

Epigenetic Markers as Predictors of Lung Disease Severity in Cystic Fibrosis

Dr Shivanthan Shanthikumar

MBBS, B Med Sci, FRACP, Dip Clin Pall Med

ORCID ID: 0000-0001-6000-3180

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

December 2020

Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences,
University of Melbourne

Department of Respiratory and Sleep Medicine, The Royal Children's Hospital

Respiratory Group, Murdoch Children's Research Institute

ABSTRACT

Cystic fibrosis (CF) is a multisystem disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Lung disease which is the major source of morbidity and mortality in CF, begins in infancy and persists and progresses in the first years of life. The most sensitive measure of CF lung disease is computed tomography (CT) and the majority of children show evidence of bronchiectasis, which represents irreversible structural lung disease, by five years of age. The rate with which lung disease develops and progresses varies greatly between people with CF. Accurate prediction of future lung disease severity in people with CF would facilitate significant improvements in clinical care, as well as provide helpful information to people with CF and their families, health authorities and CF researchers. It has been demonstrated that CF lung disease severity is determined by a combination of non-*CFTR* genetic and environmental factors. However, no robust predictive biomarkers have been developed. Epigenetic mechanisms are those which regulate gene expression but do not alter the DNA sequence itself. DNA methylation is the most widely studied epigenetic mechanism, and an individual's DNA methylation profile is determined by their underlying genotype and environmental exposures (including *in utero*). DNA methylation has been studied as a predictive biomarker in other childhood diseases, and given it is determined by the same factors which determine CF lung disease severity, is well positioned to act as a biomarker in CF. Despite this, DNA methylation based biomarkers of future lung disease severity have not been explored in CF.

The thesis presents a series of projects which aimed to identify DNA methylation based biomarkers of future CF lung disease severity. The projects in this thesis utilised biospecimens and clinical data from the world leading Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF) cohort.

In order to better understand previous attempts at identifying predictive biomarkers, Chapter Two and Three describe reviews of the evidence regarding non-*CFTR* genetic modifiers of CF lung disease and DNA methylation based biomarkers of future health outcomes in children. The reviews highlight the importance of appropriate study design when trying to identify predictive biomarkers, which then informed design of the subsequent epigenome wide association study (EWAS).

Chapters Four and Five describe the development of methods that were necessary for subsequent EWAS. In Chapter Four protocols for cryopreservation, and flow cytometry based phenotyping and sorting of specific cell populations were developed and validated for use with paediatric bronchoalveolar lavage (BAL). These methods were then used in Chapter Five to obtain purified populations of the common cell types in BAL, which were then used to develop reference epigenomes for these cell types. These reference epigenomes allowed adjustment for cell composition in the subsequent EWAS, which is a crucial step in DNA methylation studies involving samples with multiple cell types.

Chapters Six and Seven describe the EWAS studies that attempted to identify predictive biomarkers. Chapter Six involved genome wide DNA methylation profiling of BAL collected at six years of age, and assessed the ability to predict the presence or absence of bronchiectasis at nine years of age. No predictive biomarkers were identified. Chapter Seven involved genome wide DNA methylation profiling of BAL collected at one year of age, and assessed the ability to predict the presence or absence of bronchiectasis at five years of age. Seven predictive biomarkers were identified, which when assessed using area under the receiver operator curve analysis had extremely good performance as predictive biomarkers. Several of the predictive biomarkers were related to genes that are relevant to CF lung disease pathophysiology and hence may represent therapeutic targets.

By following best practice for EWAS, and in particular using tissue specific samples, adjusting for cell composition and assessing outcome using the gold standard measure, this thesis succeeded in identifying DNA methylation based biomarkers of future lung disease severity in CF. If these biomarkers are validated in external cohorts they could dramatically improve the care of children with CF around the world.

N.B. There are two files that have been submitted separate to this thesis document as supplementary files. They have been highlighted as appropriate in the thesis. In addition, the code used for statistical analysis is available at:

<https://atlassian.petermac.org.au/bitbucket/pages/OS/paed-cf-methylation/master/browse/docs/index.html>

DECLARATION

I hereby declare that:

- (i) this thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface
- (ii) due acknowledgement has been made in the text to all other material used
- (iii) the thesis is fewer than the maximum work limit in length, exclusive of tables, maps, bibliographies and appendices.



Dr Shivanthan Shanthikumar

Date: 1 July 2021

PREFACE

This thesis incorporates three original manuscripts arising from the research conducted during my PhD tenure and published in international peer-reviewed journals. The research for this thesis was made possible through the collaboration of my co-authors, whose contributions are outlined below. All co-authors provided consent for the publications to be included in the thesis, and each manuscript is accompanied by a completed co-author authorisation form.

Publications Included in this Thesis

Chapter 2

Shanthikumar S, Neeland MN, Saffery R, Ranganathan S. Gene modifiers of cystic fibrosis lung disease: a systematic review. *Pediatr Pulmonol*. 2019 Sep; 54(9):1356–1366.

Shivanthan Shanthikumar was the primary author and participated in the study design, literature review, manuscript drafting, drawing of figures, editing, revision and submission. He contributed 95% of the content for this publication. Melanie Neeland, Richard Saffery and Sarath Ranganathan participated in the review of the manuscript.

Chapter 3

Shanthikumar S, Neeland MR, Maksimovic J, Ranganathan SC, Saffery R. DNA methylation biomarkers of future health outcomes in children. *Mol Cell Pediatr*. 2020 Dec; 7(1):7.

Shivanthan Shanthikumar was the primary author and participated in the study design, literature review, manuscript drafting, drawing of figures, editing, revision and submission. He contributed 95% of the content for this publication. Melanie Neeland, Jovana Maksimovic, Sarath Ranganathan and Richard Saffery participated in the review the of manuscript.

Chapter 4

Shanthikumar S, Burton M, Saffery R, Ranganathan SC, Neeland MR. Single-cell flow cytometry profiling of BAL in children. *Am J Respir Cell Mol Biol* 2020 Aug; 63(2):152–159.

Shivanthan Shanthikumar was the primary author and participated in the study design, literature review, obtaining consent from patients and families, collection of bronchoalveolar lavage, laboratory experiments, analysis, manuscript drafting, editing, revision and submission. He contributed 60% of the content for this publication. Melanie Neeland was involved in laboratory work, analysis, drawing of figures, manuscript drafting and revision, contributing 35% of the content. Matthew Burton supported the laboratory work and manuscript revision, contributing 5% of the content. Richard Saffery and Sarath Ranganathan were involved in the review of the manuscript.

ACKNOWLEDGEMENTS

I would first like to acknowledge the people with cystic fibrosis and their families with whom I have had the privilege to work. Seeing firsthand how young people with cystic fibrosis handle the challenges of this disease is truly phenomenal and was the inspiration for this thesis. The Royal Children's Hospital Cystic Fibrosis Research Trust, in particular, was the sole reason I was able to pursue studies or a PhD in paediatric respiratory medicine. Without the hard work and generosity of the families involved in funding my training and research, I would not have been able to follow my chosen career path. I hope that I can repay the investment during my career.

I would also like to thank my supervisors, Sarath Ranganathan, Melanie Neeland, Jovana Maksimovic and Richard Saffery. I am eternally grateful to Sarath as my primary supervisor. I have witnessed exactly what it takes to be a world-class clinician-scientist and hope to emulate that. Despite the demands on his time, Sarath was always available to guide my research and was an ever-present source of support for my broader career during what has been a challenging time. Without the expertise and support offered by Melanie and Jovana, the completion of this thesis would not have been possible. I am indebted to their generosity with their time and knowledge and for always being supportive of my research. Working with them both has inspired me to pursue ongoing laboratory and omic research, and I am excited to continue working together in the coming years. Richard, at times when I was unsure whether the work in this thesis would succeed, you provided the knowledge and support to get me through. I feel so fortunate to have had the supervision of such a dedicated, talented and supportive team of people, and the mentorship they have provided over the last 4 years will shape the rest of my career.

I would also like to thank my colleagues, particularly those in the Department of Respiratory and Sleep Medicine, the cystic fibrosis team and the broader hospital. It is such a pleasure to work in an environment in which one is surrounded by constant support, enthusiasm and drive to provide high-quality care. In particular, to my friends who were also navigating the ups and downs of advanced training/PhD journeys, the ability to commiserate and celebrate together has made the last few years much more enjoyable.

This thesis would not have been possible without the work undertaken by those before me as part of the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF). The foresight and vision shown by those who established the study, along with the hard work and dedication to keep it running in the years since, are truly remarkable. I was fortunate to have access to the team for this thesis, and it has been a pleasure working with them in both Melbourne and Perth. I hope that my ongoing involvement in AREST CF will allow it to continue shaping care for infants and young children with cystic fibrosis.

I would like to acknowledge that this thesis was edited by Elite Editing and editorial intervention was restricted to Standards D and E of the *Australian Standards for Editing Practice*.

I would like to thank my friends and family. I have such a diverse group of friends who have constantly supported me during the time I have worked on this thesis. To my parents, I am forever indebted to the sacrifices you have both made so that I could be in the position I am in now. In particular, the gifts of a love of learning and the belief that I can do anything have been instrumental in guiding me through this thesis.

Most importantly, to Ella: the value of your love, support and encouragement cannot be put into words. During the time it took to complete this thesis, you have made sacrifices for my career, reassured me when things were not going well, celebrated with me when things were and made sure I laughed the entire way along. Above all else, I am so fortunate to share my life with you.

TABLE OF CONTENTS

Abstract	i
Declaration	iv
Preface	v
Acknowledgements	vii
Table of Contents.....	ix
List of Tables	xiii
List of Figures	xv
List of Abbreviations.....	xix
CHAPTER 1: Introduction.....	1
1.1 Cystic Fibrosis	1
1.2 Cystic Fibrosis Lung Disease	2
1.2.1 Morbidity and mortality related to cystic fibrosis lung disease.....	2
1.2.2 Pathophysiology of cystic fibrosis lung disease	3
1.2.3 Assessment of cystic fibrosis lung disease	5
1.2.3.1 <i>Clinical markers of cystic fibrosis lung disease</i>	5
1.2.3.2 <i>Investigations in cystic fibrosis lung disease</i>	6
1.2.3.3 <i>What is the gold standard marker of cystic fibrosis lung disease?</i>	9
1.2.4 Evolution of cystic fibrosis lung disease	11
1.2.5 Clinical significance of predicting cystic fibrosis lung disease.....	14
1.2.6 Determinants/predictors of lung disease.....	16
1.2.6.1 <i>CFTR genotype/phenotype correlation</i>	17
1.2.6.2 <i>Environment v. genetics</i>	18
1.2.6.3 <i>Gene modifiers</i>	19
1.2.6.4 <i>Current state of predictors of lung disease severity</i>	19
1.3 Epigenetics.....	20
1.3.1 Definition	20
1.3.2 Epigenetic mechanisms	20
1.3.2.1 <i>Histone modifications</i>	21
1.3.2.2 <i>Non-coding RNA</i>	22
1.3.2.3 <i>Higher-order chromatin structure</i>	23
1.3.3 DNA methylation	23
1.3.3.1 <i>How does DNA methylation develop?</i>	23
1.3.3.2 <i>How are DNA methylation patterns maintained?</i>	24
1.3.3.3 <i>What determines DNA methylation?</i>	25
1.3.3.4 <i>DNA methylation is influenced by genetics</i>	25
1.3.3.5 <i>DNA methylation is influenced by the environment</i>	26
1.3.3.6 <i>Genetics and environment interact to influence DNA methylation</i>	28
1.3.3.7 <i>DNA methylation in complex disease (including role in disease prediction)</i>	29
1.3.4 Review of methylation in cystic fibrosis	29
1.4 Project Aims and Hypotheses.....	31
CHAPTER 2: Non-CFTR Gene Modifiers of Cystic Fibrosis Lung Disease	32
2.1 Introduction	32

2.2 Published Manuscript	32
2.3 Supplementary Material	44
2.4 Research in Context	100
CHAPTER 3: DNA Methylation Biomarkers of Future Health Outcomes in Children	101
3.1 Introduction	101
3.2 Published Manuscript	101
3.3 Supplementary Material	113
3.4 Research in Context	113
CHAPTER 4: Flow Cytometry Profiling of Paediatric Bronchoalveolar Lavage Samples	114
4.1 Introduction	114
4.2 Published Manuscript	114
4.3 Supplementary Material	123
4.4 Research in Context	134
CHAPTER 5: DNA Methylation Reference Panel	135
5.1 Introduction	135
5.2 Methods	137
5.2.1 Ethics	137
5.2.2 Bronchoalveolar lavage collection and processing	137
5.2.3 Purification of individual cell types	138
5.2.4 DNA extraction	138
5.2.5 DNA methylation analysis	138
5.2.6 Data processing and statistical analysis	139
5.2.6.1 <i>Quality control</i>	139
5.2.6.2 <i>Data exploration</i>	139
5.2.6.3 <i>Identification of cell-specific DNA methylation profiles</i>	140
5.2.7 Validation	140
5.2.7.1 <i>Publicly available blood dataset</i>	140
5.2.7.2 <i>Bronchoalveolar lavage validation</i>	141
5.3 Results	141
5.3.1 Bronchoalveolar lavage collection	141
5.3.2 DNA extraction and methylation profiling	141
5.3.3 Initial quality control and data exploration	142
5.3.4 Identification of cell-specific DNA methylation profiles	146
5.3.5 Validation	148
5.3.5.1 <i>Accuracy of cell proportion estimates for artificial cell mixtures</i>	148
5.3.5.2 <i>Comparison of flow cytometry and reference-based DNA methylation estimation of cell composition of bronchoalveolar lavage samples</i>	150
5.4 Discussion	153
5.5 Research in Context	158
CHAPTER 6: Epigenome-Wide Association Study of Cystic Fibrosis Lung Disease Severity at 9 Years of Age	159
6.1 Introduction	159
6.2 Methods	162
6.2.1 Ethics	162
6.2.2 Subjects	162
6.2.3 Computed tomography classification	162

6.2.4 Subject matching	162
6.2.5 DNA extraction and genome-wide methylation analysis	163
6.2.6 DNA methylation analysis.....	163
6.3 Results	166
6.3.1 Subject selection	166
6.3.2 Subject matching	168
6.3.3 DNA extraction.....	168
6.3.4 DNA methylation analysis.....	169
6.3.4.1 <i>Quality control</i>	169
6.3.4.2 <i>Estimate cell composition</i>	170
6.3.4.3 <i>Exploring variations in the data</i>	172
6.3.4.4 <i>Differential methylation analysis</i>	177
6.4 Discussion.....	179
6.5 Appendix	184
CHAPTER 7: Epigenome-Wide Association Study of Cystic Fibrosis Lung	
Disease Severity at 5 Years of Age	187
7.1 Introduction	187
7.2 Methods	188
7.2.1 Ethics	188
7.2.2 Subjects.....	189
7.2.3 Computed tomography classification	189
7.2.4 Subject matching	190
7.2.5 DNA extraction and genome-wide methylation analysis	190
7.2.6 DNA methylation analysis.....	190
7.2.7 Performance of significant differentially methylated probes as a diagnostic test.....	192
7.2.8 Assessment of the functional impact of differentially methylated probes ..	193
7.3 Results	194
7.3.1 Subject selection	194
7.3.2 DNA methylation analysis.....	195
7.3.2.1 <i>Quality control</i>	195
7.3.2.2 <i>Estimating cell composition</i>	197
7.3.2.3 <i>Exploring variations in the methylation data</i>	198
7.3.2.4 <i>Identification of differentially methylated probes</i>	200
7.3.2.5 <i>Performance of significant differentially methylated probes as a diagnostic test</i>	204
7.3.2.6 <i>Assessment of functional effect of differentially methylated probes</i>	208
7.4 Discussion.....	212
7.5 Conclusions	217
7.6 Appendix	218
CHAPTER 8: Conclusions and Future Directions	244
8.1 Overview of the Thesis.....	244
8.2 Review of Significant Findings	245
8.3 Future Directions	250
8.3.1 Validation of predictive biomarkers	250
8.3.2 Approach to novel therapeutic targets	251
8.3.3 Multiomic approach to predictive biomarkers.....	252
8.3.3.1 <i>Genomics</i>	253
8.3.3.2 <i>Epigenomics</i>	254

8.3.3.3 <i>Transcriptomics</i>	255
8.3.3.4 <i>Proteomics</i>	256
8.3.3.5 <i>Metabolomics</i>	257
8.3.3.6 <i>Polyomic risk score</i>	258
8.3.4 Relevance of predictive biomarkers in an era of highly effective modulator therapy	259
8.4 Conclusions	260
References	261

LIST OF TABLES

Table 1.1. Summary of studies evaluating the evolution of structural lung disease in people with cystic fibrosis	13
Table 1.2. Benefits of the accurate prediction of lung disease severity in cystic fibrosis	15
Chapter 2 Appendix A: Search strategy (*MEDLINE searched twice to optimise yield).....	44
Chapter 2 Supplementary Table 1: Excluded papers	46
Chapter 2 Supplementary Table 2: Included papers	60
Chapter 2 Supplementary Table 3: Summary of results by gene	61
Table 5.1. Cell composition of bronchoalveolar lavage samples determined by flow cytometry	151
Table 6.1. Results of computed tomography review	167
Table 6.2: Subject characteristics	168
Table 6.3. Initial quality control of DNA samples.....	169
Table 7.1. Subject demographics in the no bronchiectasis and bronchiectasis groups.	194
Table 7.2. <i>CFTR</i> genotypes in the no bronchiectasis and bronchiectasis groups.	194
Table 7.3. Subject demographics in the no bronchiectasis and bronchiectasis groups after exclusion of poor-quality samples.....	195
Table 7.4. <i>CFTR</i> genotypes in the no bronchiectasis and bronchiectasis groups after exclusion of poor-quality samples.	195
Table 7.5. Significant differentially methylated probes identified using RUV-4 analysis.	202
Table 7.6. AUC analysis for each of the seven significant differentially methylated probes and composite scores	205
Table 7.7. Categorisation of results if methylation level of 0.83 or above at cg04477130 in 1-year-old BAL is used to predict 5-year-old CT outcome	207
Table 7.8. Categorisation of results if methylation level of 0.86 or above at cg04477130 in 1-year-old BAL is used to predict 5-year-old CT outcome	207
Table 7.9. Categorisation of results if composite methylation score in 1-year-old BAL (determined by multiplying six CpGs with uniform direction of	

effect) is 2.5×10^{-5} or below and is used to predict 5-year-old CT	
outcome	208
Table 7.10. Summary of epigenetic markers and regulatory region definitions.....	208
Table 7.11. Top 10 ranked probes in probe-wise linear model analysis.....	218

LIST OF FIGURES

Figure 1.1. Multifactorial pathophysiology of CF lung disease.	5
Figure 1.2. Development of structural lung disease in people with cystic fibrosis.	14
Figure 1.3. Top 10 research priorities in cystic fibrosis.	16
Figure 1.4. Structure of a chromosome.	21
Figure 1.5. DNA methylation in the early embryonic period.	24
Figure 1.6. DNA methylation-based biomarkers of cystic fibrosis lung disease severity.....	30
Figure 5.1. Initial multidimensional scaling plot, showing samples clustered together by cell type with no significant batch effect.	142
Figure 5.2. Multidimensional scaling plot post initial probe filtering, showing sex is responsible for the variation in the first component.	143
Figure 5.3. Multidimensional scaling plot after filtering out probes that mapped to sex chromosomes.....	144
Figure 5.4. Multidimensional scaling plots limited to cell-specific samples.	145
Figure 5.5. Multidimensional scaling plots showing the first four principal components after filtering out poor-performing probes.	146
Figure 5.6. Example code for the <i>estimateCellCounts2</i> function.	147
Figure 5.7. Heatmap demonstrating unique DNA methylation profiles for each individual cell type.	147
Figure 5.8. Artificial cell mixtures.	148
Figure 5.9. Comparison of known proportions in artificial cell mixtures and methylation-based estimated cell proportions.	149
Figure 5.10. Bland–Altman plots comparing known artificial cell mixture proportions (‘Truth’) and methylation-based estimated cell proportions (‘Estimate’).	150
Figure 5.11. Correlation of cell composition determined via flow cytometry (‘Flow’) and methylation-based estimates (‘Meth. est.’).	152
Figure 6.1. Optimal design of epigenome-wide association studies.	160
Figure 6.2. Summary of the analysis pipeline.	164
Figure 6.3. Heatmap demonstrating cell-specific probes (identified in Chapter 5) and used as negative control probes in the RUV-4 analysis.....	166

Figure 6.4. Initial data exploration showing that sex was responsible for the variation seen in the first principal component.	170
Figure 6.5. Data exploration following quality control.	170
Figure 6.6. Estimated cell proportions of subject samples.	171
Figure 6.7. Comparison of cell composition of the no bronchiectasis and bronchiectasis groups.	171
Figure 6.8. Scree plot demonstrating that most of the variation in the data was captured by the first five principal components.	172
Figure 6.9. Correlation between sample traits and top 20 principal components.....	173
Figure 6.10. Cell composition of samples for each pair.	174
Figure 6.11. MDS plots demonstrating the effect of cell composition on the top principal components.....	175
Figure 6.12. Comparison of BMI within pairs.	176
Figure 6.13. BMI distribution of subjects grouped by disease status.	176
Figure 6.14. Data re-exploration prior to RUV-4 analysis.	177
Figure 6.15. Data re-exploration prior to RUV-4 analysis but limited to only negative control probes, which are related to cell type.	178
Figure 7.1. Analysis pipeline.	191
Figure 7.2. Heatmap showing the methylation levels of cell-specific probes used as negative control probes in RUV analysis.	192
Figure 7.3. Initial data exploration showing that sex was responsible for the variation seen in the first principal component.	196
Figure 7.4. Data exploration after quality control, showing no clustering by disease status.	196
Figure 7.5. Estimated cell proportions of subject samples.	197
Figure 7.6. Comparison of cell composition in the bronchiectasis and no bronchiectasis groups.	198
Figure 7.7. Scree plot demonstrating that most of the variation in the data was captured by the first five principal components.	199
Figure 7.8. Correlation between sample traits and top 20 principal components.....	199
Figure 7.9. MDS plots demonstrating the effect of cell composition on the top principal components.....	200
Figure 7.10. Summary of significant differentially methylated probes.	201

Figure 7.11. MDS plots post RUV-4 adjustment, demonstrating the effect of cell composition on the top principal components.....	203
Figure 7.12. <i>p</i> -value histogram of both the unadjusted and adjusted RUV analysis...	204
Figure 7.13. Receiver operator curves for significant differentially methylated positions.....	205
Figure 7.14. Receiver operator curves for the composite scores.....	206
Figure 7.15. Gene expression of <i>HMOX2</i> from human foetal lung tissue from Descartes Human Gene Expression During Development.....	219
Figure 7.16. Gene expression of <i>HMOX2</i> using the National Institutes of Health Genotype–Tissue Expression Project.	219
Figure 7.17: Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.....	220
Figure 7.18. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	221
Figure 7.19. ATAC-seq data from foetal lung tissue.....	222
Figure 7.20. Gene expression of <i>OSBP2</i> from human foetal lung tissue from Descartes Human Gene Expression During Development.....	223
Figure 7.21. Gene expression of <i>OSBP2</i> using the National Institutes of Health Genotype–Tissue Expression Project.	223
Figure 7.22: Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.....	224
Figure 7.23. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	225
Figure 7.24. ATAC-seq data from foetal lung tissue.....	226
Figure 7.25. Gene expression of <i>LCP1</i> from human foetal lung tissue from Descartes Human Gene Expression During Development.....	227
Figure 7.26. Gene expression of <i>LCP1</i> using the National Institutes of Health Genotype–Tissue Expression Project.	227
Figure 7.27. Gene expression of <i>LRRC63</i> using the National Institutes of Health Genotype–Tissue Expression Project.	228
Figure 7.28. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	228
Figure 7.29. Data from UCSC Genome Browser, with data from Ensemble Regulatory Build, ENCODE and Blueprint added.....	229

Figure 7.30. ATAC-seq data from foetal lung tissue.....	230
Figure 7.31. Gene expression of <i>NEDD9</i> from human foetal lung tissue from Descartes Human Gene Expression During Development.....	231
Figure 7.32. Gene expression of <i>NEDD9</i> using the National Institutes of Health Genotype–Tissue Expression Project.	231
Figure 7.33. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.	232
Figure 7.34. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	233
Figure 7.35. ATAC-seq data from foetal lung tissue.....	234
Figure 7.36. Gene expression of <i>LINC01237</i> from human foetal lung tissue from Descartes Human Gene Expression During Development.....	235
Figure 7.37. Gene expression of <i>LINC01237</i> using the National Institutes of Health Genotype–Tissue Expression Project.	235
Figure 7.38. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.	236
Figure 7.39. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	237
Figure 7.40. ATAC-seq data from foetal lung tissue.....	238
Figure 7.41. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.	239
Figure 7.42. ATAC-seq data from foetal lung tissue.....	240
Figure 7.43. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added.	241
Figure 7.44. A graph of the methylation of probes located 10 kb either side of the significant DMP.....	242
Figure 7.45. ATAC-seq data from foetal lung tissue.....	243
Figure 8.1. Different omics approaches.....	253

LIST OF ABBREVIATIONS

AEC	Alveolar epithelial cell
AREST CF	Australian Respiratory Early Surveillance Team for Cystic Fibrosis
ASL	Airway surface liquid
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
AUC	Area under the receiver operating characteristics curve
BAL	Bronchoalveolar lavage
BMI	Body mass index
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CT	Computed tomography
DMP	Differentially methylated probe
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
ENCODE	Encyclopedia of DNA Elements
EWAS	Epigenome-wide association study
FACS	Fluorescence-activated cell sorting
FDR	False discovery rate
FEV	Forced expiratory volume
IL	Interleukin
LCI	Lung clearance index
MBW	Multiple breath inert gas washout
MDS	Multidimensional scaling
mQTL	Methylation quantitative trait loci
NPV	Negative predictive value
PPV	Positive predictive value
PRAGMA	Perth–Rotterdam annotated grid morphometric analysis
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RUV	Remove unwanted variation

RUV4	Remove unwanted variation, 4-step
RVRTC	Raised volume rapid thoracoabdominal compression
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
scRNA-seq	Single-cell RNA sequencing
SNP	Single-nucleotide polymorphism
SQN	Subset quantile normalisation
UCSC	University of California, Santa Cruz
UHRF	Ubiquitin-like, containing PHD and Ring finger domains
UK	United Kingdom
US	United States

Genes

<i>ABCC1</i>	Multidrug resistance-associated protein 1
<i>AURKA</i>	Aurora kinase 4
<i>CASS4</i>	Cas scaffolding protein family member 4
<i>CFTR</i>	Cystic fibrosis transmembrane conductance regulator
<i>HMOX1</i>	Heme oxygenase 1
<i>HMOX2</i>	Heme oxygenase 2
<i>HOXD</i>	Homeobox D
<i>LCP1</i>	Lymphocyte cytosolic protein 1
<i>LINC01237</i>	Long intergenic non-protein coding RNA 1237
<i>LRRC63</i>	Leucine-rich repeat-containing 63
<i>NEDD9</i>	Neural precursor cell-expressed developmentally downregulated protein 9
<i>OSBP2</i>	Oxysterol-binding protein 2

CHAPTER 1: Introduction

1.1 CYSTIC FIBROSIS

Cystic fibrosis (CF) was first described by Anderson¹ in 1938 and is the most common life-shortening autosomal recessive condition in the Caucasian population, affecting 1 in 2,500–3,500 live births. In Australia, approximately 3,300 people currently have CF.^{2 3}

CF is caused by a mutation in the CF transmembrane conductance regulator gene (*CFTR*), with over 1,000 different disease-causing mutations described.⁴ The *CFTR* gene is located on chromosome 7, is 250 kB in length and contains 27 coding exons.⁵ It encodes CF transmembrane conductance regulator (CFTR), a protein consisting of 1,480 amino acids that is expressed on the apical surface of multiple epithelial cells, including in the lungs, pancreas and sweat glands.⁶ CFTR regulates cellular chloride content, which in turn affects the movement of sodium and water, determining the consistency and viscosity of cell surface liquid.⁶

The consequences of defective CFTR function vary between organs. Clinical manifestations result from altered cell surface liquid with increased viscosity, causing obstruction of epithelium-lined ducts (i.e. the airways and pancreatic glands) and leading to irreversible tissue damage.⁶ Clinically, this causes exocrine pancreatic insufficiency, elevated sweat chloride and suppurative lung disease.⁷ Traditionally, the management of CF has consisted of multiple treatments aimed at ameliorating the consequences of defective CFTR function, including antibiotics and airway clearance for suppurative lung disease and enzyme replacement therapy for exocrine pancreatic insufficiency.⁸ Novel treatments aimed at correcting CFTR function have been developed; however, they are not available at this stage for all people with CF and do not obviate the need for traditional CF management.⁹ Currently, people with CF and their families carry a significant burden of therapy (an average of 2 hours per day for adult patients¹⁰) aimed at slowing the consequences of defective CFTR function. Despite this, people with CF suffer irreversible damage to organs such as the lungs and liver, which progresses over time and results in reduced life expectancy. The current median predicted survival for people with CF in the United Kingdom (UK) and United States (US) is 45.1 years and 41.6 years, respectively.^{11 12} The survival of a child born with CF today is difficult to accurately

predict, but in the US it is estimated to be 45.2 years, which is still well below the life expectancy of a child born without CF (78.9 years).^{11 13}

1.2 CYSTIC FIBROSIS LUNG DISEASE

Lung disease is the most significant complication of CF and has a multifactorial pathophysiology. The gold standard method for measuring lung disease is computed tomography (CT) assessment of lung structure, and studies using this method have clearly demonstrated that CF lung disease begins in infancy and persists and progresses throughout early life. However, the rate at which CF lung disease progresses varies greatly between individuals; thus, the ability to predict future lung disease severity in an individual with CF would be of significant benefit. Extensive work has been undertaken in an attempt to identify factors that can predict CF lung disease severity, with a particular focus on non-*CFTR* genetic modifiers; however, to date, no robust predictive biomarkers have been identified. This section provides further context for early-life CF lung disease and informs the design and conduct of the subsequent work of this thesis.

1.2.1 Morbidity and mortality related to cystic fibrosis lung disease

Lung disease is the main source of morbidity and mortality in people with CF. In addition to completing multiple daily treatments targeting lung disease,¹⁰ morbidity results from the use of intravenous antibiotics to treat pulmonary symptoms. In 2015, almost 50% of people in the UK with CF required treatment with intravenous antibiotics, with a median of 28 days of intravenous antibiotics per year.¹² Data from the US are similar, with 35% of people with CF requiring treatment with intravenous antibiotics and a mean of 31 days of intravenous antibiotics per year.¹¹ Australian data show that 45.8% of people with CF required admission to hospital for the treatment of pulmonary symptoms, with a mean of 20 days of treatment per year.²

Another marker of morbidity from CF lung disease is lung transplantation, which is done in people with end-stage lung disease. Worldwide, CF is the leading cause of lung transplantation in the paediatric age group¹⁴ and the third most common indication in adults.¹⁵ The major source of mortality in people with CF is lung disease, with registry data showing it causes 65–70% of CF-related deaths.^{2 11}

1.2.2 Pathophysiology of cystic fibrosis lung disease

The pathophysiology of CF lung disease is not fully understood, but is thought to be multifactorial (see Figure 1.1). Defective CFTR function not only affects the chloride concentration of airway surface liquid (ASL) but, given its regulatory actions on other channels (e.g. epithelial sodium channel, outwardly rectifying chloride channel and calcium-activated chloride channel), it also affects the movement of sodium, bicarbonate and water.⁶ The net effect of these changes is lung disease arising from a cycle of repeated infection and inflammation.¹⁶ Factors leading to repeated infection and inflammation include impaired mucociliary clearance (attributable to the ‘low volume’ hypothesis, mucus tethering and impaired mucin expansion), impaired innate bactericidal activity of ASL, immune dysregulation and congenital airway abnormalities.

It is generally accepted that impaired mucociliary clearance of mucus predisposes the lungs to infection.¹⁷ However, the exact cause of impaired mucociliary clearance is unclear. The low volume hypothesis proposes that ASL has a normal ion concentration but is dehydrated because of excessive sodium and water absorption, resulting in chloride absorption via passive paracellular pathways and viscous ASL with impaired clearance.¹⁸ ¹⁹ Using a pig model of CF, Hoegger et al.²⁰ demonstrated that mucus was tethered to submucosal glands, impairing mucociliary clearance. Others have hypothesised that low bicarbonate in ASL impairs mucin expansion, hence mucociliary clearance.²¹

Another factor predisposing the CF lung to repeated infection is defective CFTR, impairing the innate bactericidal properties of ASL. In contrast to the low volume hypothesis, the high salt hypothesis suggests that ASL has a high sodium and chloride content, which has been shown to impair its bactericidal properties.^{22 23} Pezullo et al.²⁴ argue that pH rather than sodium/chloride concentration is the primary issue, demonstrating in a porcine model that ASL is more acidic in CF than in controls and that low pH impairs the bactericidal action of ASL. However, in direct contradiction of this hypothesis, more recent data from human CF subjects suggests that ASL is not acidic.²⁵

Inflammation independent of infection is a major contributor to CF lung disease. Although its cause is unclear, immune dysregulation²⁶ and a baseline proinflammatory state, even in the absence of infection,^{16 27 28} is evident in people with CF. Tirouvanziam et al.²⁷ implanted human foetal trachea into mice, finding that compared with controls,

CF-affected trachea had a baseline inflammatory imbalance (i.e. higher interleukin (IL)-8 levels in airway liquid and abnormal localisation of leukocytes) prior to infection and a more exaggerated response to infection with *Pseudomonas aeruginosa*. This finding is further supported by analysis of bronchoalveolar lavage (BAL) fluid, which shows that people with CF have an altered cytokine profile²⁶ and elevated neutrophil count.²⁸ In addition, data from the CF ferret model showed that even in the absence of infection there was evidence of mucus plugging and inflammation which results in structural lung disease.²⁹ However, data from pig models of CF contradict the notion of a baseline proinflammatory state, with CF piglets having normal IL-8 levels in ASL and no evidence of inflammation prior to infection.³⁰ Stoltz et al.³¹ found that piglets with CF had impaired ability to eradicate *Staphylococcus aureus* compared with unaffected piglets, hypothesising that inflammation in CF is caused by low-grade bacterial stimulation. The balance of evidence, comprising studies involving human tissues and models and a more rigorous assessment of inflammation, appears to support the presence of a baseline proinflammatory state.

There is evidence that congenital airway abnormalities, likely caused by defective CFTR in utero, are more common in those with CF. Data from both pig and human studies show that in newborns with CF, the trachea is narrow and abnormally shaped.³⁰ It is unknown whether congenital tracheal abnormalities contribute to CF lung disease, but tracheal abnormalities could exacerbate lung disease by worsening airway obstruction and impairing cough clearance, predisposing people with CF to infection. Fischer et al.³² established an association between tracheomalacia and increased lung disease severity, evidenced by increased rates of *Pseudomonas aeruginosa* infection and more significant airway obstruction. However, a causative relationship has not been shown.

The multifactorial predisposition towards infection and inflammation causes lung damage, predisposing the lung to further infection and inflammation and establishing a vicious cycle that results in progressive lung disease.

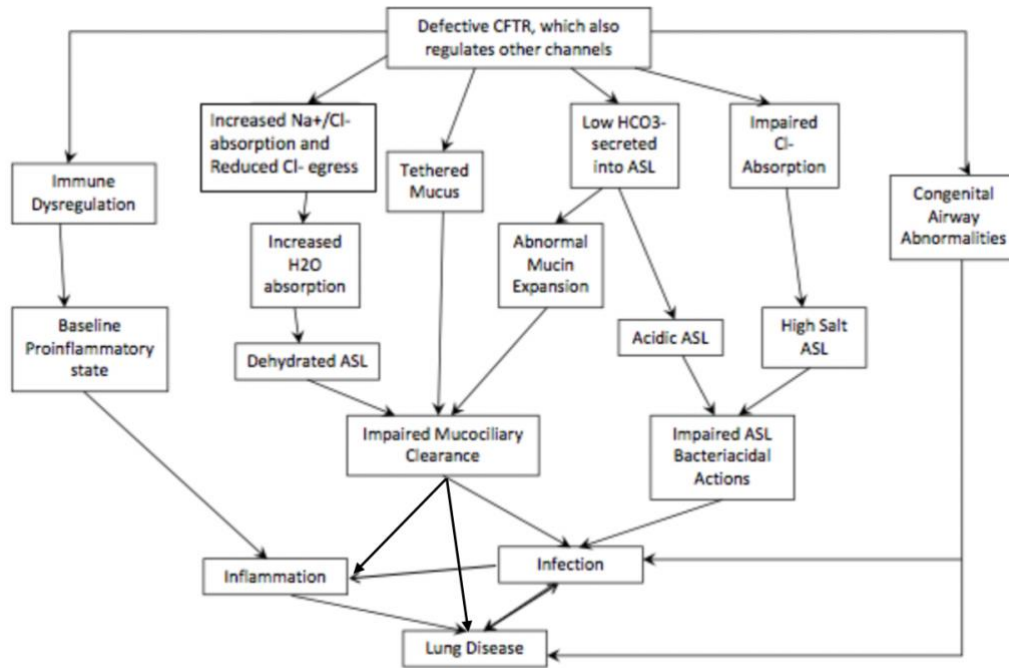


Figure 1.1. Multifactorial pathophysiology of CF lung disease. Multiple suggested mechanisms all lead to a cycle of infection and inflammation, leading to irreversible lung disease.

1.2.3 Assessment of cystic fibrosis lung disease

This section describes and compares the various methods used to assess the severity of CF lung disease, concluding that CT assessment of lung structure is the gold standard measure of CF lung disease.

1.2.3.1 Clinical markers of cystic fibrosis lung disease

CF lung disease can be measured using clinical markers such as daily symptoms, clinical signs, pulmonary exacerbation rate and quality of life. Daily symptoms such as a regular cough, increasing sputum production and reduced exercise tolerance and clinical signs such as digital clubbing and lung auscultation abnormalities are signs of advanced CF lung disease. In particular, people with CF suffer from periods of increased symptoms and signs, known as pulmonary exacerbations.³³ Although there is no consensus definition of pulmonary exacerbations, these episodes commonly involve symptoms such as increased cough, increased sputum production and reduced exercise tolerance and are managed with more intensive treatment such as hospital admission for intravenous antibiotics.³⁴ The number of pulmonary exacerbations is correlated with the rate of

decline in lung function and is used as an end point in intervention trials. Given that survival has improved for people with CF, health-related quality of life has emerged as an important marker of CF lung disease, and CF-specific assessment tools have been developed.^{35 36} A recent systematic review found a correlation between quality of life and lung function, confirming that health-related quality of life adds valid information when assessing CF lung disease.³⁷

1.2.3.2 Investigations in cystic fibrosis lung disease

Investigations of CF lung disease severity involve the evaluation of either lung structure (using histological and radiological techniques) or function (using pulmonary function tests).

1.2.3.2.1 Lung structure

Histologically, CF lung disease causes multiple abnormalities. In a report from 1976, Bedrossian et al.³⁸ reviewed autopsies of 82 people who had died from CF between 1954 and 1972. Common findings included bronchiectasis (pathological dilation of airways), mucus plugging and inflammatory infiltrates. Currently, histology plays only a small role because information about lung structure can be ascertained using non-invasive radiological techniques such as CT. In CF lung disease, CT findings (such as bronchiectasis) have been shown to correlate with histological findings at the time of lung lobe resection, reinforcing the validity of CT in CF lung disease.³⁹

When assessing CF lung disease, the changes seen on a CT scan can be described in general terms or using a variety of validated CF-specific scoring systems.⁴⁰ The main advantage of CF-specific scoring systems for CT scans is that they provide standardised results that enable the monitoring of CT scan changes over time as well as valid comparisons between subjects.

CT changes are correlated with a number of important clinical end points, including mortality, pulmonary exacerbations, quality of life and *Pseudomonas aeruginosa* infection. In a study of 366 children and adults awaiting lung transplantation, CT scores were able to predict mortality.⁴¹ Multiple studies have demonstrated a correlation between CT scores and pulmonary exacerbations. Brody et al.⁴² evaluated children aged 6–10 years, showing that baseline CT scores predicted pulmonary exacerbation rates, and

changes in CT scores over 2 years correlated with the number of pulmonary exacerbations. In another study of 72 children and adolescents, baseline CT scores were correlated with the number of pulmonary exacerbations and quality of life over a 12-month period.⁴³ These findings were replicated by another study involving 115 subjects aged 5–20 years.⁴⁴ Bortoluzzi et al.⁴⁵ monitored 68 children for 6 years and found that the baseline CT score was the best predictor of exacerbation rates. They also found that CT scores may correlate with chronic *Pseudomonas aeruginosa* infection.

The longitudinal progression of CF lung disease has also been studied using CT scans. This has primarily been done by the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF), which recruited a group of children with CF from Melbourne and Perth in Australia. The majority of children in the cohort were diagnosed at newborn screening and have since been followed up intensively with annual assessments of pulmonary infection and inflammation (via BAL) and lung function and regular assessments of structural lung disease via CT. The researchers have shown that CT changes are usually detectable early in life and persist and progress over time, especially in the presence of lung infection and inflammation.⁴⁶ de Jong et al.⁴⁷ have also shown that CT changes, in particular bronchial wall thickening, persist and worsen over time. CT scans have been used as an end point in several studies assessing the efficacy of treatments aimed at ameliorating CF lung disease. These include studies examining the efficacy of inhaled mucolytic therapy,⁴⁸ CFTR modulators,^{49 50} intravenous antibiotics, airway clearance for pulmonary exacerbation⁵¹ and BAL-directed therapy.⁵²

1.2.3.2.2 Lung function

Lung function in CF lung disease can be assessed using a number of techniques. One method of analysing lung function is by examining forced expiratory manoeuvres. This is usually undertaken with spirometry, particularly the measurement of forced expiratory volume (FEV) in 1 second (FEV₁), the most commonly used test in the assessment of lung function in CF.⁵³ Compared with healthy controls, people with CF have reduced FEV₁, which may be tracked from childhood to adulthood.⁵⁴ In addition, FEV₁ has been used to predict mortality.⁵⁵ Advantages of spirometry include that it is widely available, and technical standards have been developed to ensure comparable results between centres.⁵⁶ However, it is limited by its sensitivity, which has been shown to be inferior to other measures of lung function in preschool, child, adolescent and adult age groups.⁵⁷⁻⁵⁹

Another limitation of spirometry is the age at which it can be performed. Traditionally, spirometry was not performed in children under 6 years of age because of issues with patient technique and compliance; however, more recent evidence has shown that it can be reliably performed in some subjects as young as 3 years of age.⁶⁰

To examine forced expiratory manoeuvres in patients less than 3 years of age, who are unable to perform traditional spirometry, raised volume rapid thoracoabdominal compression (RVRTC) can be used. This involves inflating the lungs of sedated infants to a given pressure (usually 30 cm H₂O), then rapidly inflating a fitted vest, causing forced expiration of the lungs to residual volume.⁵⁴ This process enables the measurement of a full FEV curve. Given the shorter emptying time of infant lungs, FEV in half a second (FEV_{0.5}) rather than FEV₁ is often calculated. FEV_{0.5} has been shown to decline in infants with CF, particularly when associated with infection^{16 61} and pulmonary inflammation measured in BAL fluid,¹⁶ and reductions in lung function as measured by RVRTC persist and predict reduced lung function in school-age children.^{54 62} Strengths of the RVRTC technique include published standardised protocols and high sensitivity, while its limitations include the need for sedation (and associated side effects), expensive equipment and technical demands.⁵³

Another method of assessing lung function is to evaluate the efficiency of ventilation distribution within the lungs. This is most commonly done via the multiple breath inert gas washout (MBW) technique⁵⁴ in which the subject tidally breathes a marker gas, either a resident gas (e.g. nitrogen), which is washed out using 100% oxygen, or a non-resident inert gas (e.g. sulphur hexafluoride or helium), which is washed out with room air. During MBW, the lung clearance index (LCI) can be used to assess global abnormalities of ventilation distribution by measuring the number of lung volume turnovers required to reduce the marker gas to 1/40th of its starting value. An elevated LCI result indicates impaired ventilation efficiency, also known as ventilation inhomogeneity.⁵⁴ More specific information regarding the relative contributions of convection and diffusion to abnormalities can be obtained through analysis of the phase III slope.⁵³ An increased LCI has been demonstrated in preschoolers, school-age children, adolescents and adults with CF.^{57-59 63} LCI is correlated with signs of structural lung disease in school-age children^{58 64} as well as pulmonary infection and inflammation.⁶³ Vermeulen et al.⁶⁵ found that LCI predicted time to next pulmonary exacerbation and pulmonary exacerbation rate over the next 12 months as well as correlating with quality of life measures. LCI also improves

with treatment, as demonstrated in the use of CFTR modulators,⁶⁶⁻⁶⁸ nebulised hypertonic saline⁶⁹ and intravenous antibiotics for pulmonary exacerbations.⁷⁰ Longitudinal studies using LCI are limited. Fuchs et al.⁷¹ performed serial LCI measurements, CT and spirometry in 36 subjects across a wide age range (6–53 years), finding that LCI at baseline predicted structural lung disease status 3 years later. Stanojevic et al.⁷² obtained serial LCI measurements over 12 months from healthy children and children with CF aged between 2.5 and 6 years. The LCI values in children with CF were higher at baseline and deteriorated over the 12 months, while they were stable in controls. Given that phase III slope analysis is complicated and difficult to perform, it is not used as widely as LCI. Advantages of MBW with LCI are that it is performed during tidal breathing, can be used in all age groups, has a high sensitivity and can be benchmarked against international standards.⁵³ Disadvantages include the relative lack of commercially available equipment, meaning that most CF centres are unable to perform the test.⁵³ Moreover, a relative lack of longitudinal studies of MBW and LCI in CF limit its use.^{53 73}

Other pulmonary function tests such as the forced oscillation technique and interrupter technique have been performed in infants and children with CF; however, because they have limited use, they will not be discussed further.⁷⁴

1.2.3.3 What is the gold standard marker of cystic fibrosis lung disease?

Given the multiple methods of assessing lung disease in CF, including clinical markers, CT and lung function, which is the best? The ideal marker is easily and inexpensively collected and has no risks or side effects while identifying lung disease as early as possible. Clinical markers of lung disease such as daily symptoms, pulmonary exacerbation rate and health-related quality of life can be easily collected and have no side effects but are not as sensitive as other measures of lung disease, hence often fail to identify lung disease. One of the first illustrations of this was work by Long et al.,⁷⁵ who compared a group of 34 infants with CF and minimal clinical symptoms or signs with a control group. They found that those with CF had evidence of structural lung disease with bronchial wall thickening and airway dilation. Sly et al.⁷⁶ further demonstrated the poor sensitivity of clinical symptoms, finding that in a group of 57 infants newly diagnosed with CF, 84% had no symptoms, but 80% had evidence of structural lung disease. Ranganathan et al.⁷⁷ demonstrated that infants newly diagnosed with CF had reduced lung function, even in the absence of respiratory symptoms or signs. Given these findings,

although clinical markers may yield useful information, they are too insensitive to be used in isolation, particularly in infants and children with mild disease.

In terms of lung function tests, MBW with calculation of LCI has been shown to be the most sensitive.⁵⁷⁻⁵⁹ Aurora et al.⁵⁷ found that 22 of 30 children with CF aged 2–5 years had an abnormal LCI, but only four had abnormal spirometry. Gustafsson et al.⁵⁸ examined 44 subjects aged 5–19 years and found that LCI was more sensitive than FEV₁, with 31 subjects having abnormal LCI but only seven having reduced FEV₁. Horsley et al.⁵⁹ tested 33 adults with CF, finding that in all cases of low FEV₁, LCI was also abnormal, but in 10 subjects with normal FEV₁, nine had an elevated LCI.

