28. Mercado N, Ito K, Barnes PJ. Accelerated ageing of the lung in COPD: new concepts. Thorax. 2015;70(5):482-9.

29. Meiners S, Eickelberg O, Königshoff M. Hallmarks of the ageing lung. Eur Respir J. 2015;45(3):807-27.

30. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194-217.

31. Brandsma C-A, de Vries M, Costa R, Woldhuis RR, Königshoff M, Timens W. Lung ageing and COPD: is there a role for ageing in abnormal tissue repair? European Respiratory Review. 2017;26(146).

32. Tuder RM, Kern JA, Miller YE. Senescence in chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2012;9(2):62-3.

33. Antony VB, Thannickal VJ. Cellular Senescence in Chronic Obstructive Pulmonary Disease: Multifaceted and Multifunctional. Am J Respir Cell Mol Biol. 2018;59(2):135-6.

34. Todisco T, de Benedictis FM, Iannacci L, Baglioni S, Eslami A, Todisco E, et al. Mild prematurity and respiratory functions. European journal of pediatrics. 1993;152(1):55-8.

35. Barker DJ, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. Bmj. 1991;303(6804):671-5.

36. Lawlor DA, Ebrahim S, Davey Smith G. Association of birth weight with adult lung function: findings from the British Women's Heart and Health Study and a meta-analysis. Thorax. 2005;60(10):851-8.

37. Stocks J, Sonnappa S. Early life influences on the development of chronic obstructive pulmonary disease. Therapeutic advances in respiratory disease. 2013;7(3):161-73.

38. Shaheen SO, Barker DJ, Holgate ST. Do lower respiratory tract infections in early childhood cause chronic obstructive pulmonary disease? Am J Respir Crit Care Med. 1995;151(5):1649-51; discussion 51-2.

39. Kohansal R, Martinez-Camblor P, Agusti A, Buist AS, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. Am J Respir Crit Care Med. 2009;180(1):3-10.

40. Yin P, Jiang CQ, Cheng KK, Lam TH, Lam KH, Miller MR, et al. Passive smoking exposure and risk of COPD among adults in China: the Guangzhou Biobank Cohort Study. Lancet. 2007;370(9589):751-7.

41. Burke H, Leonardi-Bee J, Hashim A, Pine-Abata H, Chen Y, Cook DG, et al. Prenatal and passive smoke exposure and incidence of asthma and wheeze: systematic review and meta-analysis. Pediatrics. 2012;129(4):735-44.

42. Beyer D, Mitfessel H, Gillissen A. Maternal smoking promotes chronic obstructive lung disease in the offspring as adults. European journal of medical research. 2009;14 Suppl 4(Suppl 4):27-31.

43. Gan WQ, FitzGerald JM, Carlsten C, Sadatsafavi M, Brauer M. Associations of ambient air pollution with chronic obstructive pulmonary disease hospitalization and mortality. Am J Respir Crit Care Med. 2013;187(7):721-7.

44. Ezzati M. Indoor air pollution and health in developing countries. Lancet. 2005;366(9480):104-6.

45. Zhou Y, Zou Y, Li X, Chen S, Zhao Z, He F, et al. Lung function and incidence of chronic obstructive pulmonary disease after improved cooking fuels and kitchen ventilation: a 9-year prospective cohort study. PLoS medicine. 2014;11(3):e1001621-e.

46. Gualano RC, Hansen MJ, Vlahos R, Jones JE, Park-Jones RA, Deliyannis G, et al. Cigarette smoke worsens lung inflammation and impairs resolution of influenza infection in mice. Respir Res. 2008;9(1):53-.

47. Saetta M. Airway inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;160(5 Pt 2):S17-20.

48. MacNee W. Pathogenesis of Chronic Obstructive Pulmonary Disease. Proceedings of the American Thoracic Society. 2005;2(4):258-66.

49. Gamble E, Grootendorst DC, Hattotuwa K, O'Shaughnessy T, Ram FS, Qiu Y, et al. Airway mucosal inflammation in COPD is similar in smokers and ex-smokers: a pooled analysis. Eur Respir J. 2007;30(3):467-71.

50. Louhelainen N, Rytilä P, Haahtela T, Kinnula VL, Djukanović R. Persistence of oxidant and protease burden in the airways after smoking cessation. BMC Pulm Med. 2009;9:25.

51. Mercer BA, Kolesnikova N, Sonett J, D'Armiento J. Extracellular regulated kinase/mitogen activated protein kinase is up-regulated in pulmonary emphysema and mediates matrix metalloproteinase-1 induction by cigarette smoke. J Biol Chem. 2004;279(17):17690-6.

52. Lee SH, Goswami S, Grudo A, Song LZ, Bandi V, Goodnight-White S, et al. Antielastin autoimmunity in tobacco smoking-induced emphysema. Nat Med. 2007;13(5):567-9.

53. Hassan F, Xu X, Nuovo G, Killilea DW, Tyrrell J, Da Tan C, et al. Accumulation of metals in GOLD4 COPD lungs is associated with decreased CFTR levels. Respir Res. 2014;15(1):69.

54. Sze MA, Dimitriu PA, Suzuki M, McDonough JE, Campbell JD, Brothers JF, et al. Host Response to the Lung Microbiome in Chronic Obstructive Pulmonary Disease. American journal of respiratory and critical care medicine. 2015;192(4):438-45.

55. Chung KF, Adcock IM. Multifaceted mechanisms in COPD: inflammation, immunity, and tissue repair and destruction. Eur Respir J. 2008;31(6):1334-56.

56. Di Stefano A, Capelli A, Lusuardi M, Balbo P, Vecchio C, Maestrelli P, et al. Severity of airflow limitation is associated with severity of airway inflammation in smokers. Am J Respir Crit Care Med. 1998;158(4):1277-85.

57. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. The New England journal of medicine. 2004;350(26):2645-53.

58. Hiemstra PS, van Wetering S, Stolk J. Neutrophil serine proteinases and defensins in chronic obstructive pulmonary disease: effects on pulmonary epithelium. Eur Respir J. 1998;12(5):1200-8.

59. Dubravec DB, Spriggs DR, Mannick JA, Rodrick ML. Circulating human peripheral blood granulocytes synthesize and secrete tumor necrosis factor alpha. Proc Natl Acad Sci U S A. 1990;87(17):6758-61.

60. Hoenderdos K, Condliffe A. The neutrophil in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2013;48(5):531-9.

61. Barnes PJ. The cytokine network in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2009;41(6):631-8.

62. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2016;138(1):16-27.

63. Burgel PR, Nadel JA. Roles of epidermal growth factor receptor activation in epithelial cell repair and mucin production in airway epithelium. Thorax. 2004;59(11):992-6.

64. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. Science (New York, NY). 2004;303(5663):1532-5.

65. Hoenderdos K, Condliffe A. The Neutrophil in Chronic Obstructive Pulmonary Disease. Too Little, Too Late or Too Much, Too Soon? American Journal of Respiratory Cell and Molecular Biology. 2013;48(5):531-9.

66. Sapey E, Stockley JA, Greenwood H, Ahmad A, Bayley D, Lord JM, et al. Behavioral and structural differences in migrating peripheral neutrophils from patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2011;183(9):1176-86.

67. Linden M, Rasmussen JB, Piitulainen E, Tunek A, Larson M, Tegner H, et al. Airway inflammation in smokers with nonobstructive and obstructive chronic bronchitis. Am Rev Respir Dis. 1993;148(5):1226-32.

68. Stout RD, Jiang C, Matta B, Tietzel I, Watkins SK, Suttles J. Macrophages sequentially change their functional phenotype in response to changes in microenvironmental influences. J Immunol. 2005;175(1):342-9.

69. Pesci A, Balbi B, Majori M, Cacciani G, Bertacco S, Alciato P, et al. Inflammatory cells and mediators in bronchial lavage of patients with chronic obstructive pulmonary disease. Eur Respir J. 1998;12(2):380-6.

70. McKendry RT, Spalluto CM, Burke H, Nicholas B, Cellura D, Al-Shamkhani A, et al. Dysregulation of Anti-viral Function of CD8+T Cells in the COPD Lung: Role of the PD1/PDL1 Axis. Am J Respir Crit Care Med. 2015.

71. Donnelly LE, Barnes PJ. Defective phagocytosis in airways disease. Chest. 2012;141(4):1055-62.

72. Berenson CS, Wrona CT, Grove LJ, Maloney J, Garlipp MA, Wallace PK, et al. Impaired alveolar macrophage response to Haemophilus antigens in chronic obstructive lung disease. Am J Respir Crit Care Med. 2006;174(1):31-40.

73. Shapiro SD. The macrophage in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;160(5 Pt 2):S29-32.

74. Retamales I, Elliott WM, Meshi B, Coxson HO, Pare PD, Sciurba FC, et al. Amplification of inflammation in emphysema and its association with latent adenoviral infection. Am J Respir Crit Care Med. 2001;164(3):469-73.

75. Finkelstein R, Fraser RS, Ghezzo H, Cosio MG. Alveolar inflammation and its relation to emphysema in smokers. Am J Respir Crit Care Med. 1995;152(5 Pt 1):1666-72.

76. Meshi B, Vitalis TZ, Ionescu D, Elliott WM, Liu C, Wang XD, et al. Emphysematous lung destruction by cigarette smoke. The effects of latent adenoviral infection on the lung inflammatory response. Am J Respir Cell Mol Biol. 2002;26(1):52-7.

77. Traves SL, Culpitt SV, Russell RE, Barnes PJ, Donnelly LE. Increased levels of the chemokines GROalpha and MCP-1 in sputum samples from patients with COPD. Thorax. 2002;57(7):590-5.

78. Tomita K, Caramori G, Lim S, Ito K, Hanazawa T, Oates T, et al. Increased p21(CIP1/WAF1) and B cell lymphoma leukemia-x(L) expression and reduced apoptosis in alveolar macrophages from smokers. Am J Respir Crit Care Med. 2002;166(5):724-31.

79. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. Nature reviews Immunology. 2011;11(11):723-37.

80. Barnes PJ. Alveolar Macrophages as Orchestrators of COPD. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2004;1(1):59-70.

81. Fathi M, Johansson A, Lundborg M, Orre L, Skold CM, Camner P. Functional and morphological differences between human alveolar and interstitial macrophages. Experimental and molecular pathology. 2001;70(2):77-82.

82. Duan M, Li WC, Vlahos R, Maxwell MJ, Anderson GP, Hibbs ML. Distinct macrophage subpopulations characterize acute infection and chronic inflammatory lung disease. J Immunol. 2012;189(2):946-55.

83. Laskin DL, Weinberger B, Laskin JD. Functional heterogeneity in liver and lung macrophages. J Leukoc Biol. 2001;70(2):163-70.

84. Haugen TS, Nakstad B, Lyberg T. Heterogeneity of Procoagulant Activity and Cytokine Release in Subpopulations of Alveolar Macrophages and Monocytes. Inflammation. 1999;23(1):15-23.

85. Spiteri MA, Clarke SW, Poulter LW. Isolation of phenotypically and functionally distinct macrophage subpopulations from human bronchoalveolar lavage. Eur Respir J. 1992;5(6):717-26.

86. Shapiro SD, Campbell EJ, Kobayashi DK, Welgus HG. Dexamethasone selectively modulates basal and lipopolysaccharide-induced metalloproteinase and tissue inhibitor of metalloproteinase production by human alveolar macrophages. J Immunol. 1991;146(8):2724-9.

87. Russell RE, Thorley A, Culpitt SV, Dodd S, Donnelly LE, Demattos C, et al. Alveolar macrophage-mediated elastolysis: roles of matrix metalloproteinases, cysteine, and serine proteases. Am J Physiol Lung Cell Mol Physiol. 2002;283(4):L867-73.

88. Caramori G, Romagnoli M, Casolari P, Bellettato C, Casoni G, Boschetto P, et al. Nuclear localisation of p65 in sputum macrophages but not in sputum neutrophils during COPD exacerbations. Thorax. 2003;58(4):348-51.

89. Culpitt SV, Rogers DF, Shah P, De Matos C, Russell RE, Donnelly LE, et al. Impaired inhibition by dexamethasone of cytokine release by alveolar macrophages from patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2003;167(1):24-31.

90. Invernizzi G. Persistence of systemic inflammation in COPD in spite of smoking cessation. Multidisciplinary respiratory medicine. 2011;6(4):210-1.

91. Majo J, Ghezzo H, Cosio MG. Lymphocyte population and apoptosis in the lungs of smokers and their relation to emphysema. Eur Respir J. 2001;17(5):946-53.

92. Fairclough L, Urbanowicz RA, Corne J, Lamb JR. Killer cells in chronic obstructive pulmonary disease. Clinical science (London, England : 1979). 2008;114(8):533-41.

93. Cosio MG, Saetta M, Agusti A. Immunologic aspects of chronic obstructive pulmonary disease. The New England journal of medicine. 2009;360(23):2445-54.

94. Polosukhin VV. Ultrastructural of the bronchial epithelium in chronic inflammation. Ultrastructural pathology. 2001;25(2):119-28.

95. Salazar LM, Herrera AM. Fibrotic response of tissue remodeling in COPD. Lung. 2011;189(2):101-9.

96. Lumsden AB, McLean A, Lamb D. Goblet and Clara cells of human distal airways: evidence for smoking induced changes in their numbers. Thorax. 1984;39(11):844-9.

97. Leopold PL, O'Mahony MJ, Lian XJ, Tilley AE, Harvey BG, Crystal RG. Smoking is associated with shortened airway cilia. PLoS ONE. 2009;4(12):e8157.

98. Biagioli MC, Kaul P, Singh I, Turner RB. The role of oxidative stress in rhinovirus induced elaboration of IL-8 by respiratory epithelial cells. Free radical biology & medicine. 1999;26(3-4):454-62.

99. Kaul P, Biagioli MC, Singh I, Turner RB. Rhinovirus-induced oxidative stress and interleukin-8 elaboration involves p47-phox but is independent of attachment to intercellular adhesion molecule-1 and viral replication. The Journal of infectious diseases. 2000;181(6):1885-90.

100. Gielen V, Johnston SL, Edwards MR. Azithromycin induces anti-viral responses in bronchial epithelial cells. Eur Respir J. 2010;36(3):646-54.

101. Araya J, Cambier S, Markovics JA, Wolters P, Jablons D, Hill A, et al. Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. J Clin Invest. 2007;117(11):3551-62.

102. Chung KF. Inflammatory mediators in chronic obstructive pulmonary disease. Current drug targets Inflammation and allergy. 2005;4(6):619-25.

103. Aaron SD, Angel JB, Lunau M, Wright K, Fex C, Le Saux N, et al. Granulocyte inflammatory markers and airway infection during acute exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001;163(2):349-55.

104. Bhowmik A, Seemungal TAR, Sapsford RJ, Wedzicha JA. Relation of sputum inflammatory markers to symptoms and lung function changes in COPD exacerbations. Thorax. 2000;55(2):114.