CT technology is widely available and, like MBW, can be performed in any age group. Its major disadvantages are exposure to ionising radiation and the need for general anaesthesia to obtain high-quality results in infants and preschool-age children. General anaesthesia is necessary in these age groups because lung volume has been shown to affect CT results, and children in this age group lack the coordination to perform maximal inspiration and expiration.⁷⁸ CT uses ionising radiation, which increases a patient's lifetime risk of cancer; however, the use of modern CT machines and scanning techniques and limiting scans to every 1–2 years means the risk is likely to be negligible.⁷⁹

Given that both MBW and CT have advantages and disadvantages, choosing the gold standard marker of CF lung disease requires a comparison of their sensitivity in detecting CF lung disease. Lung function abnormalities result from altered lung structure; thus CT, which directly measures structural changes such as bronchiectasis, will always be an extremely sensitive measure of CF lung disease. As discussed, given the disadvantages of CT, it may be possible to use LCI instead if it could provide a comparable assessment of CF lung disease. In school-age children and adults with CF, LCI has excellent sensitivity compared with CT. However, in preschool children and infants, it is insensitive compared with CT. In a cross-sectional study, Ramsey et al.⁸⁰ compared LCI and CT scans in 119 children aged 0–16 years and found that LCI correlated well with CT in preschool and school-age children but was insensitive in infancy. Another study of 49 infants conducted by the same group generated similar findings, with LCI having poor association with CT in early CF lung disease.⁸¹ A cross-sectional study of 60 children with CF aged 6–10 years showed that of the 53 children who had had both CT and LCI, 43 had concordant results. Of the 10 discordant results, five subjects had normal LCI with

abnormal CT, and five had normal CT with abnormal LCI.⁶⁴ Gustafsson et al.⁵⁸ showed the equivalence of LCI and CT in a retrospective study of 44 children with CF aged 5–19 years, while Ellemunter et al.⁸² showed that LCI and CT were correlated in 34 subjects aged 6–26 years.

Given that CT is a direct and sensitive measure of structural lung disease in all age groups and is supported by a larger body of evidence, including longitudinal studies, it is generally considered the gold standard marker of CF lung disease, with clinical markers and lung function assessment providing complementary information.^{64 80}

1.2.4 Evolution of cystic fibrosis lung disease

There is now strong evidence that CF lung disease develops in early life, which is in keeping with the early origins of chronic respiratory disease concept originally proposed in 1977⁸³ which has recently attracted attention.⁸⁴

Because CF is typically diagnosed shortly after birth, either as a result of newborn screening or from clinical features such as meconium ileus, this has facilitated the collection of information on the health of the lungs in early life.

Evidence shows that the onset of lung disease is in the first few months of life. In an early study of post-mortem samples, 21 neonates with meconium ileus who had passed away in the first 3 weeks of life were compared with deceased controls of the same age group, revealing no difference in the size of bronchial mucus glands.⁸⁵ However, there was a higher rate of pulmonary infection in the meconium ileus group. Given that around 80% of people with meconium ileus have CF,⁸⁶ and all meconium ileus cases in the study had the characteristic CF pancreatic changes, this study serves as evidence that the lungs of infants with CF are vulnerable to infection and lung disease from early life. More recent work has shown evidence of structural lung disease at diagnosis. In a study of subjects diagnosed with CF via newborn screening, 57 infants (median age of 3.6 months) underwent a CT scan, and 80.7% of them had evidence of structural lung disease.⁷⁶ Of note, 21.1% had evidence of bacterial infection, and 77.7% had evidence of pulmonary inflammation from BAL conducted at the same time as the CT. There is conflicting evidence regarding pulmonary function at the time of diagnosis. This question was first addressed by Ranganathan et al.,⁷³ who found that in a group of infants diagnosed clinically rather than via newborn screening, lung function was reduced. In a study of 68

infants with CF diagnosed via newborn screening and 40 controls who underwent lung function assessment using the RVRTC technique, Linnane et al.⁸⁷ found that FEV_{0.5} was significantly lower in the CF group. However, because testing was performed over a wide range of ages (6–131 weeks), the authors converted FEV_{0.5} scores into z-scores based on age at time of testing, finding no difference in lung function between CF infants and controls at birth as measured by FEV_{0.5}. In a post-hoc analysis, the authors estimated that there was no difference between the two groups up until 6 months of age, but after 6 months, FEV_{0.5} was significantly lower in CF infants. In a more recent study, 71 infants newly diagnosed with CF via newborn screening and 54 controls underwent MBW with calculation of LCI and RVRTC at an average age of 11.4 weeks and 12.2 weeks for the CF and control groups, respectively.⁸⁸ The authors found several measures of reduced lung function, including elevated LCI and reduced FEV_{0.5}. Because this study was conducted solely in newly diagnosed infants and did not utilise statistical modelling to calculate lung function, it provides stronger evidence regarding lung function at diagnosis than does the work by Linnane et al., supporting the idea that lung function is impaired.

The most robust evidence regarding the evolution of lung disease from diagnosis has been provided by the AREST CF cohort. Utilising CT, the researchers have shown that the point prevalence of bronchiectasis increased from 29.3% at 3 months to 31.5% at 12 months, 44.0% at 2 years and 61.5% at 3 years.⁸⁹ They have also shown that bronchiectasis and air trapping persist and progress in the majority of subjects over a 12-month period.⁴⁶ The AREST CF team has also assessed the progression of lung function impairment in CF lung disease. Pillarisetti et al.¹⁶ performed RVRTC at 20 weeks, 1 year and 2 years of age, finding that lung function, FEV_{0.5} in particular, declined over time. This finding correlates with initial work from the London Cystic Fibrosis Collaboration, which used RVRTC to assess children at a mean age of 28.4 weeks and 59.0 weeks, finding that mean FEV_{0.5} declined by 20% over this period.⁹⁰ When children with CF were compared with controls at initial testing, FEV_{0.5} was reduced, and there was no catch up in lung function by the second test. More recently, the collaboration assessed a group of 72 infants with CF (diagnosed via newborn screening, differing from the cohort in the first study) and 44 controls.⁹¹ Again they found that the CF infants had reduced lung function at 3 months; however, FEV_{0.5} improved in CF infants over the first year of life, and at 12 months of age, although FEV_{0.5} was still significantly lower in the CF infants, the difference between the two groups had reduced. There was no change in the difference

between CF infants and controls at 12 months for other measures of lung function such as LCI. By school age, most children with CF have established evidence of structural lung disease. Multiple studies have shown that around 60% of people with CF around the age of 5 years will have bronchiectasis, rising to 80–90% in early adolescence (see Table 1.1 and Figure 1.2).

In summary, CF lung disease evolves in early life, with evidence of abnormal structure and function soon after diagnosis, which progresses over the first few years of life. As a result, the majority of people with CF will have structural lung disease in the form of bronchiectasis by the age of 5 years, and few are free of bronchiectasis by early adolescence.

Table 1.1. Summary of studies evaluating the evolution of structural lung disease in people with cystic fibrosis

Source	N	Age in years* (range)	Abnormal CT	Bronchiectasis
Sly et al. ⁸⁹	126	0.25	68.0%	29.3%
Sly et al. ⁷⁶	57	0.3	80.7%	18.6%
Hall et al. ⁸¹	49	0.7	55%	27%
Ramsey et al. ⁸⁰	42	0.94	58%	20%
Davies et al. ⁹²	65	1	42%	24%
Sly et al. ⁸⁹	108	1.17	68.5%	31.5%
Sly et al. ⁸⁹	74	2.17	71.6%	44.0%
Long et al. ⁹³	64	2.5	87%	30%
Sly et al. ⁸⁹	78	3.20	69.2%	61.5%
Mott et al. ⁹⁴	62	3.4	85%	61%
Stiglbauer et al. ⁹⁵	23	4.8 [#]	91%	65%
Wainwright et al. ⁵²	153	5		57%
Ramsey et al. ⁸⁰	39	5.42	85%	69%
Demirkazik et al. ⁹⁶	40	6 (0.5–18)	100%	65%
Brody et al. ⁹⁷	60	(6–10)	75%	35%
Owens et al. ⁶⁴	60	7.8	85%	—
Marchant et al. ⁹⁸	16	9.3 [#] (2.8–11.9)	100%	100%
Ramsey et al. ⁸⁰	38	9.77	94%	69%
Helbich et al. ⁹⁹	117	12.2 (0.25–32)		80.3%
Gustafsson et al. ⁵⁸	44	12 (5–19)	61%	59%
Maffessanti et al. ¹⁰⁰	36	13(5–28)	92%	89%

Note: CT: computed tomography; *values expressed as mean unless otherwise stated; [#]median.

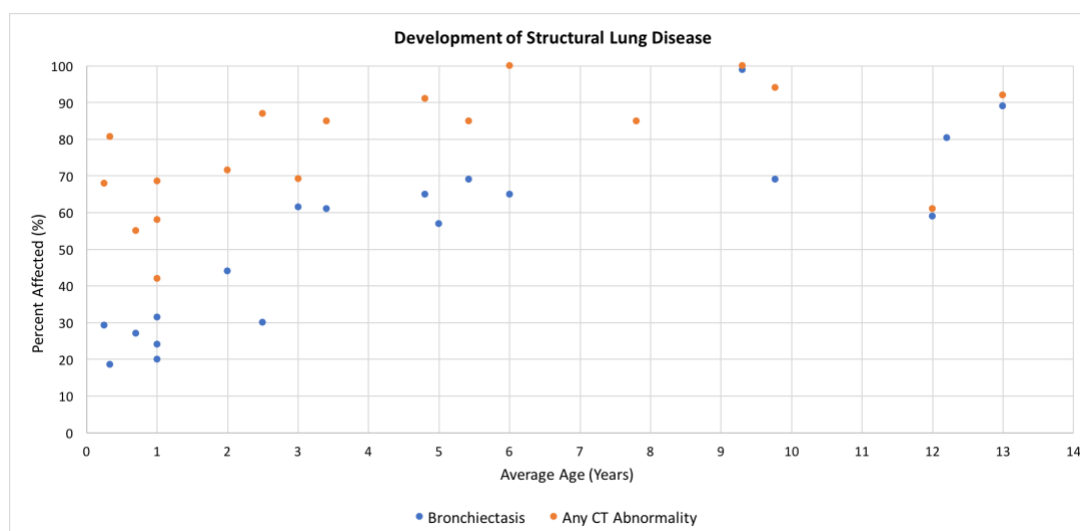


Figure 1.2. Development of structural lung disease in people with cystic fibrosis.

1.2.5 Clinical significance of predicting cystic fibrosis lung disease

As discussed, CF lung disease is a major cause of morbidity and mortality, has a multifactorial pathophysiology, can be assessed using CT and other tools, is present at birth and progresses over the first few years of life. However, there is variability in the onset and rate of deterioration of lung disease.¹⁰¹ Currently, there are no robust clinically applicable tools for predicting the onset and rate of deterioration of CF lung disease or indeed no way of predicting mild/minimal lung disease.

The ability to predict the progression of an individual's lung disease in early life would be beneficial for many reasons (see Table 1.2). For people with CF and their families, this information would allow for a more personalised understanding of the disease and its implications for the future. This may help them cope with the large treatment burden and assist with decisions such as genetic counselling and family planning. In contrast, those with minimal disease would have a better quality of life because of a reduced burden of care. For the CF multidisciplinary team, this information would enable the delivery of more personalised care, with individuals predicted to have more severe lung disease having more intensive management (i.e. earlier prescription of mucolytic therapy, lower threshold for inpatient admissions, more liberal use of antibiotics and more frequent airway clearance), and those predicted to have minimal lung disease having less intensive management (i.e. reduced frequency of airway clearance and prophylactic antibiotics). For governments or health authorities with limited resources to fund CF care, having an accurate understanding of likely morbidity and mortality from lung disease could aid in

a more tailored resource allocation. New CF therapies are expensive and struggle to satisfy current health economic assessments.¹⁰² However, if health authorities can target these new therapies at those most at risk of significant lung disease, overall costs may be reduced. Finally, for researchers, being able to predict an individual's likely lung disease course would enable the identification of subgroups who are most likely to benefit from new therapies. In addition, the underlying mechanism of a predictive tool may reveal a novel therapeutic target for the treatment of CF lung disease. For example, if the predictive biomarker was a gene or protein involved in inflammation and was associated with more severe disease, then interventions that inhibit this aspect of the inflammatory response may be developed. Likewise, if the biomarker were associated with mild lung disease, then therapies that promote a similar inflammatory response could be beneficial.

Table 1.2. Benefits of the accurate prediction of lung disease severity in cystic fibrosis

Group	Benefits of identifying risk of significant disease	Benefits of identifying risk of mild disease
Patient/family	Better understanding of the likely disease course and the need for increased care Better-informed decisions about family planning	Reduced burden of care with resultant improved quality of life Reduced exposure to iatrogenic risks Better-informed decisions about family planning
Cystic fibrosis multidisciplinary team	Implementation of earlier/more aggressive treatments Increased monitoring/surveillance	Reduced burden of care Less frequent monitoring, improving patient quality of life and allowing the healthcare team to focus their resources elsewhere
Health authorities	Allocation of resources (new treatments, more carers) to patients most likely to benefit	Avoidance of resource allocation to patients whose disease trajectory is unlikely to be altered
Researchers	Identification of subjects most likely to benefit from new interventions Potential identification of new therapeutic targets Explanation of heterogeneous responses to new treatments	Potential identification of new therapeutic targets Explanation of heterogeneous responses to new treatments

The importance of predicting lung disease severity has been highlighted by the publication of the top 10 research priorities in CF (see Figure 1.3), which were determined by an international community of healthcare professionals and lay people involved in CF.¹⁰³ The accurate prediction of lung disease severity would assist with the research of at least seven of the 10 priorities.

- 1. What are the effective ways of simplifying the treatment burden of people with cystic fibrosis (CF)?**
2. How can we relieve gastrointestinal (GI) symptoms such as stomach pain, bloating and nausea in people with cystic fibrosis?
- 3. What is the best treatment for non-tuberculous mycobacterium in people with CF (including when to start and what medication)?**
- 4. Which therapies are effective in delaying or preventing the progression of lung disease in early life in people with CF?**
5. Is there a way of preventing CF-related diabetes in people with CF?
- 6. What effective ways of motivation, support and technologies help people with CF improve and sustain adherence to treatment?**
- 7. Can exercise replace chest physiotherapy for people with CF?**
- 8. Which antibiotic combinations and dosing plans should be used for CF exacerbations, and should antibiotic combinations be rotated?**
- 9. Is there a way of reducing the negative effects of antibiotics (e.g. resistance risk and adverse symptoms) in people with CF?**
- 10. What is the best way of eradicating *Pseudomonas aeruginosa* in people with CF?**

Figure 1.3. Top 10 research priorities in cystic fibrosis.¹⁰³ Priorities highlighted in bold would be aided by the accurate prediction of lung disease severity.

Previous studies, reviewed below, have attempted to identify a predictor of CF lung disease. However, in summary, there is no tool to date that can accurately predict lung disease in patients with CF; thus, the identification of such a tool would allow for much improved care for those with CF.

1.2.6 Determinants/predictors of lung disease

Given the potential benefits of being able to predict lung disease in CF, much work has gone into identifying the determinants of lung disease. Initially, the *CFTR* mutation genotype was considered important; however, studies have shown that this is not the case. The focus of research in this area then shifted to whether environmental or non-*CFTR* genetic factors determine lung disease severity. Twin and sibling studies have been conducted to answer this question, highlighting the importance of non-*CFTR* genetic factors, which have since been extensively evaluated to identify specific genes determining CF lung disease severity. The following sections summarise the studies in this area.

1.2.6.1 *CFTR* genotype/phenotype correlation

One of the earliest hypothesised determinants of an individual's lung disease phenotype in CF was their *CFTR* genotype. As mentioned above, over 1,000 different *CFTR* mutations have been described, with the most common mutation being Phe508del (the deletion of three base pairs on chromosome 7, which causes a single amino acid deletion from the *CFTR* protein).¹⁰⁴ Traditionally, *CFTR* mutations have been divided into six functional classes based on their effects on the *CFTR* protein.¹⁰⁵ A possible relationship between *CFTR* genotype and clinical phenotype was first studied by Kerem et al.,¹⁰⁶ who divided 293 subjects into either Phe508del homozygous, Phe508del compound heterozygous (one Phe508del mutation, one other mutation) or two other mutations. They found no link between *CFTR* mutations and pulmonary phenotype as measured by spirometry. However, they did find that *CFTR* genotype was related to other clinical phenotypes such as age at diagnosis and pancreatic insufficiency. In a single-centre study, Johansen et al.¹⁰⁷ assessed 218 Danish subjects aged 4 months to 41 years for lung disease severity using spirometry and compared Phe508del homozygotes ($n = 172$, 79%) and Phe508del compound heterozygotes ($n = 46$, 20%). The primary outcome was the percentage of subjects with an FEV₁ below 70% of the predicted value, for which they found no difference between the two groups. In a further analysis, they found that a higher proportion of Phe508del homozygotes had an FEV₁ below 30% of predicted volume and an age-related decline in FEV₁, whereas compound heterozygotes did not. In another single-centre study in London, Santis et al.¹⁰⁸ found no effect of *CFTR* genotype on pulmonary disease severity. One of the most important studies on this potential relationship was conducted by the Cystic Fibrosis Genotype–Phenotype Consortium, a collaboration involving subjects from 14 countries. In a study of 798 subjects, the researchers paired 399 Phe508del homozygotes with 399 Phe508del compound heterozygotes or two other mutations. The pairs were matched for gender, age and treating centre. They found no difference between *CFTR* genotype and pulmonary phenotype (as assessed by spirometry and chest X-ray score).¹⁰⁹ More recently, Koch et al.¹¹⁰ used European registry data to review 5,589 subjects, who were divided into six subgroups based on the functional class of *CFTR* mutations they possessed. Using spirometry to assess lung disease, they found no effect of *CFTR* genotype on pulmonary phenotype.

Given that most studies, including the largest and most methodologically sound, have clearly shown the same findings, it is accepted that there is no relationship between *CFTR* genotype and pulmonary disease severity.

1.2.6.2 Environment v. genetics

As *CFTR* genotype does not determine lung disease, it has been hypothesised that a subject's pulmonary phenotype may be attributable to environmental, non-*CFTR* or stochastic (random) factors. Indeed, in 1990, the year following the identification of the *CFTR* gene, the role of non-*CFTR* gene modifiers was already being proposed.¹¹¹ Family and particularly twin studies are a well-established method of determining the relative contribution of these factors. Two major twin and sibling studies have been conducted on CF.

The European CF Twin and Sibling Study involved 636 subjects, including 277 sibling pairs, 29 monozygotic twin pairs and 12 dizygotic twin pairs.¹¹² The researchers assessed nutritional (weight predicted for height percentile) and pulmonary (CF-specific FEV₁ percentile) status. They also combined the two to form a composite outcome (referred to as DELTA) and analysed the difference between pairs. They found that DELTA was lower for monozygotic twins than for dizygotic twins, but discordance in FEV₁ between pairs was equivalent in monozygotic and dizygotic twins. They also found that age difference between sibling pairs had no effect on DELTA. Of note, discordance between monozygotic twins was observed. The findings of a lower DELTA in monozygotic twins and the fact that age difference in sibling pairs had no effect on DELTA support the role of genetic determinants of the composite outcome; however, given the discordance between monozygotic twins and lack of discordance in FEV₁ outcomes in monozygotic and dizygotic twins, environmental factors clearly influence lung disease outcomes.

The other large CF twins and sibling study was primarily conducted in the US but also involved subjects from Australia and Scotland. It has better addressed the determinants of pulmonary outcomes. An initial study involving 231 sibling pairs, 47 monozygotic twin pairs and 10 dizygotic twin pairs was limited by the relative lack of dizygotic twins.¹¹³ As a result, the authors had to use pairs of siblings who were less than 3 years apart in age as a surrogate for dizygotic twins. Although dizygotic twins and siblings both share 50% of genetic material, siblings are not a perfect surrogate because they have

different environmental exposures (particularly in utero). The authors analysed lung disease outcomes in terms of FEV₁ and trends in lung function. They found that monozygotic twins had more closely correlated lung disease outcomes than did the combined dizygotic twins/siblings group; however, there was still some discordance between monozygotic twins. The researchers gave a heritability estimate for CF lung disease of 0.68, where an estimate of 1.0 means no environmental influence.¹¹⁴ The same group further explored this area in a study involving 67 monozygotic twin pairs and 136 pairs of dizygotic twins or siblings less than 3 years apart in age.¹¹⁵ When comparing FEV₁ correlation (measured by mean difference in FEV₁ between pairs) in the monozygotic and dizygotic/sibling groups, they once again found a higher correlation in the monozygotic group. They also found that heritability estimates decreased in both monozygotic and dizygotic/sibling groups with advancing age. They concluded that genetic and environmental factors contributed equally to lung disease.

Together, these studies support the combined role of non-*CFTR* genetic modifiers and environmental factors in determining the severity of CF lung disease.

1.2.6.3 Gene modifiers

Given the potential role of non-*CFTR* genetic modifiers of CF lung disease severity and the benefit in being able to predict the severity of CF lung disease, extensive work has been undertaken in an attempt to identify genes that modify CF lung disease severity. This topic has been the subject of multiple expert reviews; however, to date, no formal systematic reviews have been done. Thus, the first project of this PhD was to systematically review the evidence regarding non-*CFTR* gene modifiers of CF lung disease severity (see Chapter 2). In short, there are currently no gene modifiers used in clinical practice to predict lung disease severity and alter treatment.

1.2.6.4 Current state of predictors of lung disease severity

Currently, despite the potential benefits of being able to predict lung disease severity, there are no predictive tools of CF lung disease severity that can be used in clinical practice. This is despite the extensive work on the *CFTR* genotype and non-*CFTR* gene modifiers; thus, the search for potential predictive tools must be widened to other potential mechanisms.

1.3 EPIGENETICS

1.3.1 Definition

The term *epigenetic* was first described by Conrad Waddington in 1942 as ‘the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being’.¹¹⁶ His concept of epigenetics was largely based on two principles: that the same gene could produce multiple phenotypes, and that there is an inherent stability of phenotypes across different genes and environments. To further explain the latter, he proposed the term *canalisation*, which refers to the idea that progenitor cells progress down developmental ‘canals’ to produce an end phenotype. An important aspect of the first definition, which persists in the current definition of epigenetics, is that epigenetic changes are cell or tissue specific.

The definition of epigenetics has evolved as new discoveries in the field have been made, although there has been some controversy about the suitability of the expanded use of the term.^{117 118} First, Nanney¹¹⁹ proposed that epigenetic changes persist during cell replication; thus, heritability was added to the definition.¹²⁰ An example of a widely used definition from this period is ‘the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence’.¹²¹

With the discovery of other mechanisms such as histone modification and non-coding RNA, the definition has been further expanded, with a widely cited modern definition being ‘the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states’.¹²² One of the most recent definitions, proposed in a recent *New England Journal of Medicine* review article, also incorporates the influence of the environment: ‘Modifications of DNA or associated factors that have information content, other than the DNA sequence itself, are maintained during cell division, are influenced by the environment, and cause stable changes in gene expression’.¹²³ The advantage of this most recent definition is that it is broad and encompasses all aspects of the current understanding of epigenetics and various epigenetic mechanisms.

1.3.2 Epigenetic mechanisms

Several epigenetic mechanisms broadly regulate gene expression through modifying the accessibility of DNA to transcription factors. These factors do not work independently of

each other; rather, they interact to determine gene expression. To understand epigenetic mechanisms, it is important to understand the structure of a chromosome (Figure 1.4). DNA exists in strands that form a double helix, and these strands wrap around eight histone proteins to form a nucleosome.¹²⁴ Nucleosomes undergo ongoing folding and wrapping to form chromatin, which undergoes further folding to become a chromatid. Chromatids join with a sister chromatid to form a chromosome. Multiple epigenetic mechanisms can act at different levels of the chromosome structure.

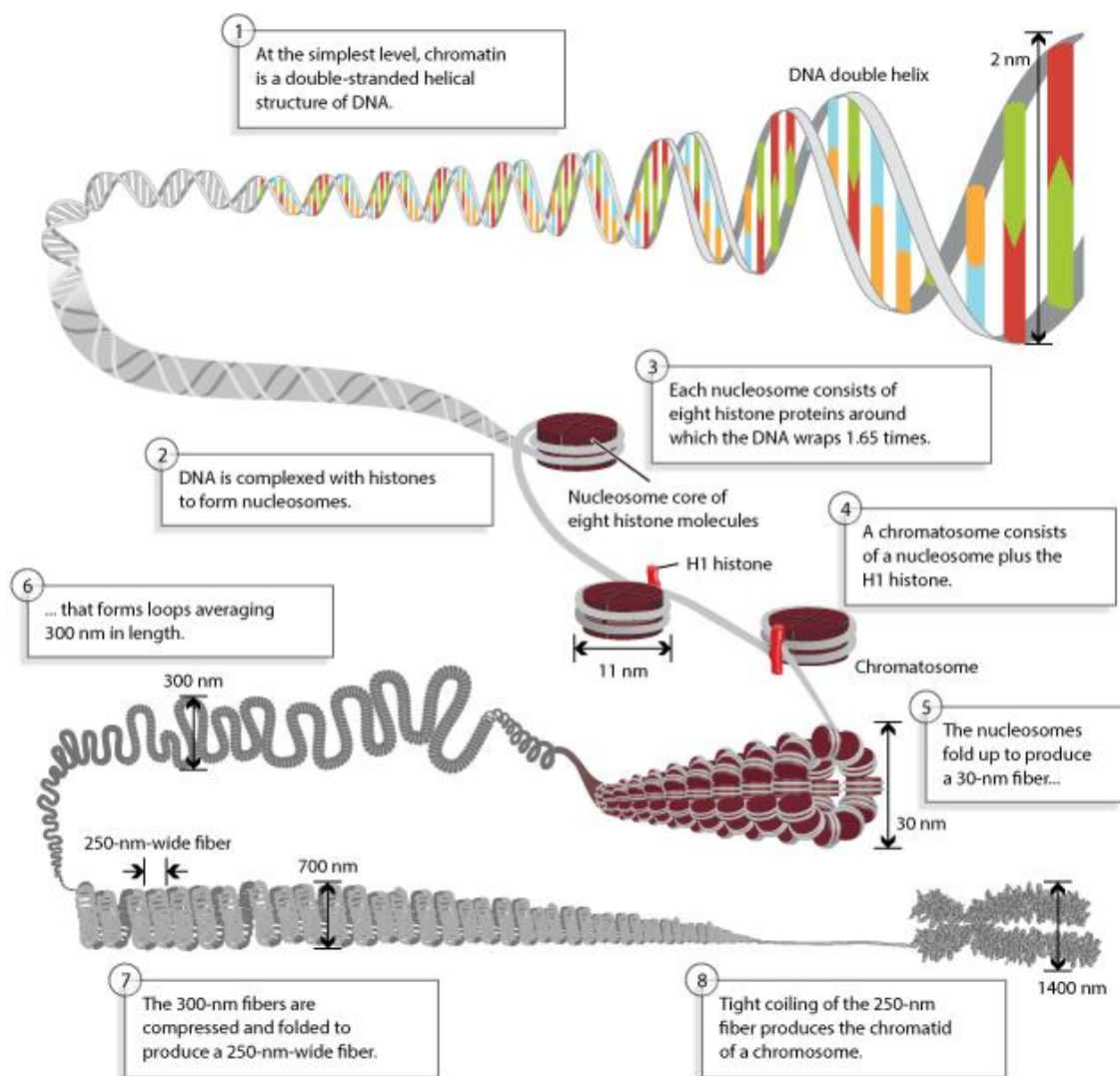


Figure 1.4. Structure of a chromosome.¹²⁵

1.3.2.1 Histone modifications

Histone modifications can also affect gene expression, with over 200 different histone modifications described.¹²³ Histone modifications determine the compactness of the nucleosome, with more compact nucleosomes less likely to be expressed. Each

nucleosome is made up of eight histone proteins, with two units each of H2A, H2B, H3 and H4.¹²⁴ Histones H3 and H4 have long N-terminal chains that sit outside of the nucleosome and are accessible to modification via the addition or removal of amino acid groups referred to as ‘histone marks’. Specific enzymes affect these histone marks, with some enzymes promoting gene expression (e.g. histone acetyl transferases, which acetylate histones, and histone demethylases) and some promoting gene repression (e.g. histone diacetyl transferases and histone methyltransferases).

1.3.2.2 Non-coding RNA

Non-coding RNA is RNA that is not translated into a protein and can exist in several forms.¹²⁶ It influences genome regulation via its effect on chromatin remodelling, epigenetic memory, genetic transcription and RNA splicing and translation. Non-coding RNA can be categorised by size into microRNAs, small interfering RNAs, small RNAs and long non-coding RNAs. MicroRNAs and small interfering RNAs are 18–25 nucleotides in size. MicroRNAs can silence genes post-transcriptionally by regulating the translation of protein. They are thought to control the translation of more than 60% of protein-coding genes.¹²⁷ Small interfering RNAs work in a similar manner to microRNAs to suppress gene expression post-transcriptionally.¹²⁸ Small RNAs are 20–200 nucleotides in size, and include small nucleolar RNA, P-element Induced Wimpy testis in *Drosophila* (PIWI)-interacting RNA and transfer RNA. Small nucleolar RNAs have multiple roles, with their main role being involvement in the regulation of ribosomal RNA.¹²⁹ PIWI-interacting RNA helps regulate elements of germ cell lines by silencing transposable elements.¹³⁰ Transfer RNAs were discovered 50 years ago and are an essential component of translation, functioning as the link between messenger RNA and the amino acid sequence.¹³¹ More recent work has shown that transfer RNAs also have other functions such as involvement in regulatory networks.¹³² Long non-coding RNAs are > 200 nucleotides in size and have a wide variety of roles in the regulation of gene expression. These include transcription of long non-coding RNAs, which act to regulate the expression of other genes, and the recruitment of histone-modifying complexes.¹³³ An example of the latter is the long non-coding RNA X-inactive-specific transcript, which recruits polycomb repressive complex 2 histone methyltransferase, which deposits a histone mark that in females leads to the silencing of one copy of the X chromosome¹³³ in a process known as X-inactivation.

1.3.2.3 Higher-order chromatin structure

Chromatin, which is made up of repeating nucleosomes, also functions in a regulatory structure. Chromatin is divided into primary, secondary and tertiary structures, and its conformation affects the accessibility of DNA for transcription factor binding and gene expression.¹³⁴ Chromatin can either exist in an accessible (active) state (termed euchromatin) or inaccessible (inactive) state (termed heterochromatin), with external influences directing chromatin to a permanent active or inactive state.

1.3.3 DNA methylation

DNA methylation is the best understood and most useful epigenetic marker with which to study human disease because of its stability over time.¹²³ Methylation refers to the addition of a methyl group, while DNA methylation refers to the addition of a methyl group to a CpG site, where a cytosine nucleotide is followed by a guanine nucleotide. When DNA methylation occurs, the methyl group is added to the fifth carbon of the cytosine nucleotide.¹³⁵ CpG sites are found at low density throughout the genome and are usually highly methylated, leading to gene inactivation. They can be found in gene regions such as the promoter (including transcription start sites and 5' untranslated regions), the gene body (exons and introns), the 3' untranslated region and intergenic regions. A minority of CpG sites are located in highly dense areas known as CpG islands, which are located in the promoter regions of around 70% of all genes.¹³⁶ CpG islands usually have a low level of DNA methylation, although this does not necessarily mean that the gene will be highly expressed because other steps such as transcription factor binding are required for gene expression.

1.3.3.1 How does DNA methylation develop?

DNA methylation undergoes a rapid change in the early embryonic period and then persists throughout life (Figure 1.5). From pre-fertilisation, sperm contain paternal DNA and methylation patterns, while oocytes contain maternal DNA and methylation patterns. At fertilisation, when a zygote or single cell is formed, paternal DNA undergoes rapid active demethylation via the ten-eleven translocation proteins, whereas maternal DNA undergoes more gradual passive demethylation because DNA methyltransferase (DNMT) 1 is not recruited.¹³⁷ The most rapid phase of demethylation occurs between fertilisation and the two-cell phase.¹³⁷ By the morula stage of embryonic development,

when the embryo is around 16 cells in size, methylation is essentially completely erased. From the blastocyst stage, consisting of around 200–300 cells, de novo methylation via DNMT3A and DNMT3B begins. The blastocyst stage is the first example of cell lineage commitment, with an inner cell mass (retaining cells that are totipotent) and a trophoblast layer. During ongoing cell division, differentiation of cell types in the embryo occurs. In the fully formed foetus, there are over 200 cell types, each with a unique epigenetic profile.

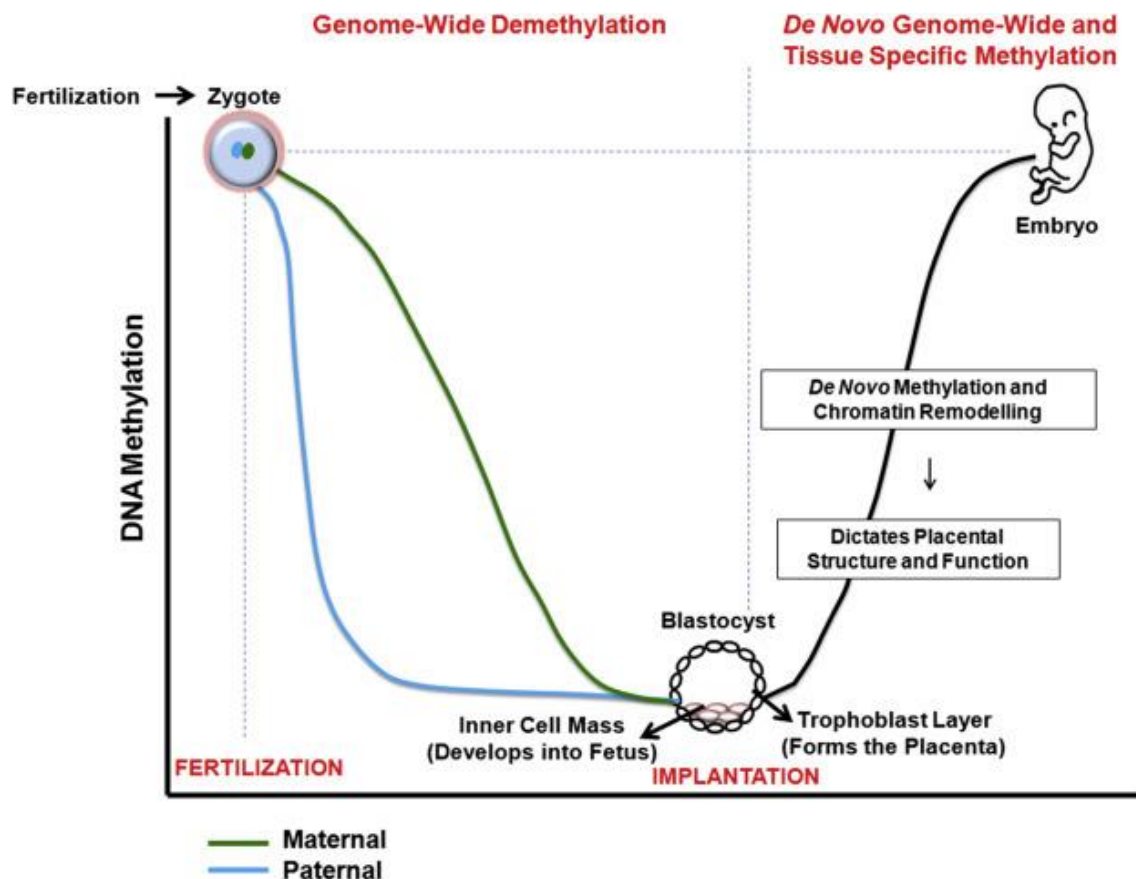


Figure 1.5. DNA methylation in the early embryonic period.¹³⁸

1.3.3.2 How are DNA methylation patterns maintained?

DNA methylation is primarily maintained by DNMT1 during cell division.¹³⁹ DNMT1 is expressed in all dividing cells and, at the time of replication, is recruited to DNA replication sites via proliferating cell nuclear antigen and UHRF1.¹⁴⁰ DNMT1 acts on hemimethylated DNA, which is DNA with a parental methylated strand and a newly synthesised unmethylated strand. DNMT1 binds to the methylated CpG sites on the parental strand and then methylates the cytosine on the newly synthesised strand, thus maintaining methylation patterns during cell division.

1.3.3.3 What determines DNA methylation?

Multiple studies have examined methylation patterns in newborn infants, finding substantial individual variations, which have been demonstrated in different cell types. One of the first studies to illustrate this was by Ollikainen et al.,¹⁴¹ who studied 91 twin pairs and investigated four differentially methylated regions in four different tissues. They found substantial variation in neonatal methylation patterns between unrelated individuals. This finding was replicated by Gordon et al.,¹⁴² who studied 34 twin pairs and analysed genome-wide methylation in three different tissue types, finding substantial variation. Genetic, in-utero environmental and stochastic factors may be responsible for variations in newborn methylation patterns, and these are discussed further. Briefly, the twin studies by Ollikainen et al.,¹⁴¹ and Gordon et al.¹⁴² and the birth cohort study by Teh et al.¹⁴³ all show that both genetic and in-utero environmental factors contribute to neonatal DNA methylation patterns. In addition, an individual's methylation pattern appears to change over time, with this phenomenon termed 'epigenetic drift'.¹⁴⁴ This highlights the role of postnatal environmental exposures in determining an individual's methylation pattern.

1.3.3.4 DNA methylation is influenced by genetics

The influence of genetic variation on the epigenetic profile has been highlighted by twin studies involving newborns, which show a higher degree of discordance in dizygotic twins versus monozygotic twins. Given that both types of twins will have identical environmental exposures, the most likely explanation for the greater discordance in dizygotic twins is genetic variation. Gordon et al.¹⁴² compared 22 monozygotic twin pairs with 12 dizygotic twin pairs, finding that monozygotic pairs showed greater similarity than did the dizygotic pairs. There are several mechanisms by which genetic variations can affect methylation. Single-nucleotide polymorphisms (SNPs) are one such mechanism—if this occurs at a CpG site (e.g. cytosine changed to adenine), the site will no longer be a CpG site, thus will no longer be able to be methylated. Genetic variations distal to CpG sites also affect methylation on either the same chromosome (*cis*) or the opposite chromosome (*trans*). The strength of the effect of the genetic variation, also known as the methylation quantitative trait loci (mQTL), on DNA methylation is related to the proximity of the gene to the site of methylation, with most mQTLs located within a few hundred base pairs of the methylation site. Evidence for the role of mQTLs is

provided by Teh et al.,¹⁴³ who performed genome-wide genotyping and methylation analysis on a birth cohort of 244 children, finding mQTLs throughout the genome.

1.3.3.5 DNA methylation is influenced by the environment

There is also clear evidence that environmental exposure, both in utero and postnatal, influences DNA methylation. Studies from the neonatal¹⁴² and postnatal¹⁴⁵ periods show that the relevant contribution of the environment outweighs the influence of the genotype. Many different environmental exposures have been studied in terms of their effects on DNA methylation, with examples of exposures relevant to CF including infection, tobacco smoke, postnatal nutrition and stress.

1.3.3.5.1 Infection

Exposure to infection can influence DNA methylation, which is of particular interest given that people with CF have a large burden of lower airway infection. The best example of bacterial infection affecting host methylation is *Helicobacter pylori* infection, which causes aberrant hypermethylation in human gastric mucosa.¹⁴⁶ Although the exact mechanism of how hypermethylation occurs is unclear, it is hypothesised that infection-related inflammation may be the cause, which again may be of relevance to CF. Other bacterial infections as well as viral and fungal infections have also been shown to affect methylation.¹⁴⁷

1.3.3.5.2 Tobacco smoke exposure

Tobacco smoke exposure, which has been shown to influence CF lung disease severity,¹⁴⁸ has been widely studied with respect to its effect on methylation. Two studies involving a total of 50 subjects demonstrated that newborns of mothers who smoked had globally reduced DNA methylation.^{149 150} In particular, Ivorra et al.¹⁵⁰ identified 25 genes with differential methylation, and the majority of differentially methylated sites were in CpG islands. For the differentially methylated sites occurring at CpG islands, in the majority of cases, the smoke-exposed newborns displayed hypermethylation. In a larger study involving a genome-wide methylation analysis of 1,062 cord blood samples, maternal smoking was found to affect the methylation of four genes: *AHRR*, *GF11*, *CYP1A1* and *MYO1G*.¹⁵¹ These genes were not differentially methylated in the study by Ivorra et al.¹⁵⁰ Further evidence has shown a dose-response relationship, with longer and more

significant smoke exposure being associated with greater change.¹⁵² In particular, it appears that persistent smoking after 18 weeks gestation affects DNA methylation, with exposure in the first trimester not appearing to be significant.¹⁵³ Studies of preschool,¹⁵⁴ adolescent¹⁵⁵ and adult¹⁵⁵ subjects have shown that alterations in methylation patterns from maternal smoking persist.

1.3.3.5.3 Postnatal nutrition

Postnatal nutrition is of vital importance in CF, with a seminal paper showing that subjects with better nutritional status have significantly improved rates of survival.¹⁵⁶ Multiple nutritional markers have been associated with altered DNA methylation. Dietary fat intake is associated with improved survival in CF, with subjects who adopt a high-fat diet having improved survival.¹⁵⁶ Thus, a high-fat diet is now part of routine CF treatment. A study involving 69 Greek preadolescent children examined whole blood methylation in relation to dietary fat intake, finding that higher dietary fat intake was associated with higher levels of DNA methylation at *TAMM41*, *MZB1* and *TXNIP* and lower levels of methylation at *GPS1* and *TAS2R13*.¹⁵⁷ People with CF also have altered lipid profiles, with reduced total cholesterol, high-density lipoprotein and low-density lipoprotein but elevated triglycerides.¹⁵⁸⁻¹⁶⁰ Lipid profiles in early life have been associated with altered DNA methylation. In a candidate gene study of 120 infants, higher total cholesterol levels were associated with *TNFA* hypomethylation, and high high-density lipoprotein levels were associated with *TNFA* and *LEP* hypomethylation.¹⁶¹ Another serum marker affected by dietary fat intake is the polyunsaturated fatty acid profile, which is associated with differential methylation. A study of infants showed that specific polyunsaturated fatty acids were associated with global DNA methylation levels,¹⁶² and a study of preadolescent children showed that the ratio of polyunsaturated fatty acid to saturated fatty acids was associated with altered DNA methylation at specific sites.¹⁵⁷

Adequate dietary folate is crucial for methylation because it provides methyl groups used in DNA methylation. Folate deficiency is common (28%) in the adult CF population,¹⁶³ although it is much less common in children with CF.^{164 165} In the general adult population, low serum folate is associated with DNA hypomethylation.^{166 167} However, findings in children are less clear. A study involving cord blood sampling showed that serum folate was associated with both global and gene-specific hypomethylation¹⁶⁸. In contrast, another birth cohort study¹⁶⁹ and a study involving children aged 5–12 years¹⁷⁰

showed no cross-sectional relationship between serum folate and global DNA methylation.

1.3.3.5.4 Psychosocial stress and mental health

Psychosocial stress and mental health issues are common in people with CF and their families. A recent international study of 6,088 people with CF and 4,102 parents of children with CF found that symptoms of depression and anxiety were increased in both patients and parents.¹⁷¹ Other smaller studies have shown that these mental health issues are associated with poorer outcomes such as reduced adherence to therapy^{172 173} and a lower health-related quality of life.^{174 175} Stress and mental health issues are also associated with altered DNA methylation. Studies have found that stress and trauma in early life are associated with differential DNA methylation in adolescence. A study examined hippocampal tissue obtained from autopsies of suicide victims with and without a history of early-life abuse compared with that of controls, finding that victims of early-life abuse had higher *NR3C1* methylation.¹⁷⁶ Esposito et al.¹⁷⁷ found that adolescents exposed to early-life adversity had altered *CFTR* methylation in blood samples. This finding was replicated by Serpeloni et al.,¹⁷⁸ who showed that children whose maternal grandmothers were exposed to trauma during pregnancy had altered DNA methylation in four genes, including *CFTR*.

1.3.3.6 Genetics and environment interact to influence DNA methylation

The effects of genetics and the environment on DNA methylation do not occur independently of each other. A study utilising subjects from the Avon Longitudinal Study of Parents and Children analysed DNA methylation at five different points over the patients' life spans, demonstrating that the effect of genotype on DNA methylation was stable, while the effect of environment increased with ageing.¹⁷⁹ This finding was reinforced by a study involving 2,603 subjects, most of whom were twins, which identified specific DNA methylation sites that were sensitive to environmental exposures.¹⁸⁰ Several human studies have identified that genotype and environment interact with each other to determine DNA methylation. In a study of 237 babies born in Singapore, the results of whole-genome methylation analysis of cord blood samples were paired with genotype data and 19 markers of in-utero environmental exposures.¹⁴³ Sites of variable methylation were identified, and analysis of genotype and environmental

exposure data showed that 25% of the variable methylation sites were explained by genotype and 75% by a combination of genotype and environment. Two notable candidate gene studies have also shown the combined influence of genotype and environment on DNA methylation. A study involving 245 teenage females from the Isle of Wight birth cohort found that the *LEPR* genotype and in-utero tobacco smoke exposure interacted to determine *LEPR* DNA methylation at 18 years, and this was also associated with body mass index (BMI).¹⁸¹ A study of 152 African Americans demonstrated that the *FKBP5* genotype (specifically at SNP rs1360780) and exposure to early-life trauma interacted to determine *FKBP5* methylation.¹⁸²

1.3.3.7 DNA methylation in complex disease (including role in disease prediction)

Given that genetic and environmental factors often interact to determine disease outcomes, DNA methylation may represent a biological marker of future health outcomes. Such predictive biomarkers are key to the goals of precision health and may predict disease progression and severity, treatment response or risk of complications. While biomarkers may be involved in disease pathogenesis, this is not a requirement for them to be clinically useful. Biomarkers may be divided into descriptive and mechanistic biomarkers—descriptive biomarkers are associated with disease but not directly involved in causal pathways, while mechanistic biomarkers are directly involved in disease pathogenesis.¹⁸³ Attempts to identify DNA methylation based on predictive biomarkers of disease severity in children are summarised in Chapter 3.

1.3.4 Review of methylation in cystic fibrosis

Given that genetics and environment interact to determine both CF lung disease severity and DNA methylation, it is possible that DNA methylation plays a role in the prediction of CF lung disease severity (see Figure 1.6). This may be through either a descriptive or a mechanistic biomarker involved in the evolution of CF lung disease.

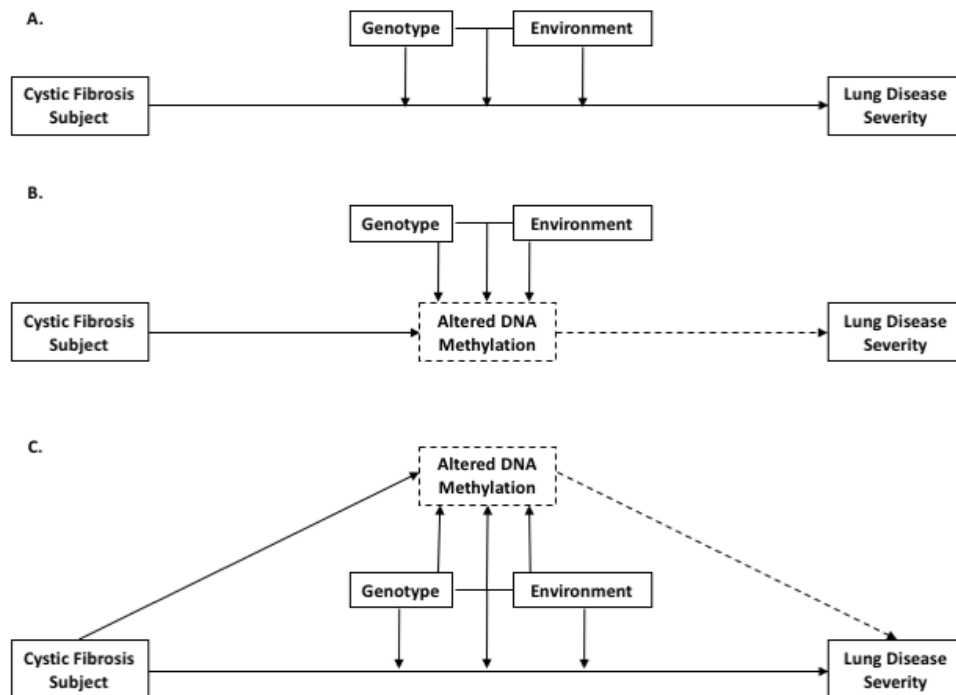


Figure 1.6. Models of how DNA methylation–based biomarkers may represent cystic fibrosis lung disease severity. A) Current understanding that genotype and environment interact to determine lung disease severity; B) A model of a mechanistic DNA methylation biomarker, whereby genotype and environment interact to alter DNA methylation, in turn affecting gene expression and lung disease severity; C) A model of a descriptive DNA methylation biomarker, whereby genotype and environment interact to modify DNA methylation but do not directly contribute to CF lung disease.