105. Berkow RL, Wang D, Larrick JW, Dodson RW, Howard TH. Enhancement of neutrophil superoxide production by preincubation with recombinant human tumor necrosis factor. The Journal of Immunology. 1987;139(11):3783.

106. Yang S, Wang Y, Mei K, Zhang S, Sun X, Ren F, et al. Tumor necrosis factor receptor 2 (TNFR2)·interleukin-17 receptor D (IL-17RD) heteromerization reveals a novel mechanism for NF-κB activation. J Biol Chem. 2015;290(2):861-71.

107. Zarubin T, Han J. Activation and signaling of the p38 MAP kinase pathway. Cell Research. 2005;15(1):11-8.

108. Russell REK, Culpitt SV, DeMatos C, Donnelly L, Smith M, Wiggins J, et al. Release and Activity of Matrix Metalloproteinase-9 and Tissue Inhibitor of Metalloproteinase-1 by Alveolar Macrophages from Patients with Chronic Obstructive Pulmonary Disease. American Journal of Respiratory Cell and Molecular Biology. 2002;26(5):602-9.

109. de Boer WI, Sont JK, van Schadewijk A, Stolk J, van Krieken JH, Hiemstra PS. Monocyte chemoattractant protein 1, interleukin 8, and chronic airways inflammation in COPD. J Pathol. 2000;190(5):619-26.

110. Barnes PJ. Mediators of Chronic Obstructive Pulmonary Disease. Pharmacological Reviews. 2004;56(4):515.

111. Takizawa H, Tanaka M, Takami K, Ohtoshi T, Ito K, Satoh M, et al. Increased Expression of Transforming Growth Factor- β 1 in Small Airway Epithelium from Tobacco Smokers and Patients with

Chronic Obstructive Pulmonary Disease (COPD). American Journal of Respiratory and Critical Care Medicine. 2001;163(6):1476-83.

112. Vignola AM, Chanez P, Chiappara G, Merendino A, Pace E, Rizzo A, et al. Transforming Growth Factor- β Expression in Mucosal Biopsies in Asthma and Chronic Bronchitis. American Journal of Respiratory and Critical Care Medicine. 1997;156(2):591-9.

113. Li MO, Wan YY, Sanjabi S, Robertson A-KL, Flavell RA. Transforming Growth Factor-B Regulation of Immune Responses. Annual review of immunology. 2006;24(1):99-146.

114. Willis BC, Borok Z. TGF-β-induced EMT: mechanisms and implications for fibrotic lung disease. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2007;293(3):L525-L34.

115. Milara J, Peiró T, Serrano A, Cortijo J. Epithelial to mesenchymal transition is increased in patients with COPD and induced by cigarette smoke. Thorax. 2013;68(5):410.

116. Domej W, Oettl K, Renner W. Oxidative stress and free radicals in COPD--implications and relevance for treatment. Int J Chron Obstruct Pulmon Dis. 2014;9:1207-24.

117. Zinellu E, Zinellu A, Fois AG, Carru C, Pirina P. Circulating biomarkers of oxidative stress in chronic obstructive pulmonary disease: a systematic review. Respir Res. 2016;17(1):150-.

118. Malhotra D, Thimmulappa R, Vij N, Navas-Acien A, Sussan T, Merali S, et al. Heightened endoplasmic reticulum stress in the lungs of patients with chronic obstructive pulmonary disease: the role of Nrf2-regulated proteasomal activity. Am J Respir Crit Care Med. 2009;180(12):1196-207.

119. Dagouassat M, Gagliolo J-M, Chrusciel S, Bourin M-C, Duprez C, Caramelle P, et al. The Cyclooxygenase-2–Prostaglandin E2 Pathway Maintains Senescence of Chronic Obstructive Pulmonary Disease Fibroblasts. American Journal of Respiratory and Critical Care Medicine. 2013;187(7):703-14.

120. Lin C-C, Lee IT, Yang Y-L, Lee C-W, Kou YR, Yang C-M. Induction of COX-2/PGE2/IL-6 is crucial for cigarette smoke extract-induced airway inflammation: Role of TLR4-dependent NADPH oxidase activation. Free Radical Biology and Medicine. 2010;48(2):240-54.

121. Sarkar P, Hayes BE. Induction of COX-2 by acrolein in rat lung epithelial cells. Molecular and cellular biochemistry. 2007;301(1-2):191-9.

122. Stockley RA. Neutrophils and protease/antiprotease imbalance. Am J Respir Crit Care Med. 1999;160(5 Pt 2):S49-52.

123. Johnson SR. Untangling the protease web in COPD: metalloproteinases in the silent zone. Thorax. 2016;71(2):105-6.

124. Ofir D, Laveneziana P, Webb KA, Lam Y-M, O'Donnell DE. Mechanisms of Dyspnea during Cycle Exercise in Symptomatic Patients with GOLD Stage I Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. 2008;177(6):622-9.

125. Ofir D, Laveneziana P, Webb KA, Lam YM, O'Donnell DE. Mechanisms of dyspnea during cycle exercise in symptomatic patients with GOLD stage I chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2008;177(6):622-9.

126. Goldklang M, Stockley R. Pathophysiology of Emphysema and Implications. Chronic obstructive pulmonary diseases (Miami, Fla). 2016;3(1):454-8.

127. Barbera JA, Roca J, Ferrer A, Felez MA, Diaz O, Roger N, et al. Mechanisms of worsening gas exchange during acute exacerbations of chronic obstructive pulmonary disease. Eur Respir J. 1997;10(6):1285-91.

128. Magee F, Wright JL, Wiggs BR, Pare PD, Hogg JC. Pulmonary vascular structure and function in chronic obstructive pulmonary disease. Thorax. 1988;43(3):183-9.

129. MacNee W. Pathophysiology of cor pulmonale in chronic obstructive pulmonary disease. Part two. Am J Respir Crit Care Med. 1994;150(4):1158-68.

130. Miller J, Edwards LD, Agusti A, Bakke P, Calverley PM, Celli B, et al. Comorbidity, systemic inflammation and outcomes in the ECLIPSE cohort. Respiratory medicine. 2013;107(9):1376-84.

131. Agusti AG. Systemic effects of chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2005;2(4):367-70; discussion 71-2.

132. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.

133. van Dijk W, Tan W, Li P, Guo B, Li S, Benedetti A, et al. Clinical relevance of fixed ratio vs lower limit of normal of FEV1/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. Ann Fam Med. 2015;13(1):41-8.

134. Guder G, Brenner S, Angermann CE, Ertl G, Held M, Sachs AP, et al. "GOLD or lower limit of normal definition? A comparison with expert-based diagnosis of chronic obstructive pulmonary disease in a prospective cohort-study". Respir Res. 2012;13(1):13.

135. Vaz Fragoso CA, McAvay G, Van Ness PH, Casaburi R, Jensen RL, MacIntyre N, et al. Phenotype of normal spirometry in an aging population. American journal of respiratory and critical care medicine. 2015;192(7):817-25.

136. Anthonisen NR, Wright EC, Hodgkin JE. Prognosis in chronic obstructive pulmonary disease. Am Rev Respir Dis. 1986;133(1):14-20.

137. Oga T, Nishimura K, Tsukino M, Sato S, Hajiro T. Analysis of the factors related to mortality in chronic obstructive pulmonary disease: role of exercise capacity and health status. Am J Respir Crit Care Med. 2003;167(4):544-9.

138. Martinez FJ, Foster G, Curtis JL, Criner G, Weinmann G, Fishman A, et al. Predictors of mortality in patients with emphysema and severe airflow obstruction. American journal of respiratory and critical care medicine. 2006;173(12):1326-34.

139. Boutou AK, Shrikrishna D, Tanner RJ, Smith C, Kelly JL, Ward SP, et al. Lung function indices for predicting mortality in COPD. Eur Respir J. 2013;42(3):616-25.

140. Jones PW, Bosh TK. Quality of life changes in COPD patients treated with salmeterol. Am J Respir Crit Care Med. 1997;155(4):1283-9.

141. Johns DP, Walters JAE, Walters EH. Diagnosis and early detection of COPD using spirometry. J Thorac Dis. 2014;6(11):1557-69.

142. Burgel PR, Bourdin A, Chanez P, Chabot F, Chaouat A, Chinet T, et al. Update on the roles of distal airways in COPD. European respiratory review : an official journal of the European Respiratory Society. 2011;20(119):7-22.

143. A clinical practice guideline for treating tobacco use and dependence: A US Public Health Service report. The Tobacco Use and Dependence Clinical Practice Guideline Panel, Staff, and Consortium Representatives. Jama. 2000;283(24):3244-54.

144. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, et al. An Official American Thoracic Society/European Respiratory Society Statement: Key Concepts and Advances in Pulmonary Rehabilitation. American Journal of Respiratory and Critical Care Medicine. 2013;188(8):e13-e64.

145. Bourne S, DeVos R, North M, Chauhan A, Green B, Brown T, et al. Online versus face-to-face pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: randomised controlled trial. BMJ Open. 2017;7(7):e014580.

146. Bekkat-Berkani R, Wilkinson T, Buchy P, Dos Santos G, Stefanidis D, Devaster JM, et al. Seasonal influenza vaccination in patients with COPD: a systematic literature review. BMC Pulm Med. 2017;17(1):79.

147. Tomczyk S, Bennett NM, Stoecker C, Gierke R, Moore MR, Whitney CG, et al. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine among adults aged ≥65 years: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morbidity and mortality weekly report. 2014;63(37):822-5.

148. Suissa S, Patenaude V, Lapi F, Ernst P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. Thorax. 2013;68(11):1029.

149. Bafadhel M, McKenna S, Terry S, Mistry V, Pancholi M, Venge P, et al. Blood eosinophils to direct corticosteroid treatment of exacerbations of chronic obstructive pulmonary disease: a randomized placebo-controlled trial. Am J Respir Crit Care Med. 2012;186(1):48-55.

150. Kerkhof M, Sonnappa S, Postma DS, Brusselle G, Agustí A, Anzueto A, et al. Blood eosinophil count and exacerbation risk in patients with COPD. Eur Respir J. 2017;50(1):1700761.

151. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. New England Journal of Medicine. 2010;363(12):1128-38.

152. Singh D, Edwards L, Tal-Singer R, Rennard S. Sputum neutrophils as a biomarker in COPD: findings from the ECLIPSE study. Respir Res. 2010;11(1):77.

153. Culpitt SV, Maziak W, Loukidis S, Nightingale JA, Matthews JL, Barnes PJ. Effect of high dose inhaled steroid on cells, cytokines, and proteases in induced sputum in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;160(5 Pt 1):1635-9.

154. Ito K, Hanazawa T, Tomita K, Barnes PJ, Adcock IM. Oxidative stress reduces histone deacetylase 2 activity and enhances IL-8 gene expression: Role of tyrosine nitration. Biochemical and biophysical research communications. 2004;315(1):240-5.

155. Ernst P, Saad N, Suissa S. Inhaled corticosteroids in COPD: the clinical evidence. Eur Respir J. 2015;45(2):525-37.

156. Rennard SI, Dale DC, Donohue JF, Kanniess F, Magnussen H, Sutherland ER, et al. CXCR2 Antagonist MK-7123. A Phase 2 Proof-of-Concept Trial for Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2015;191(9):1001-11.

157. MacNee W, Allan RJ, Jones I, De Salvo MC, Tan LF. Efficacy and safety of the oral p38 inhibitor PH-797804 in chronic obstructive pulmonary disease: a randomised clinical trial. Thorax. 2013;68(8):738-45.

158. Watz H, Barnacle H, Hartley BF, Chan R. Efficacy and safety of the p38 MAPK inhibitor losmapimod for patients with chronic obstructive pulmonary disease: a randomised, double-blind, placebo-controlled trial. The Lancet Respiratory medicine. 2014;2(1):63-72.

159. Aaron SD, Vandemheen KL, Maltais F, Field SK, Sin DD, Bourbeau J, et al. TNFα antagonists for acute exacerbations of COPD: a randomised double-blind controlled trial. Thorax. 2013;68(2):142.

160. Calverley PMA, Sethi S, Dawson M, Ward CK, Finch DK, Penney M, et al. A randomised, placebo-controlled trial of anti-interleukin-1 receptor 1 monoclonal antibody MEDI8968 in chronic obstructive pulmonary disease. Respir Res. 2017;18(1):153-.

161. Calverley PM, Rabe KF, Goehring UM, Kristiansen S, Fabbri LM, Martinez FJ. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. Lancet. 2009;374(9691):685-94.

162. Wedzicha JA, Calverley PM, Rabe KF. Roflumilast: a review of its use in the treatment of COPD. International journal of chronic obstructive pulmonary disease. 2016;11:81-90.

163. Rennard SI, Calverley PM, Goehring UM, Bredenbröker D, Martinez FJ. Reduction of exacerbations by the PDE4 inhibitor roflumilast--the importance of defining different subsets of patients with COPD. Respir Res. 2011;12(1):18.

164. Papi A, Romagnoli M, Baraldo S, Braccioni F, Guzzinati I, Saetta M, et al. Partial reversibility of airflow limitation and increased exhaled NO and sputum eosinophilia in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2000;162(5):1773-7.

165. Cosio BG, Soriano JB, López-Campos JL, Calle-Rubio M, Soler-Cataluna JJ, de-Torres JP, et al. Defining the Asthma-COPD Overlap Syndrome in a COPD Cohort. Chest. 2016;149(1):45-52.

166. Singh D, Kolsum U, Brightling CE, Locantore N, Agusti A, Tal-Singer R. Eosinophilic inflammation in COPD: prevalence and clinical characteristics. Eur Respir J. 2014;44(6):1697.

167. Siva R, Green RH, Brightling CE, Shelley M, Hargadon B, McKenna S, et al. Eosinophilic airway inflammation and exacerbations of COPD: a randomised controlled trial. Eur Respir J. 2007;29(5):906.

168. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P, et al. Acute exacerbations of chronic obstructive pulmonary disease: identification of biologic clusters and their biomarkers. Am J Respir Crit Care Med. 2011;184(6):662-71.

169. Negewo NA, McDonald VM, Baines KJ, Wark PA, Simpson JL, Jones PW, et al. Peripheral blood eosinophils: a surrogate marker for airway eosinophilia in stable COPD. Int J Chron Obstruct Pulmon Dis. 2016;11:1495-504.

170. Turato G, Semenzato U, Bazzan E, Biondini D, Tinè M, Torrecilla N, et al. Blood Eosinophilia Neither Reflects Tissue Eosinophils nor Worsens Clinical Outcomes in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2018;197(9):1216-9.

171. Eltboli O, Mistry V, Barker B, Brightling CE. Relationship between blood and bronchial submucosal eosinophilia and reticular basement membrane thickening in chronic obstructive pulmonary disease. Respirology. 2015;20(4):667-70.

172. Landis SH, Suruki R, Hilton E, Compton C, Galwey NW. Stability of Blood Eosinophil Count in Patients with COPD in the UK Clinical Practice Research Datalink. Copd. 2017;14(4):382-8.