To date, there has been scant work examining the role of DNA methylation in CF lung disease. Multidrug resistance-associated protein 1 (*ABCC1*) encodes a protein that is similar in structure and function to *CFTR*, hence has been identified as a potential gene modifier. Ideozu et al.¹⁸⁴ studied DNA methylation of an *ABCC1* promoter in 23 people with CF. They examined blood DNA, finding that the promoter was unmethylated in all subjects, hence did not influence lung disease severity. Magalhães et al.¹⁸⁵ analysed DNA methylation of 13 candidate genes in blood and nasal epithelial cells in a group of 48 subjects and a replication cohort of 30 subjects. They found that increased heme oxygenase 1 (*HMOX1*) methylation was associated with disease severity. This finding was consistent for both blood and nasal epithelial cells; however, it did not survive replication. *GSTM3* hypermethylation in nasal epithelial cells but not blood was also positively associated with disease severity, while *GSTM3* hypomethylation was more

common in those carrying a protective *GSTM3* genotype, suggesting the interplay of genotype and DNA methylation in determining disease severity. *EDNRA* hypomethylation in blood, but not nasal epithelial cells, was associated with more severe lung disease, and this finding was successfully replicated. The same group subsequently performed an epigenome-wide association study (EWAS) on nasal epithelial cells collected from 32 adults with CF and 16 controls.¹⁸⁶ They used the Illumina 450K array to assess DNA methylation and measured lung disease severity using FEV₁. They identified 50 CpG sites at which methylation was associated with CF lung disease severity. Using gene ontology, the 50 CpG sites were found to correspond to genes implicated in cell adhesion.

There are several major limitations in the existing studies on methylation in CF, including the limited number of subjects, focus on adults, lack of studies of lung tissue and classification of lung disease severity using relatively insensitive measures (e.g. FEV₁). Given the lack of studies examining the role of DNA methylation in predicting lung disease severity in CF, and their significant limitations, this thesis aimed to explore this area further using appropriate samples and end points.

1.4 PROJECT AIMS AND HYPOTHESES

Aim 1: To identify DNA methylation profiles associated with normal lung structure at 9 years of age in children with CF.

Hypothesis 1: In children with CF, specific features of DNA methylation profiles of BAL specimens at 6 years of age are predictive of bronchiectasis outcomes at 9 years of age.

Aim 2: To identify early-life DNA methylation profiles associated with normal lung structure at 5 years of age in children with CF.

Hypothesis 2: In children with CF, specific features of DNA methylation profiles of BAL specimens at 1 year of age are predictive of bronchiectasis outcomes at 5 years of age.

CHAPTER 2: Non-*CFTR* Gene Modifiers of Cystic Fibrosis Lung Disease

2.1 INTRODUCTION

The identification of factors predictive of CF lung disease severity would be extremely beneficial because these factors could serve as clinical biomarkers and inform novel therapeutic approaches. Based on the results of international twin and sibling studies, which have shown that CF lung disease is determined by a combination of non-*CFTR* genetic and environmental factors, non-*CFTR* gene modifiers have been the most widely explored potential biomarkers of CF lung disease severity.

The purpose of this project was to systematically review studies that have examined non-*CFTR* gene modifiers of CF lung disease severity. The aim was to identify validated gene modifiers and appraise the quality of the evidence. The results of this systematic review informed the design of the subsequent experiments presented in this thesis, which aimed to identify novel DNA methylation-based biomarkers of CF lung disease severity.

2.2 PUBLISHED MANUSCRIPT

The manuscript presenting the systematic review, including study methods, results and discussion, was published in *Pediatric Pulmonology* (2019), volume 54, issue 9, pages 1356–1366 (<https://doi.org/10.1002/ppul.24366>). The following is an exact electronic reprint of the manuscript as published.

REVIEW

Gene modifiers of cystic fibrosis lung disease: A systematic review

Shivanthan Shanthikumar MBBS^{1,2,3}  | Melanie N. Neeland PhD^{4,3} |
Richard Saffery PhD^{5,3} | Sarath Ranganathan PhD^{1,2,3}

¹Respiratory and Sleep Medicine Department, Royal Children's Hospital, Melbourne, Australia

²Respiratory Diseases Department, Murdoch Children's Research Institute, Melbourne, Australia

³Department of Paediatrics, The University of Melbourne, Australia

⁴Centre of Food and Allergy Research, Murdoch Children's Research Institute, Melbourne, Australia

⁵Cancer & Disease Epigenetics, Murdoch Children's Research Institute, Melbourne, Australia

Correspondence

Shivanthan Shanthikumar, Respiratory and Sleep Medicine, Royal Children's Hospital, Parkville, Victoria, 3052, Australia.
Email: shivanthan.shanthikumar@rch.org.au

Abstract

Background: Lung disease is the major source of morbidity and mortality in cystic fibrosis (CF), with large variability in severity between patients. Although accurate prediction of lung disease severity would be extremely useful, no robust methods exist. Twin and sibling studies have highlighted the importance of non-cystic fibrosis transmembrane conductance regulator (CFTR) genes in determining lung disease severity but how these impact on the severity in CF remains unclear.

Methods: A systematic review was undertaken to answer the question "In patients with CF which non-CFTR genes modify the severity of lung disease?" The method for this systematic review was based upon the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)" statement, with a narrative synthesis of results planned.

Results: A total of 1168 articles were screened for inclusion, with 275 articles undergoing detailed assessment for inclusion. One hundred and forty articles were included. Early studies focused on candidate genes, whereas more recent studies utilized genome-wide approaches and also examined epigenetic mechanisms, gene expression, and therapeutic response.

Discussion: A large body of evidence regarding non-CFTR gene modifiers of lung disease severity has been generated, examining a wide array of genes. Limitations to existing studies include heterogeneity in outcome measures used, limited replication, and relative lack of clinical impact. Future work examining non-CFTR gene modifiers will have to overcome these limitations if gene modifiers are to have a meaningful role in the care of patients with CF.

KEY WORDS

cystic fibrosis, epigenomics, gene expression, genes, modifier

1 | INTRODUCTION

Lung disease is the major source of morbidity and mortality in cystic fibrosis (CF), however, there is significant variability in severity between patients.^{1,2} Accurate prediction of severity would be extremely useful (see Table 1), and the ability to stratify patients could potentially help to address 8 of the top 10 research

priorities in CF recently determined by the stakeholders.³ For example, one priority concerns effective ways of simplifying the treatment burden in patients with CF, and it would likely be easier to simplify the burden in patients predicted to have mild lung disease.

It was originally hypothesized that cystic fibrosis transmembrane conductance regulator (CFTR) mutation status could predict lung

TABLE 1 Benefits of accurate prediction of lung disease severity in CF

Group	Benefits of identifying the risk of significant disease	Benefits of identifying mild disease
Patient/family	- A better understanding of likely disease course - Understand the necessity of burden of care	- Reduced burden of care with resultant improved quality of life - Reduced exposure to iatrogenic risks
	- Better informed decisions about family planning	- Better informed decisions about family planning
CF health care team	- Inform implementation of earlier/more aggressive treatments - Increased monitoring/surveillance	- Reduced burden of care - Less frequent monitoring, thus improving patient quality of life and allowing CF team to concentrate resources elsewhere
Health authorities	- Allocate resources (new treatments, more carers) to patients most likely to benefit	- Avoid allocation of resources to a patient whose disease trajectory is unlikely to be altered
Researchers	- Identify patients most likely to benefit from new interventions - Potentially identify new therapeutic targets - Explanation of heterogeneous response to new treatments	- Potentially identify new therapeutic targets - Explanation of heterogeneous response to new treatments

Abbreviation: CF, cystic fibrosis.

disease severity. However, multiple studies demonstrated that for the majority of disease-causing mutations, this is not the case.^{12,4} Subsequently, it was hypothesized that the severity of pulmonary phenotype may be impacted by environmental or non-CFTR genetic factors.⁵ As early as 1990, the potential role of non-CFTR gene modifiers was proposed.⁶ Family and, in particular, twin studies are a well-established method for determining the relative contribution of genetic and environmental factors and two major twin and sibling studies conducted in CF both showed that non-CFTR genetic modifiers play a significant role in determining lung disease severity.⁷⁻⁹

Numerous studies have attempted to identify which genes affect CF lung disease severity and have been the subject of multiple expert reviews¹⁰⁻¹⁴ but to our knowledge, no systematic reviews have been performed. The aim of this study was to systematically review the evidence for the effect of non-CFTR modifier genes on CF lung disease severity.

2 | METHODS

The method for this systematic review was on the basis of the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses

(PRISMA)" statement¹⁵ and the review was registered with PROSPERO (CRD42017070836).¹⁶

2.1 | Research question

"In patients with CF which non-CFTR genes modify the severity of lung disease?"

2.2 | Search strategy

Databases searched and search terms used are detailed in Appendix A. Searches were limited to the English language. Reference lists from identified articles were also searched for relevant articles.

2.3 | Study selection

Duplicate articles were removed. Titles and abstracts were reviewed and inclusion and exclusion criteria (see Table 2) were applied. For articles remaining after initial review, full text was obtained and re-reviewed for inclusion or exclusion. Where the data were only available in the abstract form they were included if they had been published in a peer-reviewed journal. Where full-text articles were unable to be obtained, the corresponding author was contacted to

TABLE 2 Inclusion/exclusion criteria

Inclusion	Exclusion
• Human studies	• Animal studies
• Patients with cystic fibrosis	• Patients with other medical conditions
• Non-CFTR gene structure, expression, or regulation studied	• Studies involving the CFTR gene only
• Measure of lung disease severity as an outcome (lung function, lung structure, microbiology, composite score, etc.)	• No measure of lung disease severity
• Original data	• Review article

Abbreviation: CFTR, non-cystic fibrosis transmembrane conductance regulator.

provide a copy, and if this was not successful the article was excluded.

2.4 | Data extraction

The data, including gene name, study method, patient number, country, age, sex, *CFTR* genotype, lung disease outcome, and results, were extracted into a template developed by the authors. The study method was classified as a candidate vs systemic, qualitative vs quantitative, and linkage vs association on the basis of previously published definitions.¹⁷

2.5 | Quality assessment and data synthesis

It was anticipated that the results of the search would not allow for a meta-analysis because of heterogeneity in terms of the study population, genes investigated, and lung disease outcomes used. As such, a narrative synthesis was planned. To assess the quality of included studies the Q-Genie tool, which is specifically designed to assess the quality of genetic association studies, was used.¹⁸ Briefly, each paper is assessed on the basis of 11 criteria. A 7 point Likert scale is used from poor (1) to excellent (7), and a score assigned for each criteria. The score from each criterion is then combined to give the overall score. The maximum score a paper can obtain is 77.

3 | RESULTS

3.1 | Summary of searches

A PRISMA diagram summarizing the results of the systematic review is shown in Figure 1. Reasons for exclusion of the papers where full-text review was undertaken are detailed in Table S1.

3.2 | Data synthesis

The data extracted from the 140 included papers are summarized in Table S2, Table S3, and Table 3. The results are briefly summarized below, with more detail regarding the findings of individual studies and the body of evidence for each gene summarized found in Table S2 and Table S3, respectively.

3.2.1 | Candidate gene/region approach

The initial studies of non-*CFTR* gene modifiers of CF disease severity focused on investigating candidate genes or regions involved in pathways implicated in the pathogenesis of CF lung disease. The majority of these studies used an association-based methodology, whereas some used linkage or a combined approach (such as transmission disequilibrium testing; see Table S3). Whilst the majority of gene modifier studies have been candidate gene studies, the highest quality evidence has been derived from larger studies, which used a genome-wide approach. As such only, the studies

regarding selected genes are described. Summaries of all candidate gene studies are found in the attached tables.

Transforming growth factor beta 1

Transforming growth factor beta 1 (*TGFβ1*), was postulated to be a gene modifier because of its role in inflammation, fibrosis, and chloride transport¹⁹ is one of the most widely investigated gene modifiers, having been investigated by 15 different studies.^{20–34} These studies have examined three different sites, with codon 10 being of most interest. The studies have yielded conflicting results, but several studies including the large Gene Modifier Study,²¹ have shown that the CC genotype is associated with more severe disease. The CC genotype is associated with increased *TGFβ1* expression, which could cause more severe disease because of increased inflammation and fibrosis.¹⁹ Studies have also demonstrated an interplay between tobacco smoke exposure and *TGFβ1* in determining disease severity,²⁸ as well as interaction with other gene modifiers.²⁴

Mannose binding lectin

Mannose-binding lectin (MBL) is part of the innate immune response, and MBL deficiency is associated with increased susceptibility to infection.³⁵ MBL has been extensively studied as a gene modifier, with a number of studies^{21,24,25,32,36–47} showing as hypothesized, a low MBL production genotype is associated with several measures of more severe lung disease, such as higher infection rates, lower lung function, and lower survival. Of note the most robust study found no effect of MBL.²¹

Macrophage migration inhibitory factor

Macrophage migration inhibitory factor (MIF) is a proinflammatory mediator and in vitro models have shown that polymorphisms with five CATT repeats at the −794 promoter are associated with reduced gene expression. The number of CATT repeats at the −794 promoter of *MIF* was associated with lung disease severity in three studies, with as expected five CATT repeats being associated with milder lung disease in terms of lung function and PA infection.^{48–50} Because of poor coverage of the *MIF* gene in some SNP microarray's this gene may not have been examined by CF GWAS studies.

Glutathione S-Transferase mu

Because of its role in protecting the lung from an oxidative lung injury, Glutathione S-Transferase mu (*GSTM1*) has been investigated in nine studies,^{21,34,51–57} and whereas some smaller studies^{51,52,54} showed a link between the null genotype and poorer outcomes, the largest study²¹ showed no effect. A possible explanation for this discrepancy is that the studies showing an effect used radiological and clinical score outcomes, whereas the large Gene Modifier study used FEV₁.

CD14

CD14 is part of the innate immune response, and presence of a C allele at position −159, is associated with reduced CD14 levels. CD14

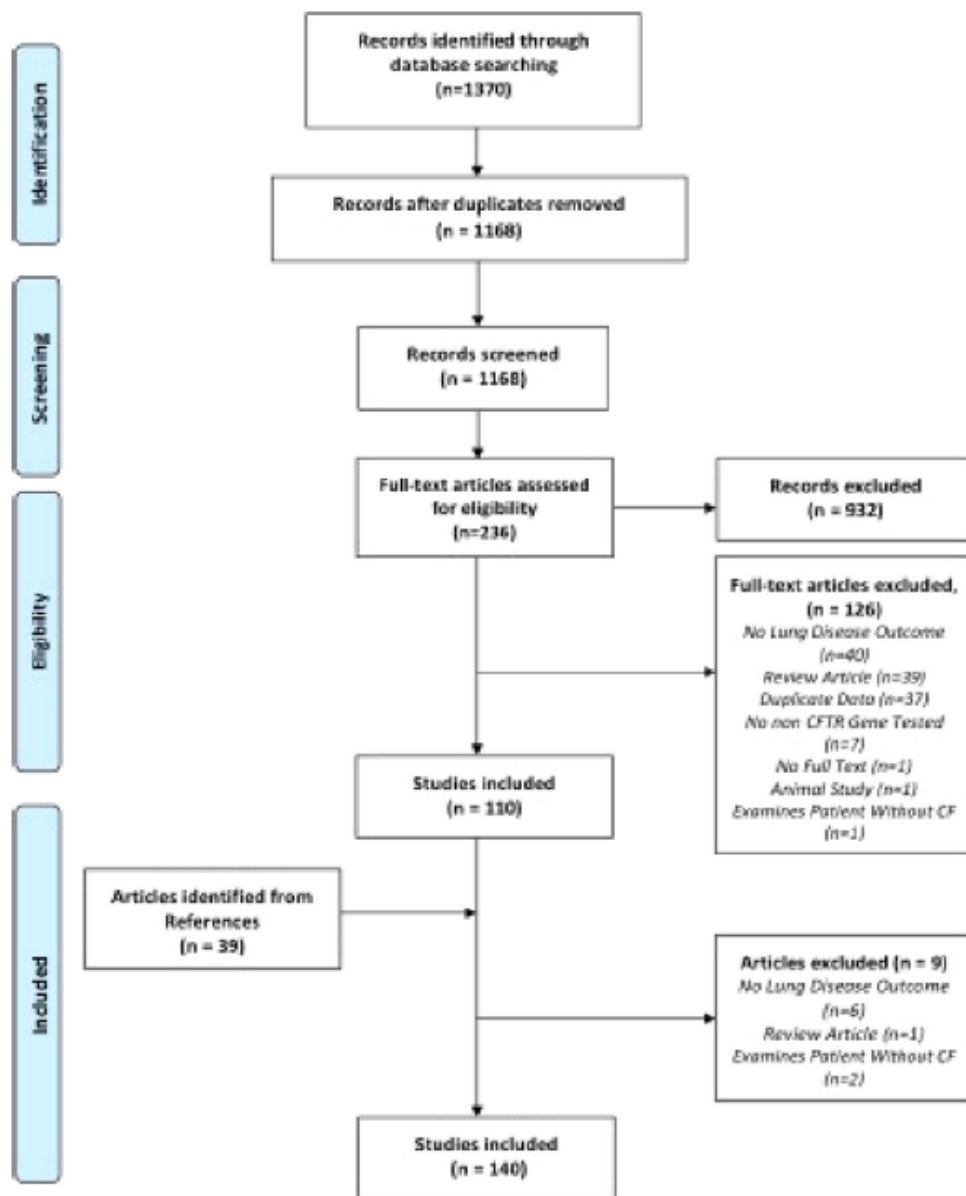


FIGURE 1 PRISMA flowchart for study selection. [Color figure can be viewed at wileyonlinelibrary.com]

genotype showed no effect on lung function in two studies.^{25,58} However, in a study using a very robust microbiological endpoint (annual bronchoalveolar lavage [BAL]) the low CD14 producing CC genotype at position -159 was associated with earlier acquisition of PA infection.⁵⁹ There has been no attempt to replicate this finding.

Human hemochromatosis protein

Higher iron levels the airway have been implicated in CF lung disease severity.^{60,61} Human hemochromatosis protein (HFE) plays a crucial

role in iron homeostasis, with mutations in HFE being associated with higher iron levels. HFE has been investigated in two separate Australian cohorts and showed that individuals with mutations had significantly lower lung function (absolute difference of 14.4% in FEV₁) and a faster rate of decline in FEV₁.^{62,63}

Endothelin receptor type A

Endothelin receptor type A (EDNRA) is a proinflammatory peptide and is also implicated in small airway constriction. One study

TABLE 3 Summary of investigated genes by the result

Hypothesized mechanism	No effect	Conflicting results	Effect shown with no attempt at replication	Multiple studies showing effect with no conflicting result
Alternate ion channel	CLC-2, SCNN1A, NEDD4L	SCNN1y, SCNN1g, SLC26A9		
Located within loci identified via genome-wide search			ATF1, DUOX2, YY1, AGTR2, CAV2, LPAR6, SLC9A3R2, TMC6, APIP, EHF, EZR	SLC9A3, HLA-CLASS II, DCTN4, SLC6A14
Bronchoconstriction		ADRB2		EDNRA
CFTR interaction	AnxA5, PPP2R5E, AHSA1, CALR, KRT18, PRSS8, ABCC2-5	ABCC1	PPP2R1A, PPP2R4, STX1A, SNAP23, KRT8, KRT19	
CFTR vicinity or haplotype	D7S495, D7S525, D7S514, XV2C, Tub20, Meth, km.19, J3.11	MP6d9	LEP1, LEP2	
Infection susceptibility	MASP-2, ABO, CHIT1, DEF84, FUT2, FUT3, Haptoglobin, MASP-1, T2R38	DEF81, CD14, CF8	C3, CXCR1, CXCR2, FAS, FCN1, FCN2, MASP-3, CHIT3-Like 1	CEACAM, HFE, HMOX1
Infection susceptibility and inflammatory response	LBP, TLR3, TLR6, TLR7, TLR8, TLR9, TLR10, IL-1A, IL-1R1, TLR4	MBL, TLR5, IL-18, IL-1RN	TLR1, TLR2	
Nitric oxide		NOS-1, NOS-3		
Inflammatory response	IFNy, IL-6, STAT3, TAX1BP1	TNF α , α -1-antitrypsin, ACE, IL-10, IFRD1, IL-8, COX2, LT- α , AGER-429, HSP70-2G	SFTPA1, SFTPA2, A20, ADRA2A, α -1-antichymotrypsin, COX1, Glucocorticoid Receptor, HLA-DQB1, HLA-DRB1, IFNyR1, MPO, Neutrophil FCyR IIA, TNFRSF1A, MC3R	TGF81, MIF
Mucin	MUC2, MUC7		MUC5AC, MUC1, MUC4, MUC20	
Oxidative lung injury		GSTM1, GSTP1, GSTT1	mEPHX	GCLC, GSTM3

Abbreviation: CFTR, non-cystic fibrosis transmembrane conductance regulator.

investigated mutations in *EDNRA* across four different cohorts and found that at SNP rs5335, the CC genotype was associated with a 5% absolute reduction in lung function.⁶⁴ Further *in vivo* experiments showed the CC genotype to be associated with increased *EDNRA* levels and hence a proinflammatory state. Methylation of *EDNRA* in blood but not nasal epithelial cells was subsequently shown to be inversely related to lung disease severity.³⁴

Carcinoembryonic antigen-related cell adhesion molecules

Carcinoembryonic antigen-related cell adhesion molecules (CEACAM) was identified as a modifier of lung disease severity in two separate studies^{65,66} using the European Twins and Siblings cohort. These studies performed a linkage analysis whereby the distribution of alleles between mildly and severely affected sibling pairs and concordant and discordant siblings was compared to find gene modifiers.

Mucin

Because of the role of altered mucin in CF lung disease, a number of mucin (*MUC*) genes have been investigated. The international GWAS meta-analysis identified that a locus at 3q29 which contains *MUC4* and *MUC20* was associated with lung disease severity.⁶⁷ A study of 762 patients found a relationship between *MUC5AC* variable number tandem repeat (VNTR) number and disease severity with the 6.4 kb VNTR being associated with more severe disease.⁶⁸ Of note, all mucin genes have only been investigated in single studies with no attempted replication.

CFTR vicinity

Several genes have been investigated as potential gene modifiers because of their proximity to *CFTR*. One study found that leptin 1 and 2 (*LEP1*, *LEP2*) were associated with increased disease severity.⁶⁹ Another study analyzed the region 7q31 to 7qtel and found a locus on 7q34 that showed differential onset of parental recombination and differential imprinting between mildly and severely affected sibling pairs.⁷⁰ However, both studies used composite measure of severity, which encompassed both respiratory and nutritional parameters. The authors of both studies hypothesized that the effect was most likely via an influence on nutrition.

3.2.2 | Systematic genome wide approach

As technology has evolved genome-wide searches, rather than targeted approaches, have become more common. Genome-wide searches examine either the entire exome (the part of the genome which is responsible for coding proteins) via whole exome sequencing or all SNPs via GWAS.

Whole exome sequencing

Two studies^{71,72} using whole exome sequencing identified genes that modify the risk of PA infection. Mutations in dynactin subunit 4 (*DCTN4*) and transmembrane channel-like protein 6 (*TMC6*) were

associated with early onset of PA infection and a caveolin 2 (*CAV2*) variant was shown to be protective. These studies also showed an effect of *TMC6* and *CAV2* on lung function, with variants associated with earlier PA acquisition being associated with lower FEV₁. One study attempted to replicate the findings with regard to *DCTN4* and found it only affected the risk of PA infection in a subgroup of males who had two class 2 *CFTR* mutations.⁷³

GWAS

Three GWAS studies^{67,74,75} have been published to date, one of which pooled GWAS results from North America and France to perform a meta-analysis.⁶⁷ A major strength of these GWAS studies when compared with previous candidate gene studies is the significantly larger number of subjects involved. For example, the GWAS meta-analysis included data from SNP microarrays from 6365 patients, as compared with the largest candidate gene study, which involved an initial cohort of 808 patients, with replication in 498 patients. This strength is reflected in the GWAS studies having the highest Q-Genie scores. The GWAS meta-analysis used FEV₁ to measure lung disease severity. Five loci that were associated with lung disease severity were identified;

1. Chromosome 3q29, which contains *MUC4* and *MUC20*, however, no subsequent studies have investigated this region or the hypothesized genes.
2. Chromosome 5p15.3 contains solute carrier family 9A3 (*SLC9A3*), a gene which was also shown in three other studies^{33,76,77} to associate with lung disease severity.
3. Chromosome 6p21.3, contains HLA Class II which was shown in other studies^{67,78,79} to associate with lung disease severity as measured by lung function and PA infection.
4. Chromosome Xq22-q23 contains angiotensin II receptor type 2 (*AGTR2*) and solute carrier family 6A14 (*SLC6A14*). No other studies have investigated *AGTR2*. *SLC6A14* appeared to associate with both lung function and risk of PA infection in a large Canadian study.⁷⁶
5. Chromosome 11p12-p13 contains ETS homologous factor (*EHF*) and together with *APIP*, identified by a previous GWAS,⁷⁴ was included in the meta-analysis. A subsequent study used targeted resequencing of this area a group of 377 patients and identified multiple SNPs associated with lung disease severity.⁸⁰

While a number of the genes of interest in the five identified loci have not been further investigated and hence replicated, the GWAS meta-analysis did include the analysis (via forest plots and Cochran's Q test) where the patients were analyzed in 13 subpopulations (on the basis of original cohort and SNP microarray platform used) to assess for heterogeneity. Using Cochran's Q test only the chromosome 3q29 loci showed evidence of heterogeneity and using a forest plot most of the subpopulations showed a significant effect of 3q29. This analysis serves as a form of internal replication showing that the

5 loci of interest were still significant even when results were analyzed in the 13 subpopulations.

A previous GWAS⁷⁴ used linkage analysis to identify 20q13.2 as a modifier locus in a subset of 486 sibling pairs. The linkage analysis uses different methodology to that used in a traditional GWAS analysis so the meta-analysis could not replicate this finding.

A separate publication used a novel analysis technique on data from GWAS performed in Canadian and French cohorts to analyze a set of genes, which affect apical plasma membrane epithelia.⁷⁷ This identified SLC9A3 (which had already been identified by the international GWAS), SLC9A3R2 and EZR as modifying lung disease severity. To date, no attempts at replicating the findings regarding SLC9A3R2 and EZR have been made.

3.2.3 | Epigenetic

Epigenetics refers to the phenomenon of heritable changes in gene expression that do not involve changes to the underlying DNA sequence.⁸¹ The most commonly investigated epigenetic mechanism is DNA methylation and four studies have investigated this. Ideozu et al⁸² analyzed methylation at ABCC1-5 and found no relationship between DNA methylation and disease severity. Magalhaes et al⁸⁴ analyzed DNA methylation of 13 candidate genes, in blood and nasal epithelial cells in a group of 48 patients, with a replication cohort of 30. They found heme oxygenase 1 (HMOX1), EDNRA, and GSTM3 DNA methylation were associated with lung disease severity. Another epigenetic mechanism is imprinting, where one copy of a gene is silenced depending on which parent it was inherited from. Two separate studies from the European Twins and Siblings studies have hypothesized a relationship between imprinting of 7q34 and disease severity.^{69,70}

3.2.4 | Gene expression

There has been a recent increase in studies examining gene expression and CF lung disease severity. Two studies examined the effect of ABCC1 gene expression, with Hurbain et al⁸³ finding that low expression in nasal epithelial cells was associated with more severe clinical disease scores. However, a subsequent study⁸² did not find a relationship. Kelly et al⁸⁴ identified that increased A20 gene expression in nasal epithelial cells was associated with milder lung disease. Magalhaes et al⁸⁴ investigated gene expression of 13 candidate genes in blood and found no effect. Only one study examined genome-wide expression in nasal epithelial cells and identified genes related to oxidoreductase activity, the ubiquitin cycle and lipid metabolism to be associated with lung disease severity.⁸⁵ A subsequent study, which examined gene expression in blood, identified gene pathways involved in HLA Class I, Phe508del processing and endoplasmic reticulum stress response to be associated with lung disease severity.⁷⁹ In particular, lysophosphatidic acid receptor 6 (LPAR6) was found to be more highly expressed in those with severe lung disease.

3.2.5 | Theratype

A theratype refers to factors which predict a patient's response to treatment, and in CF there is scope to move beyond the existing focus on theratypes related to CFTR genotype.⁸⁶ Several studies have examined gene modifiers of therapeutic response. Four studies⁸⁷⁻⁹⁰ examined whether polymorphisms of β -2 Adrenergic Receptor 2 (ADRB2), which is involved in bronchoconstriction, would affect whether patients responded to inhaled bronchodilator, and found there was no significant relationship. In a more recent, and highly relevant study, Strug et al⁹¹ examined the effect of gene modifier SLC26A9 (an alternate chloride channel identified via the GWAS meta-analysis) on the response to treatment with the CFTR modulator ivacaftor. They studied 24 Canadian patients with at least one gating mutation and measured FEV₁ before and after treatment with ivacaftor. They genotyped patients at the SNP rs7512462, and found that for each C allele (as compared with T) there was a 9.8% improvement in FEV₁ response to ivacaftor.

4 | DISCUSSION

A significant body of evidence has been generated regarding the potential modifiers of CF lung disease severity. The wide array of genes that has been studied is in keeping with the numerous mechanisms implicated in the pathogenesis of CF lung disease. A key strength of the literature is that subjects from 27 different countries have been studied, and thus, a wide cross-section of the CF population is represented. This is countered by the majority of subjects, especially in genome-wide studies, being from North America and Western Europe. There are, however, significant limitations within the evidence including variability in outcome measures, lack of replication, and limited data on subsequent clinical impact.

One issue limiting the comparability of papers is the large variability in outcome measures used. Fifty-three different outcome measures were used, with the most common measure being FEV₁. For FEV₁, at least 15 different reference equations were used, and often the selection and timing of FEV₁ measurement was unclear. This limited the comparison of genome-wide studies with each other and candidate gene studies. In the adult population, one way to overcome this is for all studies to use KNORMA, the FEV₁ measure developed by the international GWAS consortium. A limitation of this approach is that it would not be applicable to children too young to perform spirometry,¹⁴ and given there are differences in lung disease severity observed in the first years of life, there remains great interest in gene modifier studies in this age group. In fact, a National Heart Lung Blood Institute workshop in primary prevention of chronic lung disease in CF called for identification of gene modifiers that affect lung disease severity in early life.⁹² Studies involving children should use the most robust markers of lung disease severity in this age group, but they have been rarely used in gene modifier studies to date. Only two papers^{93,94} used the most sensitive

measure of lung function (lung clearance index [LCI]),⁹⁵⁻⁹⁷ and one paper⁵⁹ the most sensitive measure of lower respiratory infection (BAL).⁹⁸ Computed Tomography (CT) scan results were used more widely, however not in younger age groups where they are very sensitive markers of lung disease severity.^{99,100}

While over 100 individual genes have been investigated as potential gene modifiers, the majority of genes (76) have only been investigated in a single study. In order for any genotype-phenotype relationship to be thoroughly established, replication is key.¹⁷ However, for 48 genes, where a relationship was identified, there has as yet been no attempt at replication. This is more relevant for genes identified in candidate gene studies than for those identified in GWAS, as the GWAS meta-analysis including over 6000 patients used statistical tests of heterogeneity to replicate the findings in the subpopulations of patients included in the meta-analysis.

A relative limitation of evidence generated by studies investigating gene modifiers of CF is the limited clinical impact. The primary aim of gene modifier studies has been to identify therapeutic targets. Gene modifier studies in CF have identified some therapeutic targets.¹⁰¹ To date no clinical interventions have resulted, however, early trials are underway, which will hopefully result in efficacious new treatments.¹⁰¹ Another potential application of identifying gene modifiers, that has not been the primary aim, but perhaps represents a missed opportunity, is using them to stratify patients into different risk categories (ie mild lung disease, severe lung disease). These risk categories could in turn influence clinical management. Using gene modifier genes to stratify patient risk and inform clinical management is well established in oncology and has led to improved outcomes.¹⁰² To our knowledge, no CF centers use modifier genes to stratify patients and modify treatment. Utilizing modifier genes in risk stratification would be a step towards a more personalized approach to treatment. Identification of gene modifiers, which influence the magnitude of response to treatments, would also represent a potential clinical application. The discovery that *SLC26A9* modified the magnitude of response to ivacaftor is an example of this.⁹¹ This has subsequently been replicated in a further recent study.¹⁰³ Much more comprehensive validation is needed however before this discovery could be used in routine clinical practice. As more treatments are developed that ameliorate the basic CFTR defect, permitting multiple treatment options for patients with the same genotype, being able to predict which treatment will be of most benefit to the individual patient will be crucial and gene modifiers may be of use when determining this.

The strength of this systematic review is the broad search strategy and methodology, which followed the best practice guidelines.¹⁰⁴ The authors feel that compared with the multiple expert reviews in the area, this present an unbiased and comprehensive review of the published literature. A limitation of the systematic review is the exclusion of animal studies and studies, which purely examine *in vitro* gene expression (with no correlation to clinical outcomes). These studies interrogate potential mechanisms by which gene modifiers exert an effect and hence are very important, however, given the already large size of the body of evidence to be

reviewed, it was beyond the scope to include them. Another relative limitation is that the broad search strategy has led to the inclusion of a large number of articles, which means that each individual article is perhaps interrogated less than in the expert reviews, which have less included articles.

Studies investigating gene modifiers continue to be published^{105,106} and moving forward, to aid robust discovery of genetic modifiers of disease, it would be ideal if future studies utilized gold standard measures of lung disease severity. In adult patients that might be universal use of KNORMA, and in children that could be a combination of CT, BAL, and LCI. Whilst such testing may seem invasive, they are already being conducted regularly in cohorts such as AREST CF¹⁰⁷ and the Toronto Sick Kids CF Cohort.¹⁰⁸ Using insensitive measures of lung disease may result in an inaccurate assessment of the contribution of gene-modifiers to lung disease severity. Moving forward, a challenge in phenotyping lung disease severity will be accounting for the effect of CFTR modulating therapies, given differences in lung disease severity may be related to eligibility and access to these therapies rather than true differences in lung disease severity. The magnitude of response to these treatments may represent an important phenotype worthy of further studies, as already highlighted in the studies examining the response to ivacaftor.^{91,103} Any significant findings should then be replicated in separate cohorts using similar endpoints. For future work examining DNA methylation and gene expression, it will be important that initial gene discovery is based on samples from biologically relevant tissue (BAL, nasal epithelial cells) with subsequent replication in more easily obtained samples such as blood.⁸¹ Once genetic modifiers have been validated, the next step will be to assess whether they can be used in clinical management of patients, with those associated with increased risk of more severe disease facilitating more aggressive intervention, and those associated with mild disease opportunistically a reduced treatment burden. In this way, genetic modifiers of CF lung disease could inform more personalized and precise care for patients with CF.

ACKNOWLEDGMENTS

The authors would like to thank Poh Chua for her invaluable assistance with developing and refining the search strategy.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING

Dr Shivanthan Shanthikumar was supported by the Thoracic Society of Australia and New Zealand Paediatric Cystic Fibrosis Fellowship.

ORCID

Shivanthan Shanthikumar  <http://orcid.org/0000-0001-6000-3180>

REFERENCES

- Kerem E, Corey M, Kerem B, et al. The relation between genotype and phenotype in cystic fibrosis—analysis of the most common mutation (delta F508). *N Engl J Med*. 1990;323(22):1517-1522.
- Cystic Fibrosis Genotype-Phenotype Consortium. Correlation between genotype and phenotype in patients with cystic fibrosis. *N Engl J Med*. 1993;329(18):1308-1313.
- Rowbotham NJ, Smith S, Leighton PA, et al. The top 10 research priorities in cystic fibrosis developed by a partnership between people with CF and healthcare providers. *Thorax*. 2017.
- Koch C, Cuppens H, Rainisio M, et al. European epidemiologic registry of cystic fibrosis (ERCF): comparison of major disease manifestations between patients with different classes of mutations. *Pediatr Pulmonol*. 2001;31(1):1-12.
- Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet*. 2015;16(1):45-56.
- Santis G, Osborne L, Knight RA, Hodson ME. Independent genetic determinants of pancreatic and pulmonary status in cystic fibrosis. *Lancet*. 1990;336(8723):1081-1084.
- Collaco JM, Blackman SM, McGready J, Naughton KM, Cutting GR. Quantification of the relative contribution of environmental and genetic factors to variation in cystic fibrosis lung function. *J Pediatr*. 2010;157(5):802-807.
- Vanscoy LL, Blackman SM, Collaco JM, et al. Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med*. 2007;175(10):1036-1043.
- Mekus F, Ballmann M, Bronsveld I, Bijman J, Veeze H, Tummler B. Categories of delta F508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin Res*. 2000;3(4):277-293.
- Weiler CA, Drumm ML. Genetic influences on cystic fibrosis lung disease severity. *Frontiers in Pharmacology*. 2013;4 APR (no pagination)(Article 40).
- Cutting GR. Modifier genes in mendelian disorders: the example of cystic fibrosis. *Ann NY Acad Sci*. 2010;1214:57-69.
- Guillot L, Beauchamp J, Tabary O, Le Rouzic P, Clement A, Corvol H. Lung disease modifier genes in cystic fibrosis. *Int J Biochem Cell Biol*. 2014;52:83-93.
- Drumm ML, Ziyade AG, Davis PB. Genetic variation and clinical heterogeneity in cystic fibrosis. *Annu Rev Pathol Mech Dis*. 2012;7:267-282.
- O'Neal WK, Knowles MR. Cystic fibrosis disease modifiers: complex genetics defines the phenotypic diversity in a monogenic disease. *Annu Rev Genomics Hum Genet*. 2018;19:201-222.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;6(7):339.
- National Institute for Health Research. 2017 PROSPERO. <https://www.crd.york.ac.uk/prosperto/display_record.php?RecordID=70836>
- Genin E, Feingold J, Clerget-Darpoux F. Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*. 2008;124(4):357-368.
- Sohani ZN, Meyre D, de Souza RJ, et al. Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. *BMC Genet*. 2015;16(1):50.
- Kramer EL, Clancy JP. TGF β as a therapeutic target in cystic fibrosis. *Expert Opin Ther Targets*. 2018;22(2):177-189.
- Oueslati S, Hadj Fredj S, Siala H, Messaoud T. Study of the TGF beta1 codon 10 in cystic fibrosis tunisian patients. *Clin Chem Lab Med*. 2015;53:5837.
- Drumm ML, Konstan MW, Schluchter MD, et al. Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med*. 2005;353(14):1443-1453.
- Levy H, Linse G, Schneck K, et al. Integration of genomics with genetics-molecular phenotypes for cystic fibrosis (CF) lung disease. *American Journal of Respiratory and Critical Care Medicine*. Conference: American Thoracic Society International Conference, ATS 2010;181(1 MeetingAbstracts).
- Arkwright PD, Pravica V, Geraghty PJ, et al. End-organ dysfunction in cystic fibrosis: association with angiotensin I converting enzyme and cytokine gene polymorphisms. *Am J Respir Crit Care Med*. 2003;167(3):384-389.
- Dorfman R, Sandford A, Taylor C, et al. Complex two-gene modulation of lung disease severity in children with cystic fibrosis. *J Clin Invest*. 2008;118(3):1040-1049.
- Faria EJ, Faria IC, Ribeiro JD, Ribeiro AF, Hessel G, Bertuzzo CS. Association of MBL2, TGF-beta1 and CD14 gene polymorphisms with lung disease severity in cystic fibrosis. *Jornal Brasileiro De Pneumologia: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisiologia*. 2009;35(4):334-342.
- Eickmeier O, Boom L, Schreiner F, et al. Transforming growth factor β 1 genotypes in relation to TGF β 1, interleukin-8, and tumor necrosis factor alpha in induced sputum and blood in cystic fibrosis. *Mediators Inflamm*. 2013;2013:913135-913138.
- Bremer LA, Blackman SM, Vanscoy LL, et al. Interaction between a novel TGF β 1 haplotype and CFTR genotype is associated with improved lung function in cystic fibrosis. *Hum Mol Gen*. 2008;17(14):2228-2237.
- Collaco JM, Vanscoy L, Bremer L, et al. Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease. *JAMA*. 2008;299(4):417-424.
- Brazova J, Sismova K, Vavrova V, et al. Polymorphisms of TGF-beta1 in cystic fibrosis patients. *Clin Immunol*. 2006;121(3):350-357.
- Corvol H, Boelle PY, Brouard J, et al. Genetic variations in inflammatory mediators influence lung disease progression in cystic fibrosis. *Pediatr Pulmonol*. 2008;43(12):1224-1232.
- Arkwright PD, Laurie S, Super M, Pravica V, Schwarz MJ, Webb AK, Hutchinson IV. TGF-beta(1) genotype and accelerated decline in lung function of patients with cystic fibrosis. *Thorax*. 2000;55(6):459-462.
- Buranawuti K, Boyle MP, Cheng S, et al. Variants in mannose-binding lectin and tumour necrosis factor α affect survival in cystic fibrosis. *J Med Genet*. 2007;44(3):209-214.
- Dorfman R, Taylor C, Lin F, et al. Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis. *Pediatr Pulmonol*. 2011;46(4):385-392.
- Magalhaes M, Rivals I, Claustres M, et al. DNA methylation at modifier genes of lung disease severity is altered in cystic fibrosis. *Clin Epigenetics*. 2017;9(1):19.
- Turner MW. The role of mannose-binding lectin in health and disease. *Mol Immunol*. 2003;40(7):423-429.
- Olesen HV, Jensenius JC, Steffensen R, Thiel S, Schiøtz PO. The mannose-binding lectin pathway and lung disease in cystic fibrosis—dysfunction of mannose-binding lectin-associated serine protease 2 (MASP-2) may be a major modifier. *Clin Immunol*. 2006;121(3):324-331.
- Yarden J, Radojkovic D, De Boeck K, et al. Polymorphisms in the mannose binding lectin gene affect the cystic fibrosis pulmonary phenotype. *J Med Genet*. 2004;41(8):629-633.
- Gabolde M, Guilloud-Bataille M, Feingold J, Besmond C. Association of variant alleles of mannose binding lectin with severity of pulmonary disease in cystic fibrosis: cohort study. *BMJ*. 1999;319(7218):1166-1167.
- Garred P, Pressler T, Madsen HO, et al. Association of mannose-binding lectin gene heterogeneity with severity of lung disease and survival in cystic fibrosis. *J Clin Invest*. 1999;104(4):431-437.
- Gravina LP, Crespo C, Giugno H, et al. Mannose-binding lectin gene as a modifier of the cystic fibrosis phenotype in Argentinean pediatric patients. *J Cyst Fibros*. 2015;14(1):78-83.
- Tarova ET, Polakova H, Kayserova H, Celec P, Zuzulova M, Kadasi L. Study of the effect of DNA polymorphisms in the mannose-binding

- lectin gene (MBL2) on disease severity in Slovak cystic fibrosis patients. *Gen Physiol Biophys*. 2011;30(4):373-378.
42. Davies JC, Turner MW, Klein N, et al. Impaired pulmonary status in cystic fibrosis adults with two mutated MBL-2 alleles. *Eur Respir J*. 2004;24(5):798-804.
 43. Choi EH, Ehrmantraut M, Foster CB, Moss J, Chanock SJ. Association of common haplotypes of surfactant protein A1 and A2 (SFTPA1 and SFTPA2) genes with severity of lung disease in cystic fibrosis. *Pediatr Pulmonol*. 2006;41(3):255-262.
 44. Carlsson M, Sjöholm AG, Eriksson L, et al. Deficiency of the mannan-binding lectin pathway of complement and poor outcome in cystic fibrosis: bacterial colonization may be decisive for a relationship. *Clin Exp Immunol*. 2005;139(2):306-313.
 45. Trevisiol C, Boniotti M, Giglio L, Poli F, Morgutti M, Crovella S. MBL2 polymorphisms screening in a regional Italian CF center. *J Cyst Fibros*. 2005;4(3):189-191.
 46. McDougal KE, Green DM, Vanscoy LL, et al. Use of a modeling framework to evaluate the effect of a modifier gene (MBL2) on variation in cystic fibrosis. *Eur J Human Genet*. 2010;18(6):680-684.
 47. Haerynck F, Van Steen K, Cattaert T, et al. Polymorphisms in the lectin pathway genes as a possible cause of early chronic *Pseudomonas aeruginosa* colonization in cystic fibrosis patients. *Hum Immunol*. 2012;73(11):1175-1183.
 48. Plant BJ, Gallagher CG, Bucala R, et al. Cystic fibrosis, disease severity, and a macrophage migration inhibitory factor polymorphism. *Am J Respir Crit Care Med*. 2005;172(11):1412-1415.
 49. Adamali H, Armstrong ME, McLaughlin AM, et al. Macrophage migration inhibitory factor enzymatic activity, lung inflammation, and cystic fibrosis. *Am J Respir Crit Care Med*. 2012;186(2):162-169.
 50. Melotti P, Maffioli A, Lebecque P, et al. Impact of MIF gene promoter polymorphism on F508del cystic fibrosis patients. *PLoS ONE [Electronic Resource]*. 2014;9(12):e114274.
 51. de Lima Marson FA, Bertuzzo CS, Secolin R, Ribeiro AF, Ribeiro JD. Genetic interaction of GSH metabolic pathway genes in cystic fibrosis. *BMC Med Genet*. 2013;14:60.
 52. Marson FA, Bertuzzo CS, Ribeiro AF, Ribeiro JD. Polymorphisms in the glutathione pathway modulate cystic fibrosis severity: a cross-sectional study. *BMC Med Genet*. 2014;15:27.
 53. Lima CS, Ortega MM, Marson FA, Zulli R, Ribeiro AF, Bertuzzo CS. Cystic fibrosis transmembrane conductance regulator gene mutations and glutathione S-transferase null genotypes in cystic fibrosis patients in Brazil. *Jornal Brasileiro De Pneumologia: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisiologia*. 2012;38(1):50-56.
 54. Hull J, Thomson AH. Contribution of genetic factors other than CFTR to disease severity in cystic fibrosis. *Thorax*. 1998;53(12):1018-1021.
 55. Korytina G, Ialabaeva D, Viktorova T. Polymorphism of glutathione S-transferase M1 and P1 genes in patients with cystic fibrosis and chronic respiratory tract diseases. *Genetika*. 2004;40(3):401-408.
 56. Feuillet-Fleux MN, Nguyen-Khoa T, Lorient MA, et al. Glutathione S-transferases related to *P. aeruginosa* lung infection in cystic fibrosis children: preliminary study. *Clin Biochem*. 2009;42(1-2):57-63.
 57. Flamant C, Hermon-Claude A, Boelle PY, et al. Glutathione S-transferase M1, M3, P1 and T1 polymorphisms and severity of lung disease in children with cystic fibrosis. *Pharmacogenetics*. 2004;14(5):295-301.
 58. Haerynck F, Mahachie John JM, Van Steen K, et al. Genetic variations in toll-like receptor pathway and lung function decline in Cystic fibrosis patients. *Hum Immunol*. 2013;74(12):1649-1655.
 59. Martin AC, Laing IA, Zhang G, et al. CD14 C-159T and early infection with *Pseudomonas aeruginosa* in children with cystic fibrosis. *Respir Res*. 2005;6:63.
 60. Reid DW, Lam QT, Schneider H, Walters EH. Airway iron and iron-regulatory cytokines in cystic fibrosis. *Eur Respir J*. 2004;24(2):286-291.
 61. Stites SW, Plautz MW, Bailey K, O'Brien-Ladner AR, Wessell LJ. Increased concentrations of iron and isoferritins in the lower respiratory tract of patients with stable cystic fibrosis. *Am J Respir Crit Care Med*. 1999;160(3):796-801.
 62. Smith D, Bell S, Anderson G, Reid D. HFE 282y mutations are associated with severity of lung disease and rate of lung function decline in cystic fibrosis. *Respirology*. 2014;19:48.
 63. Pratap U, Quinn S, Blizzard LB, Reid DW. Population-based study of cystic fibrosis disease severity and haemochromatosis gene mutations. *Respirology*. 2010;15(1):141-149.
 64. Darrah R, McKone E, O'Connor C, et al. EDNRA variants associate with smooth muscle mRNA levels, cell proliferation rates, and cystic fibrosis pulmonary disease severity. *Physiol Genomics*. 2010;41(1):71-77.
 65. Becker T, Jansen S, Tamm S, Wienker TF, Tummeler B, Stanke F. Transmission ratio distortion and maternal effects confound the analysis of modulators of cystic fibrosis disease severity on 19q13. *Eur J Human Genet*. 2007;15(7):774-778.
 66. Stanke F, Becker T, Hedtfeld S, Tamm S, Wienker TF, Tummeler B. Hierarchical fine mapping of the cystic fibrosis modifier locus on 19q13 identifies an association with two elements near the genes CEACAM3 and CEACAM6. *Hum Genet*. 2010;127(4):383-394.
 67. Corvol H, Blackman SM, Boelle PY, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun*. 2015;6:8382.
 68. Guo X, Pace RG, Stonebraker JR, et al. Mucin variable number tandem repeat polymorphisms and severity of cystic fibrosis lung disease: significant association with MUC5AC. *PLoS ONE [Electronic Resource]*. 2011;6(10):e25452.
 69. Mekus F, Laabs U, Veeze H, Tummeler B. Genes in the vicinity of CFTR modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pairs. *Hum Genet*. 2003;112(1):1-11.
 70. Stanke F, Davenport C, Hedtfeld S, Tummeler B. Differential decay of parent-of-origin-specific genomic sharing in cystic fibrosis-affected sib pairs maps a paternally imprinted locus to 7q34. *Eur J Human Genet*. 2010;18(5):553-559.
 71. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of phenotypic extremes identifies CAV2 and TMC6 as interacting modifiers of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *PLoS Genet*. 2015;11(6):e1005273.
 72. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nature Genet*. 2012;44(8):886-889.
 73. Viel M, Hubert D, Burgel PR, et al. DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Clin Respir J*. 2016;10(6):777-783.
 74. Wright FA, Strug LJ, Doshi VK, et al. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nature Genet*. 2011;43(6):539-546.
 75. Gu Y, Harley IT, Henderson LB, et al. Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. *Nature*. 2009;458(7241):1039-1042.
 76. Li W, Soave D, Miller MR, et al. Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. *Hum Genet*. 2014;133(2):151-161.
 77. Soave D, Corvol H, Panjwani N, et al. A joint location-scale test improves power to detect associated SNPs, gene sets, and pathways. *Am J Hum Genet*. 2015;97(1):125-138.
 78. Aron Y, Polla BS, Biennu T, Dallava J, Dusser D, Hubert D. HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype. *American Journal of Respiratory & Critical Care Medicine*. 1999;159(5 Pt 1):1464-1468.
 79. O'Neal WK, Gallins P, Pace RG, et al. Gene expression in transformed lymphocytes reveals variation in endomembrane and HLA pathways modifying cystic fibrosis pulmonary phenotypes. *Am J Hum Genet*. 2015;96(2):318-328.