173. Bafadhel M, Saha S, Siva R, McCormick M, Monteiro W, Rugman P, et al. Sputum IL-5 concentration is associated with a sputum eosinophilia and attenuated by corticosteroid therapy in COPD. Respiration. 2009;78(3):256-62.

174. Costa C, Rufino R, Traves SL, Lapa ESJR, Barnes PJ, Donnelly LE. CXCR3 and CCR5 chemokines in induced sputum from patients with COPD. Chest. 2008;133(1):26-33.

175. Kearley J, Silver JS, Sanden C, Liu Z, Berlin AA, White N, et al. Cigarette smoke silences innate lymphoid cell function and facilitates an exacerbated type I interleukin-33-dependent response to infection. Immunity. 2015;42(3):566-79.

176. Tworek D, Majewski S, Szewczyk K, Kiszałkiewicz J, Kurmanowska Z, Górski P, et al. The association between airway eosinophilic inflammation and IL-33 in stable non-atopic COPD. Respir Res. 2018;19(1):108.

177. Ying S, O'Connor B, Ratoff J, Meng Q, Fang C, Cousins D, et al. Expression and cellular provenance of thymic stromal lymphopoietin and chemokines in patients with severe asthma and chronic obstructive pulmonary disease. J Immunol. 2008;181(4):2790-8.

178. Pavord ID, Chanez P, Criner GJ, Kerstjens HAM, Korn S, Lugogo N, et al. Mepolizumab for Eosinophilic Chronic Obstructive Pulmonary Disease. The New England journal of medicine. 2017;377(17):1613-29.

179. Brightling CE, Bleecker ER, Panettieri RA, Jr., Bafadhel M, She D, Ward CK, et al. Benralizumab for chronic obstructive pulmonary disease and sputum eosinophilia: a randomised, double-blind, placebo-controlled, phase 2a study. The Lancet Respiratory medicine. 2014;2(11):891-901.

180. Mesnil C, Raulier S, Paulissen G, Xiao X, Birrell MA, Pirottin D, et al. Lung-resident eosinophils represent a distinct regulatory eosinophil subset. J Clin Invest. 2016;126(9):3279-95.

181. Parker CM, Voduc N, Aaron SD, Webb KA, O'Donnell DE. Physiological changes during symptom recovery from moderate exacerbations of COPD. Eur Respir J. 2005;26(3):420-8.

182. Seemungal TA, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998;157(5 Pt 1):1418-22.

183. Burrows B, Bloom JW, Traver GA, Cline MG. The course and prognosis of different forms of chronic airways obstruction in a sample from the general population. The New England journal of medicine. 1987;317(21):1309-14.

184. Donaldson G, Seemungal T, Bhowmik A, Wedzicha J. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 2002;57(10):847-52.

185. Ball P. Epidemiology and treatment of chronic bronchitis and its exacerbations. Chest. 1995;108(2 Suppl):43s-52s.

186. Sunyer J, Saez M, Murillo C, Castellsague J, Martinez F, Anto JM. Air pollution and emergency room admissions for chronic obstructive pulmonary disease: a 5-year study. American journal of epidemiology. 1993;137(7):701-5.

187. Connors AF, Jr., Dawson NV, Thomas C, Harrell FE, Jr., Desbiens N, Fulkerson WJ, et al. Outcomes following acute exacerbation of severe chronic obstructive lung disease. The SUPPORT investigators (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments). Am J Respir Crit Care Med. 1996;154(4 Pt 1):959-67.

188. Eccles R. An explanation for the seasonality of acute upper respiratory tract viral infections. Acta oto-laryngologica. 2002;122(2):183-91.

189. Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. Eur Respir J Suppl. 2003;41:46s-53s.

190. Roberts CM, Lowe D, Bucknall CE, Ryland I, Kelly Y, Pearson MG. Clinical audit indicators of outcome following admission to hospital with acute exacerbation of chronic obstructive pulmonary disease. Thorax. 2002;57(2):137-41.

191. Garcia-Aymerich J, Monso E, Marrades RM, Escarrabill J, Felez MA, Sunyer J, et al. Risk factors for hospitalization for a chronic obstructive pulmonary disease exacerbation. EFRAM study. Am J Respir Crit Care Med. 2001;164(6):1002-7.

192. Soler-Cataluna JJ, Martinez-Garcia MA, Roman Sanchez P, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. Thorax. 2005;60(11):925-31.

193. Hurst JR, Donaldson GC, Perera WR, Wilkinson TM, Bilello JA, Hagan GW, et al. Use of plasma biomarkers at exacerbation of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2006;174(8):867-74.

194. Zhu A, Ge D, Zhang J, Teng Y, Yuan C, Huang M, et al. Sputum myeloperoxidase in chronic obstructive pulmonary disease. European journal of medical research. 2014;19:12.

195. O'Neil SE, Lundback B, Lotvall J. Proteomics in asthma and COPD phenotypes and endotypes for biomarker discovery and improved understanding of disease entities. J Proteomics. 2011;75(1):192-201.

196. Telenga ED, Hoffmann RF, Ruben tK, Hoonhorst SJ, Willemse BW, van Oosterhout AJ, et al. Untargeted lipidomic analysis in chronic obstructive pulmonary disease. Uncovering sphingolipids. Am J Respir Crit Care Med. 2014;190(2):155-64.

197. Hodgkin PD, Rush J, Gett AV, Bartell G, Hasbold J. The logic of intercellular communication in the immune system. Immunology and cell biology. 1998;76(5):448-53.

198. Roy S, Hochberg FH, Jones PS. Extracellular vesicles: the growth as diagnostics and therapeutics; a survey. Journal of extracellular vesicles. 2018;7(1):1438720-.

199. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654-9.

200. Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, et al. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. Nat Commun. 2011;2:180.

201. Robbins PD, Morelli AE. Regulation of immune responses by extracellular vesicles. Nat Rev Immunol. 2014;14(3):195-208.

202. Turpin D, Truchetet ME, Faustin B, Augusto JF, Contin-Bordes C, Brisson A, et al. Role of extracellular vesicles in autoimmune diseases. Autoimmunity reviews. 2016;15(2):174-83.

203. Buzas EI, Gyorgy B, Nagy G, Falus A, Gay S. Emerging role of extracellular vesicles in inflammatory diseases. Nat Rev Rheumatol. 2014;10(6):356-64.

204. Hosseini HM, Fooladi AA, Nourani MR, Ghanezadeh F. The role of exosomes in infectious diseases. Inflamm Allergy Drug Targets. 2013;12(1):29-37.

205. Van Giau V, An SSA. Emergence of exosomal miRNAs as a diagnostic biomarker for Alzheimer's disease. Journal of the Neurological Sciences. 2016;360:141-52.

206. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer. 1972;26(4):239-57.

207. Thery C, Boussac M, Veron P, Ricciardi-Castagnoli P, Raposo G, Garin J, et al. Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. J Immunol. 2001;166(12):7309-18.

208. Holmgren L, Szeles A, Rajnavolgyi E, Folkman J, Klein G, Ernberg I, et al. Horizontal transfer of DNA by the uptake of apoptotic bodies. Blood. 1999;93(11):3956-63.

209. Bergsmedh A, Szeles A, Henriksson M, Bratt A, Folkman MJ, Spetz AL, et al. Horizontal transfer of oncogenes by uptake of apoptotic bodies. Proc Natl Acad Sci U S A. 2001;98(11):6407-11.

210. Bellone M, Iezzi G, Rovere P, Galati G, Ronchetti A, Protti MP, et al. Processing of engulfed apoptotic bodies yields T cell epitopes. J Immunol. 1997;159(11):5391-9.

211. Cocca BA, Cline AM, Radic MZ. Blebs and apoptotic bodies are B cell autoantigens. J Immunol. 2002;169(1):159-66.

212. Holme PA, Solum NO, Brosstad F, Roger M, Abdelnoor M. Demonstration of platelet-derived microvesicles in blood from patients with activated coagulation and fibrinolysis using a filtration technique and western blotting. Thromb Haemost. 1994;72(5):666-71.

213. Hess C, Sadallah S, Hefti A, Landmann R, Schifferli JA. Ectosomes released by human neutrophils are specialized functional units. J Immunol. 1999;163(8):4564-73.

214. Cocucci E, Racchetti G, Meldolesi J. Shedding microvesicles: artefacts no more. Trends Cell Biol. 2009;19(2):43-51.

215. Gyorgy B, Szabo TG, Pasztoi M, Pal Z, Misjak P, Aradi B, et al. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. Cell Mol Life Sci. 2011;68(16):2667-88.

216. Burger D, Schock S, Thompson CS, Montezano AC, Hakim AM, Touyz RM. Microparticles: biomarkers and beyond. Clinical science (London, England : 1979). 2013;124(7):423-41.

217. Badimon L, Suades R, Fuentes E, Palomo I, Padró T. Role of Platelet-Derived Microvesicles As Crosstalk Mediators in Atherothrombosis and Future Pharmacology Targets: A Link between Inflammation, Atherosclerosis, and Thrombosis. Frontiers in Pharmacology. 2016;7:293.

218. Nielsen MH, Beck-Nielsen H, Andersen MN, Handberg A. A flow cytometric method for characterization of circulating cell-derived microparticles in plasma. Journal of Extracellular Vesicles. 2014;3:10.3402/jev.v3.20795.

219. Leroyer AS, Tedgui A, Boulanger CM. Role of microparticles in atherothrombosis. Journal of internal medicine. 2008;263(5):528-37.

220. MacKenzie A, Wilson HL, Kiss-Toth E, Dower SK, North RA, Surprenant A. Rapid secretion of interleukin-1beta by microvesicle shedding. Immunity. 2001;15(5):825-35.

221. Antonyak MA, Li B, Boroughs LK, Johnson JL, Druso JE, Bryant KL, et al. Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. Proc Natl Acad Sci U S A. 2011;108(12):4852-7.

222. Fernandez-Messina L, Gutierrez-Vazquez C, Rivas-Garcia E, Sanchez-Madrid F, de la Fuente H. Immunomodulatory role of microRNAs transferred by extracellular vesicles. Biology of the cell. 2015;107(3):61-77.

223. Thery C, Ostrowski M, Segura E. Membrane vesicles as conveyors of immune responses. Nat Rev Immunol. 2009;9(8):581-93.

224. Harding C, Heuser J, Stahl P. Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. J Cell Biol. 1983;97(2):329-39.

225. Pan BT, Teng K, Wu C, Adam M, Johnstone RM. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. J Cell Biol. 1985;101(3):942-8.

226. Raposo G, Nijman HW, Stoorvogel W, Liejendekker R, Harding CV, Melief CJ, et al. B lymphocytes secrete antigen-presenting vesicles. J Exp Med. 1996;183(3):1161-72.

227. Thery C, Regnault A, Garin J, Wolfers J, Zitvogel L, Ricciardi-Castagnoli P, et al. Molecular characterization of dendritic cell-derived exosomes. Selective accumulation of the heat shock protein hsc73. J Cell Biol. 1999;147(3):599-610.

228. Thery C, Duban L, Segura E, Veron P, Lantz O, Amigorena S. Indirect activation of naive CD4+ T cells by dendritic cell-derived exosomes. Nat Immunol. 2002;3(12):1156-62.

229. Wolfers J, Lozier A, Raposo G, Regnault A, Thery C, Masurier C, et al. Tumor-derived exosomes are a source of shared tumor rejection antigens for CTL cross-priming. Nat Med. 2001;7(3):297-303.

230. Karlsson M, Lundin S, Dahlgren U, Kahu H, Pettersson I, Telemo E. "Tolerosomes" are produced by intestinal epithelial cells. Eur J Immunol. 2001;31(10):2892-900.

231. Andreola G, Rivoltini L, Castelli C, Huber V, Perego P, Deho P, et al. Induction of lymphocyte apoptosis by tumor cell secretion of FasL-bearing microvesicles. J Exp Med. 2002;195(10):1303-16.

232. Admyre C, Grunewald J, Thyberg J, Gripenback S, Tornling G, Eklund A, et al. Exosomes with major histocompatibility complex class II and co-stimulatory molecules are present in human BAL fluid. Eur Respir J. 2003;22(4):578-83.

233. Kulshreshtha A, Ahmad T, Agrawal A, Ghosh B. Proinflammatory role of epithelial cell-derived exosomes in allergic airway inflammation. J Allergy Clin Immunol. 2013;131(4):1194-203, 203 e1-14.

234. Kesimer M, Scull M, Brighton B, DeMaria G, Burns K, O'Neal W, et al. Characterization of exosome-like vesicles released from human tracheobronchial ciliated epithelium: a possible role in innate defense. Faseb J. 2009;23(6):1858-68.

235. Aliotta JM, Pereira M, Sears EH, Dooner MS, Wen S, Goldberg LR, et al. Lung-derived exosome uptake into and epigenetic modulation of marrow progenitor/stem and differentiated cells. J Extracell Vesicles. 2015;4:26166.

236. Moon H-G, Kim S-H, Gao J, Quan T, Qin Z, Osorio JC, et al. CCN1 secretion and cleavage regulate the lung epithelial cell functions after cigarette smoke. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2014;307(4):L326-L37.

237. Fujita Y, Araya J, Ito S, Kobayashi K, Kosaka N, Yoshioka Y, et al. Suppression of autophagy by extracellular vesicles promotes myofibroblast differentiation in COPD pathogenesis. J Extracell Vesicles. 2015;4:28388.

238. Li CJ, Liu Y, Chen Y, Yu D, Williams KJ, Liu ML. Novel proteolytic microvesicles released from human macrophages after exposure to tobacco smoke. Am J Pathol. 2013;182(5):1552-62.

239. Li M, Yu D, Williams KJ, Liu ML. Tobacco smoke induces the generation of procoagulant microvesicles from human monocytes/macrophages. Arterioscler Thromb Vasc Biol. 2010;30(9):1818-24.

240. Cordazzo C, Petrini S, Neri T, Lombardi S, Carmazzi Y, Pedrinelli R, et al. Rapid shedding of proinflammatory microparticles by human mononuclear cells exposed to cigarette smoke is dependent on Ca2+ mobilization. Inflamm Res. 2014;63(7):539-47.

241. Soni S, Wilson MR, Dea KP, Yoshida M, Katbeh U, Woods SJ, et al. Alveolar macrophage-derived microvesicles mediate acute lung injury. Thorax. 2016;71(11):1020.

242. Bourdonnay E, Zaslona Z, Penke LRK, Speth JM, Schneider DJ, Przybranowski S, et al. Transcellular delivery of vesicular SOCS proteins from macrophages to epithelial cells blunts inflammatory signaling. J Exp Med. 2015;212(5):729-42.

243. Schneider DJ, Smith KA, Latuszek CE, Wilke CA, Lyons DM, Penke LR, et al. Alveolar macrophage-derived extracellular vesicles inhibit endosomal fusion of influenza virus. The EMBO Journal. 2020;39(16):e105057.

244. Héliot A, Landkocz Y, Roy Saint-Georges F, Gosset P, Billet S, Shirali P, et al. Smoker extracellular vesicles influence status of human bronchial epithelial cells. International Journal of Hygiene and Environmental Health. 2017;220(2, Part B):445-54.

245. Eltom S, Dale N, Raemdonck KRG, Stevenson CS, Snelgrove RJ, Sacitharan PK, et al. Respiratory infections cause the release of extracellular vesicles: implications in exacerbation of asthma/COPD. PLoS ONE. 2014;9(6):e101087-e.

246. Kim HJ, Kim Y-S, Kim K-H, Choi J-P, Kim Y-K, Yun S, et al. The microbiome of the lung and its extracellular vesicles in nonsmokers, healthy smokers and COPD patients. Exp Mol Med. 2017;49(4):e316-e.

247. Lacedonia D, Carpagnano GE, Trotta T, Palladino GP, Panaro MA, Zoppo LD, et al. Microparticles in sputum of COPD patients: a potential biomarker of the disease? International journal of chronic obstructive pulmonary disease. 2016;11:527-33.

248. Gordon C, Gudi K, Krause A, Sackrowitz R, Harvey BG, Strulovici-Barel Y, et al. Circulating endothelial microparticles as a measure of early lung destruction in cigarette smokers. Am J Respir Crit Care Med. 2011;184(2):224-32.

249. Soni S, Wilson MR, O'Dea KP, Yoshida M, Katbeh U, Woods SJ, et al. Alveolar macrophage-derived microvesicles mediate acute lung injury. Thorax. 2016;71(11):1020-9.

250. Thomashow MA, Shimbo D, Parikh MA, Hoffman EA, Vogel-Claussen J, Hueper K, et al. Endothelial microparticles in mild chronic obstructive pulmonary disease and emphysema. The Multi-Ethnic Study of Atherosclerosis Chronic Obstructive Pulmonary Disease study. Am J Respir Crit Care Med. 2013;188(1):60-8.

251. Strulovici-Barel Y, Staudt MR, Krause A, Gordon C, Tilley AE, Harvey BG, et al. Persistence of circulating endothelial microparticles in COPD despite smoking cessation. Thorax. 2016.

252. Serban KA, Rezania S, Petrusca DN, Poirier C, Cao D, Justice MJ, et al. Structural and functional characterization of endothelial microparticles released by cigarette smoke. Scientific Reports. 2016;6(1):31596.

253. Liu H, Ding L, Zhang Y, Ni S. Circulating endothelial microparticles involved in lung function decline in a rat exposed in cigarette smoke maybe from apoptotic pulmonary capillary endothelial cells. J Thorac Dis. 2014;6(6):649-55.

254. Tan DBA, Armitage J, Teo TH, Ong NE, Shin H, Moodley YP. Elevated levels of circulating exosome in COPD patients are associated with systemic inflammation. Respiratory medicine. 2017;132:261-4.

255. He S, Chen D, Hu M, Zhang L, Liu C, Traini D, et al. Bronchial epithelial cell extracellular vesicles ameliorate epithelial–mesenchymal transition in COPD pathogenesis by alleviating M2 macrophage polarization. Nanomedicine: Nanotechnology, Biology and Medicine. 2019;18:259-71.

256. Chen Y-WR, Leung JM, Sin DD. A Systematic Review of Diagnostic Biomarkers of COPD Exacerbation. PLoS ONE. 2016;11(7):e0158843.

257. Makiguchi T, Yamada M, Yoshioka Y, Sugiura H, Koarai A, Chiba S, et al. Serum extracellular vesicular miR-21-5p is a predictor of the prognosis in idiopathic pulmonary fibrosis. Respir Res. 2016;17(1):110-.

258. Yoshioka Y, Kosaka N, Konishi Y, Ohta H, Okamoto H, Sonoda H, et al. Ultra-sensitive liquid biopsy of circulating extracellular vesicles using ExoScreen. Nature Communications. 2014;5:3591.

259. Fujita Y, Yoshioka Y, Ochiya T. Extracellular vesicle transfer of cancer pathogenic components. Cancer science. 2016;107(4):385-90.

260. Takahashi T, Kobayashi S, Fujino N, Suzuki T, Ota C, He M, et al. Increased circulating endothelial microparticles in COPD patients: a potential biomarker for COPD exacerbation susceptibility. Thorax. 2012;67(12):1067-74.

261. De Smet EG, Mestdagh P, Vandesompele J, Brusselle GG, Bracke KR. Non-coding RNAs in the pathogenesis of COPD. Thorax. 2015;70(8):782-91.

262. Levanen B, Bhakta NR, Paredes PT, Barbeau R, Hiltbrunner S, Pollack JL, et al. Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients. Journal of Allergy and Clinical Immunology. 2013;131(3):894-+.

263. Kishore A, Navratilova Z, Kolek V, Novosadova E, Čépe K, du Bois RM, et al. Expression analysis of extracellular microRNA in bronchoalveolar lavage fluid from patients with pulmonary sarcoidosis. Respirology. 2018;23(12):1166-72.

264. Stolk J, Broekman W, Mauad T, Zwaginga JJ, Roelofs H, Fibbe WE, et al. A phase I study for intravenous autologous mesenchymal stromal cell administration to patients with severe emphysema. QJM : monthly journal of the Association of Physicians. 2016;109(5):331-6.

265. Weiss DJ, Casaburi R, Flannery R, LeRoux-Williams M, Tashkin DP. A placebo-controlled, randomized trial of mesenchymal stem cells in COPD. Chest. 2013;143(6):1590-8.

266. Broekman W, Khedoe PPSJ, Schepers K, Roelofs H, Stolk J, Hiemstra PS. Mesenchymal stromal cells: a novel therapy for the treatment of chronic obstructive pulmonary disease? Thorax. 2018;73(6):565.

267. Porro C, Lepore S, Trotta T, Castellani S, Ratclif L, Battaglino A, et al. Isolation and characterization of microparticles in sputum from cystic fibrosis patients. Respir Res. 2010;11(1):94-.

268. Huang L, Ma W, Ma Y, Feng D, Chen H, Cai B. Exosomes in mesenchymal stem cells, a new therapeutic strategy for cardiovascular diseases? International journal of biological sciences. 2015;11(2):238-45.

269. Escudier B, Dorval T, Chaput N, André F, Caby M-P, Novault S, et al. Vaccination of metastatic melanoma patients with autologous dendritic cell (DC) derived-exosomes: results of thefirst phase I clinical trial. Journal of Translational Medicine. 2005;3(1):10.

270. Morse MA, Garst J, Osada T, Khan S, Hobeika A, Clay TM, et al. A phase I study of dexosome immunotherapy in patients with advanced non-small cell lung cancer. Journal of translational medicine. 2005;3(1):9-.

271. Besse B, Charrier M, Lapierre V, Dansin E, Lantz O, Planchard D, et al. Dendritic cell-derived exosomes as maintenance immunotherapy after first line chemotherapy in NSCLC. Oncoimmunology. 2015;5(4):e1071008-e.

272. Kordelas L, Rebmann V, Ludwig AK, Radtke S, Ruesing J, Doeppner TR, et al. MSC-derived exosomes: a novel tool to treat therapy-refractory graft-versus-host disease. Leukemia. 2014;28(4):970-3.

273. Mu J, Zhuang X, Wang Q, Jiang H, Deng ZB, Wang B, et al. Interspecies communication between plant and mouse gut host cells through edible plant derived exosome-like nanoparticles. Molecular nutrition & food research. 2014;58(7):1561-73.

274. Wang B, Zhuang X, Deng ZB, Jiang H, Mu J, Wang Q, et al. Targeted drug delivery to intestinal macrophages by bioactive nanovesicles released from grapefruit. Molecular therapy : the journal of the American Society of Gene Therapy. 2014;22(3):522-34.

275. Zhang D, Lee H, Zhu Z, Minhas JK, Jin Y. Enrichment of selective miRNAs in exosomes and delivery of exosomal miRNAs in vitro and in vivo. Am J Physiol Lung Cell Mol Physiol. 2017;312(1):L110-l21.

276. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. Proceedings of the National Academy of Sciences. 2014;111(41):14888.

277. Wiklander OPB, Brennan MÁ, Lötvall J, Breakefield XO, El Andaloussi S. Advances in therapeutic applications of extracellular vesicles. Science translational medicine. 2019;11(492):eaav8521.

278. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. Journal of Extracellular Vesicles. 2018;7(1):1535750.

279. Thery C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006;Chapter 3:Unit 3 22.

280. Sokolova V, Ludwig AK, Hornung S, Rotan O, Horn PA, Epple M, et al. Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. Colloids and surfaces B, Biointerfaces. 2011;87(1):146-50.

281. Ismail N, Wang Y, Dakhlallah D, Moldovan L, Agarwal K, Batte K, et al. Macrophage microvesicles induce macrophage differentiation and miR-223 transfer. Blood. 2013;121(6):984-95.

282. de Menezes-Neto A, Saez MJ, Lozano-Ramos I, Segui-Barber J, Martin-Jaular L, Ullate JM, et al. Size-exclusion chromatography as a stand-alone methodology identifies novel markers in mass spectrometry analyses of plasma-derived vesicles from healthy individuals. J Extracell Vesicles. 2015;4:27378.

283. Gyorgy B, Modos K, Pallinger E, Paloczi K, Pasztoi M, Misjak P, et al. Detection and isolation of cell-derived microparticles are compromised by protein complexes resulting from shared biophysical parameters. Blood. 2011;117(4):e39-48.

284. Boing AN, van der Pol E, Grootemaat AE, Coumans FA, Sturk A, Nieuwland R. Single-step isolation of extracellular vesicles by size-exclusion chromatography. J Extracell Vesicles. 2014;3.

285. Oeyen E, Van Mol K, Baggerman G, Willems H, Boonen K, Rolfo C, et al. Ultrafiltration and size exclusion chromatography combined with asymmetrical-flow field-flow fractionation for the isolation and characterisation of extracellular vesicles from urine. J Extracell Vesicles. 2018;7(1):1490143.

286. Taylor DD, Shah S. Methods of isolating extracellular vesicles impact down-stream analyses of their cargoes. Methods. 2015;87:3-10.

287. Andreu Z, Rivas E, Sanguino-Pascual A, Lamana A, Marazuela M, González-Alvaro I, et al. Comparative analysis of EV isolation procedures for miRNAs detection in serum samples. Journal of Extracellular Vesicles. 2016;5(1):31655.

288. Torri A, Carpi D, Bulgheroni E, Crosti MC, Moro M, Gruarin P, et al. Extracellular MicroRNA Signature of Human Helper T Cell Subsets in Health and Autoimmunity. J Biol Chem. 2017;292(7):2903-15.

289. Francisco-Garcia AS, Garrido-Martín EM, Rupani H, Lau LCK, Martinez-Nunez RT, Howarth PH, et al. Small RNA Species and microRNA Profiles are Altered in Severe Asthma Nanovesicles from Broncho Alveolar Lavage and Associate with Impaired Lung Function and Inflammation. Non-coding RNA. 2019;5(4).

290. Alvarez ML, Khosroheidari M, Kanchi Ravi R, DiStefano JK. Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers. Kidney Int. 2012;82(9):1024-32.

291. Witwer KW, Buzas EI, Bemis LT, Bora A, Lasser C, Lotvall J, et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. J Extracell Vesicles. 2013;2.

292. Clayton A, Court J, Navabi H, Adams M, Mason MD, Hobot JA, et al. Analysis of antigen presenting cell derived exosomes, based on immuno-magnetic isolation and flow cytometry. J Immunol Methods. 2001;247(1-2):163-74.

293. Kim G, Yoo CE, Kim M, Kang HJ, Park D, Lee M, et al. Noble polymeric surface conjugated with zwitterionic moieties and antibodies for the isolation of exosomes from human serum. Bioconjugate chemistry. 2012;23(10):2114-20.

294. Yoo CE, Kim G, Kim M, Park D, Kang HJ, Lee M, et al. A direct extraction method for microRNAs from exosomes captured by immunoaffinity beads. Analytical biochemistry. 2012;431(2):96-8.

295. Enderle D, Spiel A, Coticchia CM, Berghoff E, Mueller R, Schlumpberger M, et al. Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. PLoS ONE. 2015;10(8):e0136133.

296. Dragovic RA, Gardiner C, Brooks AS, Tannetta DS, Ferguson DJ, Hole P, et al. Sizing and phenotyping of cellular vesicles using Nanoparticle Tracking Analysis. Nanomedicine : nanotechnology, biology, and medicine. 2011;7(6):780-8.

297. Gardiner C, Ferreira YJ, Dragovic RA, Redman CW, Sargent IL. Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis. J Extracell Vesicles. 2013;2.

298. Lacroix R, Robert S, Poncelet P, Kasthuri RS, Key NS, Dignat-George F. Standardization of platelet-derived microparticle enumeration by flow cytometry with calibrated beads: results of the International Society on Thrombosis and Haemostasis SSC Collaborative workshop. J Thromb Haemost. 2010;8(11):2571-4.

299. Logozzi M, De Milito A, Lugini L, Borghi M, Calabrò L, Spada M, et al. High Levels of Exosomes Expressing CD63 and Caveolin-1 in Plasma of Melanoma Patients. PLoS ONE. 2009;4(4):e5219.

300. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

301. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature. 2005;433(7027):769-73.

302. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome research. 2009;19(1):92-105.

303. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.

304. Care A, Catalucci D, Felicetti F, Bonci D, Addario A, Gallo P, et al. MicroRNA-133 controls cardiac hypertrophy. Nat Med. 2007;13(5):613-8.

305. Fiore R, Siegel G, Schratt G. MicroRNA function in neuronal development, plasticity and disease. Biochimica et biophysica acta. 2008;1779(8):471-8.

306. Krutzfeldt J, Stoffel M. MicroRNAs: a new class of regulatory genes affecting metabolism. Cell metabolism. 2006;4(1):9-12.

307. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834-8.

308. Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, MacDonald PE, et al. A pancreatic islet-specific microRNA regulates insulin secretion. Nature. 2004;432:226.

309. Van Pottelberge GR, Mestdagh P, Bracke KR, Thas O, van Durme YM, Joos GF, et al. MicroRNA expression in induced sputum of smokers and patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2011;183(7):898-906.

310. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415-9.

311. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science (New York, NY). 2001;293(5531):834-8.

312. Tolia NH, Joshua-Tor L. Slicer and the Argonautes. Nature Chemical Biology. 2006;3:36.

313. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281-97.

314. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010;466(7308):835-40.

315. Rupani H, Sanchez-Elsner T, Howarth P. MicroRNAs and respiratory diseases. Eur Respir J. 2013;41(3):695-705.

316. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proceedings of the National Academy of Sciences. 2009;106(7):2319-24.

317. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome biology. 2007;8(9):R201.

318. Izzotti A, Larghero P, Longobardi M, Cartiglia C, Camoirano A, Steele VE, et al. Dose-responsiveness and persistence of microRNA expression alterations induced by cigarette smoke in mouse lung. Mutation research. 2011;717(1-2):9-16.