80. Dang H, Gallins PJ, Pace RG, et al. Novel variation at chr11p13 associated with cystic fibrosis lung disease severity. *Hum Genome Var*. 2016;3:201620.
81. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. *N Engl J Med*. 2018;378(14):1323-1334.
82. Ideazu JE, Zhang X, Pan A, et al. Increased expression of plasma-induced ABCC1 mRNA in cystic fibrosis. *Int J Mol Sci*. 2017;18(8):1752.
83. Hurbain I, Semet-Gaudelus I, Vallee B, et al. Evaluation of MRP1-5 gene expression in cystic fibrosis patients homozygous for the delta F508 mutation. *Pediatr Res*. 2003;54(5):627-634.
84. Kelly C, Williams MT, Elborn JS, Ennis M, Schock BC. Expression of the inflammatory regulator A20 correlates with lung function in patients with cystic fibrosis. *J Cyst Fibros*. 2013;12(4):411-415.
85. Wright JM, Merlo CA, Reynolds JB, et al. Respiratory epithelial gene expression in patients with mild and severe cystic fibrosis lung disease. *Am J Respir Cell Mol Biol*. 2006;35(3):327-336.
86. Amaral MD, Balch WE. Hallmarks of therapeutic management of the cystic fibrosis functional landscape. *J Cyst Fibros*. 2015;14(6):687-699.
87. Hart MA, Konstan MW, Darrah RJ, et al. Beta 2 adrenergic receptor polymorphisms in cystic fibrosis. *Pediatr Pulmonol*. 2005;39(6):544-550.
88. Marson FA, Bertuzzo CS, Ribeiro AF, Ribeiro JD. Polymorphisms in ADRB2 gene can modulate the response to bronchodilators and the severity of cystic fibrosis. *BMC Pulm Med*. 2012;12:50.
89. Steagall WK, Barrow BJ, Glasgow CG, et al. Beta-2-adrenergic receptor polymorphisms in cystic fibrosis. *Pharmacogenet Genomics*. 2007;17(6):425-430.
90. Buscher R, Eilmes KJ, Grasemann H, et al. beta2 adrenoceptor gene polymorphisms in cystic fibrosis lung disease. *Pharmacogenetics*. 2002;12(5):347-353.
91. Strug LJ, Gonska T, He G, et al. Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. *Hum Mol Gen*. 2016;25(20):4590-4600.
92. Pittman JE, Cutting G, Davis SD, Ferkol T, Boucher R. Cystic fibrosis: NHLBI workshop on the primary prevention of chronic lung diseases. *Ann Am Thorac Soc*. 2014;11(Suppl 3):S161-S168.
93. von Kanel T, Stanke F, Weber M, et al. Clinical and molecular characterization of the potential CF disease modifier syntaxin 1A. *Eur J Human Genet*. 2013;21(12):1462-1466.
94. Gislis FM, von Kanel T, Kraemer R, Schaller A, Gallati S. Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis. *Eur J Human Genet*. 2013;21(4):397-403.
95. Horsley AR, Gustafsson PM, Macleod KA, et al. Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax*. 2008;63(2):135-140.
96. Gustafsson PM, De Jong PA, Tiddens HA, Lindblad A. Multiple-breath inert gas washout and spirometry versus structural lung disease in cystic fibrosis. *Thorax*. 2008;63(2):129-134.
97. Aurora P, Bush A, Gustafsson P, et al. Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis. *Am J Respir Crit Care Med*. 2005;171(3):249-256.
98. Rosenfeld M, Emerson J, Accurso F, et al. Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis. *Pediatr Pulmonol*. 1999;28(5):321-328.
99. Hall GL, Logie KM, Parsons F, et al. Air trapping on chest CT is associated with worse ventilation distribution in infants with cystic fibrosis diagnosed following newborn screening. *PLoS One*. 2011;6(8):e23932.
100. Ramsey KA, Rosenow T, Turkovic L, et al. Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *Am J Respir Crit Care Med*. 2016;193(1):60-67.
101. Darrah RJ, Jacono FJ, Joshi N, et al. AGTR2 absence or antagonism prevents cystic fibrosis pulmonary manifestations. *J Cyst Fibros*. 2018.
102. Tung NM, Garber JE. BRCA1/2 testing: therapeutic implications for breast cancer management. *Br J Cancer*. 2018;119(2):141-152.
103. Corvol H, Mesinele J, Doukseh IH, Strug LJ, Boelle PY, Guillot L. SLC26A9 gene is associated with lung function response to ivacaftor in patients with cystic fibrosis. *Front Pharmacol*. 2018;9:828.
104. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;6:e1000100.
105. Polineni D, Dang H, Gallins PJ, et al. Airway mucosal host defense is key to genomic regulation of cystic fibrosis lung disease severity. *Am J Respir Crit Care Med*. 2018;197(1):79-93.
106. Smith DJ, Klein K, Hartel G, et al. Mutations in the HFE gene can be associated with increased lung disease severity in cystic fibrosis. *Gene*. 2019;683:12-17.
107. Sly PD, Gangell CL, Chen L, et al. Risk factors for bronchiectasis in children with cystic fibrosis. *N Engl J Med*. 2013;368(21):1963-1970.
108. Eckford PDW, McCormack J, Munsie L, et al. The CF Canada-Sick Kids Program in individual CF therapy: A resource for the advancement of personalized medicine in CF. *J Cyst Fibros*. 2018.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Shanthikumar S, Neeland MN, Saffery R, Ranganathan S. Gene modifiers of cystic fibrosis lung disease: A systematic review. *Pediatric Pulmonology*. 2019;1-11. <https://doi.org/10.1002/ppul.24366>

2.3 SUPPLEMENTARY MATERIAL

The published manuscript contains supplementary material, attached as published and available online (<https://doi.org/10.1002/ppul.24366>).

Chapter 2 Appendix A: Search strategy (*MEDLINE searched twice to optimise yield)

Database	Search strategy
MEDLINE 1 (1996–present)	1. Genes, Modifier/ 2. *cystic fibrosis/ge 3. “severity of illness index”/ 4. genotype/ or genetic predisposition to disease/ or haplotypes/ or phenotype/ or genetic markers/ 5. Mutation/ 6. Polymorphism, Genetic/ 7. (diversity or severity or (candidate adj gene*1)).tw,kf 8. exp animals/ not human*.sh 9. 2 and (1 or (*genotype/ or *genetic predisposition to disease/ or *haplotypes/ or *phenotype/ or *genetic markers/) or *Mutation/ or *Polymorphism, Genetic/ or 7 or 3) 10. 9 not 8 11. limit 10 to English language
MEDLINE 2* (1996–present)	1. *cystic fibrosis/ge 2. (cystic adj fibrosis).tw,kf,hw 3. (lung or pulmonary).tw,kf,hw 4. “severity of illness index”/ 5. (severity or progression or accelerat* or decline).tw,kf,hw 6. *genotype/ or *genetic predisposition to disease/ or *haplotypes/ or *phenotype/ or *genetic markers/ 7. *Mutation/ 8. Genes, Modifier/ 9. *Polymorphism, Genetic/ 10. (diversity or (candidate adj gene*1) or (modif* adj4 gene*1)).tw,kf 11. (1 or 2) and 3 and (4 or 5) and (6 or 7 or 8 or 9 or 10) 12. exp animals/ not human*.sh 13. 11 not 12 14. limit 13 to english language
EMBASE (1976–present)	1. *cystic fibrosis/ 2. (cystic adj fibrosis).tw,kw,hw 3. (lung or pulmonary).tw,kw,hw 4. “severity of illness index”/ 5. (severity or progression or accelerat* or decline).tw,kw,hw 6. modifier gene/ 7. *genotype phenotype correlation/ 8. *genotype/ 9. exp *genetic predisposition/ 10. *haplotype/

	11. *phenotype/ 12. exp *genetic marker/ 13. exp *mutation/ 14. exp *genetic polymorphism/ 15. *genetic association/ 16. (diversity or (candidate adj gene*1) or (modif* adj4 gene*1)).tw,kw. 17. exp animal/ not human*.sh. 18. (1 or 2) and 3 and (4 or 5) and (6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16) 19. 18 not 17 20. limit 19 to English
PUBMED	(Cystic Fibrosis) AND (Lung OR Pulmonary) AND (Sever* OR Progression OR Decline OR Accelerat*) AND ("modifier gene" OR "modifier genes" OR "gene modifier" OR "gene modifiers" OR "genetic modifier" OR "genetic modifiers" OR "candidate gene" OR "candidate genes" OR "gene candidate" OR "gene candidates" OR "genetic candidate" OR "genetic candidates" OR "genotype" OR "genetic predisposition" OR haplotype* OR phenotype* OR "gene marker" OR "gene markers" OR "genetic marker" OR "genetic markers" OR mutation* OR "gene polymorphism" OR "gene polymorphisms" OR "genetic polymorphism" OR "genetic polymorphisms" OR "gene association" OR "gene associations" OR "genetic association" OR "genetic associations") AND (NOTNLM OR publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb] OR indatareview[sb] OR pubstatusaheadofprint)

Chapter 2 Supplementary Table 1: Excluded papers

Authors	Title	Journal	Reason
Johannesson et al	Polymorphic expression of multidrug resistance mRNA in lung parenchyma of nonpregnant and pregnant rats: a comparison to cystic fibrosis mRNA expression.	Biochemical & Biophysical Research Communications. 239(2):606-11, 1997 Oct 20	Animal Study
Ideozu et al	Correlation of ABCC1 promoter polymorphism to cystic fibrosis lung disease.	Pediatric Pulmonology. Conference: 30th Annual North American Cystic Fibrosis Conference. United States. 51 (pp 247), 2016. Date of Publication: October 2016.	Duplicate Data
Corvol H. et al	Genome-wide association meta-analysis in ~6,400 CF patients confirms previous findings and identifies new loci contributing to variation in lung disease.	Pediatric Pulmonology. Conference: 28th Annual North American Cystic Fibrosis Conference. Atlanta, GA United States. Conference Publication: (var.pagings). 49 (pp 271-272), 2014. Date of Publication: September 2014.	Duplicate Data
O'Neal W.K. et al	Differential gene expression identifies genetic modifiers associated With lung disease severity and age of onset of persistent pseudomonas infection in cystic fibrosis.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2014. San Diego, CA United States. Conference Publication: (var.pagings). 189 (no pagination), 2014. Date of Publication: 2014.	Duplicate Data
O'Neal W.K. et al	Differential gene expression identifies CF genetic modifiers associated with severity of lung disease.	Pediatric Pulmonology. Conference: 27th Annual North American Cystic Fibrosis Conference. Salt Lake City, UT United States. Conference Publication: (var.pagings). 48 (pp 264), 2013. Date of Publication: October 2013.	Duplicate Data
Corvol H. et al	A european meta-analysis of ancestral haplotype 8.1 and lung disease severity in cystic fibrosis.	Pediatric Pulmonology. Conference: 24th Annual North American Cystic Fibrosis Conference. Baltimore, MD United States. Conference Publication: (var.pagings). 45 (pp 282-283), 2010. Date of Publication: September 2010.	Duplicate Data
Furlan L.L. et al	IL-8 gene: A potential modifier gene of cystic fibrosis severity.	Pediatric Pulmonology. Conference: 11th SOLANEP International Congress, 15th Cystic Fibrosis Latinamerican Congress and 15th Brazilian Congress of Pediatric Pulmonology. Brazil. 51 (pp S8), 2016. Date of Publication: April 2016.	Duplicate Data

Authors	Title	Journal	Reason
Crespo C. et al	Mannose binding lectin gene as a modifier of cystic fibrosis phenotype in Argentinian pediatric patients.	Journal of Cystic Fibrosis. Conference: 35th European Cystic Fibrosis Conference. Dublin Ireland. Conference Publication: (var.pagings). 11 (pp S58), 2012. Date of Publication: June 2012.	Duplicate Data
Saavedra M.T. et al	Whole blood mRNA expression of a 10-gene panel predicts disease progression and clinical outcomes in CF: A 5 year follow-up.	Pediatric Pulmonology. Conference: 28th Annual North American Cystic Fibrosis Conference. Atlanta, GA United States. Conference Publication: (var.pagings). 49 (pp 248-249), 2014. Date of Publication: September 2014.	Duplicate Data
Guo X. et al	Relating respiratory mucin genes to CF lung disease severity: Obtaining more accurate genetic maps and identification of functional variants.	Pediatric Pulmonology. Conference: 25th Annual North American Cystic Fibrosis Conference. Anaheim, CA United States. Conference Publication: (var.pagings). 46 (pp 270), 2011. Date of Publication: October 2011.	Duplicate Data
Kelly C. et al	Altered expression of the NF-kappaB inhibitor a20 reflects CFTR mutation and FEV ₁ .	Pediatric Pulmonology. Conference: 25th Annual North American Cystic Fibrosis Conference. Anaheim, CA United States. Conference Publication: (var.pagings). 46 (pp 259), 2011. Date of Publication: October 2011.	Duplicate Data
Marson F.A. et al	Genetic interaction of NOS-1 and CFTR gene and its association with cystic fibrosis severity.	Journal of Cystic Fibrosis. Conference: 36th European Cystic Fibrosis Conference. Lisbon Portugal. Conference Publication: (var.pagings). 12 (pp S42)	Duplicate Data
Knowles M.R. et al	New findings from the international CF gene modifier consortium.	Pediatric Pulmonology. Conference: 29th Annual North American Cystic Fibrosis Conference. Phoenix, AZ United States. Conference Publication: (var.pagings). 50 (pp 171-172), 2015. Date of Publication: October 2015.	Duplicate Data
Knowles M.R. et al	Gene modifiers of CF lung disease: Update on mechanistic insights.	Pediatric Pulmonology. Conference: 25th Annual North American Cystic Fibrosis Conference. Anaheim, CA United States. Conference Publication: (var.pagings). 46 (pp 143-144), 2011. Date of Publication: October 2011.	Duplicate Data
Strug L.J. et al	SLC26A9 modulates airway response to CFTR directed therapeutics.	Pediatric Pulmonology. Conference: 30th Annual North American Cystic Fibrosis Conference. United States. 51 (pp 282-283), 2016. Date of Publication: October 2016.	Duplicate Data

Authors	Title	Journal	Reason
Von Kanel T. et al	Clinical and molecular characterization of the CF disease modifier syntaxin 1A (STX1A).	Journal of Cystic Fibrosis. Conference: 33rd European Cystic Fibrosis Conference. Valencia Spain. Conference Publication: (var.pagings). 9 (pp S1), 2010. Date of Publication: June 2010.	Duplicate Data
Shmarina G. et al	Genetic polymorphisms in TNF genes and tuberculosis in cystic fibrosis patients.	European Respiratory Journal. Conference: European Respiratory Society Annual Congress 2011. Amsterdam Netherlands. Conference Publication: (var.pagings). 38 (no pagination), 2011. Date of Publication: 01 Sep 2011.	Duplicate Data
Robinson P. et al	Cytokine genotypes and cystic fibrosis phenotype. Cytokine gene polymorphisms and severity of cystic fibrosis lung disease.	Pediatric Pulmonology. Conference: 27th Annual North American Cystic Fibrosis Conference. Salt Lake City, UT United States. Conference Publication: (var.pagings). 48 (pp 243), 2013. Date of Publication: October 2013.	Duplicate Data
Furlan LL, et al	IL8 gene as modifier of cystic fibrosis: unraveling the factors which influence clinical variability.	Hum Genet. 2016 Aug;135(8):881-94. doi: 10.1007/s00439-016-1684-4. Epub 2016 May 21. Erratum in: Hum Genet. 2017 Feb 3;.. Hum Genet. 2017 Mar;136(3):347-348.	Duplicate Data
Coutinho CA et al	TNF-alpha polymorphisms as a potential modifier gene in the cystic fibrosis.	Int J Mol Epidemiol Genet. 2014 May 29;5(2):87-99. eCollection 2014.	Duplicate Data
Marson FA et al	Polymorphisms in ADRB2 gene can modulate the response to bronchodilators and the severity of cystic fibrosis.	BMC Pulmonary Medicine. 12:50, 2012 Sep 05	Duplicate Data
Hart MA et al	Beta 2 adrenergic receptor polymorphisms in cystic fibrosis.	Pediatric Pulmonology. 39(6):544-50, 2005 Jun	Duplicate Data
Beucher J et al	AGER -429T/C is associated with an increased lung disease severity in cystic fibrosis.	PLoS ONE [Electronic Resource]. 7(7):e41913, 2012	Duplicate Data
Frangolias DD et al	Alpha 1-antitrypsin deficiency alleles in cystic fibrosis lung disease.	American Journal of Respiratory Cell & Molecular Biology. 29(3 Pt 1):390-6, 2003 Sep	Duplicate Data
Henry MT et al	An alpha1-antitrypsin enhancer polymorphism is a genetic modifier of pulmonary outcome in cystic fibrosis.	European Journal of Human Genetics. 9(4):273-8, 2001 Apr	Duplicate Data
Vankeerberghen A et al	Distribution of human beta-defensin polymorphisms in various control and cystic fibrosis populations.	Genomics. 85(5):574-81, 2005 May	Duplicate Data

Authors	Title	Journal	Reason
Taylor-Cousar JL et al	Histo-blood group gene polymorphisms as potential genetic modifiers of infection and cystic fibrosis lung disease severity.	PLoS ONE [Electronic Resource]. 4(1):e4270, 2009	Duplicate Data
Saferali A et al	Polymorphisms associated with expression of BPIFA1/BPIFB1 and lung disease severity in cystic fibrosis.	American Journal of Respiratory Cell & Molecular Biology. 53(5):607-14, 2015 Nov	Duplicate Data
Darrah R et al	EDNRA variants associate with smooth muscle mRNA levels, cell proliferation rates, and cystic fibrosis pulmonary disease severity.	Physiological Genomics. 41(1):71-7, 2010 Mar 03	Duplicate Data
de Lima Marson FA et al	Genetic interaction of GSH metabolic pathway genes in cystic fibrosis.	BMC Medical Genetics. 14:60, 2013 Jun 10	Duplicate Data
Wright FA et al	Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2.	Nature Genetics. 43(6):539-46, 2011 Jun	Duplicate Data
Corvol H et al	Ancestral haplotype 8.1 and lung disease severity in European cystic fibrosis patients.	Journal of Cystic Fibrosis. 11(1):63-7, 2012 Jan	Duplicate Data
Gu Y et al	Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease.	Nature. 458(7241):1039-42, 2009 Apr 23	Duplicate Data
Hillian AD et al	Modulation of cystic fibrosis lung disease by variants in interleukin-8.	Genes & Immunity. 9(6):501-8, 2008 Sep	Duplicate Data
Stanke F et al	Genes that determine immunology and inflammation modify the basic defect of impaired ion conductance in cystic fibrosis epithelia.	Journal of Medical Genetics. 48(1):24-31, 2011 Jan	Duplicate Data
Dorfman R et al	Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis.	Pediatric Pulmonology. 46(4):385-92, 2011 Apr	Duplicate Data
Drumm ML et al Dunn JM	Genetic modifiers of lung disease in cystic fibrosis.	New England Journal of Medicine. 353(14):1443-53, 2005 Oct 06	Duplicate Data

Authors	Title	Journal	Reason
Darrah RJ Dorfman R Sandford AJ Corey M Zielenski J Durie P Goddard K Yankaskas JR Wright FA Knowles MR Gene Modifier Study Group			
Yarden J et al	Association of tumour necrosis factor alpha variants with the CF pulmonary phenotype.	Thorax. 60(4):320-5, 2005 Apr	Duplicate Data
Azad AK et al	Mutations in the amiloride-sensitive epithelial sodium channel in patients with cystic fibrosis-like disease.	Human Mutation. 30(7):1093-103, 2009 Jul	Examines Patient Without CF
Ozbas-Gerceker F et al	Analysis of the modifying effects of TAP 1/2 genes on cystic fibrosis phenotype.	Turkish Journal of Pediatrics. 44(2):91-7, 2002 Apr-Jun	No full text
Polineni D et al	Interleukin 8 (IL8) expression in nasal epithelia is correlated with a genetic variant in the IL8 promoter associated with cystic fibrosis lung disease severity.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2012. San Francisco, CA United States. Conference Publication: (var.pagings). 185 (no pagination), 2012. Date of Publication: 2012.	No Lung Disease Outcome
Polineni D et al	Interleukin 8 (IL8) expression in cystic fibrosis (CF) nasal epithelia is correlated with a genetic variant in the IL8 promoter associated with CF lung disease severity.	Pediatric Pulmonology. Conference: 26th Annual North American Cystic Fibrosis Conference. Orlando, FL United States. Conference Publication: (var.pagings). 47 (pp 282-283), 2012. Date of Publication: September 2012.	No Lung Disease Outcome
Noel S et al	Micromnas profiling in cystic fibrosis bronchial epithelial cells: A pilot study to identify new therapeutic targets.	Pediatric Pulmonology. Conference: 24th Annual North American Cystic Fibrosis Conference. Baltimore, MD United States.	No Lung Disease Outcome

Authors	Title	Journal	Reason
		Conference Publication: (var.pagings). 45 (pp 284), 2010. Date of Publication: September 2010.	
Saavedra M et al	Identification of cystic fibrosis endotypes through peripheral blood transcriptional profiling.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2014. San Diego, CA United States. Conference Publication: (var.pagings). 189 (no pagination), 2014. Date of Publication: 2014.	No Lung Disease Outcome
Dang H et al	Filling the genomic gap in MUC5AC: Identification of genetic variation that modifies cystic fibrosis lung disease phenotype.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2013. Philadelphia, PA United States. Conference Publication: (var.pagings). 187 (no pagination), 2013. Date of Publication: 2013.	No Lung Disease Outcome
Bezzetti V et al	PLCB3 in pro-inflammatory signaling in bronchial epithelial cells.	Pediatric Pulmonology. Conference: 25th Annual North American Cystic Fibrosis Conference. Anaheim, CA United States. Conference Publication: (var.pagings). 46 (pp 266), 2011. Date of Publication: October 2011.	No Lung Disease Outcome
Saferali A et al	Pluncs as potential modifier genes in cystic fibrosis.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2012. San Francisco, CA United States. Conference Publication: (var.pagings). 185 (no pagination), 2012. Date of Publication: 2012.	No Lung Disease Outcome
Tummler B et al	Modifiers of CF: Experience of the european cystic fibrosis twin and sibling study.	Pediatric Pulmonology. Conference: 28th Annual North American Cystic Fibrosis Conference. Atlanta, GA United States. Conference Publication: (var.pagings). 49 (pp 150-151), 2014. Date of Publication: September 2014.	No Lung Disease Outcome
Karamysheva Z et al	Regulation and function of a modifier of cystic fibrosis severity.	Pediatric Pulmonology. Conference: 29th Annual North American Cystic Fibrosis Conference. Phoenix, AZ United States. Conference Publication: (var.pagings). 50 (pp 252), 2015. Date of Publication: October 2015.	No Lung Disease Outcome
Saferali A et al	Decreased expression of the SPLUNC1 gene is associated with increased pulmonary disease severity in cystic fibrosis patients.	Pediatric Pulmonology. Conference: 26th Annual North American Cystic Fibrosis Conference. Orlando, FL United States. Conference Publication: (var.pagings). 47 (pp 280-281), 2012. Date of Publication: September 2012.	No Lung Disease Outcome

Authors	Title	Journal	Reason
Saferali A et al	The rs1078761 polymorphism is associated with cystic fibrosis severity and modulates expression of the SPLUNC1 gene.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2014. San Diego, CA United States. Conference Publication: (var.pagings). 189 (no pagination), 2014. Date of Publication: 2014.	No Lung Disease Outcome
Saferali A et al	The rs1078761 polymorphism is associated with reduced SPLUNC1 expression and increased pulmonary disease severity in cystic fibrosis patients.	Journal of Cystic Fibrosis. Conference: 36th European Cystic Fibrosis Conference. Lisbon Portugal. Conference Publication: (var.pagings). 12 (pp S41)	No Lung Disease Outcome
Polineni D et al	Transforming growth factor beta 1(TGFB1) expression in vivo is correlated with interleukin 8 (IL8) expression in CF airway (nasal) epithelia.	Pediatric Pulmonology. Conference: 27th Annual North American Cystic Fibrosis Conference. Salt Lake City, UT United States. Conference Publication: (var.pagings). 48 (pp 267), 2013. Date of Publication: October 2013.	No Lung Disease Outcome
Snodgrass S.M. et al	DAB2 mediates the TGFbeta-dependent downregulation of CFTR in human airwayepithelial cells.	Pediatric Pulmonology. Conference: 24th Annual North American Cystic Fibrosis Conference. Baltimore, MD United States. Conference Publication: (var.pagings). 45 (pp 227-228), 2010. Date of Publication: September 2010.	No Lung Disease Outcome
Fanen P. et al	CFTR partners as potential modifiers of CF airway inflammation.	Pediatric Pulmonology. Conference: 28th Annual North American Cystic Fibrosis Conference. Atlanta, GA United States. Conference Publication: (var.pagings). 49 (pp 151-152), 2014. Date of Publication: September 2014.	No Lung Disease Outcome
Harris W.T. et al	Analyzying the immune repertoire in CF and Non-CF lungs.	Pediatric Pulmonology. Conference: 29th Annual North American Cystic Fibrosis Conference. Phoenix, AZ United States. Conference Publication: (var.pagings). 50 (pp 226-227), 2015. Date of Publication: October 2015.	No Lung Disease Outcome
Polineni D. et al	Characterizing inflammation in the CF nose: Complementary genomic & gene expression studies to define genetic modifiers of cystic fibrosis lung disease severity.	Pediatric Pulmonology. Conference: 24th Annual North American Cystic Fibrosis Conference. Baltimore, MD United States. Conference Publication: (var.pagings). 45 (pp 292), 2010. Date of Publication: September 2010.	No Lung Disease Outcome
Polineni D. et al	In vivo human nasal epithelia inform transcriptomic studies in cystic fibrosis (CF) complementary to genomic variation studies.	Pediatric Pulmonology. Conference: 28th Annual North American Cystic Fibrosis Conference. Atlanta, GA United States. Conference	No Lung Disease Outcome

Authors	Title	Journal	Reason
Publication: (var.pagings). 49 (pp 272), 2014. Date of Publication: September 2014.			
Polineni D. et al	Correlation of inflammatory markers in human nasal epithelia in vivo informs transcriptomic studies in cystic fibrosis (CF) complementary to genomic variation studies.	American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS 2014. San Diego, CA United States. Conference Publication: (var.pagings). 189 (no pagination), 2014. Date of Publication: 2014.	No Lung Disease Outcome
Stolzenburg LR et al	Regulatory dynamics of 11p13 suggest a role for EHF in modifying CF lung disease severity.	Nucleic Acids Res. 2017 May 26. doi: 10.1093/nar/gkx482. [Epub ahead of print]	No Lung Disease Outcome
Ramos MD et al	Extensive sequence analysis of CFTR, SCNN1A, SCNN1B, SCNN1G and SERPINA1 suggests an oligogenic basis for cystic fibrosis-like phenotypes.	Clinical Genetics. 86(1):91-5, 2014 Jul	No Lung Disease Outcome
Watt AP et al	Frequency of cytokine gene promoter polymorphisms in the Northern Ireland cystic fibrosis population.	Journal of Cystic Fibrosis. 6(5):363-5, 2007 Sep	No Lung Disease Outcome
Stanke F et al	The CF-modifying gene EHF promotes p.Phe508del-CFTR residual function by altering protein glycosylation and trafficking in epithelial cells.	European Journal of Human Genetics. 22(5):660-6, 2014 May	No Lung Disease Outcome
Huber R et al	Functional characterization of a partial loss-of-function mutation of the epithelial sodium channel (ENaC) associated with atypical cystic fibrosis.	Cellular Physiology & Biochemistry. 25(1):145-58, 2010	No Lung Disease Outcome
Endale Ahanda ML et al	The hsa-miR-125a/hsa-let-7e/hsa-miR-99b cluster is potentially implicated in Cystic Fibrosis pathogenesis.	Journal of Cystic Fibrosis. 14(5):571-9, 2015 Sep	No Lung Disease Outcome
Nagy B Jr et al	Human Epididymis Protein 4: A Novel Serum Inflammatory Biomarker in Cystic Fibrosis.	Chest. 150(3):661-72, 2016 Sep	No Lung Disease Outcome
Voisin G et al	Oxidative stress modulates the expression of genes involved in cell survival in DELTAF508 cystic fibrosis airway epithelial cells.	Physiological Genomics. 46(17):634-46, 2014 Sep 01	No Lung Disease Outcome
Adriaanse MP et al	HLA frequencies and associations in cystic fibrosis.	Tissue Antigens. 83(1):27-31, 2014 Jan	No Lung Disease Outcome

Authors	Title	Journal	Reason
Nicolis E et al	The GCC repeat length in the 5'UTR of MRP1 gene is polymorphic: a functional characterization of its relevance for cystic fibrosis.	BMC Medical Genetics. 7:7, 2006 Feb 07	No Lung Disease Outcome
Sharma P et al	MUC5B and MUC7 are differentially expressed in mucous and serous cells of submucosal glands in human bronchial airways.	American Journal of Respiratory Cell & Molecular Biology. 19(1):30-7, 1998 Jul	No Lung Disease Outcome
Sanchez-Dominguez CN et al	The tumor necrosis factor alpha (-308 A/G) polymorphism is associated with cystic fibrosis in Mexican patients.	PLoS ONE [Electronic Resource]. 9(3):e90945, 2014	No Lung Disease Outcome
Buscher R et al	P2Y2 receptor polymorphisms and haplotypes in cystic fibrosis and their impact on Ca ²⁺ influx.	Pharmacogenetics and Genomics. 16(3):199-205, 2006 Mar	No Lung Disease Outcome
Tomati V et al	Genetic Inhibition Of The Ubiquitin Ligase Rnf5 Attenuates Phenotypes Associated To F508del Cystic Fibrosis Mutation.	Scientific Reports. 5:12138, 2015 Jul 17	No Lung Disease Outcome
Mekus F et al	Genes, environment, ion transport, and cystic fibrosis.	American Journal of Respiratory & Critical Care Medicine. 169(6):770; author reply 770, 2004 Mar 15	No Lung Disease Outcome
Stolzenburg L.R. et al	Characterizing regions of open chromatin at a cystic fibrosis modifier locus.	Pediatric Pulmonology. Conference: 29th Annual North American Cystic Fibrosis Conference. Phoenix, AZ United States. Conference Publication: (var.pagings). 50 (pp 252), 2015. Date of Publication: October 2015.	No Lung Disease Outcome
Saferali A. et al	Polymorphisms associated with expression of bpifa1/bpifb1 and lung disease severity in cystic fibrosis.	American Journal of Respiratory Cell and Molecular Biology. 53 (5) (pp 607-614), 2015. Date of Publication: November 2015.	No Lung Disease Outcome
Clain J et al	A neutral variant involved in a complex CFTR allele contributes to a severe cystic fibrosis phenotype.	Human Genetics. 116(6):454-60, 2005 May	No Lung Disease Outcome
Oueslati S et al	Preliminary study of haplotypes linked to the rare cystic fibrosis E1104X mutation.	Acta Physiologica Hungarica. 102(1):86-93, 2015 Mar	No Lung Disease Outcome
Cordovado SK et al	CFTR mutation analysis and haplotype associations in CF patients.	Molecular Genetics & Metabolism. 105(2):249-54, 2012 Feb	No Lung Disease Outcome

Authors	Title	Journal	Reason
Makowski GS et al	Single tube multiplex PCR detection of 27 cystic fibrosis mutations and 4 polymorphisms using neonatal blood samples collected on Guthrie cards.	Annals of Clinical & Laboratory Science. 33(3):243-50, 2003 Summer	No Lung Disease Outcome
Bacci G. et al	Taxonomic and functional metagenomic analysis of sputum samples from stable CF patients with lower and higher pulmonary function decline.	Pediatric Pulmonology. Conference: 29th Annual North American Cystic Fibrosis Conference. Phoenix, AZ United States. Conference Publication: (var.pagings). 50 (pp 306-307), 2015. Date of Publication: October 2015.	No Non-CFTR Gene Tested
Thomassen J.C. et al	TGFbeta¹-a possible biomarker for disease progression in cystic fibrosis?.	Journal of Cystic Fibrosis. Conference: 39th European Cystic Fibrosis Conference. Switzerland. 15 (pp S7), 2016. Date of Publication: June 2016.	No Non-CFTR Gene Tested
Harris WT et al	Constrictive Bronchiolitis in Cystic Fibrosis Adolescents with Refractory Pulmonary Decline.	Ann Am Thorac Soc. 2016 Dec;13(12):2174-2183.	No Non-CFTR Gene Tested
Vanscoy LL et al	Heritability of lung disease severity in cystic fibrosis.	American Journal of Respiratory & Critical Care Medicine. 175(10):1036-43, 2007 May 15	No Non-CFTR Gene Tested
Mekus F et al	Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics.	Twin Research. 3(4):277-93, 2000 Dec	No Non-CFTR Gene Tested
Lucarelli M et al	A Genotypic-Oriented View of CFTR Genetics Highlights Specific Mutational Patterns Underlying Clinical Macrocategories of Cystic Fibrosis.	Molecular Medicine. 21:257-75, 2015 Apr 21	No Non-CFTR Gene Tested
Maurya N, Awasthi S, Dixit P	Association of CFTR gene mutation with bronchial asthma. [Review]	Indian Journal of Medical Research. 135(4):469-78, 2012 Apr	No Non-CFTR Gene Tested
Chanock SJ, Foster CB	SNPing away at innate immunity. [Review] [21 refs]	Journal of Clinical Investigation. 104(4):369-70, 1999 Aug	Review Article
Stanke F, Tummler B, Becker T	Genetic modifiers in cystic fibrosis.	New England Journal of Medicine. 354(1):88-90; author reply 88-90, 2006 Jan 05	Review Article
Cutting G.R.	Cystic Fibrosis: Using genetic association to identify modifiers of disease variability in cystic fibrosis.	European Journal of Human Genetics. 14 (8) (pp 890-891), 2006. Date of Publication: August 2006.	Review Article

Authors	Title	Journal	Reason
Cutting GR	Causes of variation in the cystic fibrosis phenotype.	Annales Nestle. 64 (3) (pp 111-117), 2006. Date of Publication: November 2006.	Review Article
Dorfman R	Modifier gene studies to identify new therapeutic targets in cystic fibrosis.	Current Pharmaceutical Design. 18 (5) (pp 674-682), 2012. Date of Publication: January 2012.	Review Article
Drumm ML et al	Genetic variation and clinical heterogeneity in cystic fibrosis.	Annual Review of Pathology: Mechanisms of Disease. 7 (pp 267-282), 2012. Date of Publication: 2012.	Review Article
Gardet A et al	Genetic architecture of human fibrotic diseases: Disease risk and disease progression.	Frontiers in Pharmacology. 4 DEC (no pagination), 2013. Article Number: Article 159. Date of Publication: 2013.	Review Article
Knowles M	Gene modifiers in cystic fibrosis: A consortium approach.	Medizinische Genetik. Conference: 21. Jahrestagung der Deutschen Gesellschaft fur Humangenetik. Hamburg Germany. Conference Publication: (var.pagings). 22 (1) (pp 65), 2010. Date of Publication: February 2010.	Review Article
Soferman R.	Immunopathophysiologic mechanisms of cystic fibrosis lung disease.	Israel Medical Association Journal. 8 (1) (pp 44-48), 2006. Date of Publication: January 2006.	Review Article
Weiler CA et al	Genetic influences on cystic fibrosis lung disease severity.	Frontiers in Pharmacology. 4 APR (no pagination), 2013. Article Number: Article 40. Date of Publication: 2013.	Review Article
Cantor CR et al	Haplotyping in biomedicine-practical challenges.	Nature Biotechnology. 23(1):21-2, 2005 Jan	Review Article
Castaldo G et al	What is the role of the non-coding regions of the CFTR gene in cystic fibrosis?.	Expert Review of Respiratory Medicine. 7(4):327-9, 2013 Aug	Review Article
Egan ME	Genetics of Cystic Fibrosis: Clinical Implications. [Review]	Clinics in Chest Medicine. 37(1):9-16, 2016 Mar	Review Article
Chang EH et al	Precision Genomic Medicine in Cystic Fibrosis. [Review]	Clinical and translational science. 8(5):606-10, 2015 Oct	Review Article
Guillot L et al	Lung disease modifier genes in cystic fibrosis. [Review]	International Journal of Biochemistry & Cell Biology. 52:83-93, 2014 Jul	Review Article
Cutting GR	Cystic fibrosis genetics: from molecular understanding to clinical application. [Review]	Nature Reviews Genetics. 16(1):45-56, 2015 Jan	Review Article

Authors	Title	Journal	Reason
Knowles MR et al	The influence of genetics on cystic fibrosis phenotypes.	Cold Spring Harbor Perspectives in Medicine. 2(12):a009548, 2012 Dec 01	Review Article
Cutting GR	Modifier genes in Mendelian disorders: the example of cystic fibrosis. [Review]	Annals of the New York Academy of Sciences. 1214:57-69, 2010 Dec	Review Article
Wolfenden LL et al	Genetic and non-genetic determinants of outcomes in cystic fibrosis. [Review] [79 refs]	Paediatric Respiratory Reviews. 10(1):32-6, 2009 Mar	Review Article
Collaco JM et al	Update on gene modifiers in cystic fibrosis. [Review] [94 refs]	Current Opinion in Pulmonary Medicine. 14(6):559-66, 2008 Nov	Review Article
Buscher R et al	Disease modifying genes in cystic fibrosis: therapeutic option or one-way road?. [Review] [131 refs]	Naunyn-Schmiedebergs Archives of Pharmacology. 374(2):65-77, 2006 Nov	Review Article
Boyle MP	Strategies for identifying modifier genes in cystic fibrosis. [Review] [62 refs]	Proceedings of the American Thoracic Society. 4(1):52-7, 2007 Jan	Review Article
Knowles MR	Gene modifiers of lung disease. [Review] [38 refs]	Current Opinion in Pulmonary Medicine. 12(6):416-21, 2006 Nov	Review Article
Cutting GR	Modifier genetics: cystic fibrosis. [Review] [133 refs]	Annual Review of Genomics & Human Genetics. 6:237-60, 2005	Review Article
Landau LI	A monogenetic disorder yet multiple and varied clinical manifestations.	Archives of Disease in Childhood. 90(10):992-3, 2005 Oct	Review Article
Davies JC et al	Modifier genes in cystic fibrosis. [Review] [64 refs]	Pediatric Pulmonology. 39(5):383-91, 2005 May	Review Article
Davies J et al	Cystic fibrosis modifier genes. [Review] [62 refs]	Journal of the Royal Society of Medicine. 98 Suppl 45:47-54, 2005	Review Article
Davies JC	Modifier genes in cystic fibrosis. [Review] [9 refs]	Pediatric Pulmonology - Supplement. 26:86-7, 2004	Review Article
Merlo CA et al	Modifier genes in cystic fibrosis lung disease. [Review] [51 refs]	Journal of Laboratory & Clinical Medicine. 141(4):237-41, 2003 Apr	Review Article
Acton JD et al	Phenotype of CF and the effects of possible modifier genes. [Review] [77 refs]	Paediatric Respiratory Reviews. 2(4):332-9, 2001 Dec	Review Article
Drumm ML	Modifier genes and variation in cystic fibrosis. [Review] [20 refs]	Respiratory Research. 2(3):125-8, 2001	Review Article

Authors	Title	Journal	Reason
Mickle JE et al	Genotype-phenotype relationships in cystic fibrosis. [Review] [72 refs]	Medical Clinics of North America. 84(3):597-607, 2000 May	Review Article
Zielenski J	Genotype and phenotype in cystic fibrosis. [Review] [119 refs]	Respiration. 67(2):117-33, 2000	Review Article
Luisetti M	Influence of a secondary genetic factor in cystic fibrosis genotype-phenotype correlations. [Review] [6 refs]	Monaldi Archives for Chest Disease. 52(4):390-1, 1997 Aug	Review Article
Kerem B et al	The molecular basis for disease variability in cystic fibrosis. [Review] [69 refs]	European Journal of Human Genetics. 4(2):65-73, 1996	Review Article
Estivill X	Complexity in a monogenic disease.	Nature Genetics. 12(4):348-50, 1996 Apr	Review Article
Ramsey, Bonnie W et al	Future directions in early cystic fibrosis lung disease research: an NHLBI workshop report.	American Journal of Respiratory & Critical Care Medicine. 185(8):887-92, 2012 Apr 15	Review Article
Hampton, Thomas H et al	A novel approach to analyze gene expression data demonstrates that the DeltaF508 mutation in CFTR downregulates the antigen presentation pathway. [Review] [52 refs]	American Journal of Physiology - Lung Cellular & Molecular Physiology. 298(4):L473-82, 2010 Apr	Review Article
Sagel, Scott D et al	Monitoring inflammation in CF. Cytokines. [Review] [97 refs]	Clinical Reviews in Allergy & Immunology. 23(1):41-57, 2002 Aug	Review Article
Gasparini P et al	The genotype of a new linked DNA marker, MP6d-9, is related to the clinical course of cystic fibrosis	J Med Genet 1990;27(1):17-20. PMID: 1968514	No Lung Disease Outcome
Baranov VS et al	Proportion of the GSTM1 0/0 genotype in some Slavic populations and its correlation with cystic fibrosis and some multifactorial diseases.	Hum Genet 1996;97:516-20	No Lung Disease Outcome
Sterpone S et al	DNA damage and related modifier genes in Italian Cystic fibrosis patients	Biol Res 42: 477-486, 2009	No Lung Disease Outcome
Aron Y et al	HLA-DR polymorphism in allergic bronchopulmonary aspergillosis.	J Allergy Clin Immunol ; 1999a;104:891-2.	Examines Patient Without CF

Authors	Title	Journal	Reason
Chalmers JD et al	Impact of mannose-binding lectin insufficiency on the course of cystic fibrosis: a review and meta-analysis	Glycobiology 2011;21:271–82.	Review Article
Grasemann H et al	Effects of gender and gene variants in constitutive nitric oxide synthases on exhaled nitric oxide.	Am J Respir Crit Care Med 2003a.	Examines Patient Without CF
Muhlebach MS et al	Association between mannan-binding lectin and impaired lung function in cystic fibrosis may be age-dependent.	Clin Exp Immunol 2006;145: 302–7.	No Lung Disease Outcome
Rohlfes EM et al	Is the hemochromatosis gene a modifier locus for cystic fibrosis?	Genet Test 1998;2:85–88	No Lung Disease Outcome
Devaney J et al	HFE alleles in an Irish cystic fibrosis population.	Genet Test 2003;7:155–158	No Lung Disease Outcome

Chapter 2 Supplementary Table 2: Included papers

See the supplementary file: *Chapter 2 Supplementary Table 2;_Included Papers_Clean.xlsx*.

Chapter 2 Supplementary Table 3: Summary of results by gene

Candidate genes

Gene	Reason	No.	Articles	Results
<i>A20</i>	Inflammation	1	(Kelly, Williams et al. 2013)	In a study of 13 Irish adult patients, increased <i>A20</i> expression from nasal epithelial cells was associated with better FEV ₁ .
<i>ABCC1-5</i>	Member of ATP Binding Cassete Superfamily (as is CFTR)	3	(Hurbain, Sermet-Gaudelus et al. 2003, Mafficini, Ortombina et al. 2011, Ideozu, Zhang et al. 2017)	In an early study looking at gene expression from nasal epithelial cells, Hurbain et al. found low <i>ABCC1</i> expression was associated with a lower Shwachman score, indicating more severe lung disease. They found <i>ABCC 2- 5</i> expression was not related to lung disease severity. Mafficini et al. found that at SNP rs504348 CC was associated with earlier PA infection and decline in lung function. Ideozu et al. analysed mutation, methylation and gene expression at rs504348 and did not replicate either of the prior findings, although this was the only study which was not limited to Phe508del homozygous patients.
<i>ABO</i>	Infection susceptibility	1	(Taylor-Cousar, Zariwala et al. 2009)	In a large group of North American Phe508del homozygous patients, no effect demonstrated on lung function of PA infection
<i>ACE</i>	Inflammation	6	(Arkwright, Pravica et al. 2003, Jaffe, Prasad et al. 2003, Drumm, Konstan et al. 2005, Marson, Ribeiro et al. 2010, Marson, Bertuzzo et al. 2012, Sahli, Fredj et al. 2014)	Arkwright et al, Sahli et al, Marson et al, Marson et al. found D/D genotype associated with more severe disease, whilst Jaffe et al. found a suggestion of a protective effect of D/D in older children. The most robust study, the Gene Modifier Study, found no effect.
<i>ADRA2A</i>	Inflammation, Airway Hyper-responsiveness	1	(Marson, Rezende et al. 2013)	Authors report several significant results such, as rs553668 GA having higher Shwachman-Kulczycki scores, however the clinical significance of these findings if limited.
<i>ADRB2</i>	Therapy Target (Bronchodilator)	5	(Buscher, Eilmes et al. 2002, Drumm, Konstan et al. 2005, Hart, Konstan et al. 2005, Steagall, Barrow	Multiple studies have examined the effect of <i>ADRB2</i> , mainly at codon 16 and codon 27, with contradictory results. The most robust study, the Gene Modifier Study, found genotype at codon 16 or 27 had no effect on lung disease severity. At codon 16, Buscher et al. found that arg/arg was associated with milder lung disease, as did Steagall et al. in a subgroup of

Gene	Reason	No.	Articles	Results
			et al. 2007, Marson, Bertuzzo et al. 2012)	Phe508del homozygous patients. Marson et al, however found arg/arg was associated with more severe disease. At codon 27 the three smaller studies found contradictory results again. Only one study investigated Codon 164, and identified 4 patients with the the/ile genotype all of whom had unusually severe lung disease.
<i>AGER -429</i>	Inflammation	3	(Laki, Laki et al. 2006, Beucher, Boelle et al. 2012, Corvol, Beucher et al. 2012)	In an initial study of patients who were Phe508del homozygous or heterozygous, carrying a C allele as part of 8.1 ancestral MHC haplotype was associated with later colonisation with PA and SA. This is in contrast to 2 later studies, one of Phe508del homozygous patients and none including all genotypes, which found that carrying C allele was associated more severe disease.
<i>AGTR2</i>	Genome Wide	1	(Corvol, Blackman et al. 2015)	In a meta-analysis involving 6,365 patients from North America and France who underwent SNP microarray, <i>AGTR2</i> was identified as a modifier of lung disease severity.
<i>AHSAI</i>	CFTR Trafficking	1	(Gisler, von Kanel et al. 2013)	In a 95 Swiss Phe508del homozygous patients <i>AHSAI</i> did not influence lung disease severity, which was measured with multiple measures of lung function including spirometry and multibreath washout.
<i>α 1 antitrypsin</i>	Inflammation	8	(Doring, Krogh-Johansen et al. 1994, Mahadeva, Stewart et al. 1998, Mahadeva, Westerbeek et al. 1998, Henry, Cave et al. 2001, Meyer, Braun et al. 2002, Frangolias, Ruan et al. 2003, Drumm, Konstan et al. 2005, Courtney, Plant et al. 2006)	Multiple studies have examined the effect of deficient (PIMM) and Deficient (PIMS, PIMZ) genotype, as well as the effect of polymorphism at 1237. Although Doring et al. reported earlier PA infection in deficient genotypes whilst Mahadeva et al. reported better lung function in deficient patients, the majority of studies including the Gene Modifier Study have found no difference in lung disease severity between sufficient and deficient genotypes. A group from Ireland have published two studies showing presence of an A allele at 1237 is associated with milder lung disease, however this has not been replicated in other more robust studies.
<i>α – 1-antichymotrypsin</i>	Inflammation	1	(Mahadeva, Sharples et al. 2001)	One study found deficient patients had milder lung disease, as measured by higher FEV ₁ , better CXR scores and less PA infection. It also reported that at -15 signal peptide the GG genotype was also associated with milder disease.
<i>AnxA5</i>	Directly interacts with CFTR	1	(Gisler, von Kanel et al. 2013)	In a 95 Swiss Phe508del homozygous patients <i>AnxA5</i> did not influence lung disease severity, which was measured with multiple measures of lung function including spirometry and multibreath washout.

Gene	Reason	No.	Articles	Results
<i>APIP</i>	Genome Wide	1	(Corvol, Blackman et al. 2015)	In a meta-analysis involving 6,365 patients from North America and France who underwent SNP microarray, <i>APIP</i> was identified as a modifier of lung disease severity.
<i>ATF1</i>	Genome Wide	2	(Wright, Merlo et al. 2006, Magalhaes, Rivals et al. 2017)	In a study that examined genome wide gene expression, <i>ATF1</i> was differentially expressed between patients with mild and severe lung disease. No link between methylation at <i>ATF1</i> and lung disease severity was found in a subsequent study. Both studies examined nasal epithelial cells.
<i>C3</i>	PA Infection Susceptibility	1	(Park, Yung et al. 2011)	In one study which examined 2 different cohorts found that C3 may influence lung disease severity. At rs393770 the G allele was associated with longer time to PA infection and better FEV ₁ in adults. At rs11569393 the T allele was associated with worse lung disease. Observations did not survive pooling of results of primary and secondary cohorts
<i>CALR</i>	CFTR Trafficking	1	(Gisler, von Kanel et al. 2013)	In 95 Swiss Phe508del homozygous patients <i>CALR</i> did not influence lung disease severity, which was measured with multiple measures of lung function including spirometry and multibreath washout.
<i>CAV2</i>	Genome Wide	1	(Emond, Louie et al. 2015)	In a large study, which initially used whole exome sequencing and then replicated the findings in another cohort, the minor allele at rs8940 was associated with later onset of PA infection and better FEV ₁ .
<i>CEACAM</i>	Infection Susceptibility	2	(Becker, Jansen et al. 2007, Stanke, Becker et al. 2010)	Two separate studies from the European Twins and Siblings group which interrogated 19q13 have identified <i>CEACAM</i> as a modifier of lung disease severity based on distribution of alleles when mildly and severely affected pairs are compared, and when discordant siblings are compared. The two studies have studied different SNPs.
<i>CD14</i>	PA Infection Susceptibility	3	(Martin, Laing et al. 2005, Faria, Faria et al. 2009, Haerynck, Mahachie John et al. 2013)	Faria et al. and Haerynck et al. found no effect on lung function, with neither study looking at microbiological end points. In a very robust study looking at annual BAL results in children Martin et al. found that at position -159, CC genotype associated with earlier PA infection.
<i>CLC-2</i>	Alternate chloride channel	1	(Blaisdell, Howard et al. 2004)	In a small study from USA which compared patients with mild or severe lung disease, based on FEV ₁ , no relationship with <i>CLC-2</i> was demonstrated.
<i>CHIT1</i>	Fungal Infection Susceptibility	1	(Livnat, Bar-Yoseph et al. 2014)	In a small study from Israel, <i>CHIT1</i> duplication was not associated with ABPA.

Gene	Reason	No.	Articles	Results
<i>Chitinase 3-like 1</i>	Fungal Infection Susceptibility	1	(Coriati, Bouvet et al. 2016)	In an abstract describing preliminary data rs4950928 C/C associated with increased pulmonary exacerbations, and those with lowest FEV ₁ more commonly express rs871799 C/G
<i>Complement Factor B</i>	PA Infection Susceptibility	1	(Park, Yung et al. 2011)	In a study that used 2 cohorts, in the initial cohort <i>CFB</i> was associated with lung disease severity however this was not replicated in the larger second cohort or in a pooled analysis.
<i>COX1</i>	Inflammation	1	(Czerska, Sobczynska-Tomaszewska et al. 2010)	One study of Polish patients found protective effect of 2 polymorphisms but in very limited subgroup Phe508del homozygous patients between 8-13years.
<i>COX2</i>	Inflammation	3	(Czerska, Sobczynska-Tomaszewska et al. 2010, Marcellino, Marson et al. 2013, Marson, Marcelino et al. 2013)	One study of Polish patients found protective effect of GC at 765 on FEV ₁ in Phe508del homozygous patients between 8-13years, and TC at 8473 on FEV ₁ , PA infection, Disease Severity Score and Brasfield score in Phe508del homozygous patients between 4-13years. These findings were not replicated other studies.
<i>CXCR1 and CXCR2</i>	Infection Susceptibility	1	(Kormann, Hector et al. 2012)	One study has examined <i>CXCR1</i> and 2 and found that certain SNPs (<i>CXCR1</i> rs16858841, rs16858816, rs16858811, <i>CXCR2</i> rs3890157, rs11574750, rs13306441) were associated longitudinal FEV ₁ predicted at 20yrs of age, and that being homozygous for haplotype <i>CXCR1_Ha</i> and <i>CXCR2_Ha</i> is protective for lung function and this is additive (more protective if homozygous for both)
<i>D7S525, D7S514, D7S495</i>	In Vicinity of CFTR	1	(Mekus, Laabs et al. 2003)	In a study utilising patients from the European Twins and Siblings cohort which interrogated the effect of genes in the vicinity of CFTR and their effect on lung disease, no relationship was found.
<i>DCTN4</i>	Exome Wide	2	(Emond, Louie et al. 2012, Viel, Hubert et al. 2016)	In the first whole exome study involving CF patients, Emond et al. found missense mutations at rs11954652 and rs35772018 in <i>DCTN4</i> were associated with earlier first PA infection and earlier chronic PA infection in a group of 91 patients. In the same paper they replicated their findings in 1322 patients. Viel et al. attempted to replicate this findings in a smaller group and found missense variant of <i>DCTN4</i> only increased chronic PA infection rates in the subgroup of males who were CFTR II/II.
<i>DEFB1</i>	Infection Susceptibility	4	(Vankeerberghen, Scudiero et al. 2005, Tesse, Cardinale et al. 2008, Segat, Morgutti et	Tesse et al. found that at position -52, carrying an A allele increased risk of chronic PA infection. Similarly carrying G allele at position -20, and being A-C-G for haplotype -52, -44, -20 increased risk of PA infection. There was no effect on lung function. Vankeerberghen et al. found no effect although this may be due to a population limited to Phe508del

Gene	Reason	No.	Articles	Results
			al. 2010, Crovella, Segat et al. 2011)	homozygotes and they examined time of first PA infection rather than chronic infection. In a smaller study, using a smaller but similar patient group to Tesse et al. and the same outcome measures, Segat et al. found no effect on PA infection or FEV1. In a study of 92 adults, Crovella et al. found that at -20, AA genotype more commonly associated with milder lung disease, which may be in keeping with Tesse et al. findings given they found at presence of G at -20 was associated with increased infection risk.
<i>DEFβ4</i>	Infection Susceptibility	1	(Hollox, Davies et al. 2005)	In a study of 255 patients from the United Kingdom there was no relationship between copy number variants at <i>DEFβ4</i> and multiple measures of lung disease severity
<i>DUOX2</i>	Genome Wide	2	(Wright, Merlo et al. 2006, Magalhaes, Rivals et al. 2017)	In a study that examined genome wide gene expression, <i>DUOX2</i> was differentially expressed between patients with mild and severe lung disease. No link between methylation at <i>DUOX2</i> and lung disease severity was found in a subsequent study. Both studies examined nasal epithelial cells.
<i>EDNRA</i>	Broncho-constriction	2	(Darrah, McKone et al. 2010, Magalhaes, Rivals et al. 2017)	Across four different cohorts from USA and Ireland, rs5335 C/C was associated with reduced lung function (approximately 5% absolute less) than G/G. A subsequent study showed and inverse relationship between methylation at <i>EDNRA</i> in blood and disease severity. No link with methylation in nasal epithelial cells was found.
<i>EHF</i>	Genome Wide	1	(Corvol, Blackman et al. 2015)	In a meta-analysis involving 6,365 patients from North America and France who underwent SNP microarray, <i>EHF</i> was identified as a modifier of lung disease severity.
<i>EZR</i>	Apical Plasma Membrane Constituent	1	(Soave, Corvol et al. 2015)	In a large study using patients from the Canadian Gene Modifier Study, and replicated in patients from the French Gene Modifier study, Soave et al. found <i>EZR</i> was related to lung disease severity
<i>FAS</i>	PA Infection Susceptibility	1	(Kumar, Becker et al. 2008)	In a group of Phe508del homozygous patients, variation in SNP rs7901656 at intron 2 associated with disease severity and gene expression. This SNP is representative of a wider haplotype. The direction of effect on disease severity is unclear.
<i>FCN1 and FCN2</i>	Infection Susceptibility	1	(Haerynck, Van Steen et al. 2012)	One study has examined <i>FCN1</i> and <i>FCN2</i> in 96 Belgian patients. In <i>FCN1</i> they found 2 SNPs associated with risk of PA infection; Carriage of a G allele at rs2989727 or carriage of an A allele at rs1071583 is associated with earlier PA infection. In <i>FCN2</i> they found 2 SNPs associated with risk of PA infection; At rs7865453 AC, vs. AA is associated with earlier PA infection and at rs7851696 GT, vs. GG is also associated with earlier PA infection.

Gene	Reason	No.	Articles	Results
<i>FUT2, FUT3</i>	Infection Susceptibility	1	(Taylor-Cousar, Zariwala et al. 2009)	In a large group of North American Phe508del homozygous patients, no effect demonstrated on lung function of PA infection
<i>GCLC</i>	Oxidative Lung Injury	3	(McKone, Shao et al. 2006, de Lima Marson, Bertuzzo et al. 2013, Marson, Bertuzzo et al. 2014)	The most robust study showed that in patients with mild CFTR mutations, increasing <i>GCLC</i> GAG trinucleotide repeat number was associated with increased FEV ₁ . Two other studies found significant effects, with 129 CC (vs. TC, TT) associated with higher rates of mucoid PA, and -3506 AA (vs. AG, GG) associated with higher rates of non-mucoid PA and higher Bhalla score.
<i>Glucocorticoid Receptor</i>	Inflammation	1	(Corvol, Nathan et al. 2007)	In a large French study at the Bcll polymorphism, GG associated with more rapid decline in FEV1 and FVC than CC or CG and this is more pronounced in Phe508del homozygous patients
<i>GSTM1</i>	Oxidative Stress	9	(Hull and Thomson 1998, Flamant, Henrion-Cau de et al. 2004, Korytina, Iaibaeva et al. 2004, Drumm, Konstan et al. 2005, Feuillet-Fieux, Nguyen-Khoa et al. 2009, Lima, Ortega et al. 2012, de Lima Marson, Bertuzzo et al. 2013, Marson, Bertuzzo et al. 2014, Magalhaes, Rivals et al. 2017)	<i>GSTM1</i> has been evaluated by multiple studies and whilst some small studies (Hull et al, Marson et al, Marson et al) have shown a link between the null genotype and outcomes such as worse radiological and clinical scores, the largest study (Drumm et al) found no relationship. In a more recent study no link between methylation and lung disease severity was found.
<i>GSTM3</i>	Oxidative Stress	2	(Flamant, Henrion-Cau de et al. 2004, Magalhaes, Rivals et al. 2017)	A study involving 146 French patients found that carriage of a B allele at <i>GSTM3</i> was associated with better lung function at 9 years of age. A subsequent study found that methylation in nasal epithelial cells correlated with lung disease severity but direction of effect unclear.
<i>GSTP1</i>	Oxidative Stress	6	(Flamant, Henrion-Cau de et al. 2004, Korytina, Iaibaeva et al. 2004, Drumm, Konstan et al. 2005, Feuillet-Fieux,	In a study of 60 Russian patients, Korytina et al. assessed lung disease severity using the Schwachman-Bracefield score and found that at position 105 Val/Val was found more often in the moderate disease group, and Ile/Val more often in the severe group. A small study of French children by Feuillet-Fieux et al. found that AA genotype (as compared to AG, GG) is

Gene	Reason	No.	Articles	Results
			Nguyen-Khoa et al. 2009, de Lima Marson, Bertuzzo et al. 2013, Marson, Bertuzzo et al. 2014)	protective against PA infection. No effect seen in other studies, including the larger and more robust Gene Modifier Study.
<i>GSTT1</i>	Oxidative Stress	5	(Flamant, Henrion-Cau de et al. 2004, Feuillet-Fieux, Nguyen-Khoa et al. 2009, Lima, Ortega et al. 2012, de Lima Marson, Bertuzzo et al. 2013, Marson, Bertuzzo et al. 2014)	One study in a Brazilian population by Marson et al. found the null genotype associated with worse Bhalla score and oxygen saturations, whilst other studies were did not show an effect.
<i>Haptoglobin</i>	Infection susceptibility	1	(Shteinberg, Rivlin et al. 2015)	In a study of 142 Irish and Israeli patients, no effect was observed
<i>HFE</i>	Infection susceptibility	2	(Pratap, Quinn et al. 2010, Smith, Bell et al. 2014)	One article and one abstract from separate Australian centres show that carrying a <i>HFE</i> mutation is associated with more severe lung disease. Pratap et al. showed carriers of at least one <i>HFE</i> mutation (H63D, C282Y) had lower FEV ₁ (absolute difference 14.4%), Lower FVC, (absolute difference 12.2%) and faster annual rate of decline (-3.37% vs. -1.19%) whilst Smith et al. showed C282Y carriers had a lower FEV ₁ than patients with no mutation or the H63D mutation.
<i>HLA – Class II</i>	Atopy Susceptibility and Genome Wide	3	(Aron, Polla et al. 1999, Corvol, Blackman et al. 2015, O’Neal, Gallins et al. 2015)	An initial study involving 98 French adults showed that DR7 was associated with higher serum IgE and higher PA infection, where as DR4 was associated with lower rates of PA with no effect on lung function. In two large subsequent genome wide studies, HLA-Class II has been associated with lung disease severity. In a meta-analysis of genome wide SNP microarrays performed on patients from North America and France, HLA-DRA was identified as a modifier of lung disease severity. In a study involving 754 patients from USA which examined genome wide gene expression data from lymphoblasts, HLA-Class II genes (including HLA-DRA) was associated with age at persistent PA infection.
<i>HLA-DQB1, HLA-DRB1</i>	Inflammation	1	(Laki, Laki et al. 2006)	In a study of Hungarian patients, when carried as part of 8.1 Ancestral MHC haplotype, HLA-DQB1*0201 and HLA-DRB1*0301 associated with delayed PA and SA colonisation

Gene	Reason	No.	Articles	Results
<i>HMOX1</i>	Infection Susceptibility	2	(Park, Yung et al. 2011, Magalhaes, Rivals et al. 2017)	In a large study of Canadian patients, when the paediatric group was analysed presence of A allele at rs2071749 was associated with higher FEV ₁ , and in adults presence of T allele at rs2071746 was associated with worse FEV ₁ . In a study of French patients with Increasing methylation in blood cells was associated there was increasing severity of lung disease, but this was not shown in nasal epithelial cells.
<i>HSP70-2G</i>	Inflammation	2	(Laki, Laki et al. 2006, Corvol, Beucher et al. 2012)	In an initial study Laki et al. found that G allele at 1267, as part of 8.1 Ancestral MHC haplotype, was associated with delayed PA and SA colonisation. This is in contrast to a subsequent larger study which found the 8.1 Ancestral MHC haplotype was associated with more severe lung disease, as measured by lower FEV ₁ .
<i>IFN- γ</i>	Inflammation	1	(Arkwright, Pravica et al. 2003)	In a study of 261 children and adults from the UK no link between <i>IFN- γ</i> and lung disease severity was found.
<i>IFN- γ-R1</i>	Inflammation	1	(Labenski, Hedtfeld et al. 2011)	Using patients from the European Twins and Siblings study, Labenski et al. identified 2 SNPs and a haplotype which were associated with lung disease severity.
<i>IFRD1</i>	Inflammation	4	(Gu, Harley et al. 2009, de Lima Marson, Bariani Marcelino et al. 2013, Marcelliino, Marson et al. 2013, Magalhaes, Rivals et al. 2017)	In a GWAS, rs7817 CT genotype (as opposed to CC or TT) was associated with lower lung function. In 2 subsequent studies in a Brazilian population no link was found. In addition no link between methylation of IFRD1 and lung disease severity was identified.
<i>IL-1A</i>	PA Infection Susceptibility, Inflammation	1	(Levy, Murphy et al. 2009)	In a large study of North American patients, in which an initial investigation of 808 Phe508del homozygous patients was performed, followed by replication in a smaller group encompassing all CFTR genotypes, there was no relationship between IL-1A and lung disease severity.
<i>IL-1β</i>	PA Infection Susceptibility, Inflammation	4	(Corvol, Boelle et al. 2008, Levy, Murphy et al. 2009, Labenski, Hedtfeld et al. 2011, de Vries, Griffiths et al. 2014)	Multiple studies have analysed different SNPs with contrasting results. In the first study Corvol et al. found no association between lung disease severity and rs1143627 and rs16944, whereas a later study by De Vries et al. found that carrying T (either TT or CT) at rs16944 was associated with slightly accelerated decline in FEV ₁ over 13 years (-0.41% per year). Labenski et al. found rs3917356 and rs4848306 was associated with disease severity, and Levy et al. found rs1143634 (AG/AA worse than GG) and rs1143939 (CT/TT, worse than CC) associated with cross sectional and longitudinal lung disease severity.

Gene	Reason	No.	Articles	Results
<i>IL-1R1</i>	PA Infection Susceptibility, Inflammation	1	(Levy, Murphy et al. 2009)	In a large study of North American patients, in which an initial investigation of 808 Phe508del homozygous patients was performed, followed by replication in a smaller group encompassing all CFTR genotypes, there was no relationship between IL-R1 and lung disease severity.
<i>IL-1RN</i>	PA Infection Susceptibility, Inflammation	3	(Corvol, Boelle et al. 2008, Levy, Murphy et al. 2009, Levy, Linse et al. 2010)	Whilst 2 large studies from North America and Europe found no effect, an abstract found that A allele at rs451578 (not studied previously) associated with lower median FEV ₁ , with A homozygous worse than compound heterozygous.
<i>IL-6</i>	Inflammation	1	(Corvol, Boelle et al. 2008)	A large European study examining all CFTR genotypes found that rs1800795 was not associated with lung disease severity.
<i>IL-8</i>	Inflammation	4	(Corvol, Boelle et al. 2008, Hillian, Londono et al. 2008, de Vries, Griffiths et al. 2014, Furlan, Marson et al. 2016)	Multiple studies have found conflicting results with the most robust evidence coming from Corvol et al. and Hillian et al. At -251 TT (vs. AA, AT) has been associated with later PA infection in a study by Corvol et al. and slower decline in FEV ₁ in a study by de Vries et al. However Furlan et al. found AT had higher FEV ₁ and Hillian et al. found TT was more commonly found in severely affected patients. At -396 Corvol et al. found TT (vs. GG, GT) associated with later PA infection, however Furlan et al. found TT was associated with more PA infection and lower FEV ₁ . At +781 Corvol et al. found CC (vs. TT, TC) associated with later PA infection, however once again Furlan et al. found that CC was associated with more PA infection, and also found TT was associated with lower FEV ₁ .
<i>IL-10</i>	Inflammation	6	(Arkwright, Pravica et al. 2003, Brouard, Knauer et al. 2005, Drumm, Konstan et al. 2005, Corvol, Boelle et al. 2008, Tesse, Cardinale et al. 2008, de Vries, Griffiths et al. 2014)	The only study to find a relationship between <i>IL-10</i> and lung function was by De Vries et al. who examined -1082 and found AG, AA (vs. GG) was associated with a faster rate of decline in FEV ₁ over 13 years (0.24% additional decline per year). Finding were not replicated in other studies including large European (Corvol) and North American Studies (Drumm). When examining microbiological endpoints Tesse et al. found that -1082 GG associated with trend to higher PA infection that AA, and that G-C-C haplotype at -1082, -819, -592 haplotype was associated with higher risk of PA infection. De Vries and Arkwright found no effect on PA infection. In the only study to examine Aspergillus infection and ABPA as an endpoint Brouard found that at -1082 GG associated with increased ABPA and Aspergillus infection.
<i>J3.11</i>	Proximity to CFTR, CFTR Haplotype	2	(Santis, Osborne et al. 1990, Mekus, Laabs et al. 2003)	Two relatively early studies both failed to show a relationship between <i>J3.11</i> and lung disease severity.

Gene	Reason	No.	Articles	Results
<i>KM.19</i>	CFTR Haplotype	1	(Santis, Osborne et al. 1990)	In an early study which examined haplotypes which included CFTR, <i>KM.19</i> was not associated with lung disease severity.
<i>KRT8</i>	CFTR trafficking and processing	1	(Stanke, Hedtfeld et al. 2011)	In a small group of well phenotyped patients who are part of the European Twins and Siblings study, haplotype 2211 at rs1907671, rs4300473, rs2035878 and rs2035875 was associated with milder clinical disease, and the contrasting haplotype 1122 associated with more severe clinical disease and no residual intestinal chloride secretion
<i>KRT18</i>	CFTR trafficking and processing	1	(Stanke, Hedtfeld et al. 2011)	In a small group of well phenotyped patients who are part of the European Twins and Siblings study, no effect was demonstrated
<i>KRT19</i>	CFTR trafficking and processing	1	(Gisler, von Kanel et al. 2013)	In a study of 95 Swiss Phe508del homozygous patients, 3 SNPS were found to be associated with lung disease severity measured using multibreath washout; c.90T>C, c.179G>C, c.471C>T
<i>Lewis</i>	Infection Susceptibility	1	(Taylor-Cousar, Zariwala et al. 2009)	In a large group of North American Phe508del homozygous patients, no effect demonstrated on lung function of PA infection
<i>Lipopolysachari de Binding protein (LBP)</i>	Infection Susceptibility and Inflammation	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who were either Phe508del homozygous or heterozygous no effect was demonstrated.
<i>LPAR6</i>	Genome Wide	1	(O'Neal, Gallins et al. 2015)	In a large study of Phe508del homozygous patients which examined genome wide gene expression from lymphoblastoid cells lines, higher <i>LPAR6</i> expression was associated with more severe lung disease.
<i>LT- α</i>	Inflammation	3	(Corvol, Boelle et al. 2008, Corvol, Beucher et al. 2012, Shmarina, Pukhalsky et al. 2013)	In a group of Russian patients, +252AA when carried together with TNF α -308GA had a higher FEV ₁ and FVC. Further reinforced by patients +252AA having a lower sputum neutrophil elastase. In support of this a large French study found that carriage of a G allele at +252, as part of 8.1 Ancestral MHC haplotype, was associated with more severe lung disease. In a separate study by the same French group, no link was found.
<i>LEP1, LEP2</i>	Proximity to CFTR	1	(Mekus, Laabs et al. 2003)	In a study utilising the European Twins and Siblings cohort found <i>LEP</i> polymorphisms where associated with disease severity, but a composite measure incorporating lung function and weight was used, and <i>LEP</i> genes thought to influence nutrition more than respiratory status.
<i>MASP-1</i>	PA Infection Susceptibility	1	(Haerynck, Van Steen et al. 2012)	In a study of 96 Belgian patients with at least one Phe508del mutation, no effect was found.

Gene	Reason	No.	Articles	Results
<i>MASP-2</i>	PA Infection Susceptibility	3	(Carlsson, Sjöholm et al. 2005, Olesen, Jensenius et al. 2006, Haerynck, Van Steen et al. 2012)	No effect found in 3 studies which together examined 317 patients. Olesen et al. did report one patient homozygous for GG at 359 who had extremely severe lung disease
<i>MASP-3</i>	PA Infection Susceptibility	1	(Haerynck, Van Steen et al. 2012)	One study of patients with at least one Phe508del mutation has examined <i>MASP-3</i> and has found that at rs850312 carriage of an A allele (as opposed to G homozygotes) is associated with earlier PA infection.
<i>MBL</i>	Infection Susceptibility and Inflammation	16	(Gabolde, Guilloud-Bataille et al. 1999, Garred, Pressler et al. 1999, Davies, Turner et al. 2004, Yarden, Radojkovic et al. 2004, Carlsson, Sjöholm et al. 2005, Drumm, Konstan et al. 2005, Trevisiol, Boniotto et al. 2005, Choi, Ehrmantraut et al. 2006, Olesen, Jensenius et al. 2006, Buranawuti, Boyle et al. 2007, Dorfman, Sandford et al. 2008, Faria, Faria et al. 2009, McDougal, Green et al. 2010, Tarova, Polakova et al. 2011, Haerynck, Van Steen et al. 2012, Gravina, Crespo et al. 2015)	<i>MBL</i> is one of the most widely studied potential gene modifiers in CF. In general studies have categorised mutations as either causing low or high MBL production. Whilst some studies, including the large North American Gene Modifier Study, have found no effect, the majority have shown that genotypes associated with low or deficient MBL production are associated with more significant lung function impairment, earlier PA infection, and poorer long term survival.
<i>MC3R</i>	Inflammation	1	(Park, Sharma et al. 2012)	In data presented in abstract form, siblings with concordant disease shared the same <i>MC3R</i> status, whereas those with discordant disease were more likely to have different <i>MC3R</i> mutations.

Gene	Reason	No.	Articles	Results
<i>mEPHX</i>	Oxidative Stress	1	(Rohovyk, Vishtak et al. 2014)	In data presented in abstract form examining 35 Phe508del homozygous patients from Ukraine, 415 AA was more commonly found in cases of mild CF than severe CF. However how mild vs. severe was classified was not stated.
<i>MetH</i>	Proximity to CFTR	1	(Mekus, Laabs et al. 2003)	In a study of 68 Phe50del homozygous patients who were part of the European Twins and Siblings cohort, <i>MetH</i> was not associated with lung disease severity.
<i>MIF</i>	Inflammation	4	(Plant, Gallagher et al. 2005, Adamali, Armstrong et al. 2012, Melotti, Mafficini et al. 2014, De Lima Marson, Bertuzzo et al. 2016)	Three studies examining number of CATT repeats at the -794 promoter showed that 5 CATT repeats (either heterozygous or homozygous) is associated with milder disease (later PA infection and later age at FEV ₁ <60%, slower rate of decline in FEV ₁). Importantly this mutation may not have been examined in GWAS studies. In an abstract De Lima Marson found no association between mutation at -173 and disease severity.
<i>MP6d9</i>	Proximity to CFTR	2	(Santis, Osborne et al. 1990, Sahli, Hadj-Fredj et al. 2016)	In an early study of patients from the United Kingdom no link was found. In a subsequent study of Tunisian patients, with poorly defined clinical measures of disease severity, the 2/2 genotype was associated with more severe disease.
<i>MPO</i>	Inflammation	1	(Reynolds, Sermet-Gaudelus et al. 2006)	In a study of French patients with any CFTR genotype, patients not infected with SA or PA with GG (vs. GA) genotype at -463 had much more severe disease with worse FEV ₁ (73.5 vs. 95.7%), FVC, CXR and disease severity scores.
<i>MUC1</i>	Mucin	1	(Guo, Pace et al. 2011)	In a large study of American patients who were Phe508del homozygous, larger VNTR was non significantly associated with milder lung disease
<i>MUC2</i>	Mucin	1	(Guo, Pace et al. 2011)	In a large study of American patients who were Phe508del homozygous, no effect was demonstrated.
<i>MUC4</i>	Mucin	1	(Corvol, Blackman et al. 2015)	In the largest gene modifier study, which examined 6365 patients from USA, Canada and France, <i>MUC4</i> along with <i>MUC20</i> were genes of interest at a locus that was found to be associated with lung disease severity.
<i>MUC5AC</i>	Mucin	2	(Guo, Pace et al. 2011, Magalhaes, Rivals et al. 2017)	In a large group of American, Phe508del homozygous patients, Guo et al. found the 6.4kb VNTR was more commonly found in patients with severe lung disease. In a separate study no link between methylation and lung disease severity was found.

Gene	Reason	No.	Articles	Results
<i>MUC7</i>	Mucin	1	(Guo, Pace et al. 2011)	In a large study of American patients who were Phe508del homozygous, no effect was demonstrated
<i>MUC20</i>	Mucin	1	(Corvol, Blackman et al. 2015)	In the largest gene modifier study, which examined 6365 patients from USA, Canada and France, <i>MUC20</i> along with <i>MUC4</i> were genes of interest at a locus that was found to be associated with lung disease severity.
<i>Neutrophil FC γ IIA</i>	Inflammation	1	(De Rose, Arduino et al. 2005)	In a study of Italian patients, in the subgroup of Phe508del homozygous patients, those carrying R allele had increased risk of chronic PA infection
<i>NEDD4L</i>	Interact with alternate chloride channel	1	(Gisler, von Kanel et al. 2013)	In a study of Swiss Phe508del homozygous patients, no effect was seen
<i>NOS-1</i>	Infection Susceptibility, Inflammation, Broncho-constriction	4	(Grasemann, Knauer et al. 2000, Grasemann, Storm van's Gravesande et al. 2002, Texereau, Marullo et al. 2004, Marson, Bertuzzo et al. 2014)	In 2 separate studies Grasemann et al. found that having high AAT repeats at intron 20 is associated with higher rates of PA infection. Texereau et al. identified having 2 alleles with a high number of GT repeats was associated with a slower annual rate of decline in FEV ₁ (0.8% vs. 3.2-3.3%). Marson tested multiple mutations against numerous outcomes and found contradictory results.
<i>NOS-3</i>	Infection Susceptibility, Inflammation, Broncho-constriction	3	(Grasemann, Storm van's Gravesande et al. 2003, Drumm, Konstan et al. 2005, Marson, Bertuzzo et al. 2014)	In an initial study it was found that females carrying T allele at NOS3 894, had higher FEno and lower PA infection. There was no effect in males. The large and robust Gene modifier study found no effect, however this used FEV ₁ as an endpoint, and analysed both genders. Subsequently data presented in abstract form analysing both genders in a Brazilian CF population found presence of the T allele at 894 was associated with earlier lung disease.
<i>PPP2R1A</i>	Interact with CFTR	1	(Gisler, von Kanel et al. 2013)	In a study of Phe508del homozygous patients from Switzerland <i>PPP2R1A</i> , SNP c.*465T>A, AA (vs. TT, AT) associated with milder course for FRC.
<i>PPP2R4</i>	Interact with CFTR	1	(Gisler, von Kanel et al. 2013)	In a study of Phe508del homozygous patients from Switzerland SNP c-185A>C, AA (vs. AC, CC) associated with more significant decline in FEV ₁ , sREFF and LCI.
<i>PPP2R5E</i>	Interact with CFTR	1	(Gisler, von Kanel et al. 2013)	In a study of Phe508del homozygous patients from Switzerland no effect was seen
<i>PRSS8</i>	Interact with alternate chloride channel	1	(Gisler, von Kanel et al. 2013)	In a study of Phe508del homozygous patients from Switzerland no effect was seen

Gene	Reason	No.	Articles	Results
<i>SCNN1A</i>	Inflammation	1	(Stanke, Becker et al. 2006)	In a study of Phe508del homozygous patients who were part of the European Twins and Siblings study no effect was seen.
<i>SCNN1β</i>	Inflammation	3	(Stanke, Becker et al. 2006, Viel, Leroy et al. 2008, Stanke, Becker et al. 2011)	Two studies using Phe508del homozygous patients from the European Twins and Siblings study found <i>SCNN1β</i> affected lung disease severity, however a study from a French cohort found it had no effect.
<i>SCNN1γ</i>	Inflammation	4	(Stanke, Becker et al. 2006, Viel, Leroy et al. 2008, Stanke, Becker et al. 2011, Magalhaes, Rivals et al. 2017)	Two studies using Phe508del homozygous patients from the European Twins and Siblings study found <i>SCNN1γ</i> affected lung disease severity. However, two studies from separate French groups showed that <i>SCNN1γ</i> mutation or methylation status did not affect lung disease severity.
<i>SLC26A9</i>	Alternate Ion Channel	2	(Li, Soave et al. 2014, Strug, Gonska et al. 2016)	In a large study of Canadian patients Li et al. found no link with lung function or PA infection rate was found. In a subsequent study of 70 patients Strug et al. found that in patients with one copy of G551D, the number of C alleles at rs7512464, was positively correlated with FEV ₁ and response to ivacaftor
<i>SLC9A3</i>	Genome Wide	4	(Dorfman, Taylor et al. 2011, Li, Soave et al. 2014, Corvol, Blackman et al. 2015, Soave, Corvol et al. 2015)	Three studies using patients from the Canadian gene modifier study have studied the role of <i>SLC9A3</i> and found it affects lung disease severity. Dorfman et al. found that the number of T alleles at rs4957061 was associated with more severe lung disease (earlier onset PA, worse predicted FEV ₁ at 10years). Li et al. found that the T allele at rs17563161 was associated with worse CF specific FEV ₁ . Soave et al. included a replication cohort using patients from the French Gene Modifier study and also found SNPs that were related to lung disease severity. The effect of <i>SLC9A3</i> is further reinforced by the international GWAS meta-analysis which found it is located within a locus that modifies lung disease severity.
<i>SLC9A3R2</i>	Genome Wide	1	(Soave, Corvol et al. 2015)	In a large study using patients from the Canadian Gene Modifier Study, and replicated in patients from the French Gene Modifier study, Soave et al. found SNPs from <i>SLC9A3R2</i> which were related to lung disease severity

Gene	Reason	No.	Articles	Results
<i>SLC6A14</i>	Genome Wide	2	(Li, Soave et al. 2014, Corvol, Blackman et al. 2015)	A large Canadian study found that at rs3788766 T allele is associated with worse lung function and earlier PA infection, and at rs5905283 C allele associated with worse FEV ₁ . Further reinforced by the subsequent international GWAS meta-analysis which found that <i>SLC6A14</i> is located within a locus that is associated with lung disease severity.
<i>SNAP23</i>	Interacts with CFTR	1	(Gisler, von Kanel et al. 2013)	In a study of Phe508del homozygous patients SNP c267-9T>C, CT associated with milder progression in hyperinflation, measured with multiple lung function measures, when compared to CC or TT
<i>STAT3</i>	Inflammation	1	(Labenski, Hedtfeld et al. 2011)	In a subgroup of patients from the European Twins and Siblings cohort no effect was seen.
<i>STFPA1</i> , <i>STFPA2</i>	Immune Response	1	(Choi, Ehrmantraut et al. 2006)	In a study of 137 American adults, presence of at least one 6A3 allele on <i>STFPA1</i> was associated with lower FEV ₁ and ATS score. Presence of 1A1 on <i>STFPA2</i> was also associated with lower FEV ₁ and ATS score. Presence of the 6A3/1A1 haplotype was associated with lower FEV ₁ and multiple scores.
<i>STX1A</i>	Interacts with CFTR	1	(von Kanel, Stanke et al. 2013)	In a study of 56 Phe508del homozygous patients presence of G allele at rs4363087 was associated with poorer lung function (FEV ₁ , LCI, sREFF), with G homozygous patients worse than compound heterozygous patients. This was replicated in the European twins and siblings study using the CONS outcome measure. Functional work showed that rs4363087 is in linkage disequilibrium with aberrant splicing of mRNA leading to nonsense mediated mRNA decay.
<i>T2R38</i>	Infection susceptibility	1	(Turnbull, Lund-Palau et al. 2016)	In a study of 271 patients from the United Kingdom no effect was seen.
<i>TAX1BP1</i>	Inflammation	1	(Kelly, Williams et al. 2013)	In a small study of Irish patients, <i>TAX1BP1</i> gene expression measured on nasal epithelia, was not associated with lung disease severity.
<i>TGFB1</i>	Inflammation, Lung Fibrosis Risk	15	(Arkwright, Laurie et al. 2000, Arkwright, Pravica et al. 2003, Drumm, Konstan et al. 2005, Brazova, Sismova et al. 2006, Buranawuti, Boyle et al. 2007, Bremer, Blackman et al. 2008,	<i>TGFB1</i> is one of the most widely studied gene modifiers. At Codon 10, the CC genotype (as opposed to TT or TC) has been shown to be associated with more severe lung disease in multiple studies, including the large North American Gene Modifier Study. It has also been shown to modulate the effect of other modifiers of lung disease severity such as other genes (MBL2) and environmental factors (smoke exposure). At -509 the data is conflicting with some studies showing no effect, and others showing the TT genotype is associated with more

Gene	Reason	No.	Articles	Results
			Collaco, Vanscoy et al. 2008, Corvol, Boelle et al. 2008, Dorfman, Sandford et al. 2008, Faria, Faria et al. 2009, Levy, Linse et al. 2010, Dorfman, Taylor et al. 2011, Eickmeier, Boom et al. 2013, Oueslati, Hadj Fredj et al. 2015, Magalhaes, Rivals et al. 2017)	severe disease and that is modulates the effect of smoke exposure. At Codon 25 one study has shown an effect whilst two have shown no effect.
<i>TLR1</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation, Rs5743551 AA associated with greater rate of FEV ₁ decline over 5 years than AG
<i>TLR2</i>	Inflammation, Infection Susceptibility	2	(Haerynck, Mahachie John et al. 2013, Magalhaes, Rivals et al. 2017)	Haerynck et al. found that in a Belgian population of patients with at least one copy of Phe508del, rs1898830 GG (vs. AA, GA), rs3804100 TT (vs. TC), and rs5743708 GA (vs. GG) all associated with more severe decline in FEV ₁ over 5 years. No link between methylation at <i>TLR2</i> and lung disease severity
<i>TLR3</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TLR4</i>	Inflammation, Infection Susceptibility	3	(Urquhart, Allen et al. 2006, Park, Yung et al. 2011, Haerynck, Mahachie John et al. 2013)	After finding a relationship with age at diagnosis, Park et al. found no association with FEV ₁ or time to PA in a larger secondary cohort. No effect seen in other studies.

Gene	Reason	No.	Articles	Results
<i>TLR5</i>	Inflammation, Infection Susceptibility	3	(Blohmke, Park et al. 2010, Haerynck, Mahachie John et al. 2013, Magalhaes, Rivals et al. 2017)	Blohmke et al. examined one SNP (rs5744168) in 1078 Canadian patients and found no effect. Haerynck et al. examined 4 SNPs in 98 Belgian patients and also found no effect at rs5744168 but did find that rs5744174 CC was associated with more severe decline in lung function over 5 years than TC. No link between methylation at TLR5 and lung disease severity.
<i>TLR6</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TLR7</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TLR8</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TLR9</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TLR10</i>	Inflammation, Infection Susceptibility	1	(Haerynck, Mahachie John et al. 2013)	In a study of 96 Belgian patients who carried at least one Phe508del mutation no effect demonstrated
<i>TMC6</i>	Whole Exome	1	(Emond, Louie et al. 2015)	In a large study which initially used Whole Exome Sequencing and included a replication cohort, the minor allele of rs34712518 was associated with earlier onset PA infection and worse FEV ₁
<i>TNF α</i>	Inflammation	13	(Hull and Thomson 1998, Arkwright, Pravica et al. 2003, Drumm, Konstan et al. 2005, Yarden, Radojkovic et al. 2005, Laki, Laki et al. 2006, Schmitt-Grohe, Stuber et	TNFα has been studied extensively, with multiple different alleles studied. Polymorphism at position 308 is the most extensively studied with the majority of studies showing no effect. However, 5 studies did show an effect, with 3 studies showing carriage of the A allele (vs. G) associated with more severe disease including 2 where it was part of MHC Ancestral Haplotype 8.1. Position 1031 and 857 have been studied by one study each, neither of which showed an effect. Buranawuti et al. showed carriage of GA (vs. GG) at position 238 was associated with a survival advantage, however Coutinho et al. and Yarden et al. showed no

Gene	Reason	No.	Articles	Results
			al. 2006, Buranawuti, Boyle et al. 2007, Corvol, Boelle et al. 2008, Corvol, Beucher et al. 2012, Shmarina, Pukhalsky et al. 2013, Coutinho, Marson et al. 2014, de Vries, Griffiths et al. 2014, Khorrami, Bonyadi et al. 2015)	effect, although they used clinical, microbiological and lung function end points rather than survival. Yarden et al. is the only paper to analyse position 691 and they found ins/del genotype to be associated with later PA infection and higher lung function.
<i>TNFRSF1A</i>	Inflammation	1	(Stanke, Becker et al. 2006)	In a study using Phe508del homozygous patients from the European Twins and Siblings cohort, haplotype D12S889-rs767455 (10)G showed elevated transmission in family transmission studies and was found more commonly in mildly affected sibling pairs suggesting it is associated with mild disease
<i>Tub20</i>	Proximity to CFTR	1	(Mekus, Laabs et al. 2003)	Using Phe508del homozygous patients from the European Twins and Siblings cohort no link with lung disease severity was found.
<i>XV2C</i>	Proximity of CFTR	1	(Mekus, Laabs et al. 2003)	Using Phe508del homozygous patients from the European Twins and Siblings cohort no link with lung disease severity was found.
<i>YY1</i>	Genome Wide	2	(Wright, Merlo et al. 2006, Magalhaes, Rivals et al. 2017)	A study of 12 patients from USA identified YY1 as being differentially expressed between patients with mild and severe disease as measured by FEV ₁ . No link between methylation at YY1 and lung disease severity

Regions/Whole Exome/Whole Genome

Region	No.	Articles	
Whole Exome Sequencing	2	(Emond, Louie et al. 2012, Emond, Louie et al. 2015)	The same authors have performed 2 studies using WES to look at risk of PA infection, and have identified missense mutations in DCTN4, and minor allele of TMC6 rs34712518 as being associated with early onset PA infection. Minor allele of CAV2 rs8940 is protective. Of note TMC6 rs34712518 is associated with significantly lower FEV1, whilst CAV2 res8940 is associated with higher FEV1.

Region	No.	Articles	
7q31 to 7qtel	1	(Stanke, Davenport et al. 2010)	Using siblings and parents identified a locus on 7q34 which showed differential onset of paternal recombination in severely vs. mildly affected pairs. Also found evidence of enriched imprinting associated elements in the locus and allele frequencies in the locus differed between mildly and severely affected pairs. Importantly disordered imprinting in this region is implicated in conditions associated poor growth and CONS (the outcome measure used) is a composite of FEV1and BMI.
11p13	3	(Wright, Strug et al. 2011, Corvol, Blackman et al. 2015, Dang, Gallins et al. 2016)	A large GWAS using patients from North America identified SNP rs12793173, at 11p13, as being linked with lung disease severity. This was confirmed in the international GWAS meta analysis, and of note a larger effect was seen when the Phe508del homozygous patients subgroup was used. Dang et al. performed targeted resequencing of this area in 377 patients and then imputed the results to 6365 patients. They found that multiple SNPs in this area were associated with lung disease severity but only in Phe508del homozygous patients
12p13	1	(Stanke, Becker et al. 2006)	In a study involving sibling pairs and parents 12p13 TNFRSF1A haplotype D12S889-rs767455 (10)G showed elevated transmission in family transmission studies and was found more commonly in mildly affected sibling pairs suggesting it is associated with mild disease
16p12	1	(Stanke, Becker et al. 2006)	In a study involving sibling pairs and parents 16p12; SCNN1 δ haplotype distribution of rs238547-rs152730-rs250563 different in discordant sibling as opposed to concordant siblings, SCNN1 γ haplotype rs5735-rs5723 TG underrepresented (based on family transmission) suggesting CF patients with this haplotype may have severe disease and not survived
19q13	3	(Zielenski, Corey et al. 1999, Becker, Jansen et al. 2007, Stanke, Becker et al. 2010)	The first study of 19q13 found it affected risk of meconium ileus but had no effect on lung disease. Two subsequent studies using sibling pairs have occurred. Becker et al. found differential distribution between mildly and severely affected pairs at rs4802129, which implicated the CEACAM gene. A follow up study found severely affected pairs were more likely be G-C-T-C-T-C-A at haplotype rs1549960-rs3795020-rs12459454-rs4803507-rs4803508-rs28547939-rs11548735. This changes the coding sequence of CEACAM6. It also found that when comparing concordant and discordant pairs, discordant are more likely to be G-G-G-A-C at Haplotype rs6508999-rs6509000-rs28824227-rs8106505-rs10414823, which potentially affects CEACAM3.
20q13.2	2	(Wright, Strug et al. 2011, Corvol, Blackman et al. 2015)	In a large North American GWAS, a link between lung disease severity and rs6024460 (located at 20q13.2) was found. Subsequently in the International GWAS meta-analysis this relationship was not replicated.
Genome Wide SNP	3	(Gu, Harley et al. 2009, Wright, Strug et al. 2011,	The largest study of gene modifiers, which was a meta-analysis of an international cohort, by Corvol et al. examined 6365 patients using SNP microarray and found 5 loci associated with lung disease severity. One of these loci had also been identified by a previous GWAS, and in both studies was shown to have most

Region	No.	Articles	
		Corvol, Blackman et al. 2015)	significant effect in Phe508del homozygous patients. Previous GWAS had also identified 20q13.2 and IFRD1 as area of interest however these were not identified in the Corvol et al. study.
Haplotype at CFTR	1	(Santis, Osborne et al. 1990)	One of the earliest studies of potential gene modifiers found no relationship between CFTR haplotype and lung disease severity.
Genes in CFTR vicinity	1	(Mekus, Laabs et al. 2003)	In a study using Phe508del homozygous patients from the European Twins and Siblings cohort, no clear links to lung disease outcomes found. LEP polymorphisms and paternal imprinting of loci at MEST/PEG1 implicated in affecting disease severity, but likely affect growth more than lung disease.
Ancestral Haplotype 8.1	2	(Laki, Laki et al. 2006, Corvol, Beucher et al. 2012)	In a small study of Hungarian patients carriage of Ancestral Haplotype 8.1 was associated with delayed colonisation of PA and SA. A subsequent larger study found it was associated with more severe lung disease measured as a 6.4% reduction in FEV1 percent predicted.
Apical Gene Set	1	(Soave, Corvol et al. 2015)	In a study involving the large Canadian and French Gene Modifier cohorts, Soave et al. examined a 3814 SNPs, from 155 genes, whose products localised to the apical cell membrane. They found that SLC9A3, SLC9A3R2 and EZR all had SNPs that were related to severity of lung disease.
Surfactant Protein Haplotypes	1	(Choi, Ehrmantraut et al. 2006)	In a study of 137 American adults, presence of at least on 6A3 allele on STFPA1 was associated with lower FEV1 and ATS score. Presence of 1A1 on STFPA2 was also associated with lower FEV1 and ATS score. Presence of the 6A3/1A1 haplotype was associated with lower FEV1 and multiple scores.
Inflammatory Genes	2	(Hillian, Londono et al. 2008, Stanke, Becker et al. 2011)	Hillian examined 52 inflammatory genes and only 3 showed a nominal effect (IL8, IL10R, TGFβ). Only IL8 proceeded to further testing. IL8 rs4073AA was found to be protective, whilst rs4073TT was associated with severe lung disease. The effect was more pronounced in males. A different study examining 52 genes, which used Phe508del homozygous patients from the European Twins and Siblings cohort, found the following genes associated with disease severity; BEST4, S100A11, IL10, POMC, CXCR2, TLR2, SLC26A2, PON2, MBL2, BEST2, GSTT1, KDELR3, PLA2G2A, CLCA4Sat, GSTM1, PDZK1, IL1B, TLR9, IL8, CD14, ADRB2, HLA TNFa-Lta, HLA DQ - TAP2, IFNGR1, KDELR2, LEP, D7Sat3, D7S495, TLR4, SPA - SPD, CD95, MUC2, SMPD1, GSTP1, TNFR, KRT8, NHERF2, SCNN1B/G, PRSS8, NHERF1, D19S197. The majority of these findings did not retain significance after correction for multiple testing
Genome Wide Gene Expression	3	(Wright, Merlo et al. 2006, Levy, Linse et al. 2010, O'Neal, Gallins et al. 2015)	In the first study of genome wide gene expression, which is the only one to examine cells from the airway, Wright et al. identified 53 genes with differential expression in patients with mild CF lung disease compared to those with severe lung disease or controls. They also identified 569 genes which were upregulated in those with

Region	No.	Articles	
			severe compared to control/mild CF group. Gene ontology revealed these genes involved in oxidoreductase activity, ubiquitin cycle, and lipid metabolism. Levy et al. looked at serum and found that TGF β 1 and IL1RN were differentially expressed between those with CF and controls. In the largest study to date O'Neal et al. compared groups with mild and severe lung disease, and those with early and late onset of PA infection. They examined gene expression from lymphoblasts. They found that LPAR6 was more highly expressed in those with severe lung and that 35 gene pathways were identified which mainly concerned HLA class I, phe508del processing and the Endoplasmic Reticulum stress response. HLA class II was associated with age at PA infection.
Mucin	3	(Guo, Pace et al. 2011, Corvol, Blackman et al. 2015, Magalhaes, Rivals et al. 2017)	Guo et al. examined MUC1, MUC2, MUC5AC, and MUC7 and found MUC5AC VNTR size 6.4kb allele associated with severe lung disease. Also a trend for larger MUC1 to be associated with milder lung disease was observed. MUC5AC 6.30/6.35 kb alleles were more prevalent in mild and 6.40/6.45 more prevalent in severe disease groups. A study found no relationship between methylation of MUC5AC and disease severity. In the international GWAS meta analysis MUC4 and MUC20 were associated with lung disease severity.
Toll Like Receptor Genes	5	(Urquhart, Allen et al. 2006, Blohmke, Park et al. 2010, Park, Yung et al. 2011, Haerynck, Mahachie John et al. 2013, Magalhaes, Rivals et al. 2017)	Haerynck et al. study examined TLR1-10 and found that the majority of genes had no effect. They did find TLR1 Rs5743551 AA associated with greater rate of FEV1 decline over 5 years than AG. At TLR2 rs1898830 GG (vs. AA, GA), rs3804100 TT (vs. TC), and rs5743708 GA (vs. GG) all associated with more severe decline in FEV1 over 5 years. At TLR5 rs5744174 CC was associated with more severe decline in lung function over 5 years than TC. Blohmke et al. studied a different TLR5 SNP and found no effect. Park et al. examined the effect of TLR4 and found an association with age of diagnosis but not lung function severity or time to PA infection, whilst Urquhart et al. found no effect at TLR4. Magalhaes et al. examined methylation at TLR2 and TLR5 and found no effect.
Various Genes	3	(De Lima Marson, Bertuzzo et al. 2016)	In work presented at the same conference, De Lima Marson examined the effect of 256 SNPs across 125 different genes, in 126 Brazilian patients with CF. Three different abstracts covered the effect on oxygen saturations/spirometry, microbiological colonisation, and severity scores. There were multiple findings reported, however the significance of these findings is questionable.