319. Ezzie ME, Crawford M, Cho J-H, Orellana R, Zhang S, Gelinas R, et al. Gene expression networks in COPD: microRNA and mRNA regulation. Thorax. 2012;67(2):122-31.

320. Sato T, Liu X, Nelson A, Nakanishi M, Kanaji N, Wang X, et al. Reduced miR-146a increases prostaglandin E(2)in chronic obstructive pulmonary disease fibroblasts. Am J Respir Crit Care Med. 2010;182(8):1020-9.

321. Stanczyk J, Pedrioli DM, Brentano F, Sanchez-Pernaute O, Kolling C, Gay RE, et al. Altered expression of MicroRNA in synovial fibroblasts and synovial tissue in rheumatoid arthritis. Arthritis Rheum. 2008;58(4):1001-9.

322. Nakasa T, Miyaki S, Okubo A, Hashimoto M, Nishida K, Ochi M, et al. Expression of microRNA-146 in rheumatoid arthritis synovial tissue. Arthritis Rheum. 2008;58(5):1284-92.

323. Izzotti A, Calin GA, Arrigo P, Steele VE, Croce CM, De Flora S. Downregulation of microRNA expression in the lungs of rats exposed to cigarette smoke. Faseb J. 2009;23(3):806-12.

324. D'hulst AI, Bracke KR, Maes T, De Bleecker JL, Pauwels RA, Joos GF, et al. Role of tumour necrosis factor-α receptor p75 in cigarette smoke-induced pulmonary inflammation and emphysema. Eur Respir J. 2006;28(1):102-12.

325. Barh D, Malhotra R, Ravi B, Sindhurani P. MicroRNA let-7: an emerging next-generation cancer therapeutic. Current oncology (Toronto, Ont). 2010;17(1):70-80.

326. Blischak J. Differential epxression analysis with edgeR 2016 [Available from: https://gist.github.com/jdblischak/11384914.

327. Conickx G, Avila Cobos F, van den Berge M, Faiz A, Timens W, Hiemstra PS, et al. microRNA profiling in lung tissue and bronchoalveolar lavage of cigarette smoke-exposed mice and in COPD patients: a translational approach. Sci Rep. 2017;7(1):12871.

328. Savarimuthu Francis SM, Davidson MR, Tan ME, Wright CM, Clarke BE, Duhig EE, et al. MicroRNA-34c is associated with emphysema severity and modulates SERPINE1 expression. BMC genomics. 2014;15:88.

329. Baker J, Colley T, Ito K, Barnes P. The key role of microRNA-34a in the reduction of sirtuin-1 in COPD. Eur Respir J. 2016;48(suppl 60):OA4977.

330. Mizuno S, Bogaard HJ, Gomez-Arroyo J, Alhussaini A, Kraskauskas D, Cool CD, et al. MicroRNA-199a-5p is associated with hypoxia-inducible factor-1alpha expression in lungs from patients with COPD. Chest. 2012;142(3):663-72.

331. Hassan F, Nuovo GJ, Crawford M, Boyaka PN, Kirkby S, Nana-Sinkam SP, et al. MiR-101 and miR-144 Regulate the Expression of the CFTR Chloride Channel in the Lung. PLoS ONE. 2012;7(11):e50837.

332. Christenson SA, Brandsma C-A, Campbell JD, Knight DA, Pechkovsky DV, Hogg JC, et al. miR-638 regulates gene expression networks associated with emphysematous lung destruction. Genome medicine. 2013;5(12):114-.

333. Kim WJ, Lim JH, Hong Y, Hong S-H, Bang CY, Lee JS, et al. Altered miRNA expression in lung tissues of patients with chronic obstructive pulmonary disease. Molecular & Cellular Toxicology. 2017;13(2):207-12.

334. Faiz A, Steiling K, Roffel MP, Postma DS, Spira A, Lenburg ME, et al. Effect of long-term corticosteroid treatment on microRNA and gene-expression profiles in COPD. Eur Respir J. 2019;53(4):1801202.

335. Lewis A, Riddoch-Contreras J, Natanek SA, Donaldson A, Man WD-C, Moxham J, et al. Downregulation of the serum response factor/miR-1 axis in the quadriceps of patients with COPD. Thorax. 2012;67(1):26-34.

336. Donaldson A, Natanek SA, Lewis A, Man WD, Hopkinson NS, Polkey MI, et al. Increased skeletal muscle-specific microRNA in the blood of patients with COPD. Thorax. 2013;68(12):1140-9.

337. Pinkerton M, Chinchilli V, Banta E, Craig T, August A, Bascom R, et al. Differential expression of microRNAs in exhaled breath condensates of patients with asthma, patients with chronic obstructive pulmonary disease, and healthy adults. J Allergy Clin Immunol. 2013;132(1):217-9.

338. Soeda S, Ohyashiki JH, Ohtsuki K, Umezu T, Setoguchi Y, Ohyashiki K. Clinical relevance of plasma miR-106b levels in patients with chronic obstructive pulmonary disease. International journal of molecular medicine. 2013;31(3):533-9.

339. Leuenberger C, Schuoler C, Bye H, Mignan C, Rechsteiner T, Hillinger S, et al. MicroRNA-223 controls the expression of histone deacetylase 2: a novel axis in COPD. Journal of molecular medicine (Berlin, Germany). 2016;94(6):725-34.

340. Akbas F, Coskunpinar E, Aynaci E, Oltulu YM, Yildiz P. Analysis of serum micro-RNAs as potential biomarker in chronic obstructive pulmonary disease. Exp Lung Res. 2012;38(6):286-94.

341. Wang M, Huang Y, Liang Z, Liu D, Lu Y, Dai Y, et al. Plasma miRNAs might be promising biomarkers of chronic obstructive pulmonary disease. The clinical respiratory journal. 2016;10(1):104-11.

342. Molina-Pinelo S, Pastor MD, Suarez R, Romero-Romero B, Gonzalez De la Pena M, Salinas A, et al. MicroRNA clusters: dysregulation in lung adenocarcinoma and COPD. Eur Respir J. 2014;43(6):1740-9.

343. Puig-Vilanova E, Aguiló R, Rodríguez-Fuster A, Martínez-Llorens J, Gea J, Barreiro E. Epigenetic Mechanisms in Respiratory Muscle Dysfunction of Patients with Chronic Obstructive Pulmonary Disease. PLoS ONE. 2014;9(11):e111514.

344. Xie L, Wu M, Lin H, Liu C, Yang H, Zhan J, et al. An increased ratio of serum miR-21 to miR-181a levels is associated with the early pathogenic process of chronic obstructive pulmonary disease in asymptomatic heavy smokers. Molecular bioSystems. 2014;10(5):1072-81.

345. Velasco-Torres Y, Ruiz-López V, Pérez-Bautista O, Buendía-Roldan I, Ramírez-Venegas A, Pérez-Ramos J, et al. miR-34a in serum is involved in mild-to-moderate COPD in women exposed to biomass smoke. BMC Pulm Med. 2019;19(1):227.

346. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nature Reviews Drug Discovery. 2017;16:203.

347. György B, Hung ME, Breakefield XO, Leonard JN. Therapeutic Applications of Extracellular Vesicles: Clinical Promise and Open Questions. Annual Review of Pharmacology and Toxicology. 2015;55(1):439-64.

348. Monsel A, Zhu Y-G, Gennai S, Hao Q, Hu S, Rouby J-J, et al. Therapeutic Effects of Human Mesenchymal Stem Cell-derived Microvesicles in Severe Pneumonia in Mice. American journal of respiratory and critical care medicine. 2015;192(3):324-36.

349. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J. 2005;26(2):319-38.

350. Gevenois PA, De Vuyst P, de Maertelaer V, Zanen J, Jacobovitz D, Cosio MG, et al. Comparison of computed density and microscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med. 1996;154(1):187-92.

351. Gevenois PA, de Maertelaer V, De Vuyst P, Zanen J, Yernault JC. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. Am J Respir Crit Care Med. 1995;152(2):653-7.

352. Bommart S, Marin G, Bourdin A, Molinari N, Klein F, Hayot M, et al. Relationship between CT air trapping criteria and lung function in small airway impairment quantification. BMC Pulmonary Medicine. 2014;14(1):29.

353. Lötvall J, Hill AF, Hochberg F, Buzás EI, Di Vizio D, Gardiner C, et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. Journal of extracellular vesicles. 2014;3:26913-.

354. Almqvist N, Lonnqvist A, Hultkrantz S, Rask C, Telemo E. Serum-derived exosomes from antigen-fed mice prevent allergic sensitization in a model of allergic asthma. Immunology. 2008;125(1):21-7.

355. Prado N, Marazuela EG, Segura E, Fernandez-Garcia H, Villalba M, Thery C, et al. Exosomes from bronchoalveolar fluid of tolerized mice prevent allergic reaction. J Immunol. 2008;181(2):1519-25.

356. Shin TS, Kim JH, Kim YS, Jeon SG, Zhu Z, Gho YS, et al. Extracellular vesicles are key intercellular mediators in the development of immune dysfunction to allergens in the airways. Allergy. 2010;65(10):1256-65.

357. Torregrosa Paredes P, Esser J, Admyre C, Nord M, Rahman QK, Lukic A, et al. Bronchoalveolar lavage fluid exosomes contribute to cytokine and leukotriene production in allergic asthma. Allergy. 2012;67(7):911-9.

358. Gregson AL, Hoji A, Injean P, Poynter ST, Briones C, Palchevskiy V, et al. Altered Exosomal RNA Profiles in Bronchoalveolar Lavage from Lung Transplants with Acute Rejection. Am J Respir Crit Care Med. 2015;192(12):1490-503.

359. Qazi KR, Torregrosa Paredes P, Dahlberg B, Grunewald J, Eklund A, Gabrielsson S. Proinflammatory exosomes in bronchoalveolar lavage fluid of patients with sarcoidosis. Thorax. 2010;65(11):1016-24.

360. Maemura T, Fukuyama S, Sugita Y, Lopes TJS, Nakao T, Noda T, et al. Lung-Derived Exosomal miR-483-3p Regulates the Innate Immune Response to Influenza Virus Infection. The Journal of infectious diseases. 2018;217(9):1372-82.

361. Gardiner C, Di Vizio D, Sahoo S, Théry C, Witwer KW, Wauben M, et al. Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. Journal of extracellular vesicles. 2016;5:32945-.

362. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proceedings of the National Academy of Sciences. 2011;108(12):5003-8.

363. Page AM. The cytoskeleton architechture of trpanosomes: University of London; 1999.

364. Burke H, Spalluto CM, Cellura D, Staples KJ, Wilkinson TMA. Role of exosomal microRNA in driving skeletal muscle wasting in COPD. Eur Respir J. 2015;46(suppl 59).

365. Hill AF, Pegtel DM, Lambertz U, Leonardi T, O'Driscoll L, Pluchino S, et al. ISEV position paper: extracellular vesicle RNA analysis and bioinformatics. Journal of Extracellular Vesicles. 2013;2(1):22859.

366. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011;17(1):3.

367. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc 2010 [Version 11.7:[

368. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

369. Molecular Biology Select. Cell. 2006;126(2):223-5.

370. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nature reviews Molecular cell biology. 2011;12(4):246-58.

371. Maniatis T, Reed R. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature. 1987;325(6106):673-8.

372. Christov CP, Gardiner TJ, Szuts D, Krude T. Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Mol Cell Biol. 2006;26(18):6993-7004.

373. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

374. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research. 2016;5:1408-.

375. Love M. rnaseqGene: RNA-seq workflow: gene-level exploratory analysis and differential expression. R package version 1.4.0 2018 [Available from: https://github.com/mikelove/rnaseqGene/. .

376. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology. 2010;11(3):R25.

377. Chen Y, Lun, A.T.L., and Smyth, G.K. . Differential expression analysis of complex RNA-seq experiments using edgeR. Statistical Analysis of Next Generation Sequence Data. 2014:51-74.

378. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, et al. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. 2013;59(1):S1-6.

379. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245-50.

380. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.

381. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155-D62.

382. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res. 2014;42(17):e133.

383. Takahashi T, Kubo H. The role of microparticles in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2014;9:303-14.

384. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292-4.

385. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

386. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

387. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.

388. Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics. 2018;19(1):536.

389. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods. 2017;14(4):417-9.

390. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nature Biotechnology. 2017;35(4):316-9.

391. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nature methods. 2018;15(7):475-6.

392. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

393. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35(12):2084-92.

394. Liu H, Brannon AR, Reddy AR, Alexe G, Seiler MW, Arreola A, et al. Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. BMC Syst Biol. 2010;4:51.

395. Van der Auwera I, Limame R, van Dam P, Vermeulen PB, Dirix LY, Van Laere SJ. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. Br J Cancer. 2010;103(4):532-41.

396. da Silveira WA, Renaud L, Simpson J, Glen WB, Jr., Hazard ES, Chung D, et al. miRmapper: A Tool for Interpretation of miRNA¯mRNA Interaction Networks. Genes. 2018;9(9).

397. León LE, Calligaris SD. Visualization and Analysis of MiRNA–Targets Interactions Networks. In: Rani S, editor. MicroRNA Profiling: Methods and Protocols. New York, NY: Springer New York; 2017. p. 209-20.

398. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

399. Wang J, Zhong J, Chen G, Li M, Wu FX, Pan Y. ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network. IEEE/ACM transactions on computational biology and bioinformatics. 2015;12(4):815-22.

400. Gao T, Qian J. EAGLE: An algorithm that utilizes a small number of genomic features to predict tissue/cell type-specific enhancer-gene interactions. PLOS Computational Biology. 2019;15(10):e1007436.

401. Zhang Y, Hou J, Ge F, Cao F, Li H, Wang P, et al. Integrating microRNA and mRNA expression profiles of acute promyelocytic leukemia cells to explore the occurrence mechanisms of differentiation syndrome. Oncotarget. 2016;7(45):73509-24.

402. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research. 2016;45(D1):D331-D8.

403. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005;21(16):3448-9.

404. Tao L, Shi J, Huang X, Hua F, Yang L. Identification of a lncRNA-miRNA-mRNA network based on competitive endogenous RNA theory reveals functional lncRNAs in hypertrophic cardiomyopathy. Exp Ther Med. 2020;20(2):1176-90.

405. Rivals I, Personnaz L, Taing L, Potier M-C. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics. 2006;23(4):401-7.

406. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE. 2010;5(11):e13984-e.

407. Bafadhel M, Peterson S, De Blas MA, Calverley PM, Rennard SI, Richter K, et al. Predictors of exacerbation risk and response to budesonide in patients with chronic obstructive pulmonary disease: a post-hoc analysis of three randomised trials. The Lancet Respiratory medicine. 2018;6(2):117-26.

408. Hogg JC, McDonough JE, Suzuki M. Small airway obstruction in COPD: new insights based on micro-CT imaging and MRI imaging. Chest. 2013;143(5):1436-43.