Note: ABPA: Allergic bronchopulmonary aspergillosis, BAL: Bronchoalveolar lavage, BMI: Body mass index, CFTR: Cystic fibrosis transmembrane conductance regulator, FEno: Fraction exhaled nitric oxide, FEV₁: Forced expiratory volume in 1 second, FVC: Forced vital capacity, GWAS: Genome-wide association study; IgE: Immunoglobulin E, LCI: Lung clearance index, MHC: Major histocompatibility complex, PA: *Pseudomonas aeruginosa*, SA: *Staphylococcus aureus*, SNP: Single nucleotide polymorphism, sREFF: Specific effective airway resistance.

References

- Adamali, H., M. E. Armstrong, A. M. McLaughlin, G. Cooke, E. McKone, C. M. Costello, C. G. Gallagher, L. Leng, J. A. Baugh, G. Fingerle-Rowson, R. J. Bucala, P. McLoughlin and S. C. Donnelly (2012). "Macrophage migration inhibitory factor enzymatic activity, lung inflammation, and cystic fibrosis." Am J Respir Crit Care Med **186**(2): 162-169.
- Arkwright, P. D., S. Laurie, M. Super, V. Pravica, M. J. Schwarz, A. K. Webb and I. V. Hutchinson (2000). "TGF-beta(1) genotype and accelerated decline in lung function of patients with cystic fibrosis." Thorax **55**(6): 459-462.
- Arkwright, P. D., V. Pravica, P. J. Geraghty, M. Super, A. K. Webb, M. Schwarz and I. V. Hutchinson (2003). "End-organ dysfunction in cystic fibrosis: association with angiotensin I converting enzyme and cytokine gene polymorphisms." American Journal of Respiratory & Critical Care Medicine **167**(3): 384-389.
- Aron, Y., B. S. Polla, T. Bienvenu, J. Dall'ava, D. Dusser and D. Hubert (1999). "HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype." American Journal of Respiratory & Critical Care Medicine **159**(5 Pt 1): 1464-1468.
- Becker, T., S. Jansen, S. Tamm, T. F. Wienker, B. Tummler and F. Stanke (2007). "Transmission ratio distortion and maternal effects confound the analysis of modulators of cystic fibrosis disease severity on 19q13." European Journal of Human Genetics **15**(7): 774-778.
- Beucher, J., P. Y. Boelle, P. F. Busson, C. Muselet-Charlier, A. Clement, H. Corvol and C. F. M. G. S. I. French (2012). "AGER -429T/C is associated with an increased lung disease severity in cystic fibrosis." PLoS ONE [Electronic Resource] **7**(7): e41913.
- Blaisdell, C. J., T. D. Howard, A. Stern, P. Bamford, E. R. Bleeker and O. C. Stine (2004). "CLC-2 single nucleotide polymorphisms (SNPs) as potential modifiers of cystic fibrosis disease severity." BMC Medical Genetics **5**: 26.
- Blohmke, C. J., J. Park, A. F. Hirschfeld, R. E. Victor, J. Schneiderman, D. Stefanowicz, M. A. Chilvers, P. R. Durie, M. Corey, J. Zielenski, R. Dorfman, A. J. Sandford, D. Daley and S. E. Turvey (2010). "TLR5 as an anti-inflammatory target and modifier gene in cystic fibrosis." J Immunol **185**(12): 7731-7738.

- Brazova, J., K. Sismova, V. Vavrova, J. Bartosova, M. Macek, Jr., H. Lauschman and A. Sediva (2006). "Polymorphisms of TGF-beta1 in cystic fibrosis patients." Clinical Immunology **121**(3): 350-357.
- Bremer, L. A., S. M. Blackman, L. L. Vanscoy, K. E. McDougal, A. Bowers, K. M. Naughton, D. J. Cutler and G. R. Cutting (2008). "Interaction between a novel TGFB1 haplotype and CFTR genotype is associated with improved lung function in cystic fibrosis." Human Molecular Genetics **17**(14): 2228-2237.
- Brouard, J., N. Knauer, P. Y. Boelle, H. Corvol, A. Henrion-Caude, C. Flamant, F. Bremont, B. Delaisi, J. F. Duhamel, C. Marguet, M. Roussey, M. C. Miesch, K. Chadelat, M. Boule, B. Fauroux, F. Ratjen, H. Grasemann and A. Clement (2005). "Influence of interleukin-10 on *Aspergillus fumigatus* infection in patients with cystic fibrosis." J Infect Dis **191**(11): 1988-1991.
- Buranawuti, K., M. P. Boyle, S. Cheng, L. L. Steiner, K. McDougal, M. D. Fallin, C. Merlo, P. L. Zeitlin, B. J. Rosenstein and P. J. Mogayzel (2007). "Variants in mannose-binding lectin and tumour necrosis factor α affect survival in cystic fibrosis." Journal of medical genetics **44**(3): 209-214.
- Buscher, R., K. J. Eilmes, H. Grasemann, B. Torres, N. Knauer, K. Sroka, P. A. Insel and F. Ratjen (2002). "beta2 adrenoceptor gene polymorphisms in cystic fibrosis lung disease." Pharmacogenetics **12**(5): 347-353.
- Carlsson, M., A. G. Sjolholm, L. Eriksson, S. Thiel, J. C. Jensenius, M. Segelmark and L. Truedsson (2005). "Deficiency of the mannan-binding lectin pathway of complement and poor outcome in cystic fibrosis: bacterial colonization may be decisive for a relationship." Clin Exp Immunol **139**(2): 306-313.
- Choi, E. H., M. Ehrmantraut, C. B. Foster, J. Moss and S. J. Chanock (2006). "Association of common haplotypes of surfactant protein A1 and A2 (SFTPA1 and SFTPA2) genes with severity of lung disease in cystic fibrosis." Pediatric Pulmonology **41**(3): 255-262.
- Collaco, J. M., L. Vanscoy, L. Bremer, K. McDougal, S. M. Blackman, A. Bowers, K. Naughton, J. Jennings, J. Ellen and G. R. Cutting (2008). "Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease." JAMA **299**(4): 417-424.

- Coriati, A., G. F. Bouvet, A. Lavoie-Pilote and Y. Berthiaume (2016). "Role of specific genetic variants in Chitinase 3-like 1 as a potential marker of cystic fibrosis disease severity." Pediatric Pulmonology **51**: 245.
- Corvol, H., J. Beucher, P. Y. Boelle, P. F. Busson, C. Muselet-Charlier, A. Clement, F. Ratjen, H. Grasemann, J. Laki, C. N. Palmer, J. S. Elborn and A. Mehta (2012). "Ancestral haplotype 8.1 and lung disease severity in European cystic fibrosis patients." Journal of Cystic Fibrosis **11**(1): 63-67.
- Corvol, H., S. M. Blackman, P. Y. Boelle, P. J. Gallins, R. G. Pace, J. R. Stonebraker, F. J. Accurso, A. Clement, J. M. Collaco, H. Dang, A. T. Dang, A. Franca, J. Gong, L. Guillot, K. Keenan, W. Li, F. Lin, M. V. Patrone, K. S. Raraigh, L. Sun, Y. H. Zhou, O. n. W.K, M. K. Sontag, H. Levy, P. R. Durie, J. M. Rommens, M. L. Drumm, F. A. Wright, L. J. Strug, G. R. Cutting and M. R. Knowles (2015). "Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis." Nature Communications **6 (no pagination)**(8382).
- Corvol, H., P. Y. Boelle, J. Brouard, N. Knauer, K. Chadelat, A. Henrion-Caude, C. Flamant, C. Muselet-Charlier, M. Boule, B. Fauroux, C. Vallet, J. Feingold, F. Ratjen, H. Grasemann and A. Clement (2008). "Genetic variations in inflammatory mediators influence lung disease progression in cystic fibrosis." Pediatric Pulmonology **43**(12): 1224-1232.
- Corvol, H., N. Nathan, C. Charlier, K. Chadelat, P. Le Rouzic, O. Tabary, B. Fauroux, A. Henrion-Caude, J. Feingold, P. Y. Boelle and A. Clement (2007). "Glucocorticoid receptor gene polymorphisms associated with progression of lung disease in young patients with cystic fibrosis." Respiratory Research **8**: 88.
- Courtney, J. M., B. J. Plant, K. Morgan, J. Rendall, C. Gallagher, M. Ennis, N. Kalsheker, S. Elborn and C. M. O'Connor (2006). "Association of improved pulmonary phenotype in Irish cystic fibrosis patients with a 3' enhancer polymorphism in alpha-1-antitrypsin." Pediatric Pulmonology **41**(6): 584-591.
- Coutinho, C. A., F. A. Marson, A. R. Marcelino, L. C. Bonadia, M. P. Carlin, A. F. Ribeiro, J. D. Ribeiro and C. S. Bertuzzo (2014). "TNF-alpha polymorphisms as a potential modifier gene in the cystic fibrosis." Int J Mol Epidemiol Genet **5**(2): 87-99.

- Crovella, S., L. Segat, A. Amato, E. Athanasakis, V. Bezzerri, C. Braggion, R. Casciaro, G. Castaldo, C. Colombo, A. E. Covone, V. De Rose, R. Gagliardini, C. Lanzara, L. Minicucci, M. Morgutti, E. Nicolis, F. Pardo, S. Quattrucci, V. Raia, R. Ravazzolo, M. Seia, V. Stanzial, L. Termini, L. Zazzeron, G. Cabrini and P. Gasparini (2011). "A polymorphism in the 5' UTR of the DEFB1 gene is associated with the lung phenotype in F508del homozygous Italian cystic fibrosis patients." Clinical Chemistry & Laboratory Medicine **49**(1): 49-54.
- Czerska, K., A. Sobczynska-Tomaszewska, D. Sands, A. Nowakowska, D. Bak, K. Wertheim, J. Poznanski, J. Zielenski, A. Norek and J. Bal (2010). "Prostaglandin-endoperoxide synthase genes COX1 and COX2 - novel modifiers of disease severity in cystic fibrosis patients." Journal of Applied Genetics **51**(3): 323-330.
- Dang, H., P. J. Gallins, R. G. Pace, X. L. Guo, J. R. Stonebraker, H. Corvol, G. R. Cutting, M. L. Drumm, L. J. Strug, M. R. Knowles and W. K. O'Neal (2016). "Novel variation at chr11p13 associated with cystic fibrosis lung disease severity." Human Genome Variation **3** (no pagination)(201620).
- Darrah, R., E. McKone, C. O'Connor, C. Rodgers, A. Genatossio, S. McNamara, R. Gibson, J. S. Elborn, M. Ennis, C. G. Gallagher, N. Kalsheker, M. Aitken, D. Wiese, J. Dunn, P. Smith, R. Pace, D. Londono, K. A. B. Goddard, M. R. Knowles and M. L. Drumm (2010). "EDNRA variants associate with smooth muscle mRNA levels, cell proliferation rates, and cystic fibrosis pulmonary disease severity." Physiological Genomics **41**(1): 71-77.
- Davies, J. C., M. W. Turner, N. Klein, C. Booth, M. Johnson, N. Shen, K. Fidler, S. Sharma, J. Burgess, A. Bush, D. M. Geddes and E. W. F. W. Alton (2004). "Impaired pulmonary status in cystic fibrosis adults with two mutated MBL-2 alleles." European Respiratory Journal **24**(5): 798-804.
- de Lima Marson, F., A. Bariani Marcelino, L. Rezende, A. Ribeiro and J. Ribeiro (2013). "The IFRD1 (57460C> T polymorphism) gene: a negative report in cystic fibrosis clinical severity." J Mol Genet Med **7**(058): 1747-0862.1000.
- De Lima Marson, F. A., C. S. Bertuzzo, M. A. G. De Oliveira Ribeiro, A. A. D. C. Toro, A. F. Ribeiro and J. D. Ribeiro (2016). "Bhalla, kanga and shwachman-kulczycki score on cystic fibrosis: A look by the modifier genes." Pediatric Pulmonology **51**: S35.

- De Lima Marson, F. A., C. S. Bertuzzo, M. A. G. De Oliveira Ribeiro, A. A. D. C. Toro, A. F. Ribeiro and J. D. Ribeiro (2016). "Lung function on cystic fibrosis: What explains the genetic?" Pediatric Pulmonology **51**: S27-S28.
- De Lima Marson, F. A., C. S. Bertuzzo, A. F. Ribeiro and J. D. Ribeiro (2016). "A negative report to modifier genes of the cystic fibrosis: The case of-173g>c polymorphism in the MIF gene." Pediatric Pulmonology **51**: S37-S38.
- de Lima Marson, F. A., C. S. Bertuzzo, R. Secolin, A. F. Ribeiro and J. D. Ribeiro (2013). "Genetic interaction of GSH metabolic pathway genes in cystic fibrosis." BMC Medical Genetics **14**: 60.
- De Lima Marson, F. A., C. S. Bertuzzo, A. D. C. Toro, M. A. G. De Oliveira Ribeiro, A. F. Ribeiro and J. D. Ribeiro (2016). "Performance of infection/colonization chronic of the lung in cystic fibrosis: A genetic model to establish the clinical variability." Pediatric Pulmonology **51**: S19-S20.
- De Rose, V., C. Arduino, N. Cappello, R. Piana, P. Salmin, M. Bardessono, M. Goia, R. Padoan, E. Bignamini, D. Costantini, G. Pizzamiglio, V. Bennato, C. Colombo, A. Giunta and A. Piazza (2005). "Fcgamma receptor IIA genotype and susceptibility to P. aeruginosa infection in patients with cystic fibrosis." European Journal of Human Genetics **13**(1): 96-101.
- de Vries, L., A. Griffiths, D. Armstrong and P. J. Robinson (2014). "Cytokine gene polymorphisms and severity of CF lung disease." Journal of Cystic Fibrosis **13**(6): 699-705.
- Dorfman, R., A. Sandford, C. Taylor, B. Huang, D. Frangolias, Y. Wang, R. Sang, L. Pereira, L. Sun, Y. Berthiaume, L. C. Tsui, P. D. Pare, P. Durie, M. Corey and J. Zielenski (2008). "Complex two-gene modulation of lung disease severity in children with cystic fibrosis." Journal of Clinical Investigation **118**(3): 1040-1049.
- Dorfman, R., C. Taylor, F. Lin, L. Sun, A. Sandford, P. Pare, Y. Berthiaume, M. Corey, P. Durie and J. Zielenski (2011). "Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis." Pediatric Pulmonology **46**(4): 385-392.

- Doring, G., H. Krogh-Johansen, S. Weidinger and N. Hoiby (1994). "Allotypes of alpha 1-antitrypsin in patients with cystic fibrosis, homozygous and heterozygous for deltaF508." Pediatr Pulmonol **18**(1): 3-7.
- Drumm, M. L., M. W. Konstan, M. D. Schluchter, A. Handler, R. Pace, F. Zou, M. Zariwala, D. Fargo, A. Xu, J. M. Dunn, R. J. Darrah, R. Dorfman, A. J. Sandford, M. Corey, J. Zielenski, P. Durie, K. Goddard, J. R. Yankaskas, F. A. Wright and M. R. Knowles (2005). "Genetic modifiers of lung disease in cystic fibrosis." New England Journal of Medicine **353**(14): 1443-1453.
- Eickmeier, O., L. Boom, F. Schreiner, M. J. Lentze, N. G. D, R. Schubert, S. Zielen and S. Schmitt-Grohe (2013). "Transforming growth factor beta1 genotypes in relation to TGFbeta1, interleukin-8, and tumor necrosis factor alpha in induced sputum and blood in cystic fibrosis." Mediators of Inflammation **2013**: 913135.
- Emond, M. J., T. Louie, J. Emerson, J. X. Chong, R. A. Mathias, M. R. Knowles, M. J. Rieder, H. K. Tabor, D. A. Nickerson, K. C. Barnes, R. L. Gibson and M. J. Bamshad (2015). "Exome Sequencing of Phenotypic Extremes Identifies CAV2 and TMC6 as Interacting Modifiers of Chronic Pseudomonas aeruginosa Infection in Cystic Fibrosis." PLoS Genetics **11 (6) (no pagination)**(e1005273).
- Emond, M. J., T. Louie, J. Emerson, W. Zhao, R. A. Mathias, M. R. Knowles, F. A. Wright, M. J. Rieder, H. K. Tabor, D. A. Nickerson, K. C. Barnes, L. National Heart, G. O. E. S. P. Blood Institute, G. O. Lung, R. L. Gibson and M. J. Bamshad (2012). "Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis." Nature Genetics **44**(8): 886-889.
- Faria, E. J., I. C. Faria, J. D. Ribeiro, A. F. Ribeiro, G. Hessel and C. S. Bertuzzo (2009). "Association of MBL2, TGF-beta1 and CD14 gene polymorphisms with lung disease severity in cystic fibrosis." Jornal Brasileiro De Pneumologia: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisiologia **35**(4): 334-342.
- Feuillet-Fieux, M. N., T. Nguyen-Khoa, M. A. Loriot, M. Kelly, P. de Villartay, I. Sermet, P. Verrier, J. P. Bonnefont, P. Beaune, G. Lenoir and B. Lacour (2009). "Glutathione S-transferases related to P. aeruginosa lung infection in cystic fibrosis children: preliminary study." Clin Biochem **42**(1-2): 57-63.

- Flamant, C., A. Henrion-Caude, P. Y. Boelle, F. Bremont, J. Brouard, B. Delaisi, J. F. Duhamel, C. Marguet, M. Roussey, M. C. Miesch, M. Boule, R. C. Strange and A. Clement (2004). "Glutathione-S-transferase M1, M3, P1 and T1 polymorphisms and severity of lung disease in children with cystic fibrosis." Pharmacogenetics **14**(5): 295-301.
- Frangolias, D. D., J. Ruan, P. J. Wilcox, A. G. Davidson, L. T. Wong, Y. Berthiaume, R. Hennessey, A. Freitag, L. Pedder, M. Corey, N. Sweezey, J. Zielenski, E. Tullis and A. J. Sandford (2003). "Alpha 1-antitrypsin deficiency alleles in cystic fibrosis lung disease." American Journal of Respiratory Cell & Molecular Biology **29**(3 Pt 1): 390-396.
- Furlan, L. L., F. A. Marson, J. D. Ribeiro, C. S. Bertuzzo, J. B. Salomao Junior and D. R. Souza (2016). "IL8 gene as modifier of cystic fibrosis: unraveling the factors which influence clinical variability." Hum Genet **135**(8): 881-894.
- Gabolde, M., M. Guilloud-Bataille, J. Feingold and C. Besmond (1999). "Association of variant alleles of mannose binding lectin with severity of pulmonary disease in cystic fibrosis: cohort study." BMJ **319**(7218): 1166-1167.
- Garred, P., T. Pressler, H. O. Madsen, B. Frederiksen, A. Svejgaard, N. Hoiby, M. Schwartz and C. Koch (1999). "Association of mannose-binding lectin gene heterogeneity with severity of lung disease and survival in cystic fibrosis." Journal of Clinical Investigation **104**(4): 431-437.
- Gisler, F. M., T. von Kanel, R. Kraemer, A. Schaller and S. Gallati (2013). "Identification of SNPs in the cystic fibrosis interactome influencing pulmonary progression in cystic fibrosis." European Journal of Human Genetics **21**(4): 397-403.
- Grasemann, H., N. Knauer, R. Buscher, K. Hubner, J. M. Drazen and F. Ratjen (2000). "Airway nitric oxide levels in cystic fibrosis patients are related to a polymorphism in the neuronal nitric oxide synthase gene." American Journal of Respiratory & Critical Care Medicine **162**(6): 2172-2176.
- Grasemann, H., K. Storm van's Gravesande, R. Buscher, N. Knauer, E. S. Silverman, L. J. Palmer, J. M. Drazen and F. Ratjen (2003). "Endothelial nitric oxide synthase variants in cystic fibrosis lung disease." Am J Respir Crit Care Med **167**(3): 390-394.
- Grasemann, H., K. Storm van's Gravesande, S. Gartig, M. Kirsch, R. Buscher, J. M. Drazen and F. Ratjen (2002). "Nasal nitric oxide levels in cystic fibrosis patients are associated with a neuronal NO synthase (NOS1) gene polymorphism." Nitric Oxide **6**(2): 236-241.

- Gravina, L. P., C. Crespo, H. Giugno, L. Sen, L. Chertkoff, A. Mangano and C. Castanos (2015). "Mannose-binding lectin gene as a modifier of the cystic fibrosis phenotype in Argentinean pediatric patients." Journal of Cystic Fibrosis **14**(1): 78-83.
- Gu, Y., I. T. Harley, L. B. Henderson, B. J. Aronow, I. Vietor, L. A. Huber, J. B. Harley, J. R. Kilpatrick, C. D. Langefeld, A. H. Williams, A. G. Jegga, J. Chen, M. Wills-Karp, S. H. Arshad, S. L. Ewart, C. L. Thio, L. M. Flick, M. D. Filippi, H. L. Grimes, M. L. Drumm, G. R. Cutting, M. R. Knowles and C. L. Karp (2009). "Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease." Nature **458**(7241): 1039-1042.
- Guo, X., R. G. Pace, J. R. Stonebraker, C. W. Commander, A. T. Dang, M. L. Drumm, A. Harris, F. Zou, D. M. Swallow, F. A. Wright, W. K. O'Neal and M. R. Knowles (2011). "Mucin variable number tandem repeat polymorphisms and severity of cystic fibrosis lung disease: significant association with MUC5AC." PLoS ONE [Electronic Resource] **6**(10): e25452.
- Haerynck, F., J. M. Mahachie John, K. Van Steen, P. Schelstraete, S. Van daele, B. Loeys, M. Van Thielen, I. De Canck, L. Nuytinck and F. De Baets (2013). "Genetic variations in toll-like receptor pathway and lung function decline in Cystic fibrosis patients." Human Immunology **74**(12): 1649-1655.
- Haerynck, F., K. Van Steen, T. Cattaert, B. Loeys, P. Schelstraete, K. Claes, M. Van Thielen, I. De Canck, J. M. John and F. De Baets (2012). "Polymorphisms in the lectin pathway genes as a possible cause of early chronic *Pseudomonas aeruginosa* colonization in cystic fibrosis patients." Human immunology **73**(11): 1175-1183.
- Hart, M. A., M. W. Konstan, R. J. Darrah, M. D. Schluchter, A. Storfer-Isser, L. Xue, D. Londono, K. A. Goddard and M. L. Drumm (2005). "Beta 2 adrenergic receptor polymorphisms in cystic fibrosis." Pediatric Pulmonology **39**(6): 544-550.
- Henry, M. T., S. Cave, J. Rendall, C. M. O'Connor, K. Morgan, M. X. FitzGerald and N. Kalsheker (2001). "An alpha1-antitrypsin enhancer polymorphism is a genetic modifier of pulmonary outcome in cystic fibrosis." European Journal of Human Genetics **9**(4): 273-278.
- Hillian, A. D., D. Londono, J. M. Dunn, K. A. B. Goddard, R. G. Pace, M. R. Knowles and M. L. Drumm (2008). "Modulation of cystic fibrosis lung disease by variants in interleukin-8." Genes and Immunity **9**(6): 501-508.

- Hollox, E. J., J. Davies, U. Griesenbach, J. Burgess, E. W. Alton and J. A. Armour (2005). "Beta-defensin genomic copy number is not a modifier locus for cystic fibrosis." Journal of negative results in biomedicine **4**(1): 9.
- Hull, J. and A. H. Thomson (1998). "Contribution of genetic factors other than CFTR to disease severity in cystic fibrosis." Thorax **53**(12): 1018-1021.
- Hurbain, I., I. Sermet-Gaudelus, B. Vallee, M. N. Feuillet, G. Lenoir, J. F. Bernaudin, A. Edelman and A. Fajac (2003). "Evaluation of MRP1-5 gene expression in cystic fibrosis patients homozygous for the delta F508 mutation." Pediatr Res **54**(5): 627-634.
- Ideozu, J. E., X. Zhang, A. Pan, Z. Ashrafi, K. J. Woods, M. J. Hessner, P. Simpson and H. Levy (2017). "Increased Expression of Plasma-Induced ABCC1 mRNA in Cystic Fibrosis." Int J Mol Sci **18**(8).
- Jaffe, A., A. Prasad, K. Robertson, E. Hawe, K. Fidler and H. Montgomery (2003). "Angiotensin converting enzyme (ACE) insertion/deletion polymorphism in children with cystic fibrosis. ." Pediatric Pulmonology **36**(S25): 219.
- Kelly, C., M. T. Williams, J. S. Elborn, M. Ennis and B. C. Schock (2013). "Expression of the inflammatory regulator A20 correlates with lung function in patients with cystic fibrosis." Journal of Cystic Fibrosis **12**(4): 411-415.
- Khorrami, A., M. Bonyadi, M. Rafeey and O. Omrani (2015). "Association of TNF-alpha gene variants with clinical manifestation of cystic fibrosis patients of Iranian Azeri Turkish ethnicity." Iranian Journal of Pediatrics **25 (1) (no pagination)**(e307).
- Kormann, M. S., A. Hector, V. Marcos, L. E. Mays, M. Kappler, T. Illig, N. Klopp, S. Zeilinger, M. Carevic, N. Rieber, O. Eickmeier, S. Zielen, A. Gaggar, B. Moepps, M. Griesse and D. Hartl (2012). "CXCR1 and CXCR2 haplotypes synergistically modulate cystic fibrosis lung disease." European Respiratory Journal **39**(6): 1385-1390.
- Korytina, G., D. Iaibaeva and T. Viktorova (2004). "Polymorphism of glutathione-S-transferase M1 and P1 genes in patients with cystic fibrosis and chronic respiratory tract diseases." Genetika **40**(3): 401-408.
- Kumar, V., T. Becker, S. Jansen, A. van Barneveld, K. Boztug, S. Wolfl, B. Tummler and F. Stanke (2008). "Expression levels of FAS are regulated through an evolutionary conserved element in intron 2, which modulates cystic fibrosis disease severity." Genes & Immunity **9**(8): 689-696.

- Labenski, H., S. Hedtfeld, T. Becker, B. Tummler and F. Stanke (2011). "Initial interrogation, confirmation and fine mapping of modifying genes: STAT3, IL1B and IFNGR1 determine cystic fibrosis disease manifestation." European Journal of Human Genetics **19**(12): 1281-1288.
- Laki, J., I. Laki, K. Nemeth, R. Ujhelyi, O. Bede, E. Endreffy, K. Bolbas, K. Gyurkovits, E. Csiszer, E. Solyom, G. Dobra, A. Halasz, E. Pozsonyi, K. Rajczy, Z. Prohaszka, G. Fekete and G. Fust (2006). "The 8.1 ancestral MHC haplotype is associated with delayed onset of colonization in cystic fibrosis." International Immunology **18**(11): 1585-1590.
- Levy, H., G. Linse, K. Schneck, M. Reske, M. Kadunski, S. Jia, C. Lambert and M. Hessner (2010). "Integration of genomics with genetics-molecular phenotypes for cystic fibrosis (CF) lung disease." American Journal of Respiratory and Critical Care Medicine. Conference: American Thoracic Society International Conference, ATS **181**(1 Meeting Abstracts).
- Levy, H., A. Murphy, F. Zou, C. Gerard, B. Klanderman, B. Schuermann, R. Lazarus, K. C. Garcia, J. C. Celedon, M. Drumm, M. Dahmer, M. Quasney, K. Schneck, M. Reske, M. R. Knowles, G. B. Pier, C. Lange and S. T. Weiss (2009). "IL1B polymorphisms modulate cystic fibrosis lung disease." Pediatric Pulmonology **44**(6): 580-593.
- Li, W., D. Soave, M. R. Miller, K. Keenan, F. Lin, J. Gong, T. Chiang, A. L. Stephenson, P. Durie, J. Rommens, L. Sun and L. J. Strug (2014). "Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities." Human Genetics **133**(2): 151-161.
- Lima, C. S., M. M. Ortega, F. A. Marson, R. Zulli, A. F. Ribeiro and C. S. Bertuzzo (2012). "Cystic fibrosis transmembrane conductance regulator gene mutations and glutathione S-transferase null genotypes in cystic fibrosis patients in Brazil." Jornal Brasileiro De Pneumologia: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisiologia **38**(1): 50-56.
- Livnat, G., R. Bar-Yoseph, A. Mory, E. Dagan, N. Elias, R. Gershoni and L. Bentur (2014). "Duplication in CHIT1 gene and the risk for Aspergillus lung disease in CF patients." Pediatr Pulmonol **49**(1): 21-27.

- Mafficini, A., M. Ortombina, I. Sermet-Gaudelius, P. Lebecque, T. Leal, P. Iansa, G. Reyckler, K. Dahan, X. Pepermans, G. Lenoir, A. Leonard, C. Sorio, B. Assael and P. Melotti (2011). "Impact of polymorphism of Multidrug Resistance-associated Protein 1 (ABCC1) gene on the severity of cystic fibrosis." Journal of Cystic Fibrosis **10**(4): 228-233.
- Magalhaes, M., I. Rivals, M. Claustres, J. Varilh, M. Thomasset, A. Bergougnoux, L. Mely, S. Leroy, H. Corvol, L. Guillot, M. Murris, E. Beyne, D. Caimmi, I. Vachier, R. Chiron and A. De Sario (2017). "DNA methylation at modifier genes of lung disease severity is altered in cystic fibrosis." Clinical Epigenetics **9** (1) (no pagination)(19).
- Mahadeva, R., L. Sharples, R. I. Ross-Russell, A. K. Webb, D. Bilton and D. A. Lomas (2001). "Association of alpha(1)-antichymotrypsin deficiency with milder lung disease in patients with cystic fibrosis." Thorax **56**(1): 53-58.
- Mahadeva, R., S. Stewart, D. Bilton and D. A. Lomas (1998). "Alpha-1 antitrypsin deficiency alleles and severe cystic fibrosis lung disease." Thorax **53**(12): 1022-1024.
- Mahadeva, R., R. C. Westerbeek, D. J. Perry, J. U. Lovegrove, D. B. Whitehouse, N. R. Carroll, R. I. Ross-Russell, A. K. Webb, D. Bilton and D. A. Lomas (1998). "Alpha1-antitrypsin deficiency alleles and the Taq-I G-->A allele in cystic fibrosis lung disease." Eur Respir J **11**(4): 873-879.
- Marceliino, A. R., F. A. Marson, C. S. Bertuzzo and J. D. Ribeiro (2013). "The -765G>C and 8473T>C polymorphism in COX2 gene and 57460C>T polymorphism in IFRD1 gene as modifiers of cystic fibrosis severity." Journal of Cystic Fibrosis. Conference: 36th European Cystic Fibrosis Conference. Lisbon Portugal. Conference Publication: **12**(pp S51).
- Marson, F., C. Bertuzzo, A. Ribeiro and J. Ribeiro (2014). "Polymorphisms in NOS-1 and NOS-3 genes may influence the cystic fibrosis severity?" European Respiratory Journal. Conference: European Respiratory Society Annual Congress **44**(no pagination).
- Marson, F., A. Marcelino, J. Ribeiro, A. Ribeiro and C. Bertuzzo (2013). "COX-2 gene polymorphisms: genetic determinants of cystic fibrosis comorbidities." International Journal of Genetics **5**(1): 132.

- Marson, F., L. Rezende, D. Furgeri, A. Ribeiro, J. Ribeiro and C. Bertuzzo (2013). "ADRA2A is a cystic fibrosis modifier gene." International Journal of Genetics **5**(1): 125.
- Marson, F. A., C. S. Bertuzzo, T. D. Hortencio, J. D. Ribeiro, L. C. Bonadia and A. F. Ribeiro (2012). "The ACE gene D/I polymorphism as a modulator of severity of cystic fibrosis." BMC Pulmonary Medicine **12**: 41.
- Marson, F. A., C. S. Bertuzzo, A. F. Ribeiro and J. D. Ribeiro (2012). "Polymorphisms in ADRB2 gene can modulate the response to bronchodilators and the severity of cystic fibrosis." BMC Pulmonary Medicine **12**: 50.
- Marson, F. A., C. S. Bertuzzo, A. F. Ribeiro and J. D. Ribeiro (2014). "Polymorphisms in the glutathione pathway modulate cystic fibrosis severity: a cross-sectional study." BMC Medical Genetics **15**: 27.
- Marson, F. A. L., J. D. Ribeiro, C. S. Bertuzzo and A. F. Ribeiro (2010). "Association of clinical severity of cystic fibrosis (CF) with polymorphism D/I in ACE gene." Journal of Cystic Fibrosis **9**: S4.
- Martin, A. C., I. A. Laing, G. Zhang, S. Brennan, K. Winfield, P. D. Sly, S. M. Stick, J. Goldblatt and P. N. LeSouef (2005). "CD14 C-159T and early infection with *Pseudomonas aeruginosa* in children with cystic fibrosis." Respir Res **6**: 63.
- McDougal, K. E., D. M. Green, L. L. Vanscoy, M. D. Fallin, M. Grow, S. Cheng, S. M. Blackman, J. M. Collaco, L. B. Henderson and K. Naughton (2010). "Use of a modeling framework to evaluate the effect of a modifier gene (MBL2) on variation in cystic fibrosis." European Journal of Human Genetics **18**(6): 680-684.
- McKone, E. F., J. Shao, D. D. Frangolias, C. L. Keener, C. A. Shephard, F. M. Farin, M. R. Tonelli, P. D. Pare, A. J. Sandford, M. L. Aitken and T. J. Kavanagh (2006). "Variants in the glutamate-cysteine-ligase gene are associated with cystic fibrosis lung disease." American Journal of Respiratory & Critical Care Medicine **174**(4): 415-419.
- Mekus, F., U. Laabs, H. Veeze and B. Tummler (2003). "Genes in the vicinity of CFTR modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pairs." Human Genetics **112**(1): 1-11.

- Melotti, P., A. Mafficini, P. Lebecque, M. Ortombina, T. Leal, E. Pintani, X. Pepermans, C. Sorio and B. M. Assael (2014). "Impact of MIF gene promoter polymorphism on F508del cystic fibrosis patients." PLoS ONE [Electronic Resource] **9**(12): e114274.
- Meyer, P., A. Braun and A. A. Roscher (2002). "Analysis of the two common alpha-1-antitrypsin deficiency alleles PiMS and PiMZ as modifiers of *Pseudomonas aeruginosa* susceptibility in cystic fibrosis." Clinical Genetics **62**(4): 325-327.
- O'Neal, W. K., P. Gallins, R. G. Pace, H. Dang, W. E. Wolf, L. C. Jones, X. Guo, Y. H. Zhou, V. Madar, J. Huang, L. Liang, M. F. Moffatt, G. R. Cutting, M. L. Drumm, J. M. Rommens, L. J. Strug, W. Sun, J. R. Stonebraker, F. A. Wright and M. R. Knowles (2015). "Gene expression in transformed lymphocytes reveals variation in endomembrane and HLA pathways modifying cystic fibrosis pulmonary phenotypes." American Journal of Human Genetics **96**(2): 318-328.
- Olesen, H. V., J. C. Jensenius, R. Steffensen, S. Thiel and P. O. Schiotz (2006). "The mannan-binding lectin pathway and lung disease in cystic fibrosis-dysfunction of mannan-binding lectin-associated serine protease 2 (MASP-2) may be a major modifier." Clinical Immunology **121**(3): 324-331.
- Oueslati, S., S. Hadj Fredj, H. Siala and T. Messaoud (2015). "Study of the TGF beta1 codon 10 in cystic fibrosis tunisian patients." Clinical Chemistry and Laboratory Medicine **53**: S837.
- Park, J., N. Sharma and G. Cutting (2012). "Evaluation of the melanocortin-3-receptor as a potential modifier of CF lung disease." Pediatric Pulmonology **47**: 281.
- Park, J. E., R. Yung, D. Stefanowicz, K. Shumansky, L. Akhabir, P. R. Durie, M. Corey, J. Zielenski, R. Dorfman, D. Daley and A. J. Sandford (2011). "Cystic fibrosis modifier genes related to *Pseudomonas aeruginosa* infection." Genes & Immunity **12**(5): 370-377.
- Plant, B. J., C. G. Gallagher, R. Bucala, J. A. Baugh, S. Chappell, L. Morgan, C. M. O'Connor, K. Morgan and S. C. Donnelly (2005). "Cystic fibrosis, disease severity, and a macrophage migration inhibitory factor polymorphism." American Journal of Respiratory & Critical Care Medicine **172**(11): 1412-1415.

- Pratap, U., S. Quinn, L. B. Blizzard and D. W. Reid (2010). "Population-based study of cystic fibrosis disease severity and haemochromatosis gene mutations." Respirology **15**(1): 141-149.
- Reynolds, W. F., I. Sermet-Gaudelus, V. Gausson, M. N. Feuillet, J. P. Bonnefont, G. Lenoir, B. Descamps-Latscha and V. Witko-Sarsat (2006). "Myeloperoxidase promoter polymorphism -463G is associated with more severe clinical expression of cystic fibrosis pulmonary disease." Mediators of Inflammation **2006**(2): 36735.
- Rohovyk, N., N. Vishtak, H. Makuh, L. Bober and O. Lyha (2014). "Alleles polymorphism of microsomal epoxide hydrolase gene (mEPHX) as modifier factor in cystic fibrosis." Journal of Cystic Fibrosis **13**: S45.
- Sahli, C., S. H. Fredj, H. Siala, A. Bibi and T. Messaoud (2014). "First study of angiotensin converting enzyme in cystic fibrosis Tunisian patients." Clinical Chemistry and Laboratory Medicine **52**(9): e211-e215.
- Sahli, C., S. Hadj-Fredj, R. Dabboubi, K. Bousseta, A. Mehrzi and T. Messaoud (2016). "Association of MP6d9 polymorphism with clinical variability in cystic fibrosis patients." Clinical Laboratory **62**(11): 2139-2143.
- Santis, G., L. Osborne, R. A. Knight and M. E. Hodson (1990). "Linked marker haplotypes and the DELTAF⁵⁰⁸ mutation in adults with mild pulmonary disease and cystic fibrosis." Lancet **335**(8703): 1426-1429.
- Schmitt-Grohe, S., F. Stuber, M. Book, J. Bargon, T. O. Wagner, C. Naujoks, R. Schubert, M. J. Lentze and S. Zielen (2006). "TNF-alpha promoter polymorphism in relation to TNF-alpha production and clinical status in cystic fibrosis." Lung **184**(2): 99-104.
- Segat, L., M. Morgutti, E. Athanasakis, C. Trevisiol, A. Amaddeo, F. Poli and S. Crovella (2010). "Analysis of DEFB1 regulatory SNPs in cystic fibrosis patients from North-Eastern Italy." Int J Immunogenet **37**(3): 169-175.
- Shmarina, G., A. Pukhalsky, N. Petrova, E. Zakharova, L. Avakian, N. Kapranov and V. Alioshkin (2013). "TNF gene polymorphisms in cystic fibrosis patients: contribution to the disease progression." Journal of Translational Medicine **11**: 19.

- Shteinberg, M., J. Rivlin, M. Gur, M. Konopnicki, N. Stein, M. M. Tunney, J. S. Elborn, D. G. Downey, E. Johnston, H. Shalom and A. Levy (2015). "Lack of Association Between Haptoglobin Phenotype and Cystic Fibrosis Outcomes." Lung **193**(6): 1017-1021.
- Smith, D., S. Bell, G. Anderson and D. Reid (2014). "HFE 282y mutations are associated with severity of lung disease and rate of lung function decline in cystic fibrosis." Respirology **19**: 48.
- Soave, D., H. Corvol, N. Panjwani, J. Gong, W. Li, P. Y. Boelle, P. R. Durie, A. D. Paterson, J. M. Rommens, L. J. Strug and L. Sun (2015). "A Joint Location-Scale Test Improves Power to Detect Associated SNPs, Gene Sets, and Pathways." American Journal of Human Genetics **97**(1): 125-138.
- Stanke, F., T. Becker, H. Cuppens, V. Kumar, J. J. Cassiman, S. Jansen, D. Radojkovic, B. Siebert, J. Yarden, D. W. Ussery, T. F. Wienker and B. Tummmler (2006). "The TNFalpha receptor TNFRSF1A and genes encoding the amiloride-sensitive sodium channel ENaC as modulators in cystic fibrosis." Human Genetics **119**(3): 331-343.
- Stanke, F., T. Becker, S. Hedtfeld, S. Tamm, T. F. Wienker and B. Tummmler (2010). "Hierarchical fine mapping of the cystic fibrosis modifier locus on 19q13 identifies an association with two elements near the genes CEACAM3 and CEACAM6." Human Genetics **127**(4): 383-394.
- Stanke, F., T. Becker, V. Kumar, S. Hedtfeld, C. Becker, H. Cuppens, S. Tamm, J. Yarden, U. Laabs, B. Siebert, L. Fernandez, M. Macek, Jr., D. Radojkovic, M. Ballmann, J. Greipel, J. J. Cassiman, T. F. Wienker and B. Tummmler (2011). "Genes that determine immunology and inflammation modify the basic defect of impaired ion conductance in cystic fibrosis epithelia." Journal of Medical Genetics **48**(1): 24-31.
- Stanke, F., C. Davenport, S. Hedtfeld and B. Tummmler (2010). "Differential decay of parent-of-origin-specific genomic sharing in cystic fibrosis-affected sib pairs maps a paternally imprinted locus to 7q34." European Journal of Human Genetics **18**(5): 553-559.
- Stanke, F., S. Hedtfeld, T. Becker and B. Tummmler (2011). "An association study on contrasting cystic fibrosis endophenotypes recognizes KRT8 but not KRT18 as a modifier of cystic fibrosis disease severity and CFTR mediated residual chloride secretion." BMC Medical Genetics **12**: 62.

- Steagall, W. K., B. J. Barrow, C. G. Glasgow, J. W. Mendoza, M. Ehrmantraut, J. P. Lin, P. A. Insel and J. Moss (2007). "Beta-2-adrenergic receptor polymorphisms in cystic fibrosis." Pharmacogenetics and Genomics **17**(6): 425-430.
- Strug, L. J., T. Gonska, G. He, K. Keenan, W. Ip, P. Y. Boelle, F. Lin, N. Panjwani, J. Gong, W. Li, D. Soave, B. Xiao, E. Tullis, H. Rabin, M. D. Parkins, A. Price, P. C. Zuberbuhler, H. Corvol, F. Ratjen, L. Sun, C. E. Bear and J. M. Rommens (2016). "Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics." Human Molecular Genetics **25**(20): 4590-4600.
- Tarova, E. T., H. Polakova, H. Kayserova, P. Celec, M. Zuzulova and L. Kadasi (2011). "Study of the effect of DNA polymorphisms in the mannose-binding lectin gene (MBL2) on disease severity in Slovak cystic fibrosis patients." General Physiology & Biophysics **30**(4): 373-378.
- Taylor-Cousar, J. L., M. A. Zariwala, L. H. Burch, R. G. Pace, M. L. Drumm, H. Calloway, H. Fan, B. W. Weston, F. A. Wright, M. R. Knowles and G. Gene Modifier Study (2009). "Histo-blood group gene polymorphisms as potential genetic modifiers of infection and cystic fibrosis lung disease severity." PLoS ONE [Electronic Resource] **4**(1): e4270.
- Tesse, R., F. Cardinale, T. Santostasi, A. Polizzi, A. Manca, L. Mappa, G. Iacoviello, F. De Robertis, V. Logrillo and L. Armenio (2008). "Association of β -defensin-1 gene polymorphisms with *Pseudomonas aeruginosa* airway colonization in cystic fibrosis." Genes and immunity **9**(1): 57-60.
- Tesse, R., F. Cardinale, T. Santostasi, A. Polizzi, L. Mappa, A. Manca, F. De Robertis, O. Silecchia and L. Armenio (2008). "Association of interleukin-10 gene haplotypes with *Pseudomonas aeruginosa* airway colonization in cystic fibrosis." J Cyst Fibros **7**(4): 329-332.
- Texereau, J., S. Marullo, D. Hubert, J. Coste, D. J. Dusser, J. Dall'Ava-Santucci and A. T. Dinh-Xuan (2004). "Nitric oxide synthase 1 as a potential modifier gene of decline in lung function in patients with cystic fibrosis." Thorax **59**(2): 156-158.
- Trevisiol, C., M. Boniotto, L. Giglio, F. Poli, M. Morgutti and S. Crovella (2005). "MBL2 polymorphisms screening in a regional Italian CF Center." Journal of Cystic Fibrosis **4**(3): 189-191.

- Turnbull, A., H. Lund-Palau, R. Murphy, A. Simbo, A. Shoemark, K. Wong, A. Bush, E. Alton and J. Davies (2016). "The T2R38 bitter taste receptor as a modifier of host response to pseudomonas aeruginosa in cystic fibrosis: Does T2R38 genotype impact on clinical infection?" Thorax **71**: A44.
- Urquhart, D. S., J. Allen, M. Elrayess, K. Fidler, N. Klein and A. Jaffé (2006). "Modifier effect of the Toll-like receptor 4 D299G polymorphism in children with cystic fibrosis." Archivum immunologiae et therapiae experimentalis **54**(4): 271-276.
- Vankeerberghen, A., O. Scudiero, K. De Boeck, M. Macek Jr, P. F. Pignatti, N. Van Hul, H. Nuytten, F. Salvatore, G. Castaldo, D. Zemkova, V. Vavrova, J. J. Cassiman and H. Cuppens (2005). "Distribution of human beta-defensin polymorphisms in various control and cystic fibrosis populations." Genomics **85**(5): 574-581.
- Viel, M., D. Hubert, P. R. Burgel, E. Genin, I. Honore, B. Martinez, N. Gaïtch, J. Chapron, R. Kanaan, D. Dusser, E. Girodon and T. Bienvenu (2016). "DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis." The clinical respiratory journal **10**(6): 777-783.
- Viel, M., C. Leroy, D. Hubert, I. Fajac and T. Bienvenu (2008). "ENaCbeta and gamma genes as modifier genes in cystic fibrosis." Journal of Cystic Fibrosis **7**(1): 23-29.
- von Kanel, T., F. Stanke, M. Weber, A. Schaller, J. Racine, R. Kraemer, M. Chanson, B. Tummler and S. Gallati (2013). "Clinical and molecular characterization of the potential CF disease modifier syntaxin 1A." European Journal of Human Genetics **21**(12): 1462-1466.
- Wright, F. A., L. J. Strug, V. K. Doshi, C. W. Commander, S. M. Blackman, L. Sun, Y. Berthiaume, D. Cutler, A. Cojocaru, J. M. Collaco, M. Corey, R. Dorfman, K. Goddard, D. Green, J. W. Kent Jr, E. M. Lange, S. Lee, W. Li, J. Luo, G. M. Mayhew, K. M. Naughton, R. G. Pace, P. Pare, J. M. Rommens, A. Sandford, J. R. Stonebraker, W. Sun, C. Taylor, L. L. Vanscoy, F. Zou, J. Blangero, J. Zielenski, W. K. O'Neal, M. L. Drumm, P. R. Durie, M. R. Knowles and G. R. Cutting (2011). "Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2." Nature Genetics **43**(6): 539-546.

- Wright, J. M., C. A. Merlo, J. B. Reynolds, P. L. Zeitlin, J. G. Garcia, W. B. Guggino and M. P. Boyle (2006). "Respiratory epithelial gene expression in patients with mild and severe cystic fibrosis lung disease." American Journal of Respiratory Cell & Molecular Biology **35**(3): 327-336.
- Yarden, J., D. Radojkovic, K. De Boeck, M. Macek Jr, D. Zemkova, V. Vavrova, R. Vlietinck, J. J. Cassiman and H. Cuppens (2005). "Association of tumour necrosis factor alpha variants with the CF pulmonary phenotype." Thorax **60**(4): 320-325.
- Yarden, J., D. Radojkovic, K. De Boeck, M. Macek, Jr., D. Zemkova, V. Vavrova, R. Vlietinck, J. J. Cassiman and H. Cuppens (2004). "Polymorphisms in the mannose binding lectin gene affect the cystic fibrosis pulmonary phenotype." Journal of Medical Genetics **41**(8): 629-633.
- Zielenski, J., M. Corey, R. Rozmahel, D. Markiewicz, I. Aznarez, T. Casals, S. Larriba, B. Mercier, G. R. Cutting, A. Krebsova, M. Macek, Jr., E. Langfelder-Schwind, B. C. Marshall, J. DeCelle-Germana, M. Claustres, A. Palacio, J. Bal, A. Nowakowska, C. Ferec, X. Estivill, P. Durie and L. C. Tsui (1999). "Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13." Nat Genet **22**(2): 128-129.

2.4 RESEARCH IN CONTEXT

This project systematically reviewed a large body of evidence, with multiple non-*CFTR* gene modifiers identified. However, the negligible clinical impact of this research is disappointing. As highlighted in the discussion, the factors contributing to this include the inconsistent phenotyping of subjects using insensitive outcome measures and the limited replication of positive results.

The findings of this review informed the subsequent work presented in this thesis in several ways. First, given the relationship between genotype and DNA methylation, validated gene modifiers (shown in Table 3 in the manuscript) represent candidate sites for targeted evaluation as potential DNA methylation-based biomarkers of CF lung disease severity. In addition, the experiments conducted as part of this thesis have endeavoured to use the most sensitive and robust tools to phenotype subjects. To enable the replication of any positive findings, two separate experiments were conducted. The thesis also presents two methodological papers to assist in the external validation of any positive findings. The systematic review may have broader implications than the thesis alone. The key findings, particularly regarding limitations of the evidence to date and consequent recommendations for study design, should inform the ongoing search for predictive biomarkers in CF, which may significantly improve clinical care.

CHAPTER 3: DNA Methylation Biomarkers of Future Health Outcomes in Children

3.1 INTRODUCTION

The identification of predictive biomarkers could transform the care delivered to subjects with chronic disease. By predicting disease severity, the development of complications or response to therapy, biomarkers would allow clinicians to recommend more precise care, maximising the potential benefits while minimising the risk of iatrogenic harm. Given that disease severity is frequently determined by the interaction of genetics and the environment, epigenetic biomarkers are well positioned to act as predictive biomarkers. In particular, DNA methylation-based predictive biomarkers have been assessed in multiple paediatric conditions.

The purpose of this project was to review studies that have examined DNA methylation-based predictive biomarkers in chronic childhood diseases. The aim was to appraise and summarise the evidence, with a particular focus on study design. The results of this project then informed the design of the experiments included in this thesis, which aimed to identify DNA methylation-based predictive biomarkers of CF lung disease severity.

3.2 PUBLISHED MANUSCRIPT

The narrative review was published in *Molecular and Cellular Paediatrics* (2020), volume 7, issue 7 (<https://doi.org/10.1186/s40348-020-00099-0>). The following is an exact electronic reprint of the manuscript as published.

REVIEW

Open Access

DNA methylation biomarkers of future health outcomes in children



Shivanthan Shanthikumar^{1,2,3*} , Melanie R. Neeland^{3,4}, Jovana Maksimovic^{2,3,5}, Sarath C. Ranganathan^{1,2,3} and Richard Saffery^{3,4}

Abstract

Biomarkers which predict future health outcomes are key to the goals of precision health. Such biomarkers do not have to be involved in the causal pathway of a disease, and their performance is best assessed using statistical tests of clinical performance and evaluation of net health impact. DNA methylation is the most commonly studied epigenetic process and represents a potential biomarker of future health outcomes. We review 25 studies in non-oncological paediatric conditions where DNA methylation biomarkers of future health outcomes are assessed. Whilst a number of positive findings have been described, the body of evidence is severely limited by issues with outcome measures, tissue-specific samples, accounting for sample cell type heterogeneity, lack of appropriate statistical testing, small effect sizes, limited validation, and no assessment of net health impact. Future studies should concentrate on careful study design to overcome these issues, and integration of DNA methylation data with other 'omic', clinical, and environmental data to generate the most clinically useful biomarkers of paediatric disease.

Keywords: Pediatrics, DNA methylation, Biomarker, Epigenetics, Precision medicine

Introduction

Precision health (also referred to as precision medicine and personalised medicine) can be defined as interventions "targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations" [1]. It is a rapidly growing field [2] and has been described as an "ongoing revolution in medicine, moving it from a reactive to a proactive discipline, where ultimately the objective is to maximize wellness for each individual rather than simply to treat disease" [3]. Accordingly, significant resources are being allocated to the development of precision approaches across a wide range of healthcare areas with the aim of making discoveries which can be incorporated into routine care [4].

Biomarkers that allow prediction of future health outcomes are key foci of precision health [5]. These may identify those who are at high risk of developing a disease or alternatively may predict disease progression and severity, treatment response, or risk of complications. Advances in genomics in particular, along with bioinformatics, proteomics, and metabolomics, are all contributing to the identification of biomarkers in the precision health field [1]. Whilst biomarkers may be involved in disease pathogenesis, this is not a requirement for them to be clinically useful. Biomarkers have been divided into 'descriptive' vs. 'mechanistic' biomarkers, where descriptive biomarkers are associated with disease but not directly involved in the causal pathway, and mechanistic biomarkers are directly involved in disease pathogenesis [6]. Either type of biomarker is relevant to the goals of precision health. The utility of biomarkers can be assessed using statistical tests traditionally used to assess clinical investigations such as sensitivity, specificity, positive predictive value, negative predictive value, and area under a receiver operator curve [7]. In addition, if

* Correspondence: shivanthan.shanthikumar@rch.org.au

¹Respiratory and Sleep Medicine, Royal Children's Hospital, Flemington Road, Parkville, Melbourne, Victoria 3052, Australia

²Respiratory Diseases, Murdoch Children's Research Institute, Melbourne, Australia

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

a biomarker is to be used in routine healthcare, a positive net health impact must be demonstrated, including assessment of factors such as cost-effectiveness and impact on quality of life and disease severity [7].

Epigenetic processes regulate gene expression, without changes in the underlying DNA sequence [8]. These are tissue-specific (i.e. the epigenetic profile of leucocytes will be different to that of bronchial epithelial cells), generally persist during cell replication (are hence heritable), and maybe influenced by the environment. A recent all-encompassing definition of epigenetics is “modifications of DNA or associated factors that have information content, other than the DNA sequence itself, are maintained during cell division, are influenced by the environment, and cause stable changes in gene expression” [8]. There are several different epigenetic mechanisms (DNA methylation, histone modification, non-coding RNA, and higher-order chromatin structure) which regulate gene expression via modifying the accessibility of DNA to transcription and other regulatory factors (see Fig. 1) [8].

Epigenetic processes do not work independently; rather, they interact to determine gene expression. DNA methylation is the best understood and most useful epigenetic marker to study human disease due to stability over time and ease of measurement [8]. The most abundant form of DNA methylation in the genome involves the addition of a methyl group (CH_3) to the fifth carbon of a cytosine nucleotide within a cytosine-phosphate-guanine dinucleotide (CpG) [10]. CpG sites are found throughout the genome in low density and are usually highly methylated leading to gene inactivation. They can be found in gene regions such as the promoter (including transcription start sites (TSS), 5' untranslated region (UTR)), gene body (exons, introns), 3' UTR, and intergenic regions. A minority of CpG sites are located in highly dense areas called CpG islands, and these are located in the promoter regions of around 70% of all genes [11]. CpG islands usually have a low level of DNA methylation although this does not necessarily mean the gene will be highly expressed as other steps such as transcription factor binding are required for

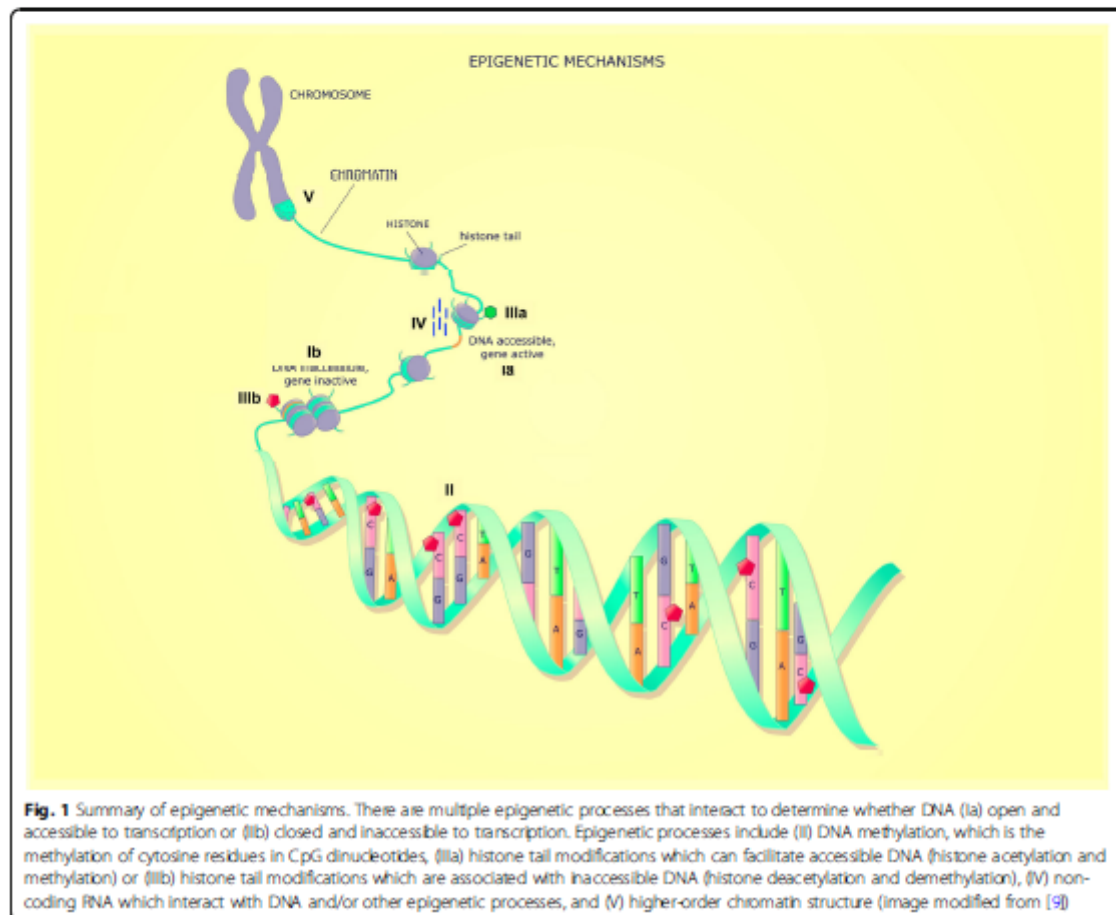


Fig. 1 Summary of epigenetic mechanisms. There are multiple epigenetic processes that interact to determine whether DNA (Ia) open and accessible to transcription or (Ib) closed and inaccessible to transcription. Epigenetic processes include (II) DNA methylation, which is the methylation of cytosine residues in CpG dinucleotides, (IIIa) histone tail modifications which can facilitate accessible DNA (histone acetylation and methylation) or (IIIb) histone tail modifications which are associated with inaccessible DNA (histone deacetylation and demethylation), (IV) non-coding RNA which interact with DNA and/or other epigenetic processes, and (V) higher-order chromatin structure (image modified from [9])

gene expression. Whilst at the level of an individual allele CpG methylation is a binary state (i.e. the CpG is either methylated or not), the DNA methylation level of a sample will range from 0–100%. This is because samples will contain a number of different alleles and will frequently involve multiple different cell types each of which has a unique methylation pattern. Hence, if a sample has a 50% methylation level, this could be because 50% of the alleles in the sample are methylated or because one cell type makes up 50% of the sample, and 100% of the alleles in that cell type are methylated and 0% of the alleles in other cell types are methylated. An individual's DNA methylation profile is influenced by their underlying genetic makeup and both antenatal and postnatal environmental exposures [12, 13].

As disease outcomes are often thought to be due to a combination of genotype, environmental exposures, and the interaction of the two [14], DNA methylation patterns are potentially ideal biomarkers. Studies attempting to identify DNA methylation biomarkers compare methylation patterns between groups of patients who differ based on the outcome of interest. DNA methylation can be assessed at specific CpG sites of interest or using an epigenome-wide association study (EWAS) approach. The most commonly used product for EWAS is the Illumina Infinium Methylation Array which has evolved from assessing 27,000 (27K array) CPG sites to 450,000 (450k) and now over 850,000 (referred to as the EpicArray). In either design, the average methylation at each CpG site is compared with the aim of identifying statistically significant differential methylation between the two groups, referred to as a differentially methylated position (DMP).

The aim of this review is to summarise the evidence regarding the use of DNA methylation as predictive biomarkers in children with a diagnosed disease. This includes prediction of disease severity, treatment response, or development of complications. Studies examining cancer have not been included as reviews in this area have been recently published [15–17]. In addition, cross-sectional association studies which attempt to link DNA methylation to current disease state have been excluded, as this review focuses on studies which predict future outcomes in children as these offer greater insight as to

the predictive ability of DNA methylation profiles in development of disease.

Evidence Reviewed

Articles were identified via the search strategy in Table 1. A number of studies using DNA methylation as a biomarker in children have been completed (summarised in Supplementary Table 1). Included studies examine whether DNA methylation can predict disease severity [18–26], treatment response [27–34], or development of a complication [35–41].

Neurocognitive impairment in preterm infants

Premature infants are known to have an increased risk of later life neurocognitive impairment. Two studies have attempted to identify DNA methylation biomarkers of future cognitive impairment in premature infants. Tilley et al. [35] examined genome-wide methylation, using the 450K array, in a homogenised sample of chorionic placental tissue of 84 extremely premature infants (born at < 28 weeks gestation). At 10 years of age, participants underwent a cognitive assessment using the School-Age Differential Ability Scales–II. Using logistic regression analysis, they identified 17 probe sites, corresponding to 16 genes involved in neuronal development and function, where increased DNA methylation was associated with cognitive impairment at 10 years. Ten of the probes were located in the gene body, 3 in the promoter region, 3 in the 5'UTR, and 1 in the 3'UTR. The effect size was modest with odds ratios (OR) for these associations between 1.04 and 1.09. For each of the genes, a 1% increase in methylation at their respective probe site resulted in a 4–7% increase in the odds of cognitive impairment. There was evidence of a dose-response relationship, with higher levels of methylation being associated with greater impairment. Arpon et al. [36] examined 22 full-term infants and 24 premature infants (< 34 weeks gestation). They performed an EWAS using a 450K array on whole blood samples collected at 12 months of age. Neurodevelopmental outcome at 2–3 years of age was assessed using the Bayley Scale of Infant Development. Correlations were assessed using Pearson or Spearman tests, and linear regression was performed where significant correlations existed. Correction for multiple testing

Table 1 Search strategy to identify articles

Database searched	Pubmed
Dates	All articles until 5.12.2019
Search terms	DNA-methylation* AND (prognosis OR survival OR outcome* OR progression OR deterioration OR pathophysiology OR physiopathology OR morbidit* OR mortalit*) AND (predict* OR response OR risk) AND (newborn* OR baby OR babies OR neonat* OR infan* OR toddler* OR pre-schooler* OR preschooler* OR kindergarten OR boy OR boys OR girl OR girls OR child OR children OR childhood OR adolescen* OR pediatric* OR paediatric* OR youth* OR teen OR teens OR teenage*) AND (NOTNLM OR publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb] OR indatereview[sb] OR pubstatusaheadofprint)

was undertaken using the Benjamini-Hochberg procedure. They found methylation of a single CpG site located in the 5'UTR of *SLC6A3* showed the greatest difference between preterm and term infants (increased methylation in preterm) and that increased methylation in both term and preterm infants was associated with significantly poorer motor function ($r = -0.55$, $p = 0.0001$) and mental function ($r = -0.44$, $p = 0.0028$). The sites of differential methylation identified in each of these studies did not overlap, which may reflect differences in sample size, the biospecimen used for methylation analysis, outcome assessment tool, and age at both methylation analysis and outcome assessment.

Asthma and allergy

In a large study examining 65 patients with food allergy, Martino et al. [18] assessed DNA methylation in naive CD4 T cells at enrolment and then 2–4 years later. Subjects were aged between 11 and 15 months at enrolment and defined as being food allergic based on a positive skin prick test and clinical reactivity. At reassessment, patients were classified as having persistent food allergy (based on oral food challenge) or resolved allergy. The investigators examined whether the change in DNA methylation profiles from baseline to reassessment could predict the likelihood of persistent allergy, and they focused on 26 loci which were demonstrated to be differentially methylated between patients with and without food allergy. They found that in 24 of 26 loci, there was a significant change (adjusted p value < 0.05) in methylation from baseline to follow-up in patients with persistent allergy, whereas methylation was stable in patients whose allergy resolved. In the group with persistent allergy, the maximum change in methylation at an individual locus between baseline and follow-up was 17.2%. Fu et al. [42] studied methylation in the 5'UTR of $\beta 2$ adrenoceptor (*ADRB2*) in 182 children with asthma aged between 5 and 12 years. Blood was collected at enrolment, and patients were followed up for 12 months. Parent-reported symptoms were used to calculate an asthma severity score, which was used to classify patients into mild or severe disease. Multivariate logistic regression (adjusted for age and gender) was used. DNA methylation was classified as 'low,' 'intermediate,' and 'high,' and asthma as 'mild' or 'severe.' In a dose-dependent manner, increasing methylation of *ADRB2* was associated with more severe asthma, with those with intermediate methylation having an OR 4.11 (95% CI 1.58–10.73) for severe asthma and those with high methylation having an OR 7.63 (95% CI 3.02–19.26). The median methylation level in the severe asthma group was 1.14% (range 0.2 to 3.6%) and in the mild asthma group was 0.81% (0 to 2.4%). They also examined nitrous oxide exposure and found that in the highly

exposed group, those with *ADRB2* hypermethylation had an OR of 4.59 (95% CI 1.03–20.55) of having severe asthma.

Two studies have investigated whether DNA methylation markers can predict response to corticosteroids during an acute asthma exacerbation. Xiao et al. [27] studied Vanin-1 (*VNN1*) methylation in nasal epithelial cells, a surrogate for lower airway epithelium, obtained in 18 children aged 5–18 years at presentation to hospital with an acute asthma exacerbation and again after 12–24 h. *VNN1* was identified as a gene of interest via an initial gene expression investigation of $> 20,000$ genes. Patients were treated with a standardised asthma protocol which included criteria for discharge. Good corticosteroid response was defined as a length of hospital stay ≤ 24 h, whereas poor response was defined as a length of stay > 24 h. Those who had a good response to corticosteroids had an increase in methylation at a single CpG, located in the promoter region of *VNN1*, whereas those that did not respond to corticosteroids had evidence of decreased methylation, although the raw figures were not reported. When assessed using Fischer's exact test, the differences were significant ($p = 0.003$). At the same hospital, Zhang et al. [28] used the same design but assessed genome-wide methylation using the 450K array. Their study involved 33 children presenting with acute asthma exacerbations aged 5–18 years. They used linear regression and surrogate variable analysis to correct for confounders, as well as correcting for multiple testing. They found hypermethylation at 1 site in the promoter of lactate dehydrogenase C (*LDHC*) at presentation was associated with poor treatment response. The mean (standard error) methylation level was 87.7% (2.1%) in the poor response group, and 78.6% (2.4%) in the good responders ($p = 0.007$). A major strength of the two studies looking at treatment effect was the use of nasal epithelial cells to assess DNA methylation, which is a better surrogate for bronchial epithelial cells than peripheral blood [43].

Psychiatric disorders

Another area which has been widely studied in psychiatric disorders. In a group of 164 patients with 22q11 microdeletion syndrome, who are at an increased risk of psychiatric disorders, Starnawska et al. [38] performed an EWAS using the 450K array on neonatal blood spot samples. The later development of a psychiatric disorder was assessed by examining the Danish national psychiatric registry at a mean age of 14.7 years. Linear regression was used to adjust for confounders such as sex and age. A total of 9 differentially methylated probes were significantly associated with development of a psychiatric disorder (unadjusted p value $< 10^{-6}$). The effect size of each individual CpG was not described. The probes were

located in the promoter region of associated genes (4/9), intergenic regions (3/9), and the gene body (2/9). A further 9 probes associated with intellectual disability, 3 with schizophrenia, 2 with behavioural disorders, and 8 with development disorders. The individual effect size was not detailed for any of these associations. The most common location of probes associated with intellectual disability was the promoter (4/9), and for development disorders, it was the body (4/9). Several studies have examined whether DNA methylation biomarkers can predict response to treatment of psychiatric conditions. Gasso et al. [29] examined whether methylation at 7 CpG sites in the promoter region of 5-hydroxytryptamine receptor 1B (*HTR1B*), measured in blood, could predict treatment response to 12 weeks of the selective serotonin reuptake inhibitor medication, fluoxetine. They studied 83 children aged between 10 and 17 years who were treated with fluoxetine due to major depressive disorder, obsessive-compulsive disorder, or generalised anxiety disorder. Subjects completed multiple validated scales which measure disease severity, and the change in these scales across 12 weeks was used to define treatment response. Pearson's test was used to assess correlation, and a Bonferroni correction for multiple testing was performed. When the methylation level at the 7 CpG sites was averaged, they found a moderately strong inverse relationship with higher methylation level associated with reduced fluoxetine response ($r = -0.335$; $p = 0.004$). The methylation level for the two groups was not reported. In a study of 116 children aged 6–13 years with an anxiety disorder, the response to cognitive behaviour therapy (CBT) was assessed [31]. Using buccal swab samples, targeted analysis of methylation of 6 CpG sites in the promoter region of the serotonin transporter (*SLC6A4*) was performed pre- and post-treatment with CBT. Treatment response was based on whether the anxiety disorder had resolved. Subjects with a sustained CBT response (measured 6 months post-treatment) were more likely to have increased methylation at a single CpG during treatment (+ 3.48%), whereas non-responders were more likely to decrease (– 5.44%), and this relationship was significant when corrected for multiple testing ($p = 0.002$). In another study also examining buccal swabs from children ($n = 98$) with anxiety disorders, and their response to cognitive behaviour therapy, DNA methylation of CpG sites in the promoter region of FK506 binding protein 5 (*FKBP5*) and glucocorticoid receptor (*GR*) was assessed before and immediately after treatment [30]. The children were aged between 5 and 18 years. Treatment response was defined as a change in primary anxiety disorder severity. No significant effect of *FKBP5* or *GR* methylation was found. In a study of 111 children, mean age 9.25 years, with attention deficit

hyperactivity disorder (ADHD) who were treated with methylphenidate, dopamine transporter (*DAT1*), and dopamine receptor D4 (*DRD4*), promoter methylation was assessed in blood samples collected pre-treatment to see if this predicted treatment response [32]. The diagnosis of ADHD was based on fulfilling international criteria, and treatment response was based on change in a validated ADHD severity questionnaire across 6 weeks. Spearman's rank correlation test was used to assess correlations. They found lower mean *DAT1* methylation was associated with greater treatment response ($\rho = -0.222$, $p = 0.019$) although the mean methylation levels were not reported.

Neonatology

Three studies have attempted to identify DNA methylation biomarkers of severity of neonatal abstinence syndrome (NAS). Wachman et al. [20] analysed opioid receptor mu 1 (*OPRM1*) promoter methylation in the saliva of 58 newborns and the relationship with NAS severity. NAS severity was primarily defined based on the need for pharmacological treatment, which was determined based on a standardised protocol. They used linear and logistic regression models as well as correction for multiple testing using the Benjamini and Hochberg method. The study found infants needing pharmacological treatment for NAS had higher methylation at several CpGs: CpG – 18 (11.4% vs 4.4%, $p = .001$), CpG – 14 (46.1% vs 24.0%, $p = .002$), and CpG + 23 (26.3% vs 12.9%, $p = .008$). The same group of authors conducted another study with the same methods and research questions in a group of 86 infants with NAS. They did not replicate the finding with regard to CpGs – 18, +23, but did find that hypermethylation of CpG – 14 was associated with greater severity of NAS measured by need for 2 medications to control symptoms (adjusted difference = 4.9% (95% CI 1.8–8.1%), $p = 0.0030$) [19]. In a study of 21 babies, McLaughlin et al. [21] also used early life saliva swabs and analysed methylation in ATP binding cassette subfamily B member 1 (*ABCB1*) promoter, cytochrome P450 family 2 subfamily D member 6 (*CYP2D6*) exon 1, and *OPRM1* promoter and exon 1. They used a standardised protocol for monitoring NAS severity. Unlike the studies by Wachman et al., they did not find any significant effect. A potential explanation is the smaller sample size of 21, as compared to 58 and 86 in the positive studies.

In a group of 20 premature infants, DNA methylation in the *GR* promoter was analysed in blood from day 4 of life [22]. The outcome of interest was need for treatment with glucocorticosteroids for treatment of lung disease of prematurity or late-onset circulatory collapse, although criteria for exactly when glucocorticosteroids were prescribed were not detailed. Using a logistic

regression model, they found a weak association between methylation at a single CpG site and need for treatment with glucocorticosteroids (likelihood ratio 3.889, $p = 0.0486$). Dhas et al. [37] investigated 51 newborns with sepsis which was defined as clinical features and presence of biochemical markers. They analysed global DNA methylation in whole blood and using independent samples t tests found those who developed renal failure as a complication of sepsis had slightly higher global methylation levels (2.88% vs. 2.39%, $p = 0.02$).

Inflammatory bowel disease

Two studies have investigated children with inflammatory bowel disease (IBD). Howell et al. [23] used intestinal biopsy samples from 109 children with IBD and performed an EWAS to identify biomarkers which predicted disease severity over the following 18 months. Disease severity was based on clinical factors such as requirement for treatment with biological agents and time to third treatment escalation following diagnosis. They also evaluated genome-wide gene expression. Differential methylation analysis was performed using limma for DNA methylation data, with a significant false discovery rate [FDR] of < 0.01 . They identified a DNA methylation profile that was associated with increased disease severity, as well as demonstrating a gene expression profile that is associated with severe disease with significant overlap in the genes involved. A strength of this study was using samples directly from the tissue of interest. Another study used blood samples collected at diagnosis and the EpicArray to perform an EWAS on 164 children with Chron's disease [24]. They measured disease severity by considering the time to progress to complicated disease and attempted to identify DMP using linear mixed-effects models and the Bonferroni correction for multiple testing. They were unable to find any CpG sites that were associated with disease severity.

Idiopathic scoliosis

Meng et al. studied patients with idiopathic scoliosis and the risk of progression of the scoliosis curve beyond 30°, which is a marker of severity. They identified CpG sites of interest in blood, by first investigating two female twin pairs with idiopathic scoliosis who were discordant for curve progression. The top 4 DMP were chosen for replication in a cross-sectional cohort of 92 adolescents who were divided based on the severity of scoliosis at presentation, and only one CpG site (cg01374129) remained a significant DMP. This CpG site is not associated with a gene. They then tested the performance of methylation at this individual CpG in a cohort of 276 adolescent subjects who had mild scoliosis at presentation and then were divided based on whether their scoliosis progressed prior to skeletal maturation. They used

the Mann-Whitney U test to compare the mean methylation level and progression status. Mean methylation at cg01374129 was lower in the group that progressed (approximately 10% vs. 16%, $p < 0.0001$), and when a receiver operator curve was calculated, the area under the curve was 0.827 (95% CI 0.780 to 0.876), with a sensitivity of 76.4% and a specificity of 85.6% at a cutoff level of 15.1%. Two studies have attempted to identify biomarkers of development of chronic pain complications post-scoliosis surgery. In a study of 128 adolescent patients with idiopathic scoliosis undergoing scoliosis surgery, *OPRM1* promoter methylation was assessed at 22 CpG sites in blood collected during the operation [40]. Acute and chronic pain outcomes were assessed using internationally validated definitions and were dichotomised for analysis. Linear regression models were used for analysis with significance set at $p < 0.05$. There was no correction for multiple testing. After adjustment for preoperative pain score and morphine consumption over postoperative days 1 and 2, higher *OPRM1* promoter methylation at 3 CpG sites was associated with increased acute postoperative pain. The strength of the association was weak for two CpG sites with a 1% increase in DNA methylation associated with regression coefficients of 0.921 and 1.864; however, one CpG had a regression coefficient of 17.736. When chronic postsurgical pain was assessed, there was significant different methylation at two CpG sites. The OR for these associations were relatively weak: 1.069 (95% confidence interval 1.022–1.119) and 1.037 (1.000–1.075). Once again, adjustment was performed for preoperative pain and postoperative morphine consumption. The two CpG sites did not overlap with those associated with acute postoperative pain. The authors provided further evidence for the role of *OPRM1* methylation via using functional genomic analysis from publicly available databases which showed increased methylation would lead to reduced gene expression and reduced response to opioids. In a subsequent study, the same authors assessed 46 adolescents undergoing scoliosis surgery. They performed an EWAS using the EpicArray on peripheral blood samples. For this study, they used medication use and questionnaires to assess pain severity. Surrogate variance analysis was used to control for confounders, and adjusted linear regression models were used to test for DMP. They identified 637 significantly differentially methylated probes associated with chronic postsurgical pain ($p < 0.05$). The DMPs were most commonly found in the promoter region (23.31%) and gene bodies (36.34%). The 637 DMPs were associated with 310 genes and which are associated with pathways such as GABA receptor signalling, protein kinase C signalling, dopamine receptor, and cyclic adenosine monophosphate-mediated signalling.

Other conditions

A number of other conditions have been investigated by single studies. A single study has investigated 68 children with juvenile idiopathic arthritis [26]. Patients were stratified based on relapsing or sustained remission following cessation of anti-tumour necrosis alpha treatment. The 450K array was performed on CD₄ T cells collected as anti-tumour necrosis alpha treatment was stopped. Weighted gene coexpression network analysis was used to identify differential methylation. This analysis examines correlated CpGs rather than individual CpGs and reports differentially methylated regions (DMR). They identified 5 DMRs between the two groups. The DMRs were associated with genes implicated in T cell activation, and this relationship was confirmed in subsequent gene expression and protein analysis experiments, with elevated markers of T cell activation in those who did not have sustained remission post-therapy cessation. Kuo et al. [33] studied 36 patients with Kawasaki disease and assessed the relationship between Fc fragment of IgG receptor IIa (*FCGR2A*) promoter methylation and response to treatment with intravenous immunoglobulin (IVIG), based on defervescence within 48 h. The sample source for methylation analysis was not reported. Mean methylation levels between responders and non-responders were compared using the chi-square test. They identified 5 CpG sites where increased methylation was associated with non-response to IVIG. At the 5 CpG sites, the difference in mean methylation level between the two groups ranged from 9.75 to 18.94%, and at all sites, the relationship was strong (p value < 0.0001). Yang et al. [34] studied 20 patients with infantile spasms and whether global DNA methylation in blood collected prior to treatment starting was associated with response to treatment with ACTH (based on change in seizure frequency). Mean methylation between the responders and non-responders was compared using the unpaired Student's t test. No relationship between mean DNA global methylation level and ACTH treatment effect was identified. In a group of 97 patients with Beckwith-Wiedemann syndrome, Gaston et al. [41] examined the relationship between *KCNQ1OT* and *H19* methylation and risk of tumour development (a known complication of Beckwith-Wiedemann syndrome). They used blood samples, and in some patients, tissue samples from affected organs were also obtained. Using the Kaplan-Meier method, abnormal *H19* methylation (defined as a methylation index > 60%) was strongly associated with an increased risk of tumour development (34.6% vs. 4.2%, hazard ratio 10, p < 0.0001).

Discussion

Several studies have investigated the use of DNA methylation as a biomarker to predict future health outcomes in children. In support of the notion that DNA

methylation is a viable biomarker, 22 of the 25 studies report a positive finding, although that may represent publication bias [44]. Unfortunately, due to issues with chosen outcomes, tissue-specific samples, accounting for sample cell type heterogeneity, lack of appropriate statistical testing, small effect sizes, limited validation, and no assessment of net health impact, the goal of DNA methylation-based biomarkers being used in clinical care for non-oncological paediatric diseases remains distant.

One of the most important aspects of any biomarker study is to ensure that the outcome of the study is robust and clinically relevant. A number of studies have used validated and widely accepted tools for assessing the outcome, such as the use of the Bayleys Scale of Infant Development for assessing neurocognitive development. Where possible, gold standard outcome measures should be used as they are more robust than other outcome measures such as parent-reported asthma symptoms, which are open to error [45]. This is of particular relevance for studies examining severity of disease where the study uses doctor prescription of a medication as a marker of severity (i.e. corticosteroids in prematurity or biological agents in IBD). Such endpoints are valid, but only if the indications for such treatments are clearly explained, in keeping with international prescribing practices and are true surrogates for disease outcomes. If not, any change seen may reflect the alternative practice of those involved in the study and may not be generalizable to a wider population.

Another important consideration is the tissue source for DNA methylation analysis. As a large proportion of methylation variation is tissue-specific [8], it is desirable (where possible) to analyse samples from the target tissue [46], for example, using intestinal biopsies to identify biomarkers in IBD [23] rather than using blood samples which were used in the majority of studies. It can be difficult to obtain samples from the tissue of interest in living humans, and particularly children as they are less tolerant of invasive testing. Potential solutions involve using non-invasive sampling methods such as induced sputum [47] or urine samples [48] which have successfully been used for DNA methylation analysis [49, 50]. Another potential method for analysing tissue-specific DNA methylation in a relatively non-invasive manner is by analysing cell-free DNA, which is made up of small fragments of DNA that circulate in the blood and are thought to originate from apoptotic/dying cells [51]. The original cell type of cell-free DNA can often be determined through the identification of DNA methylation signatures limited to a specific tissue type [52], and in adult diseases, these may act a biomarker to identify specific health conditions [53]. To date, this approach has largely been applied to early identification of primary (or relapsed) malignancy or antenatal screening for fetal

genetic abnormalities [54]. As yet, this approach has not been assessed in children for less pathogenic conditions, but represents a method to potentially sample the tissue of interest from a peripheral blood sample.

Where samples used for DNA methylation analysis involve multiple cell types (i.e. blood, buccal swabs), it is important that cell composition of the sample is accounted for in the analysis [55]. As each cell type will have a unique DNA methylation profile, any differences seen between cases and controls may equally reflect differences in cell composition as bona fide methylation changes associated with disease. A relative weakness of most studies included in this review is that the majority of them (15 of 25) do not account for issues arising from differential cell composition. Six of 25 [24, 29, 35, 36, 38, 39] did account for cell composition, whilst several [18, 23, 26–28] avoided this issue by using a sample that only included one cell type. A differing view is that alterations in cell composition are often an important feature of a specific disease/condition (i.e. inflammation will alter the cell composition of blood) and therefore should not be ‘adjusted for’ in any methylation analysis. Rather, they may provide key insights into the cellular disruption associated with disease pathogenesis.

The statistical analyses used to date limit the clinical applicability of the evidence that has been generated. For a test to be used in clinical practice, the performance should be assessed via statistical measures such as sensitivity, specificity, positive predictive value, negative predictive value, and area under a receiver operator curve. Only one [25] of the included studies used these statistical measures, whilst the majority of studies used statistical tests that describe associations. This may be because the authors are focusing on identifying associations which may inform disease mechanisms and therapeutic targets, but it would be ideal if statistical tests of predictive ability were also used as identification of clinically relevant biomarkers would be extremely useful. Another important statistical consideration is study power. The majority of EWAS studies conducted have had a small sample size, with no power calculation performed, and are likely underpowered [46]. However, recently, a publicly available tool for EWAS power calculation has been developed making it easier to design adequately powered studies, although it is only suitable for use with certain commonly collected specimens [56]. Given current genome-wide methylation assays measure methylation at 850,000 sites, another important statistical consideration is correction for multiple testing and there are many well-established methods for this.

A major limitation to the evidence generated in this field so far is the generally reported small disease-associated effect sizes of differential methylation between groups. These likely arise from a combination of the

limited sensitivity and specificity of current methods used to assess DNA methylation, relative to the gold standard whole-genome bisulphite sequencing [57] in combination with the relatively large disease-independent inter-individual variation in methylation at many genomic sites. Weak evidence of associations is unlikely to perform well when aligned against the criteria of a clinically relevant biomarker. In addition, these findings are less likely to be replicated in external cohorts. A lack of replication is another limitation of the current evidence. As mentioned, 22 papers had a significant finding; however, for the majority of these findings, there has been no attempt at replication. One area where there has been an attempt at replication is the role of *OPRM1* methylation in NAS, with two studies finding a significant association [20] and one study finding no association [21]. A number of EWAS have yielded positive results; however, these have not been replicated in either subsequent EWAS or targeted studies.

Despite the limitations of the evidence so far, there is still potential for useful DNA methylation biomarkers to be identified. Whilst most of the issues raised can be overcome by careful study design, the small observed effect sizes may be more difficult. One approach would be to incorporate DNA methylation data into combined predictive models. These models could pool data from multiple ‘omic’ inputs such as the genome (genetic mutations), methylome (DNA methylation), transcriptome (gene expression), and proteome (protein levels). An example of a program which is using the proposed multi-omic approach to better understand disease pathology and predict disease outcomes is the Integrative Human Microbiome Project. The second phase of this project is collecting biospecimens from patients to investigate the onset of inflammatory bowel disease and type 2 diabetes and is analysing host genome, methylome, transcriptome, and proteome as well as other factors such as the microbiome [58]. The omics data could also be integrated with existing diagnostics tests, clinical information, and also information regarding environmental exposures. In a study which examined prognostic biomarkers in multiple cancer types, omic-based biomarkers (including DNA methylation) outperformed existing clinical scores, and a combination of omic and clinical data had the best performance [59]. In another study examining adult patients, the combination of clinical information (via regular clinical assessment), environmental data (obtained using wearable sensors), and proteomics (via assessment of high sensitive C reactive protein) allowed earlier detection of inflammatory diseases such as Lyme disease and also earlier detection of insulin resistance [60]. The integration of the multiple data sources, however, is not without challenges [61]. In addition to demonstrating that such a biomarker

performed well from a statistical point of view, a positive net health impact would also need to be demonstrated [7]. However, if these hurdles are overcome, biomarkers encompassing multiple data sources, including DNA methylation, will represent a great advance in health care.

Conclusions

If the goals of precision health are to be recognised, biomarkers which predict future health outcomes will need to be identified. With the expansion in 'omics' technology, there are multiple tools which may aid discovery of such biomarkers. One such tool is epigenetics, in particular DNA methylation, and multiple studies have already undertaken research in this area. Whilst several positive findings have been made, there is a long way to go before these findings can be incorporated into clinical care. To ensure the validity of any future discoveries, researchers should ensure consideration is given to disease outcomes, tissue-specific specimens, adjustment for cell-type heterogeneity, appropriate statistical tests, replication of positive findings, and incorporating DNA methylation data into combined biomarkers. If these factors are considered, DNA methylation biomarkers have the potential to improve the healthcare delivered to children.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40348-020-00099-0>.

Additional file 1: Table 1. Included articles.

Abbreviations

DNA: Deoxyribonucleic acid; RNA: Ribonucleic acid; CpG: Cytosine-phosphate-guanine dinucleotide; UTR: Untranslated region; TSS: Transcriptional start site; EWAS: Epigenome-wide association study; DMP: Differentially methylated positions; OR: Odds ratio; ADRB2: $\beta 2$ adrenoceptor; WNV: Varicella; LDHC: Lactate dehydrogenase C; HTR1B: 5-Hydroxytryptamine receptor 1B; CBT: Cognitive behavioural therapy; SLC6A4: Serotonin transporter; RBP5: FK506 binding protein 5; GR: Glucocorticoid receptor; ADHD: Attention deficit hyperactivity disorder; DAT1: Dopamine transporter; DRD4: Dopamine receptor D4; NAS: Neonatal abstinence syndrome; OPRM1: Opioid receptor mu 1; ABCB1: ATP binding cassette subfamily B member 1; CYP2D6: Cytochrome P450 family 2 subfamily D member 6; IBD: Inflammatory bowel disease; FDR: False discovery rate; DMR: Differentially methylated regions; Fc γ R2A: Fc fragment of IgG receptor 1a; IMG: Intravenous immunoglobulin

Acknowledgements

The authors would like to thank Poh Chus for her invaluable assistance with developing and refining the search strategy.

Authors' contributions

All authors were involved in developing the concept for the review. SS designed and executed the literature search and summarised the published evidence. All authors were involved in the drafting and revision of the manuscript. The author(s) read and approved the final manuscript.

Funding

No funding for the writing of the review article was obtained.

Availability of data and materials

Not applicable (review article)

Ethics approval and consent to participate

Not applicable (review article)

Consent for publication

Not applicable (review article)

Competing interests

The authors declare that they have no competing interests.

Author details

¹Respiratory and Sleep Medicine, Royal Children's Hospital, Flemington Road, Parkville, Melbourne, Victoria 3052, Australia. ²Respiratory Diseases, Murdoch Children's Research Institute, Melbourne, Australia. ³Department of Paediatrics, The University of Melbourne, Melbourne, Australia. ⁴Epigenetics, Murdoch Children's Research Institute, Melbourne, Australia. ⁵Computational Biology, Peter MacCallum Cancer Centre, Melbourne, Australia.