409. McDonough JE, Yuan R, Suzuki M, Seyednejad N, Elliott WM, Sanchez PG, et al. Small-airway obstruction and emphysema in chronic obstructive pulmonary disease. The New England journal of medicine. 2011;365(17):1567-75.

410. Bhatt SP, Soler X, Wang X, Murray S, Anzueto AR, Beaty TH, et al. Association between Functional Small Airway Disease and FEV1 Decline in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2016;194(2):178-84.

411. Turato G, Semenzato U, Bazzan E, Biondini D, Tinè M, Torrecilla N, et al. Blood Eosinophilia Neither Reflects Tissue Eosinophils nor Worsens Clinical Outcomes in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. 2017;197(9):1216-9.

412. Abdelwahab S, Gupta R, Radicioni G, Jones L, Dang H, O'Neal W, et al. Airway Epithelial Derived Exosomes on Protecting and Remodeling of the Lung. The FASEB Journal. 2015;29(1 Supplement).

413. Groot Kormelink T, Mol S, de Jong EC, Wauben MHM. The role of extracellular vesicles when innate meets adaptive. Seminars in Immunopathology. 2018;40(5):439-52.

414. Mazzeo C, Cañas JA, Zafra MP, Rojas Marco A, Fernández-Nieto M, Sanz V, et al. Exosome secretion by eosinophils: A possible role in asthma pathogenesis. Journal of Allergy and Clinical Immunology. 2015;135(6):1603-13.

415. Rossaint J, Kühne K, Skupski J, Van Aken H, Looney MR, Hidalgo A, et al. Directed transport of neutrophil-derived extracellular vesicles enables platelet-mediated innate immune response. Nat Commun. 2016;7:13464.

416. O'Donnell R, Breen D, Wilson S, Djukanovic R. Inflammatory cells in the airways in COPD. Thorax. 2006;61(5):448-54.

417. Karlicic V. Cellular composition of bronchoalveolar lavage (BAL) as inflammation indicator in patients with chronic obstructive pulmonary disease (COPD). Eur Respir J. 2014;44(Suppl 58):P707.

418. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome biology. 2016;17:13.

419. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE. 2014;9(1):e78644.

420. Francisco-Garcia A, Martinez-Nunez RT, Rupani H, Lau LC, Howarth PH, Sanchez-Elsner T. LSC Abstract – Altered small RNA cargo in severe asthma exosomes. Eur Respir J. 2016;48(suppl 60):PP101.

421. Kosaka N, Iguchi H, Hagiwara K, Yoshioka Y, Takeshita F, Ochiya T. Neutral sphingomyelinase 2 (nSMase2)-dependent exosomal transfer of angiogenic microRNAs regulate cancer cell metastasis. J Biol Chem. 2013;288(15):10849-59.

422. Chung S, Vu S, Filosto S, Goldkorn T. Src regulates cigarette smoke-induced ceramide generation via neutral sphingomyelinase 2 in the airway epithelium. American journal of respiratory cell and molecular biology. 2015;52(6):738-48.

423. Filosto S, Becker C, Ashfaq M, Tognon E, Goldkorn T. Cigarette Smoke Induces Epidermal Growth Factor Receptor Resistance To Tyrosine Kinase Inhibitors. Molecular Targets for Therapeutic Development in Lung Cancer. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2011. p. A5077-A.

424. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nat Commun. 2013;4:2980.

425. Tauler J, Mulshine JL. Lung cancer and inflammation: interaction of chemokines and hnRNPs. Current opinion in pharmacology. 2009;9(4):384-8.

426. Katsimpoula S, Patrinou-Georgoula M, Makrilia N, Dimakou K, Guialis A, Orfanidou D, et al. Overexpression of hnRNPA2/B1 in bronchoscopic specimens: a potential early detection marker in lung cancer. Anticancer Res. 2009;29(4):1373-82.

427. Koppers-Lalic D, Hackenberg M, Bijnsdorp IV, van Eijndhoven MAJ, Sadek P, Sie D, et al. Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. Cell reports. 2014;8(6):1649-58.

428. Zhang D, Lee H, Wang X, Groot M, Sharma L, Dela Cruz CS, et al. A potential role of microvesicle-containing miR-223/142 in lung inflammation. Thorax. 2019;74(9):865.

429. Guduric-Fuchs J, O'Connor A, Camp B, O'Neill CL, Medina RJ, Simpson DA. Selective extracellular vesicle-mediated export of an overlapping set of microRNAs from multiple cell types. BMC genomics. 2012;13:357.

430. Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. Cell. 2005;123(5):819-31.

431. Johnnidis JB, Harris MH, Wheeler RT, Stehling-Sun S, Lam MH, Kirak O, et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature. 2008;451(7182):1125-9.

432. Felli N, Pedini F, Romania P, Biffoni M, Morsilli O, Castelli G, et al. MicroRNA 223-dependent expression of LMO2 regulates normal erythropoiesis. Haematologica. 2009;94(4):479-86.

433. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007;129(7):1401-14.

434. Seumois G, Vijayanand P, Eisley CJ, Omran N, Kalinke L, North M, et al. An integrated nano-scale approach to profile miRNAs in limited clinical samples. American journal of clinical and experimental immunology. 2012;1(2):70-89.

435. Maes T, Cobos FA, Schleich F, Sorbello V, Henket M, De Preter K, et al. Asthma inflammatory phenotypes show differential microRNA expression in sputum. J Allergy Clin Immunol. 2016;137(5):1433-46.

436. Fulci V, Scappucci G, Sebastiani GD, Giannitti C, Franceschini D, Meloni F, et al. miR-223 is overexpressed in T-lymphocytes of patients affected by rheumatoid arthritis. Human immunology. 2010;71(2):206-11.

437. Neudecker V, Brodsky KS, Clambey ET, Schmidt EP, Packard TA, Davenport B, et al. Neutrophil transfer of miR-223 to lung epithelial cells dampens acute lung injury in mice. Science translational medicine. 2017;9(408).

438. Zhu H, Leung SW. Identification of microRNA biomarkers in type 2 diabetes: a meta-analysis of controlled profiling studies. Diabetologia. 2015;58(5):900-11.

439. Chen Q, Wang H, Liu Y, Song Y, Lai L, Han Q, et al. Inducible microRNA-223 down-regulation promotes TLR-triggered IL-6 and IL-1beta production in macrophages by targeting STAT3. PLoS ONE. 2012;7(8):e42971.

440. Liu Y, Wang R, Jiang J, Yang B, Cao Z, Cheng X. miR-223 is upregulated in monocytes from patients with tuberculosis and regulates function of monocyte-derived macrophages. Mol Immunol. 2015;67(2 Pt B):475-81.

441. Yang M, Chen J, Su F, Yu B, Su F, Lin L, et al. Microvesicles secreted by macrophages shuttle invasion-potentiating microRNAs into breast cancer cells. Molecular cancer. 2011;10:117.

442. Chen L, Lu FB, Chen DZ, Wu JL, Hu ED, Xu LM, et al. BMSCs-derived miR-223-containing exosomes contribute to liver protection in experimental autoimmune hepatitis. Mol Immunol. 2018;93:38-46.

443. Suzuki R, Amatya VJ, Kushitani K, Kai Y, Kambara T, Takeshima Y. miR-182 and miR-183 Promote Cell Proliferation and Invasion by Targeting FOXO1 in Mesothelioma. Frontiers in Oncology. 2018;8(446).

444. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. Int J Cancer. 2010;127(12):2870-8.

445. Zhu W, Zhou K, Zha Y, Chen D, He J, Ma H, et al. 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. 2016;11(4):e0153046.

446. Sun L, Lin P, Chen Y, Yu H, Ren S, Wang J, et al. miR-182-3p/Myadm contribute to pulmonary artery hypertension vascular remodeling via a KLF4/p21-dependent mechanism. Theranostics. 2020;10(12):5581-99.

447. Chen Y, Zhang Q, Zhou Y, Yang Z, Tan M. Inhibition of miR-182-5p attenuates pulmonary fibrosis via TGF-β/Smad pathway. Hum Exp Toxicol. 2020;39(5):683-95.

448. Barnes PJ. Small airway fibrosis in COPD. Int J Biochem Cell Biol. 2019;116:105598.

449. Zhao Z, Ma X, Hsiao T-H, Lin G, Kosti A, Yu X, et al. A high-content morphological screen identifies novel microRNAs that regulate neuroblastoma cell differentiation. Oncotarget. 2014;5(9):2499-512.

450. Zhao Z, Partridge V, Sousares M, Shelton SD, Holland CL, Pertsemlidis A, et al. microRNA-2110 functions as an onco-suppressor in neuroblastoma by directly targeting Tsukushi. PLoS ONE. 2018;13(12):e0208777.

451. Lyu L, Zhang X, Li C, Yang T, Wang J, Pan L, et al. Small RNA Profiles of Serum Exosomes Derived From Individuals With Latent and Active Tuberculosis. Frontiers in microbiology. 2019;10:1174.

452. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008;10(5):593-601.

453. Korpal M, Lee ES, Hu G, Kang Y. The miR-200 Family Inhibits Epithelial-Mesenchymal Transition and Cancer Cell Migration by Direct Targeting of E-cadherin Transcriptional Repressors ZEB1 and ZEB2. J Biol Chem. 2008;283(22):14910-4.

454. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AM, et al. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. Proc Natl Acad Sci U S A. 2006;103(35):13180-5.

455. Sohal SS, Mahmood MQ, Walters EH. Clinical significance of epithelial mesenchymal transition (EMT) in chronic obstructive pulmonary disease (COPD): potential target for prevention of airway fibrosis and lung cancer. Clin Transl Med. 2014;3(1):33-.

456. Chan YC, Roy S, Khanna S, Sen CK. Downregulation of endothelial MicroRNA-200b supports cutaneous wound angiogenesis by desilencing GATA binding protein 2 and vascular endothelial growth factor receptor 2. Arteriosclerosis, Thrombosis, and Vascular Biology. 2012;32(6):1372-82.

457. Rasmussen MH, Jensen NF, Tarpgaard LS, Qvortrup C, Rømer MU, Stenvang J, et al. High expression of microRNA-625-3p is associated with poor response to first-line oxaliplatin based treatment of metastatic colorectal cancer. Molecular oncology. 2013;7(3):637-46.

458. Rasmussen MH, Lyskjær I, Jersie-Christensen RR, Tarpgaard LS, Primdal-Bengtson B, Nielsen MM, et al. miR-625-3p regulates oxaliplatin resistance by targeting MAP2K6-p38 signalling in human colorectal adenocarcinoma cells. Nat Commun. 2016;7:12436.

459. Kirschner MB, Cheng YY, Badrian B, Kao SC, Creaney J, Edelman JJ, et al. Increased circulating miR-625-3p: a potential biomarker for patients with malignant pleural mesothelioma. J Thorac Oncol. 2012;7(7):1184-91.

460. Verma K, Jyotsana N, Buenting I, Luther S, Pfanne A, Thum T, et al. miR-625-3p is upregulated in CD8+ T cells during early immune reconstitution after allogeneic stem cell transplantation. PLoS ONE. 2017;12(8):e0183828.

461. Dong X, Xu M, Ren Z, Gu J, Lu M, Lu Q, et al. Regulation of CBL and ESR1 expression by microRNA-22-3p, 513a-5p and 625-5p may impact the pathogenesis of dust mite-induced pediatric asthma. International journal of molecular medicine. 2016;38(2):446-56.

462. Qian F-H, Deng X, Zhuang Q-X, Wei B, Zheng D-D. miR-625-5p suppresses inflammatory responses by targeting AKT2 in human bronchial epithelial cells. Mol Med Rep. 2019;19(3):1951-7.

463. Panganiban RP, Wang Y, Howrylak J, Chinchilli VM, Craig TJ, August A, et al. Circulating microRNAs as biomarkers in patients with allergic rhinitis and asthma. J Allergy Clin Immunol. 2016;137(5):1423-32.

464. Lacedonia D, Palladino GP, Foschino-Barbaro MP, Scioscia G, Carpagnano GE. Expression profiling of miRNA-145 and miRNA-338 in serum and sputum of patients with COPD, asthma, and asthma-COPD overlap syndrome phenotype. Int J Chron Obstruct Pulmon Dis. 2017;12:1811-7.

465. Howe JRVI, Li ES, Streeter SE, Rahme GJ, Chipumuro E, Russo GB, et al. MiR-338-3p regulates neuronal maturation and suppresses glioblastoma proliferation. PLoS ONE. 2017;12(5):e0177661.

466. Kos A, Olde Loohuis NF, Wieczorek ML, Glennon JC, Martens GJ, Kolk SM, et al. A potential regulatory role for intronic microRNA-338-3p for its host gene encoding apoptosis-associated tyrosine kinase. PLoS ONE. 2012;7(2):e31022.

467. Liang L, Gao L, Zou X-P, Huang M-L, Chen G, Li J-J, et al. Diagnostic significance and potential function of miR-338-5p in hepatocellular carcinoma: A bioinformatics study with microarray and RNA sequencing data. Mol Med Report. 2018;17(2):2297-312.

468. Lu M, Huang H, Yang J, Li J, Zhao G, Li W, et al. miR-338-3p regulates the proliferation, apoptosis and migration of SW480 cells by targeting MACC1. Exp Ther Med. 2019;17(4):2807-14.

469. Liang C-Y, Li Z-Y, Gan T-Q, Fang Y-Y, Gan B-L, Chen W-J, et al. Downregulation of hsa-microRNA-204-5p and identification of its potential regulatory network in non-small cell lung cancer: RT-qPCR, bioinformatic- and meta-analyses. Respir Res. 2020;21(1):60.

470. Liu X, Gao X, Zhang W, Zhu T, Bi W, Zhang Y. MicroRNA-204 deregulation in lung adenocarcinoma controls the biological behaviors of endothelial cells potentially by modulating Janus kinase 2-signal transducer and activator of transcription 3 pathway. IUBMB life. 2018;70(1):81-91.

471. Guo W, Zhang Y, Zhang Y, Shi Y, Xi J, Fan H, et al. Decreased expression of miR-204 in plasma is associated with a poor prognosis in patients with non-small cell lung cancer. International journal of molecular medicine. 2015;36(6):1720-6.

472. Schneider R, McKeever P, Kim T, Graff C, van Swieten JC, Karydas A, et al. Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study. Journal of Neurology, Neurosurgery & Psychiatry. 2018;89(8):851-8.

473. Chiu C-C, Yeh T-H, Chen R-S, Chen H-C, Huang Y-Z, Weng Y-H, et al. Upregulated Expression of MicroRNA-204-5p Leads to the Death of Dopaminergic Cells by Targeting DYRK1A-Mediated Apoptotic Signaling Cascade. Frontiers in Cellular Neuroscience. 2019;13(399).

474. Wang Y, Li W, Zang X, Chen N, Liu T, Tsonis PA, et al. MicroRNA-204-5p Regulates Epithelial-to-Mesenchymal Transition during Human Posterior Capsule Opacification by Targeting SMAD4. Investigative Ophthalmology & Visual Science. 2013;54(1):323-32.

475. Song N, Li P, Song P, Li Y, Zhou S, Su Q, et al. MicroRNA-138-5p Suppresses Non-small Cell Lung Cancer Cells by Targeting PD-L1/PD-1 to Regulate Tumor Microenvironment. Frontiers in cell and developmental biology. 2020;8:540.

476. Faner R, D. Morrow J, Casas-Recasens S, Cloonan S, Noell G, López-Giraldo A, et al. Do sputum or circulating blood samples reflect the pulmonary transcriptomic differences of COPD patients? A multi-tissue transcriptomic network META-analysis2019.

477. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011;39(Database issue):D163-9.

478. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. Nucleic acids research. 2018;46(D1):D239-D45.

479. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic acids research. 2013;41(Web Server issue):W169-W73.

480. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinformatics. 2007;8:69.

481. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome biology. 2003;5(1):R1.

482. Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. Bioinformatics. 2016;32(9):1316-22.

483. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nature genetics. 2007;39(10):1278-84.

484. Mu W, Zhang W. Bioinformatic Resources of microRNA Sequences, Gene Targets, and Genetic Variation. Frontiers in genetics. 2012;3:31-.

485. Peter ME. Regulating Cancer Stem Cells the miR Way. Cell Stem Cell. 2010;6(1):4-6.

486. Shalgi R, Lieber D, Oren M, Pilpel Y. Global and Local Architecture of the Mammalian microRNA–Transcription Factor Regulatory Network. PLOS Computational Biology. 2007;3(7):e131.

487. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research. 2014;42(Database issue):D68-D73.

488. Didiano D, Hobert O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. Nature structural & molecular biology. 2006;13(9):849-51.

489. Watson A ÖL, Angermann BR, Spalluto CM, Hühn M, Burke H, Cellura D, Freeman A, Muthas D, Etal D, Belfield G, Karlsson F, Nordström K, Ostridge K, Staples KJ, Wilkinson TMA. COVID-19 Related Gene Expression in the Lung -Insights into the Susceptibility to Infection and Inflammation in COPD. 2021.

490. Denecke B, Gräber S, Schäfer C, Heiss A, Wöltje M, Jahnen-Dechent W. Tissue distribution and activity testing suggest a similar but not identical function of fetuin-B and fetuin-A. Biochem J. 2003;376(Pt 1):135-45.

491. Meex RC, Hoy AJ, Morris A, Brown RD, Lo JC, Burke M, et al. Fetuin B Is a Secreted Hepatocyte Factor Linking Steatosis to Impaired Glucose Metabolism. Cell metabolism. 2015;22(6):1078-89.

492. Diao W-q, Shen N, Du Y-p, Liu B-b, Sun X-y, Xu M, et al. Fetuin-B (FETUB): a Plasma Biomarker Candidate Related to the Severity of Lung Function in COPD. Scientific Reports. 2016;6(1):30045.

493. George L, Taylor AR, Esteve-Codina A, Soler Artigas M, Thun GA, Bates S, et al. Blood eosinophil count and airway epithelial transcriptome relationships in COPD versus asthma. Allergy. 2020;75(2):370-80.

494. Green AR, Barbour S, Horn T, Carlos J, Raskatov JA, Holman TR. Strict Regiospecificity of Human Epithelial 15-Lipoxygenase-2 Delineates Its Transcellular Synthesis Potential. Biochemistry. 2016;55(20):2832-40.

495. Kristjansson RP, Benonisdottir S, Davidsson OB, Oddsson A, Tragante V, Sigurdsson JK, et al. A loss-of-function variant in ALOX15 protects against nasal polyps and chronic rhinosinusitis. Nature genetics. 2019;51(2):267-76.

496. Wilson SM, Newton R, Giembycz M. Phosphodiesterase 4 Inhibitors Suppress Chemokine Release From Human Airway Epithelial Cells By A Novel, 15-Lipoxygenase-2- And PPARg-Dependent Mechanism. Phosphodiesterase Inhibitors as Therapeutics for Lung Diseases. American Thoracic Society International Conference Abstracts: American Thoracic Society; 2012. p. A5692-A.

497. Churg A, Zhou S, Wright JL. Matrix metalloproteinases in COPD. Eur Respir J. 2012;39(1):197.

498. Baines KJ, Fu J-J, McDonald VM, Gibson PG. Airway gene expression of IL-1 pathway mediators predicts exacerbation risk in obstructive airway disease. International journal of chronic obstructive pulmonary disease. 2017;12:541-50.

499. Balk-Møller E, Windeløv JA, Svendsen B, Hunt J, Ghiasi SM, Sørensen CM, et al. Glucagon-Like Peptide 1 and Atrial Natriuretic Peptide in a Female Mouse Model of Obstructive Pulmonary Disease. J Endocr Soc. 2019;4(1):bvz034-bvz.

500. He XL, Dukkipati A, Garcia KC. Structural determinants of natriuretic peptide receptor specificity and degeneracy. Journal of molecular biology. 2006;361(4):698-714.

501. Anand-Srivastava MB. Natriuretic peptide receptor-C signaling and regulation. Peptides. 2005;26(6):1044-59.

502. Bianciotti LG, Vatta MS, Elverdin JC, di Carlo MB, Negri G, Fernandez BE. Atrial natriuretic factor-induced amylase output in the rat parotid gland appears to be mediated by the inositol phosphate pathway. Biochemical and biophysical research communications. 1998;247(1):123-8.

503. Li Y, Hashim S, Anand-Srivastava MB. Intracellular peptides of natriuretic peptide receptor-C inhibit vascular hypertrophy via Gqalpha/MAP kinase signaling pathways. Cardiovascular research. 2006;72(3):464-72.

504. Sun J-Z, Oparil S, Lucchesi P, Thompson JA, Chen Y-F. Tyrosine kinase receptor activation inhibits NPR-C in lung arterial smooth muscle cells. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2001;281(1):L155-L63.

505. Kachroo P, Morrow JD, Kho AT, Vyhlidal CA, Silverman EK, Weiss ST, et al. Co-methylation analysis in lung tissue identifies pathways for fetal origins of COPD. Eur Respir J. 2020;56(4).

506. Wang IM, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, et al. Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. Am J Respir Crit Care Med. 2008;177(4):402-11.

507. Imkamp K, Berg M, Vermeulen CJ, Heijink IH, Guryev V, Kerstjens HAM, et al. Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci. Journal of Allergy and Clinical Immunology. 2018;142(1):314-7.e15.

508. Zayed H. Novel Comprehensive Bioinformatics Approaches to Determine the Molecular Genetic Susceptibility Profile of Moderate and Severe Asthma. International journal of molecular sciences [Internet]. 2020 2020/06//; 21(11).

509. Bonnans C, Chou J, Werb Z. Remodelling the extracellular matrix in development and disease. Nature reviews Molecular cell biology. 2014;15(12):786-801.

510. Hill JM, Zalos G, Halcox JP, Schenke WH, Waclawiw MA, Quyyumi AA, et al. Circulating endothelial progenitor cells, vascular function, and cardiovascular risk. The New England journal of medicine. 2003;348(7):593-600.

511. Brittan M, Hoogenboom MM, Padfield GJ, Tura O, Fujisawa T, MacLay JD, et al. Endothelial progenitor cells in patients with chronic obstructive pulmonary disease. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2013;305(12):L964-L9.

512. Asahara T, Murohara T, Sullivan A, Silver M, van der Zee R, Li T, et al. Isolation of putative progenitor endothelial cells for angiogenesis. Science (New York, NY). 1997;275(5302):964-7.

513. Lin Y-Z, Zhong X-N, Chen X, Liang Y, Zhang H, Zhu D-L. Roundabout signaling pathway involved in the pathogenesis of COPD by integrative bioinformatics analysis. International journal of chronic obstructive pulmonary disease. 2019;14:2145-62.

514. Rekers NV, Bajema IM, Mallat MJK, Anholts JDH, de Vaal YJH, Zandbergen M, et al. Increased Metallothionein Expression Reflects Steroid Resistance in Renal Allograft Recipients. American Journal of Transplantation. 2013;13(8):2106-18.

515. Carlier FM, Dupasquier S, Ambroise J, Detry B, Lecocq M, Biétry–Claudet C, et al. Canonical WNT pathway is activated in the airway epithelium in chronic obstructive pulmonary disease. EBioMedicine. 2020;61.

516. Bermingham ML, Walker RM, Marioni RE, Morris SW, Rawlik K, Zeng Y, et al. Identification of novel differentially methylated sites with potential as clinical predictors of impaired respiratory function and COPD. EBioMedicine. 2019;43:576-86.

517. Rižner TL, Penning TM. Role of aldo-keto reductase family 1 (AKR1) enzymes in human steroid metabolism. Steroids. 2014;79:49-63.

518. Zhao SF, Wang SG, Zhao ZY, Li WL. AKR1C1-3, notably AKR1C3, are distinct biomarkers for liver cancer diagnosis and prognosis: Database mining in malignancies. Oncol Lett. 2019;18(5):4515-22.

519. Steiling K, Lenburg ME, Spira A. Airway gene expression in chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society. 2009;6(8):697-700.

520. Sheu CC, Chang WA, Tsai MJ, Liao SH, Chong IW, Kuo PL. Bioinformatic analysis of next-generation sequencing data to identify dysregulated genes in fibroblasts of idiopathic pulmonary fibrosis. International journal of molecular medicine. 2019;43(4):1643-56.

521. Soroosh A, Albeiroti S, West GA, Willard B, Fiocchi C, de la Motte CA. Crohn's Disease Fibroblasts Overproduce the Novel Protein KIAA1199 to Create Proinflammatory Hyaluronan Fragments. Cellular and Molecular Gastroenterology and Hepatology. 2016;2(3):358-68.e4.

522. Li L, Yan L-H, Manoj S, Li Y, Lu L. Central Role of CEMIP in Tumorigenesis and Its Potential as Therapeutic Target. Journal of Cancer. 2017;8(12):2238-46.

523. Qin J, Yang T, Zeng N, Wan C, Gao L, Li X, et al. Differential coexpression networks in bronchiolitis and emphysema phenotypes reveal heterogeneous mechanisms of chronic obstructive pulmonary disease. Journal of Cellular and Molecular Medicine. 2019;23(10):6989-99.

524. Arseni L, Lombardi A, Orioli D. From Structure to Phenotype: Impact of Collagen Alterations on Human Health. International journal of molecular sciences. 2018;19(5).

525. Muniategui A, Pey J, Planes FJ, Rubio A. Joint analysis of miRNA and mRNA expression data. Briefings in bioinformatics. 2013;14(3):263-78.

526. Wang L, Liang Y, Mao Q, Xia W, Chen B, Shen H, et al. Circular RNA CRIM1 inhibits invasion and metastasis in lung adenocarcinoma through the microRNA (miR)-182/miR-93-leukemia inhibitory factor receptor pathway. Cancer science. 2019;110(9):2960-72.

527. Jens Glienke AS, Andreas Menrad, Karl-Heinz Thierauch. CRIM1 is involved in endothelial cell capillary formation in vitro and is expressed in blood vessels in vivo. Mechanisms of Development. 2002;119(2):165-75.

528. Hoang TT, Sikdar S, Xu C-J, Lee MK, Cardwell J, Forno E, et al. Epigenome-wide association study of DNA methylation and adult asthma in the Agricultural Lung Health Study. Eur Respir J. 2020;56(3).

529. Shu J, Wang, L., Han, F., Chen, Y., Wang, S., & Luo, F. BTBD7 Downregulates E-Cadherin and Promotes Epithelial-Mesenchymal Transition in Lung Cancer. BioMed research international. 2019;2019: 5937635.

530. Stoll P, Wuertemberger U, Bratke K, Zingler C, Virchow JC, Lommatzsch M. Stage-dependent association of BDNF and TGF-β1with lung function in stable COPD. Respir Res. 2012;13(1):116.

531. Loza MJ, Watt R, Baribaud F, Barnathan ES, Rennard SI. Systemic inflammatory profile and response to anti-tumor necrosis factor therapy in chronic obstructive pulmonary disease. Respir Res. 2012;13(1):12.

532. Pinto-Plata V, Toso J, Lee K, Park D, Bilello J, Mullerova H, et al. Profiling serum biomarkers in patients with COPD: associations with clinical parameters. Thorax. 2007;62(7):595-601.

533. Aravamudan B, Thompson M, Pabelick C, Prakash YS. Brain-derived neurotrophic factor induces proliferation of human airway smooth muscle cells. J Cell Mol Med. 2012;16(4):812-23.

534. Lommatzsch M, Schloetcke K, Klotz J, Schuhbaeck K, Zingler D, Zingler C, et al. Brain-derived Neurotrophic Factor in Platelets and Airflow Limitation in Asthma. American Journal of Respiratory and Critical Care Medicine. 2005;171(2):115-20.

535. Takatsuki S, Nakamura R, Haga Y, Mitsui K, Hashimoto T, Shimojima K, et al. Severe pulmonary emphysema in a girl with interstitial deletion of 2q24.2q24.3 including ITGB6. American Journal of Medical Genetics Part A. 2010;152A(4):1020-5.

536. Bhaskaran V, Nowicki MO, Idriss M, Jimenez MA, Lugli G, Hayes JL, et al. The functional synergism of microRNA clustering provides therapeutically relevant epigenetic interference in glioblastoma. Nature Communications. 2019;10(1):442.

537. Manikandan P, Nagini S. Cytochrome P450 Structure, Function and Clinical Significance: A Review. Current drug targets. 2018;19(1):38-54.

538. Shimada T, Hayes CL, Yamazaki H, Amin S, Hecht SS, Guengerich FP, et al. Activation of chemically diverse procarcinogens by human cytochrome P-450 1B1. Cancer Res. 1996;56(13):2979-84.

539. Smerdová L, Neča J, Svobodová J, Topinka J, Schmuczerová J, Kozubík A, et al. Inflammatory mediators accelerate metabolism of benzo[a]pyrene in rat alveolar type II cells: the role of enhanced cytochrome P450 1B1 expression. Toxicology. 2013;314(1):30-8.

540. Kamata S, Fujino N, Yamada M, Grime K, Suzuki S, Ota C, et al. Expression of cytochrome P450 mRNAs in Type II alveolar cells from subjects with chronic obstructive pulmonary disease. Pharmacology research & perspectives. 2018;6(3):e00405.

541. Rebecchi MJ, Pentyala SN. Structure, Function, and Control of Phosphoinositide-Specific Phospholipase C. Physiological reviews. 2000;80(4):1291-335.

542. Zhu L, Ly H, Liang Y. PLC-γ1 Signaling Plays a Subtype-Specific Role in Postbinding Cell Entry of Influenza A Virus. J Virol. 2014;88(1):417-24.

543. Zhu L, Yuan C, Ding X, Xu S, Yang J, Liang Y, et al. PLC-γ1 is involved in the inflammatory response induced by influenza A virus H1N1 infection. Virology. 2016;496:131-7.

544. Zhu L, Jones C, Zhang G. The Role of Phospholipase C Signaling in Macrophage-Mediated Inflammatory Response. Journal of immunology research. 2018;2018:5201759.

545. Jia W, Feng YI, Sanders AJ, Davies EL, Jiang WG. Phosphoinositide-3-Kinase Enhancers, PIKEs: Their Biological Functions and Roles in Cancer. Anticancer Res. 2016;36(3):1103.

546. Wang Z, Li W, Guo Q, Wang Y, Ma L, Zhang X. Insulin-Like Growth Factor-1 Signaling in Lung Development and Inflammatory Lung Diseases. BioMed Research International. 2018;2018:1-27.

547. Chan SMH, Selemidis S, Bozinovski S, Vlahos R. Pathobiological mechanisms underlying metabolic syndrome (MetS) in chronic obstructive pulmonary disease (COPD): clinical significance and therapeutic strategies. Pharmacol Ther. 2019;198:160-88.

548. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. Journal of Allergy and Clinical Immunology. 2016;138(1):16-27.

549. Pascoe S, Locantore N, Dransfield MT, Barnes NC, Pavord ID. Blood eosinophil counts, exacerbations, and response to the addition of inhaled fluticasone furoate to vilanterol in patients with chronic obstructive pulmonary disease: a secondary analysis of data from two parallel randomised controlled trials. The Lancet Respiratory Medicine. 2015;3(6):435-42.

550. Pavord ID, Lettis S, Locantore N, Pascoe S, Jones PW, Wedzicha JA, et al. Blood eosinophils and inhaled corticosteroid/longacting β-2 agonist efficacy in COPD. Thorax. 2016;71(2):118-25.

551. Meyer KC, Raghu G, Baughman RP, Brown KK, Costabel U, du Bois RM, et al. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am J Respir Crit Care Med. 2012;185(9):1004-14.

552. Wang L, Liu Y, Du L, Li J, Jiang X, Zheng G, et al. Identification and validation of reference genes for the detection of serum microRNAs by reverse transcription-quantitative polymerase chain reaction in patients with bladder cancer. Mol Med Rep. 2015;12(1):615-22.

553. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells. 2020;9(2):276.

554. Solé C, Moliné T, Vidal M, Ordi-Ros J, Cortés-Hernández J. An Exosomal Urinary miRNA Signature for Early Diagnosis of Renal Fibrosis in Lupus Nephritis. Cells. 2019;8(8).

555. Xiong DD, Lv J, Wei KL, Feng ZB, Chen JT, Liu KC, et al. A nine-miRNA signature as a potential diagnostic marker for breast carcinoma: An integrated study of 1,110 cases. Oncology reports. 2017;37(6):3297-304.

556. Mompeón A, Ortega-Paz L, Vidal-Gómez X, Costa TJ, Pérez-Cremades D, Garcia-Blas S, et al. Disparate miRNA expression in serum and plasma of patients with acute myocardial infarction: a systematic and paired comparative analysis. Scientific Reports. 2020;10(1):5373.

557. Shen BY, Wu N, Yang JM, Gould SJ. Protein Targeting to Exosomes/Microvesicles by Plasma Membrane Anchors. J Biol Chem. 2011;286(16):14383-95.

558. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proceedings of the National Academy of Sciences. 2008;105(30):10513.

559. Li SC, Huang LH, Chien KJ, Pan CY, Lin PH, Lin Y, et al. MiR-182-5p enhances in vitro neutrophil infiltration in Kawasaki disease. Mol Genet Genomic Med. 2019;7(12):e990.

560. Wurm AA, Zjablovskaja P, Kardosova M, Gerloff D, Brauer-Hartmann D, Katzerke C, et al. Disruption of the C/EBPalpha-miR-182 balance impairs granulocytic differentiation. Nat Commun. 2017;8(1):46.

561. Karali M, Guadagnino I, Marrocco E, De Cegli R, Carissimo A, Pizzo M, et al. AAV-miR-204 Protects from Retinal Degeneration by Attenuation of Microglia Activation and Photoreceptor Cell Death. Mol Ther Nucleic Acids. 2020;19:144-56.

562. Li H, Wang J, Liu X, Cheng Q. MicroRNA-204-5p suppresses IL6-mediated inflammatory response and chemokine generation in HK-2 renal tubular epithelial cells by targeting IL6R. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2019;97(2):109-17.

563. Fielding CA, McLoughlin RM, McLeod L, Colmont CS, Najdovska M, Grail D, et al. IL-6 regulates neutrophil trafficking during acute inflammation via STAT3. J Immunol. 2008;181(3):2189-95.

564. Feng JS, Sun JD, Wang XD, Fu CH, Gan LL, Ma R. MicroRNA-204-5p targets SOX11 to regulate the inflammatory response in spinal cord injury. European review for medical and pharmacological sciences. 2019;23(10):4089-96.

565. Yoon J, Um H-N, Jang J, Bae Y-A, Park W-J, Kim HJ, et al. Eosinophil Activation by Toll-Like Receptor 4 Ligands Regulates Macrophage Polarization. Frontiers in cell and developmental biology. 2019;7(329).

566. Mirjam P. Roffel C-AB, Maarten Van Den Berge, Ilse Boudewijn, Ken Bracke, Tania Maes, Irene Heijink. Unraveling the role of miR-223-3p in the regulation of airway inflammation in asthma and COPD. ERJ Open Research. 2019;5.

567. Roffel MP, Brandsma C-A, Van Den Berge M, Boudewijn I, Bracke K, Maes T, et al. Unraveling the role of miR-223-3p in the regulation of airway inflammation in asthma and COPD. ERJ Open Research. 2019;5(suppl 2):OP07.

568. Nobs SP, Kayhan M, Kopf M. GM-CSF intrinsically controls eosinophil accumulation in the setting of allergic airway inflammation. Journal of Allergy and Clinical Immunology. 2019;143(4):1513-24.e2.

569. Zhou Y, Zhang T, Yan Y, You B, You Y, Zhang W, et al. MicroRNA-223-3p regulates allergic in fl ammation by targeting INPP4A. Braz J Otorhinolaryngol. 2020.

570. Marshall AJ, Hou S, Wu X, Li H. Control of B cell activation and migration by PI 3-kinase: role of inositol polyphosphate 4-phosphatases. The Journal of Immunology. 2016;196(1 Supplement):198.5.

571. Siddiqui SH, Guasconi A, Vestbo J, Jones P, Agusti A, Paggiaro P, et al. Blood Eosinophils: A Biomarker of Response to Extrafine Beclomethasone/Formoterol in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. 2015;192(4):523-5.

572. George JN, Thoi LL, McManus LM, Reimann TA. Isolation Of Human-Platelet Membrane Microparticles From Plasma And Serum. Blood. 1982;60(4):834-40.

573. Gemmell CH, Sefton MV, Yeo EL. Platelet-Derived Microparticle Formation Involves Glycoprotein-IIb-IIIa - Inhibition By RGDS and a Glanzmann Thrombasthenia Defect. J Biol Chem. 1993;268(20):14586-9.

574. Noerholm M, Balaj L, Limperg T, Salehi A, Zhu LD, Hochberg FH, et al. RNA expression patterns in serum microvesicles from patients with glioblastoma multiforme and controls. BMC Cancer. 2012;12:11.

575. Alvarez ML, Khosroheidari M, Kanchi Ravi R, DiStefano JK. Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers. Kidney Int. 2012;82(9):1024-32.

576. Konoshenko MY, Lekchnov EA, Vlassov AV, Laktionov PP. Isolation of Extracellular Vesicles: General Methodologies and Latest Trends. BioMed Research International. 2018;2018:8545347.

577. Ridder K, Keller S, Dams M, Rupp AK, Schlaudraff J, Del Turco D, et al. Extracellular vesicle-mediated transfer of genetic information between the hematopoietic system and the brain in response to inflammation. PLoS biology. 2014;12(6):e1001874.

578. Ridder K, Sevko A, Heide J, Dams M, Rupp AK, Macas J, et al. Extracellular vesicle-mediated transfer of functional RNA in the tumor microenvironment. Oncoimmunology. 2015;4(6):e1008371.

579. Zomer A, Maynard C, Verweij FJ, Kamermans A, Schäfer R, Beerling E, et al. In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior. Cell. 2015;161(5):1046-57.

580. Das S, Ansel KM, Bitzer M, Breakefield XO, Charest A, Galas DJ, et al. The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research. Cell. 2019;177(2):231-42.

581. Srinivasan S, Yeri A, Cheah PS, Chung A, Danielson K, De Hoff P, et al. Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. Cell. 2019;177(2):446-62.e16.

582. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. J Extracell Vesicles. 2018;7(1):1535750.

583. Chiang CY, Chen C. Toward characterizing extracellular vesicles at a single-particle level. Journal of biomedical science. 2019;26(1):9.

584. Nolan JP, Duggan E. Analysis of Individual Extracellular Vesicles by Flow Cytometry. Methods Mol Biol. 2018;1678:79-92.

585. Mathieu M, Martin-Jaular L, Lavieu G, Théry C. Specificities of secretion and uptake of exosomes and other extracellular vesicles for cell-to-cell communication. Nat Cell Biol. 2019;21(1):9-17.

586. Cha DJ, Franklin JL, Dou Y, Liu Q, Higginbotham JN, Demory Beckler M, et al. KRAS-dependent sorting of miRNA to exosomes. eLife. 2015;4:e07197.

587. Mukherjee K, Ghoshal B, Ghosh S, Chakrabarty Y, Shwetha S, Das S, et al. Reversible HuR-microRNA binding controls extracellular export of miR-122 and augments stress response. EMBO reports. 2016;17(8):1184-203.

588. Santangelo L, Giurato G, Cicchini C, Montaldo C, Mancone C, Tarallo R, et al. The RNA-Binding Protein SYNCRIP Is a Component of the Hepatocyte Exosomal Machinery Controlling MicroRNA Sorting. Cell reports. 2016;17(3):799-808.

589. Villarroya-Beltri C, Gutiérrez-Vázquez C, Sánchez-Cabo F, Pérez-Hernández D, Vázquez J, Martin-Cofreces N, et al. Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nat Commun. 2013;4:2980.

590. Kaur S, Elkahloun AG, Arakelyan A, Young L, Myers TG, Otaizo-Carrasquero F, et al. CD63, MHC class 1, and CD47 identify subsets of extracellular vesicles containing distinct populations of noncoding RNAs. Sci Rep. 2018;8(1):2577.

591. Willemse BWM, Postma DS, Timens W, ten Hacken NHT. The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation. Eur Respir J. 2004;23(3):464.

592. Roffel MP, Bracke KR, Heijink IH, Maes T. miR-223: A Key Regulator in the Innate Immune Response in Asthma and COPD. Front Med (Lausanne). 2020;7:196-.

593. Cheng Y, Wang X, Wang P, Li T, Hu F, Liu Q, et al. SUSD2 is frequently downregulated and functions as a tumor suppressor in RCC and lung cancer. Tumor Biology. 2016;37(7):9919-30.

594. Turner MC, Chen Y, Krewski D, Calle EE, Thun MJ. Chronic obstructive pulmonary disease is associated with lung cancer mortality in a prospective study of never smokers. Am J Respir Crit Care Med. 2007;176(3):285-90.

595. Guan P, Yin Z, Li X, Wu W, Zhou B. Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. Journal of Experimental & Clinical Cancer Research. 2012;31(1):54.

596. "2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung." Brian L. Graham, Vito Brusasco, Felip Burgos, Brendan G. Cooper, Robert Jensen, Adrian Kendrick, Neil R. MacIntyre, Bruce R. Thompson and Jack Wanger. Eur Respir J 2017; 49: 1600016. Eur Respir J. 2018;52(5).

597. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. Proc Natl Acad Sci U S A. 2014;111(41):14888-93.

598. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, et al. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. International Journal of Cancer. 2010;127(12):2870-8.

599. Shi L, Kojonazarov B, Elgheznawy A, Popp R, Dahal BK, Böhm M, et al. miR-223-IGF-IR signalling in hypoxia- and load-induced right-ventricular failure: a novel therapeutic approach. Cardiovascular research. 2016;111(3):184-93.

600. Yoshida T, Delafontaine P. Mechanisms of IGF-1-Mediated Regulation of Skeletal Muscle Hypertrophy and Atrophy. Cells. 2020;9(9).

601. Hurst JR, Dickhaus J, Maulik PK, Miranda JJ, Pastakia SD, Soriano JB, et al. Global Alliance for Chronic Disease researchers' statement on multimorbidity. The Lancet Global Health. 2018;6(12):e1270-e1.

602. Rennard SI, Drummond MB. Early chronic obstructive pulmonary disease: definition, assessment, and prevention. Lancet. 2015;385(9979):1778-88.

603. Martinez FJ, Han MK, Allinson JP, Barr RG, Boucher RC, Calverley PMA, et al. At the Root: Defining and Halting Progression of Early Chronic Obstructive Pulmonary Disease. American Journal of Respiratory and Critical Care Medicine. 2018;197(12):1540-51.

604. Allinson JP, Hardy R, Donaldson GC, Shaheen SO, Kuh D, Wedzicha JA. Combined Impact of Smoking and Early-Life Exposures on Adult Lung Function Trajectories. American Journal of Respiratory and Critical Care Medicine. 2017;196(8):1021-30.

605. El-Emir E. DG. The BLF Early COPD Development Partnership Grant. Clinical Trials.gov: Imperial College London; 2018.

606. Stockley RA, Halpin DMG, Celli BR, Singh D. Chronic Obstructive Pulmonary Disease Biomarkers and Their Interpretation. American Journal of Respiratory and Critical Care Medicine. 2019;199(10):1195-204.

607. George L, Brightling CE. Eosinophilic airway inflammation: role in asthma and chronic obstructive pulmonary disease. Therapeutic advances in chronic disease. 2016;7(1):34-51.

608. Sennels HP, Jørgensen HL, Hansen AL, Goetze JP, Fahrenkrug J. Diurnal variation of hematology parameters in healthy young males: the Bispebjerg study of diurnal variations. Scand J Clin Lab Invest. 2011;71(7):532-41.

609. Nicholas B, Staples KJ, Moese S, Meldrum E, Ward J, Dennison P, et al. A novel lung explant model for the ex vivo study of efficacy and mechanisms of anti-influenza drugs. J Immunol. 2015;194(12):6144-54.

610. Kanayama M, He YW, Shinohara ML. The lung is protected from spontaneous inflammation by autophagy in myeloid cells. J Immunol. 2015;194(11):5465-71.

611. Staples KJ, Nicholas B, McKendry RT, Spalluto CM, Wallington JC, Bragg CW, et al. Viral Infection of Human Lung Macrophages Increases PDL1 Expression via IFNβ. PLoS ONE. 2015;10(3):e0121527.

612. Wallington JC, Williams AP, Staples KJ, Wilkinson TMA. IL-12 and IL-7 synergize to control mucosal-associated invariant T-cell cytotoxic responses to bacterial infection. J Allergy Clin Immunol. 2018;141(6):2182-95.e6.