Received: 15 April 2020 Accepted: 25 June 2020

Published online: 09 July 2020

References

- Jameson JL, Longo DL (2015) Precision medicine—personalized, problematic, and promising. *N Engl J Med* 372:2229–2234
- Ginsburg GS, Phillips KA (2018) Precision medicine: from science to value. *Health Affairs (Project Hope)* 37:694–701
- Hood L, Friend SH (2011) Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat Rev Clin Oncol* 8:184–187
- Collins FS, Varmus H (2015) A new initiative on precision medicine. *N Engl J Med* 372:793–795
- Foroutan B (2015) Personalized medicine: a review with regard to biomarkers. *J Bioequivalence Bioavailability* 7:244
- Robinson WH, Lindstrom TM, Cheung RK, Sokolove J (2013) Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat Rev Rheumatol* 9:267–276
- Pletcher MJ, Pignone M (2011) Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. *Circulation* 123:1116–1124
- Feinberg AP (2018) The key role of epigenetics in human disease prevention and mitigation. *N Engl J Med* 378:1323–1334
- Shutterstock/EpiGenetics. Epigenetic mechanisms. <https://www.shutterstock.com/image-vector/epigenetic-mechanisms-methylation-acetylation-dna-can-372050701>
- Holliday R, Pugh JE (1975) DNA modification mechanisms and gene activity during development. *Science (New York, NY)* 187:226–232
- Weber M, Hellmann I, Stadler MB, Ramos L, Pao S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39:457–466
- Gervin K, Hammer M, Alselsen HE, Moe R, Nygard H, Brandt I, Gjessing HK, Harris JR, Undlien DE, Lyle R (2011) Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res* 21:1813–1821
- Teh AL, Pan H, Chen L, Ong ML, Dogra S, Wong J, MacIsaac JL, Mah SM, McBreen LM, Saw SM, Godfrey KM, Chong YS, Kwek K, Kwok CK, Soh SE, Chong MF, Barton S, Kamani N, Cheong CY, Buschdorf JP, Stunkel W, Kobor MS, Meaney MJ, Gluckman PD, Holbrook JD (2014) The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res* 24:1064–1074
- Hunter DJ (2005) Gene-environment interactions in human diseases. *Nat Rev Genet* 6:287–298
- Locke WJ, Guanais D, Ma C, Liew YJ, Duesing KR, Fung KYC, Ross JP (2019) DNA methylation cancer biomarkers: translation to the clinic. *Front Genet* 10:1150
- Leygo C, Williams M, Jin HC, Chan MWY, Chu WK, Grusch M, Cheng YY (2017) DNA methylation as a noninvasive epigenetic biomarker for the detection of cancer. *Dis Markers* 2017:3726595
- Pan Y, Liu G, Zhou F, Su B, Li Y (2018) DNA methylation profiles in cancer diagnosis and therapeutics. *Clin Exp Med* 18:1–14
- Martino D, Neeland M, Dang T, Cobb J, Ellis J, Barnett A, Tang M, Vulliamin P, Allen K, Saffery R (2018) Epigenetic dysregulation of naive CD4+ T-cell activation genes in childhood food allergy. *Nat Commun* 9:3308

19. Wachman EM, Hayes MJ, Lester BM, Terrin N, Brown MS, Nielsen DA, Davis JM (2014) Epigenetic variation in the mu-opioid receptor gene in infants with neonatal abstinence syndrome. *J Pediatr* 166:472–478
20. Wachman EM, Hayes MJ, Shrestha H, Nikita FNU, Nolin A, Hoyo L, Daigle K, Jones HE, Nielsen DA (2018) Epigenetic variation in OPRM1 gene in opioid-exposed mother-infant dyads. *Genes Brain Behav* 17:e12476
21. McLaughlin P, Mactier H, Gillis C, Hickish T, Parker A, Liang WJ, Osselson MD (2017) Increased DNA methylation of ABCB1, CYP2D6, and OPRM1 genes in newborn infants of methadone-maintained opioid-dependent mothers. *J Pediatr* 190:180–184.e181
22. Kantake M, Yoshitake H, Ishikawa H, Anki Y, Shimizu T (2014) Postnatal epigenetic modification of glucocorticoid receptor gene in preterm infants: a prospective cohort study. *BMJ Open* 4:e005318
23. Howell KJ, Kracizy J, Nayak KM, Gasparetto M, Ross A, Lee C, Mak TN, Koo BK, Kumar N, Lawley T, Sinha A, Rosenstiel P, Heuschkel R, Stegle O, Zilbauer M (2018) DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. *Gastroenterology* 154:585–598
24. Somnineni HK, Venkateswaran S, Kilian V, Marigorta UM, Mo A, Okou DT, Kellermayer R, Mondal K, Cobb D, Walters TD (2019) Blood-derived DNA methylation signatures of Crohn's disease and severity of intestinal inflammation. *Gastroenterology* 156:2254–2265. e2253
25. Meng Y, Lin T, Liang S, Gao R, Jiang H, Shao W, Yang F, Zhou X (2018) Value of DNA methylation in predicting curve progression in patients with adolescent idiopathic scoliosis. *EBioMedicine* 36:489–496
26. Spreafico R, Rossetti M, Whitaker JW, Wang W, Lovell DJ, Albani S (2016) Epipolymorphisms associated with the clinical outcome of autoimmune arthritis affect CD4+ T cell activation pathways. *Proc Natl Acad Sci U S A* 113:13845–13850
27. Xiao C, Biagini Myers JM, Ji H, Metz K, Martin LJ, Lindsey M, He H, Powers R, Ullm A, Ruff B, Erickson MB, Somnineni HK, Simmons J, Strait RT, Kercsmar CM, Khurana Hershey GK (2015) Vanin-1 expression and methylation discriminate pediatric asthma corticosteroid treatment response. *J Allergy Clin Immunol* 136:923–931. e923
28. Zhang X, Biagini Myers JM, Yadagiri VK, Ullm A, Chen X, Weirauch MT, Khurana Hershey GK, Ji H (2017) Nasal DNA methylation differentiates corticosteroid treatment response in pediatric asthma: a pilot study. *PLoS One* 12:e0186150
29. Gasso P, Rodriguez N, Blazquez A, Monteagudo A, Boloc D, Piana MT, Lafuente A, Lazaro L, Amáiz JA, Mas S (2017) Epigenetic and genetic variants in the HTR1B gene and clinical improvement in children and adolescents treated with fluoxetine. *Prog Neuro-Psychopharmacol Biol Psychiatry* 75:28–34
30. Roberts S, Keers R, Lester KJ, Coleman JR, Breen G, Arendt K, Blatter-Meunier J, Cooper P, Creswell C, Fjermestad K, Havik OE, Heron C, Hogendoom SM, Hudson JL, Krause K, Lyneham HJ, Morris T, Nauta M, Rapee RM, Rey Y, Schneider S, Schneider SC, Silverman WK, Thastum M, Thirlwall K, Waite P, Eley TC, Wong CC (2015) Hpa axis related genes and response to psychological therapies: genetics and epigenetics. *Depress Anxiety* 32:861–870
31. Roberts S, Lester KJ, Hudson JL, Rapee RM, Creswell C, Cooper P, Thirlwall KJ, Coleman JR, Breen G, Wong CC, Eley TC (2014) Serotonin transporter [corrected] methylation and response to cognitive behaviour therapy in children with anxiety disorders. *Transl Psychiatry* 4:e444
32. Ding K, Yang J, Reynolds GP, Chen B, Shao J, Liu R, Qian Q, Liu H, Yang R, Wen J, Kang C (2017) DAT1 methylation is associated with methylphenidate response on oppositional and hyperactive-impulsive symptoms in children and adolescents with ADHD. *World J Biol Psychiatry* 18:291–299
33. Kuo HC, Hsu Y-W, Wu M-S, Woon PY, Wong HS-C, Tsai L-J, Lin R-K, Kahan S, Hsieh K-S, Chang W-C (2015) FCGR2A promoter methylation and risks for intravenous immunoglobulin treatment responses in Kawasaki disease. *Mediat Inflamm* 2015:155
34. Yang G, Wang J, Shi XY, Yang XF, Ju J, Liu YJ, Li ZF, Li YF, Zou LP (2015) Detection of global DNA hypomethylation of peripheral blood lymphocytes in patients with infantile spasms. *Epilepsy Res* 109:28–33
35. Tilley SK, Martin EM, Smeester L, Joseph RM, Kuban KCK, Heeren TC, Dammann OU, O'Shea TM, Fry RC (2018) Placental CpG methylation of infants born extremely preterm predicts cognitive impairment later in life. *PLoS One* 13:e0193271
36. Arpon A, Milagro FI, Laja A, Segura V, de Pisoni MS, Riezu-Boj JI, Alfredo MJ (2018) Methylation changes and pathways affected in preterm birth: a role for SLC6A3 in neurodevelopment. *Epigenomics* 10:91–103
37. Dhas BB, Antony HA, Bhat V, Newton B, Parja SC (2015) Global DNA methylation in neonatal sepsis. *Indian J Pediatr* 82:340–344
38. Stamatou A, Hansen CS, Sparso T, Madsen W, Olsen L, Bertalan M, Buil A, Bybjerg-Grauholm J, Bækvad-Hansen M, Hougaard DM, Mortensen PB, Pedersen CB, Nyegaard M, Wege T, Weinshenker S (2017) Differential DNA methylation at birth associated with mental disorder in individuals with 22q11.2 deletion syndrome. *Transl Psychiatry* 7:e1221
39. Chidambaram V, Zhang X, Geisler K, Stubbeman BL, Chen X, Weirauch MT, Meller J, Ji H (2019) Enrichment of genomic pathways based on differential DNA methylation associated with chronic postsurgical pain and anxiety in children: a prospective, pilot study. *J Pain*
40. Chidambaram V, Zhang X, Martin LJ, Ding L, Weirauch MT, Geisler K, Stubbeman BL, Sadhasivam S, Ji H (2017) DNA methylation at the mu-1 opioid receptor gene (OPRM1) promoter predicts preoperative, acute, and chronic postsurgical pain after spine fusion. *Pharmacogenom Personal Med* 10:157–168
41. Gaston V, Le Bouc Y, Soupre V, Burglen L, Donadieu J, Oro H, Audry G, Vazquez MP, Gicquel C (2001) Analysis of the methylation status of the KCNQ1OT and H19 genes in leukocyte DNA for the diagnosis and prognosis of Beckwith-Wiedemann syndrome. *Eur J Human Genetics* 9:409–418
42. Fu A, Leaderer BP, Gent JF, Leaderer D, Zhu Y (2012) An environmental epigenetic study of ADRB2 5'-UTR methylation and childhood asthma severity. *Clin Exp Allergy* 42:1575–1581
43. Yang IV, Richards A, Davidson EJ, Stevens AD, Kolakowski CA, Martin RJ, Schwartz DA (2017) The nasal methylome: a key to understanding allergic asthma. *Am J Respir Crit Care Med* 195:829–831
44. Dickstein K (1990) The existence of publication bias and risk factors for its occurrence. *Jama* 263:1385–1389
45. Aaron SD, Boulet LP, Reddel HK, Gershon AS (2018) Underdiagnosis and overdiagnosis of asthma. *Am J Respir Crit Care Med* 198:1012–1020
46. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealis JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, Milosavljevic A, Segmund KD, Bock C, Irfay RA (2013) Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* 10:949–955
47. Gibson P, Grootendorst D, Henry R, Pin I, Rytka P, Wark P, Wilson N, Djukanovic R (2002) Sputum induction in children. *Eur Respir J* 20:44 s–46s
48. May CW (2018) Urine collection methods in children: which is the best? *Nurs Clin North America* 53:137–143
49. Hwang SH, Kim KU, Kim JE, Kim HH, Lee MK, Lee CH, Lee SY, Oh T, An S (2011) Detection of HOXA9 gene methylation in tumor tissues and induced sputum samples from primary lung cancer patients. *Clin Chem Lab Med* 49: 699–704
50. van der Heijden AG, Mengual L, Ingelmo-Torres M, Lozano JJ, van Rijn-van de Westerlo CCM, Balasuri M, Geavlete B, Moldoveanu C, Ene C, Dinney CP, Czerniak B, Schalken JA, Klemenec L, Ribal MJ, Witjes JA, Alcaraz A (2018) Urine cell-based DNA methylation classifier for monitoring bladder cancer. *Clin Epigenetics* 10:71
51. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R, Rosenfeld N (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 17:223–238
52. Moss J, Magenheimer J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, Fu KY, Kiss E, Spalding KL, Landesberg G, Zick A, Grishpun A, Shapiro AMI, Grompe M, Wittenberg AD, Glaser B, Shemer R, Kaplan T, Dor Y (2018) Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* 9:5068
53. Feng H, Jin P, Wu H (2019) Disease prediction by cell-free DNA methylation. *Brief Bioinform* 20:585–597
54. Liang D, Cram DS, Tan H, Linpeng S, Liu Y, Sun H, Zhang Y, Tian F, Zhu H, Xu M, Wang H, Yu F, Wu L (2019) Clinical utility of noninvasive prenatal screening for expanded chromosome disease syndromes. *Genet Med* 21: 1998–2006
55. Houseman EA, Kim S, Kelsey KT, Wiencke JK (2015) DNA methylation in whole blood: uses and challenges. *Curr Environ Health Reports* 2:145–154
56. Graw S, Henn R, Thompson JA, Koester DC (2019) pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics* 20:218

57. Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. *Biology* 5:3. <https://www.mdpi.com/2079-7737/5/1/3/html>
58. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host & Microbe* 2014; 16: 276–289
59. Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, Landi MT, Ghosh D, Chatterjee N, Baladandayuthapani V, Zhao H (2017) Integrating clinical and multiple omics data for prognostic assessment across human cancers. *Sci Rep* 7:16954
60. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schussler-Fiorenza Rose SM, Perelman D, Colbert E, Runge R, Rego S, Sonecha R, Datta S, McLaughlin T, Snyder MP (2017) Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol* 15: e2001402
61. Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. *Nat Rev Genet* 19:299–310

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.3 SUPPLEMENTARY MATERIAL

The published manuscript contains supplementary material published, published and available online (<https://doi.org/10.1186/s40348-020-00099-0>). To view this material please see supplementary file: *Chapter 3 Supplementary Table 1_Included Articles.xlsx*.

3.4 RESEARCH IN CONTEXT

This project reviewed 25 studies across a range of paediatric diseases. Although a number of positive results were identified, there are currently no DNA methylation-based biomarkers used in clinical practice. The discussion highlighted the strengths and limitations of the existing studies as well as outlining optimal study design.

The findings of this review informed the experiments presented in this thesis. The themes are consistent with those of Chapter 2, including the need for robust and sensitive outcome measures of disease severity and the validation of positive findings. Additional considerations include the need for tissue-specific samples accounting for cell type heterogeneity in the sample. For paediatric CF, BAL is a method used in clinical care to obtain tissue-specific samples. However, BAL samples contain multiple cell types that must be accounted for in the analysis. The findings of this project informed the design of subsequent experiments presented in this thesis.

CHAPTER 4: Flow Cytometry Profiling of Paediatric Bronchoalveolar Lavage Samples

4.1 INTRODUCTION

Chapter 3 highlighted the importance of accounting for cell type heterogeneity in EWAS. This was also a recommendation from guidelines published on the optimal methods for EWAS.¹⁸⁷ This is relevant to the experiments presented in this thesis because the ideal tissue-specific specimen for the analysis of CF lung disease is BAL fluid, which contains multiple cell types. Previous studies using microscopy to characterise the cellular composition of BAL specimens have consistently reported alveolar macrophages, granulocytes (neutrophils and eosinophils), lymphocytes and epithelial cells as the predominant cell types.¹⁸⁸⁻¹⁹³ Thus, it is important to account for the composition of these cell types in the BAL samples involved in an EWAS.

The most established method used to account for cell type heterogeneity in an EWAS was described by Houseman et al.¹⁹⁴ and involves the deconvolution of sample composition using a reference set of DNA methylation signatures for each cell type in the sample. DNA methylation signatures for cells in commonly analysed samples such as blood are publicly available¹⁹⁴; however, no such panel exists for BAL samples. The first step in developing such a reference panel is to isolate each cell type of interest for DNA methylation analysis. The best method for the isolation of cell populations is fluorescence-activated cell sorting (FACS)¹⁸⁷; however, no robust methods for FACS of paediatric BAL samples exist. The purpose of this project was to develop a methodology for FACS of paediatric BAL samples. The methodology involved all required steps, from sample collection, processing and storage to flow cytometry protocols and analysis of results. The results of this project then informed the subsequent work of this thesis, permitting the development of a BAL-specific DNA methylation reference panel.

4.2 PUBLISHED MANUSCRIPT

The manuscript describing this work was published in the *American Journal of Respiratory Cell and Molecular Biology* (2020), vol. 63, issue 2, pages 152–159 (<https://doi.org/10.1165/rcmb.2019-0453MA>). The following is an exact electronic reprint of the manuscript as published.

MAJOR TECHNICAL ADVANCES

Single-Cell Flow Cytometry Profiling of BAL in Children

Shivanthan Shanthikumar^{1,2,3}, Matthew Burton⁴, Richard Saffery^{3,5}, Sarath C. Ranganathan^{1,2,3}, and Melanie R. Neeland^{3,6}

¹Respiratory and Sleep Medicine, Royal Children's Hospital, Melbourne, Victoria, Australia; ²Respiratory Diseases, ³Flow Cytometry and Imaging, ⁴Epigenetics, and ⁵Population Allergy, Murdoch Children's Research Institute, Melbourne, Victoria, Australia; and ⁶Department of Paediatrics, The University of Melbourne, Melbourne, Victoria, Australia

ORCID IDs: 0000-0001-6000-3180 (S.S.); 0000-0002-9510-4181 (R.S.); 0000-0002-6975-0307 (S.C.R.); 0000-0001-7301-9982 (M.R.N.).

Abstract

Childhood pulmonary diseases not only cause childhood morbidity and mortality but also can cause long-term pulmonary impairment. The clinical management of many childhood pulmonary diseases is hampered by a limited understanding of the underlying pathophysiological mechanisms. Flow cytometry, which can be used to phenotype individual cell populations or isolate cells for downstream analysis, represents a crucial technology that can help to elucidate the pathophysiology of these conditions. Here, we describe a flow cytometry-based method for purification and characterization of cell populations in BAL from children.

This includes assessment of the effect of cryopreservation on cell phenotype and frequency, a knowledge gap recently identified by an American Thoracic Society report on flow cytometry in lung samples. To our knowledge, this is the first study to simultaneously quantify alveolar macrophages, T cells (CD4 and CD8), B cells, natural killer cells, dendritic cells, granulocytes, and monocytes (CD16⁺/CD16⁻) in the BAL of children. The protocols described can be used to advance investigation of the pathophysiology of childhood pulmonary diseases.

Keywords: pediatrics; BAL; flow cytometry; lung diseases; cystic fibrosis

Rapid pulmonary growth and development occurs during the childhood years (1). Pulmonary diseases during this period not only cause morbidity and mortality at the time but also impair pulmonary function development, which has consequences for later life (1). It is vitally important to understand the pathophysiological mechanisms underlying childhood pulmonary diseases to facilitate optimal management of these conditions, not only alleviating childhood morbidity but also preserving long-term pulmonary function. There are many childhood pulmonary

diseases, such as preschool wheeze, asthma, suppurative lung diseases (cystic fibrosis [CF], primary ciliary dyskinesia, and protracted bacterial bronchitis), and childhood interstitial lung diseases, which remain poorly understood. A better understanding of the individual cell populations in the lung, how they behave, and how they interact would help to develop a greater understanding of the pathophysiology of these diseases and may also reveal therapeutic targets.

A relative barrier to understanding childhood pulmonary diseases is the

availability of tissue-specific samples. As compared with adults, children are less able to expectorate sputum (2) and less likely to have invasive procedures where lung tissue is biopsied or resected (3), and so obtaining samples from the lung can be difficult. BAL, which contains epithelial and immune cells from the lung and can be collected safely in children during clinical tests, is a potential method to obtain organ-specific samples in children (4).

Flow cytometry is a well-established technology that can be used to identify individual cells and characterize their

(Received in original form December 19, 2019; accepted in final form March 24, 2020)

Supported by a Vertex Innovation Award by Vertex Pharmaceuticals and a Melbourne Children's LifeCourse Postdoctoral Fellowship (M.R.N.). Vertex did not have any input into study design, study execution, analysis of results or manuscript development.

Author Contributions: S.S., R.S., S.C.R., and M.R.N. contributed to the concept and design of the study. S.S. and S.C.R. were responsible for patient recruitment and BAL collection. S.S. and M.R.N. were responsible for BAL processing. S.S., M.B., and M.R.N. executed the flow cytometry. S.S. and M.R.N. analyzed the data. All authors were involved in the drafting and editing of the manuscript.

Correspondence and requests for reprints should be addressed to Shivanthan Shanthikumar, M.B. B.S., B.Med.Sci., F.R.A.C.P., Clin.Dip.Pall.Med., Respiratory and Sleep Medicine, Royal Children's Hospital, 50 Flemington Road, Parkville, Victoria 3052, Australia. E-mail: shivanthan.shanthikumar@rch.org.au.

This article has a data supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Cell Mol Biol Vol 63, Iss 2, pp 152–159, Aug 2020

Copyright © 2020 by the American Thoracic Society

Originally Published in Press as DOI: 10.1165/rncmb.2019-0453MA on March 24, 2020

Internet address: www.atsjournals.org

function. In relation to pulmonary disease, it has been described as "the centerpiece of cellular phenotyping" (5). In addition to cellular phenotyping, flow cytometry can be used to isolate and purify individual cell populations for downstream analysis with other technologies (epigenomics, transcriptomics, proteomics, etc.). However, performing high-quality flow cytometry on tissue-specific samples is not without its challenges. A recent paper described a method for isolating airway macrophages from sputum from adult patients with CF; however, this method is not applicable to children, given they rarely expectorate (6). The American Thoracic Society (ATS) also recently published guidelines for improving the performance of flow cytometry in samples obtained from the lung (5). The report described current best practice but also highlighted challenges for reproducibility of high-quality flow cytometry and knowledge gaps such as the effect of cryopreservation on BAL samples.

In the present study, using samples from preschool-aged children with CF, we aimed to develop a flow cytometry protocol for phenotyping the cellular composition of pediatric BAL as well as addressing the above knowledge gap regarding the effect of cryopreservation on this sample type. Achieving such a goal will assist researchers to perform reproducible high-quality flow cytometry using childhood BAL samples, with the potential to improve understanding of the mechanisms of childhood pulmonary diseases with resulting improvements in clinical care.

Methods

Subjects and Clinical Data

All subjects with CF were cared for at the Royal Children's Hospital, Melbourne, Australia. Their families gave informed consent for their involvement in the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF) research program (HREC #25054), which includes collection of BAL samples and clinical data (7).

Bronchoscopy and BAL

As part of routine care, subjects underwent annual BAL around the time of their birthday from 1 to 6 years, inclusive. BAL was performed under general anesthesia.

Each BAL aliquot consisted of 1 ml/kg (maximum 20 ml) of normal saline being inserted via the working channel of the bronchoscope and then suctioned for return.

Preparation of Samples for Flow Cytometry

Detailed methodology on cryopreservation of BAL as well as preparation of samples for flow cytometry are provided in the supplementary material and depicted in Figure 1A. For all experiments, compensation was performed at the time of sample acquisition using compensation beads (BD Biosciences), and results were analyzed (manual gating and t-distributed stochastic neighbor embedding [tSNE] analysis) using FlowJo Version 10.6 software (FlowJo LLC). All experiments were performed using a BD FACSAria Fusion (configuration provided in Table E1 in the data supplement). Antibody selection and manual gating strategies were defined based on previous literature (8–11) and are in keeping with a recommendations from the recent ATS workshop report (5).

Standard Panel and Deep Profiling Panel

Two flow cytometry panels were developed. Preliminary gating strategies to select for live single cells for both panels are shown in Figure E1. We first developed and optimized a standard protocol to quantify and isolate the four major cell types previously identified in BAL from adults (alveolar macrophages, granulocytes, alveolar epithelial cells, and lymphocytes) (11). To do this, the following antibody panel was used: CD45 APC-Cy7, CD15 PE-CF594, CD206 BV421, and EpCAM BB515 (all from BD Biosciences) (Table E2), and propidium iodide (PI) was used for assessment of viability. An augmented panel was then developed for deeper immune profiling in a smaller cohort of children, including assessment of neutrophils, CD4 T cells, CD8 T cells, B cells, natural killer (NK) cells, monocytes, and dendritic cells (DCs). This was accomplished using a 13-parameter antibody panel comprising CD45 APC-Cy7, CD15 PE-CF594, CD206 BV421, CD3 BB515, CD19 BV605, CD4 A700, CD8 BV650, HLADR V500, CD14 BV786, CD11c PE-Cy7, CD16 BV395, and CD56 BV711 (all from BD Biosciences) (Table E3), and PI was used for assessment of viability. To ensure that PI did not

compromise the staining of other channels, during initial panel development we analyzed PI staining against all other parameters to inform the gating strategies presented here (Figure E2).

Data Analysis

Results are presented as box and whisker plots (where the box represents the 25th and 75th percentiles, the line in the box represents the median, and the whiskers extend to minimum and maximum values). All figures were generated using GraphPad Prism, version 6.01 (GraphPad Software Inc.) and FlowJo Version 10.6 (FlowJo LLC).

Results

Assessment of Major Cell Populations in Pediatric BAL

Detailed demographics and clinical characteristics of the study cohort are described in Table E4. Briefly, BAL samples from 14 subjects with a mean age of 38.6 months (range, 14–61 mo) were used to develop the standard panel. The mean volume inserted for lavage was 14.6 ml, with a mean return of 9.7 ml. The range in percentage of BAL return was 45% to 84% (Table E4). Average cell viability after thaw across all samples for this panel was 78% (individual sample viability is provided in Table E4).

Alveolar macrophages (CD45⁺CD206⁺), granulocytes (CD45⁺CD206⁺CD15⁺), lymphocytes (CD45⁺CD206⁺CD15⁺ low forward scatter/side scatter [SSC]), and alveolar epithelial cells (CD45⁺EpCAM⁺) were identified by flow cytometry using the manual gating strategy outlined in Figure 1B. The median (and range) cell proportions for each cell type (expressed as percentage of live cells) were 71% (32–94%), 6.1% (1–25%), 3.1% (1.5–17%), and 0.86% (0.2–7.4%), respectively (Figure 1C). The individual frequency of each of these populations from the 14 subjects is shown in Figure 1D. Using this panel, we were able to identify an average of 84% of all cells in pediatric BAL.

Purification of Cell Populations by Cell Sorting

We also sought to determine the feasibility of FACS in purifying cell populations from BAL for downstream analysis. With an average of 9.7 ml of BAL, we were consistently able to sort (mean and [range]) alveolar

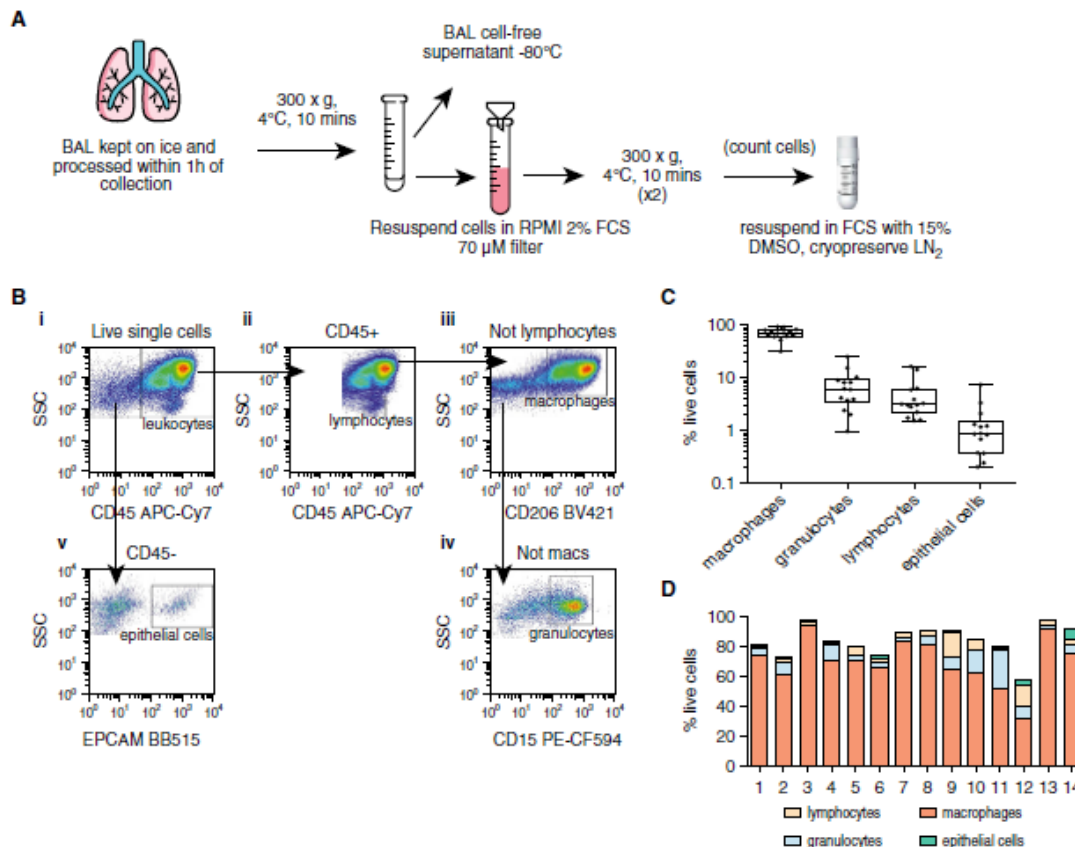


Figure 1. Single-cell flow cytometry phenotyping of major BAL cell populations in children with cystic fibrosis (CF). (A) Experimental workflow for collection and cryopreservation of BAL samples. (B) Representative flow cytometry manual gating strategy. (i) Within the single-cell live gate, CD45 is used to identify leukocytes from other nonimmune cell types. (ii) Within the CD45⁺ leukocyte gate, the CD45-positive, low side scatter (SSC) population represents lymphocytes, and the CD45-positive, high SSC population represents other leukocytes. (iii) A "not gate" removing the lymphocytes was performed, and macrophages were identified based on CD206⁺ expression. (iv) A "not gate" removing the macrophages was performed, and granulocytes were identified based on CD15⁺ expression. (v) CD45⁻ EpCAM⁺ SSC high cells are alveolar epithelial cells. (C) Cell composition of BAL in $n=14$ children with CF expressed as percentage of live cells. (D) Stacked bar graph representing cell frequency in each individual. EpCAM BB515 = epithelial cell adhesion molecule; LN₂ = liquid nitrogen; macs = macrophages.

macrophages (7.1×10^5 cells [4.6×10^4 – 2.1×10^6]), granulocytes (8.7×10^4 cells [5.9×10^3 – 1.8×10^5]), lymphocytes (6.6×10^4 cells [2.9×10^3 – 1.9×10^5]), and airway epithelial cells (1.1×10^4 cells [2.2×10^2 – 3.8×10^4]). The cell numbers purified per ml of BAL for each patient are provided in Table E5.

Assessment of Rare Immune Cells by Deep Immune Profiling

Given the success of the standard flow cytometry panel, we next developed an

augmented panel to identify rare immune cell populations, most of which have not been previously investigated simultaneously in pediatric BAL. Cryopreserved BAL samples from seven subjects were used (demographics provided in Table E6). Average cell viability after thaw across all samples for this panel was 74% (individual sample viability is provided in Table E6). The gating strategy for the deep immune profiling panel is shown in Figure 2A. Total alveolar macrophages and total granulocytes were identified as described

above. Within the CD15⁺ total granulocyte population, cells were subtyped into CD16⁺ or CD16⁻ populations, where CD15⁺CD16⁺ cells have been previously identified as neutrophils (12). CD206⁻CD15⁺ cells were then identified as CD3⁺ lymphocytes or HLADR⁺ cells. CD3⁺ lymphocytes were subtyped as CD4 or CD8 positive. Within the HLADR⁺ cell fraction, CD19⁺ B cells and CD14⁺ monocytes were identified. CD14⁺ monocytes were also investigated for expression of CD16. Total DCs were

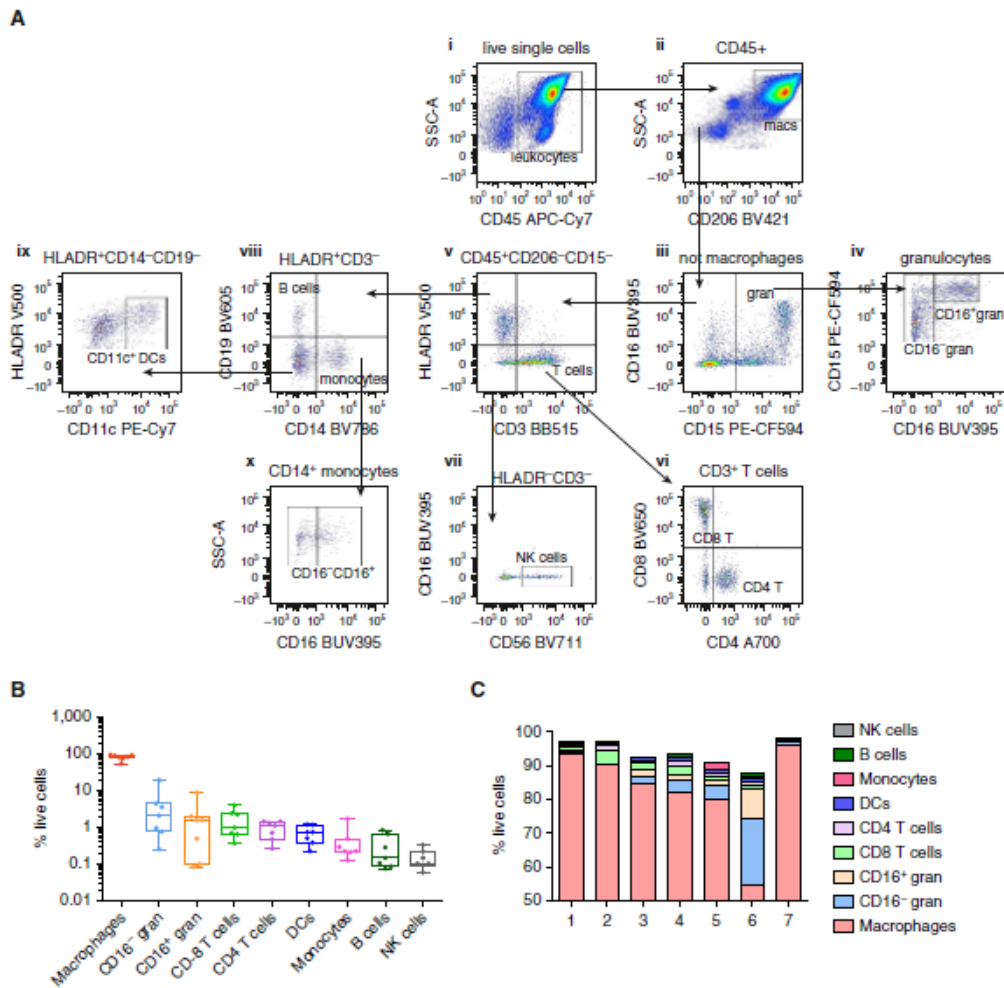


Figure 2. Deep immune-cell phenotyping of BAL in children with CF. (A) Representative flow cytometry manual gating strategy of live single cells. (i) CD45⁺ cells are defined as leukocytes. (ii) Within the CD45⁺ leukocyte population, CD206 and SSC^{high} cells are alveolar macrophages. (iii) A "not macrophage" gate was performed, and CD15 and CD16 were used to identify granulocytes (CD15⁺) among the CD45⁺CD206⁻ (not macrophage) cell fraction. (iv) Within the CD15⁺ granulocyte gate, cells positive for CD16 (neutrophils) were identified. (v) Within the CD45⁺CD206⁻CD15⁻ population, CD3 was used to identify total T cells, and (vi) CD4 and CD8 then used to identify the respective T cell subpopulations. (vii) CD45⁺CD206⁻CD15⁻CD3⁻HLADR⁻ cells were selected for expression of the natural killer (NK) cell marker CD56. (viii) CD45⁺CD206⁻CD15⁻CD3⁻HLADR⁺ cells were analyzed for expression of CD19 (B cells) and CD14 (monocytes). (ix) CD11c positive cells within the CD19⁻CD14⁻ population were identified as CD11c⁺ dendritic cells (DCs). HLADR⁺CD11c⁻ cells within this gate were not defined further. (x) Within the monocyte population, CD16 expression was also analyzed. (B) Cell composition of BAL expressed as percentage of live cells in $n = 7$ children with CF. (C) Stacked bar graph representing cell frequency for each individual.

identified within the HLADR⁺ population and were negative for CD19 and CD14 and expressed CD11c. CD56⁺ NK cells were identified within the CD206⁻CD15⁻CD3⁻HLADR⁻ cell

population. The median (and range) cell proportions (as percentage of live cells in BAL) for each cell type was: alveolar macrophages (85% [55–96%]), CD16⁻ granulocytes (2.1% [0.26–19%]), CD16⁺

granulocytes (1.6% [0.09–9%]), CD8 T cells (1% [0.37–4%]), CD4 T cells (1.1% [0.27–1.5%]), DCs (0.74% [0.23–1.2%]), monocytes (0.23% [0.13–1.7%]), B cells (0.15% [0.076–0.85%]), and NK cells (0.1%

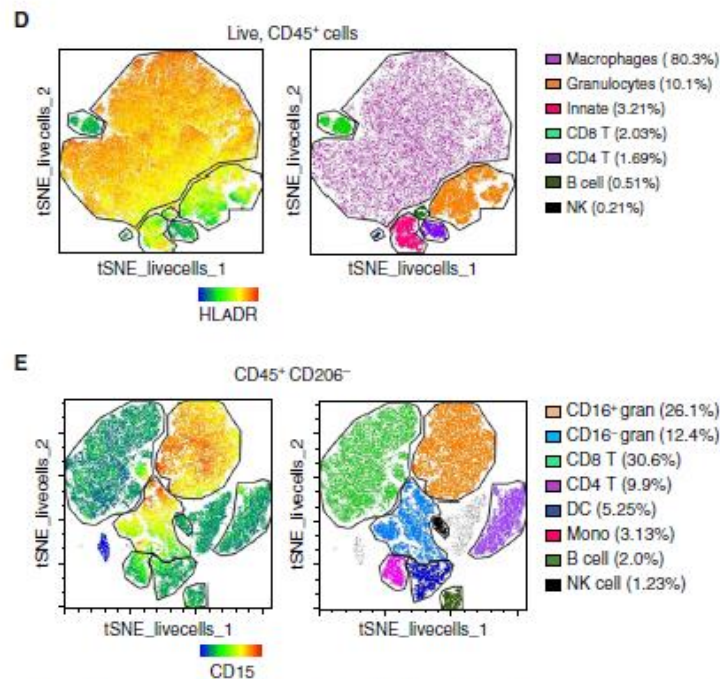


Figure 2. (Continued). (D) t-distributed stochastic neighbor embedding (tSNE) representations of live CD45⁺ leukocytes (70,000 randomly selected cells [10,000 cells per patient]), colored by intensity of HLADR expression. The generated tSNE clusters were classified into the following cell types on the basis of expression of lineage markers: macrophages (light purple), granulocytes (orange), CD8 T cells (light green), CD4 T cells (dark purple), B cells (dark green), NK cells (black), and other CD11c-expressing innate cells (pink). (E) tSNE representations of CD45⁺CD206⁻ leukocytes excluding macrophages (70,000 randomly selected cells [10,000 cells per patient]). tSNE clusters were classified into the following cell types on the basis of expression of lineage markers: CD16⁺ granulocytes (orange), CD16⁻ granulocytes (light blue), CD8 T cells (light green), CD4 T cells (purple), DCs (blue), monocytes (pink), B cells (dark green), NK cells (black), and undefined (gray). A = area; gran = granulocytes; HLADR = human leukocyte antigen DR isotype; mono = monocytes.

[0.06–0.34%]) (Figure 2B). Within the monocyte population, an average of 39.4% expressed CD16. The individual frequency of each of these populations from the seven children is shown in Figure 2C. Although our standard panel was able to characterize an average of 84% of all live cells in BAL, this deep immune profiling panel allowed characterization of an average of 94% of all live cells in BAL.

To verify our manual gating approach of the deep profiling panel and to visualize these data in two dimensions, we ran a tSNE analysis, as shown in Figures 2D and 2E. We first performed this analysis on 70,000 randomly selected (10,000 cells per patient) live CD45⁺ cells (all leukocytes) using CD206, CD15, CD16, CD14, CD3, CD4, CD8, CD19, CD56, HLADR, and

CD11c for clustering. On the basis of the expression of these lineage markers, the generated tSNE clusters were classified into seven cell types with the following frequencies (percentage CD45⁺ cells): macrophages (80.3%), granulocytes (10.5%), CD8 T cells (2.03%), CD4 T cells (1.67%), B cells (0.51%), NK cells (0.21%), and CD11c-expressing innate cell cluster (total 3.21%) (Figure 2D). To improve visualization and investigate heterogeneity of the rare cell populations, we also performed a tSNE analysis on 70,000 randomly selected (10,000 cells per patient) live CD45⁺CD206⁻ cells (all leukocytes excluding macrophages). On the basis of the expression of the above lineage markers, the tSNE islands were classified into eight cell subsets with

the following frequencies (percentage CD45⁺CD206⁻ cells): CD16⁺ granulocytes (26.1%), CD16⁻ granulocytes (12.4%), CD8 T cells (30.6%), CD4 T cells (9.9%), DCs (5.25%), monocytes (3.13%), B cells (2.0%), and NK cells (1.23%) (Figure 2E). Importantly, the phenotypes and frequencies identified by unsupervised analyses were comparable to those obtained by manual gating.

The Influence of Cryopreservation on Immune Cell Frequency and Phenotype in BAL

To determine if the cryopreservation and thaw procedure influences the phenotype or recovery of immune cells within BAL, fresh BAL samples from three children with CF were collected and divided so that half of the

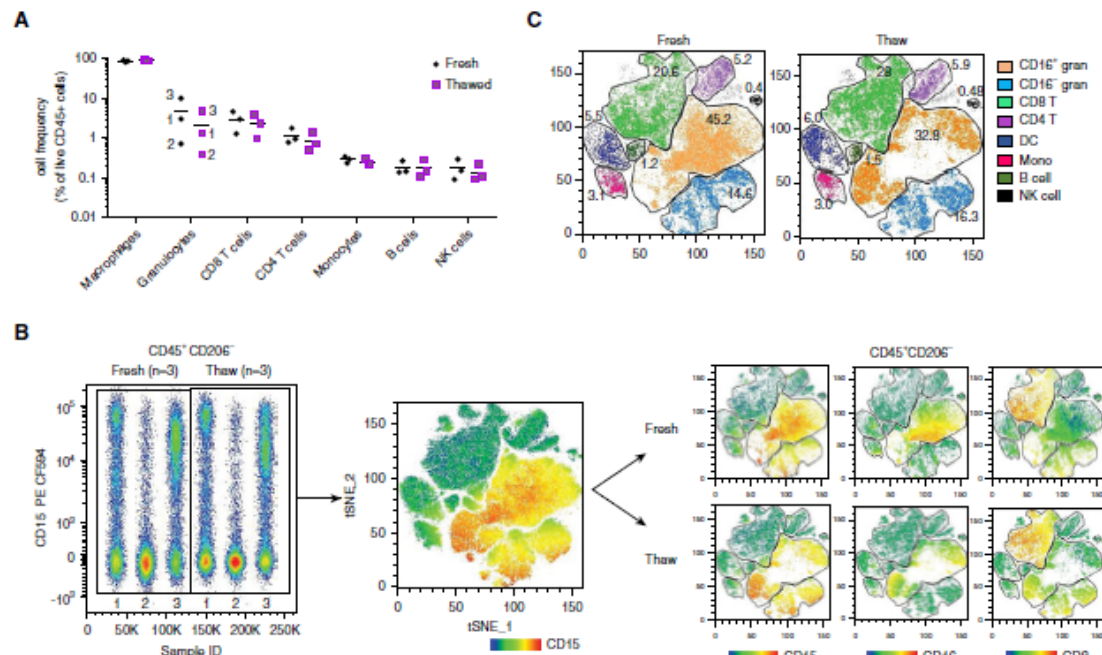


Figure 3. Comparison of cell proportion and phenotype in fresh and thawed BAL samples from $n=3$ children with CF. (A) Mean cell frequency of each cell type when processed fresh (black) and after cryopreservation and thaw (purple). (B) A total of 10,000 randomly selected CD45⁺CD206⁻ cells from each sample were concatenated into a single file containing 60,000 cells for tSNE analysis. The figure depicts intensity of CD15, CD16, and CD8 expression in the nonmacrophage leukocyte population between fresh and thawed samples. (C) tSNE clusters were classified into the following cell types on the basis of expression of lineage markers: CD16⁺ granulocytes (orange), CD16⁺ granulocytes (light blue), CD8 T cells (light green), CD4 T cells (purple), DCs (blue), monocytes (pink), B cells (dark green), and NK cells (black). The figure depicts a comparison of unsupervised cell frequency between fresh and thawed samples in a single representative patient (Sample ID 3).

sample could proceed to flow cytometry on the day of collection (fresh), and half was cryopreserved in liquid nitrogen before flow cytometry analysis (thawed). The deep immune profiling panel was used. Average cell viability for fresh and thawed samples in this data set was 71% and 74%, respectively. The frequency of the immune cell populations (macrophages, granulocytes, CD8 T cells, CD4 T cells, monocytes, B cells, and NK cells) were compared between fresh and thawed samples by manual gating analysis. Mean percentage of live cells between fresh and thawed, respectively, were the following: alveolar macrophages (88 vs. 92), granulocytes (4.5 vs. 2.2), CD8 T cells (3.0 vs. 2.6), CD4 T cells (1.2 vs. 0.9), monocytes (0.29 vs. 0.25), B cells (0.19 vs. 0.18), and NK cells (0.18 vs. 0.14) (Figure 3A). The main effect of cryopreservation was observed within the granulocyte fraction, with the most significant reduction observed in the

CD15⁺CD16⁺ neutrophil population. The alteration in expression of CD16 on CD15⁺ cells after cryopreservation is depicted in the corresponding tSNE analysis performed on nonmacrophage leukocyte populations from these fresh and thawed samples (60,000 randomly selected CD45⁺CD206⁻ cells [10,000 cells from each sample]) (Figure 3B). Figure 3C shows a direct comparison in the spatial pattern and cluster frequencies between fresh and thawed samples in a representative patient, color coded by the assigned cell population.

Discussion

This study outlines the successful application of multiparameter flow cytometry to extensively phenotype, quantify, and purify cell populations from BAL of children <5 years of age with CF. We developed both a standard flow

cytometry panel to quantify and purify the four major cell populations and an augmented panel for deeper phenotyping of immune cells in BAL. This work can serve as a protocol for reproducible, high-quality flow cytometry analysis of pediatric BAL for investigation of childhood pulmonary diseases.

The standard flow cytometry panel revealed a cellular composition of CF BAL in keeping with previous microscopy-based studies (13–15), which are generally considered less accurate than flow cytometry (11, 16). Previous studies have reported that macrophages are the predominant cell type in BAL, at 40% to 63% of the overall cell count (14, 15, 17, 18), whereas granulocytes make up 18% to 53% of the total cell count (13–15, 17–20). The lymphocyte proportion has previously been shown to be 6% to 8% (14, 15, 17), which is similar to the frequency we identified in this study. Likewise, epithelial cells have previously

been shown to be 2% of the overall cell count, comparable to our findings (17).

Deeper profiling allowed the identification of rare immune cell populations in pediatric CF BAL. To our knowledge, this is the first study to simultaneously quantify T cells (CD4 and CD8), B cells, NK cells, DCs, granulocytes, and monocytes (CD16⁺/CD16⁻) in the BAL of children. Previous use of flow cytometry on BAL from children has been limited to measurement of mean fluorescence intensity of surface markers of interest (21, 22) or phenotyping of a few cell populations (23–25), rather than deep profiling of cell composition. Other studies, however, have profiled cell composition of BAL in adults. Yu and colleagues (8) obtained BAL from healthy adult volunteers and developed a protocol that defined the following myeloid cell types: alveolar macrophages, neutrophils, mast cells, eosinophils, basophils, DCs, and monocytes (CD16⁺/CD16⁻). Cells of a lymphoid lineage were not defined further. Myeloid cells made up 81% of all immune cells, and alveolar macrophages made up 97% of all myeloid cells, meaning that all other myeloid cells isolated occurred very rarely (<1% of BAL immune cells). Freeman and colleagues (26) analyzed BAL from a cohort of adult subjects with chronic obstructive pulmonary disease, identifying varying proportion of alveolar macrophages (66.1%), neutrophils (10.4%), CD8 T cells (8.1%), CD4 T cells (1.6%), classical monocytes (CD14⁺CD16⁻; 3%), nonclassical monocytes (CD14^{dim} and CD16⁺; 0.9%), intermediate monocytes (CD14⁺CD16⁺; 0.9%), eosinophils (0.82%), and DCs (0.3%). The relatively higher proportion of lymphocytes, and therefore lymphocyte subsets, identified in the Freeman and colleagues study when compared with our data, may relate to the different age (child vs. adult) or disease (CF vs. chronic obstructive pulmonary disease) of the subjects involved (26).

To complement and validate our traditional manual gating approach, we used a dimensionality reduction technique known as tSNE, which visualizes high-dimensional cytometry data by giving each data point a location on a two-dimensional map (27). We were able to show that the BAL immune cell populations identified by discrete marker expression in conventional analysis were also well separated by tSNE dimensionality reduction. Traditional

manual gating can suffer from individual user bias when defining cell populations and quickly becomes impractical as the number of parameters increases (28). Computational approaches, such as tSNE, provide a powerful tool to explore flow cytometry data and, as described in our study, can complement traditional manual gating approaches, particularly for less-described sample types, such as pediatric BAL.

In addition to improved accuracy and high-throughput clinical application, the major advantage of flow cytometry as compared with other methods of determining cell composition is the ability to isolate individual cell populations for downstream analyses with novel “omics” technologies, such as epigenomics or transcriptomics. Many of these technologies require tissue-specific samples, and BAL allows sampling of the immune and epithelial cells in the lungs. The yield of purified cell populations derived from the clinically relevant volumes of BAL used in this study should be amenable to downstream analysis. This is especially true for new technologies such as single-cell sequencing, where samples can be run with as few as 10³ to 10⁴ cells (29).

A recent ATS report on the use of flow cytometry on lung-derived samples stated the effect of cryopreservation on BAL samples has not been explicitly studied (5); however, we performed this direct comparison here. We note that a significant proportion of the CD15⁺CD16⁺ granulocyte population was affected by the cryopreservation and thaw process; however, the CD15⁺CD16⁻ granulocyte population was recoverable. This was somewhat expected, as granulocyte cryopreservation is historically difficult and inconsistent (30, 31), and this should be considered when adopting our protocol. However, the expression patterns and cell frequencies of all other leukocyte populations (macrophages, T cells, B cells, and innate cells) and their subsets (CD4 T cells, CD8 T cells, monocytes, NK cells, and DCs) were unaffected by this procedure. In addition, given BAL samples are obtained in a clinical setting with associated unpredictable timing of collection, it is often unrealistic to expect samples will always be able to proceed directly to flow cytometry analysis. As such, the cryopreservation method is often required. An alternate approach would be paraformaldehyde fixing and staining the

cells (24); however, this precludes purification of live cell populations and is associated with other issues, such as dye degradation (5).

A practical consideration when performing flow cytometry on BAL samples is the intrinsically and highly autofluorescent macrophages, which comprise ~80% of all BAL cells. These cells can pose problems during analysis, as they spill into many channels. The live/dead gating strategy used in our study was confirmed to not exclude highly autofluorescent macrophages by careful examination of the differences in spectral signature of our viability dye and macrophage autofluorescence across all instrument channels. The addition of the macrophage-specific marker CD206 to all panels permits further identification of these macrophages, so as to not affect the resolution of other cell types during downstream gating.

A limitation of our deep profiling panel is that it does not include surface markers for unconventional lymphocyte populations, such as $\gamma\delta$ T cells, mucosal associated invariant T cells, and innate lymphoid cells, which have been shown to be present in airway samples and may likely make up the remaining 6% of cells undefined in our study (32, 33). Our panel could be modified by the addition of antibodies that allow further discrimination of these cell types. It should also be considered when interpreting our findings that all samples in this study were obtained from children with CF, as these are the samples that are clinically collected and available at our center. However, the protocols developed are based on methods that have been successfully used in other diseases in adults and will be amenable for use in other childhood pulmonary diseases (8–11).

The protocols developed in this study could be used to by all researchers who want to perform cellular phenotyping in BAL samples from children and can also be adopted for purification of cell populations for downstream analysis. This will, we hope, lead to the development and use of high-quality flow cytometry to improve our understanding of childhood pulmonary diseases. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Nadeene Clarke for organization and assistance with BAL sample collection.

References

- Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med* 2013;1: 728–742.
- Ronchetti K, Tame JD, Palsey C, Thia LP, Doull I, Howe R, et al. The CF-Sputum Induction Trial (CF-SpIT) to assess lower airway bacterial sampling in young children with cystic fibrosis: a prospective internally controlled interventional trial. *Lancet Respir Med* 2018;6:461–471.
- Shanthikumar S, Massie J, Ranganathan S, Schuhmann M, Eberhardt R, Irving LB, et al. Utility of endobronchial ultrasound in assessment of intrathoracic lesions in paediatric patients. *Respiration* 2019;98: 340–346.
- Faro A, Wood RE, Schechter MS, Leong AB, Wittkugel E, Abode K, et al.; American Thoracic Society Ad Hoc Committee on Flexible Airway Endoscopy in Children. Official American Thoracic Society technical standards: flexible airway endoscopy in children. *Am J Respir Crit Care Med* 2015;191:1066–1080.
- Tighe RM, Redente EF, Yu Y-R, Herold S, Sperling AI, Curtis JL, et al. Improving the quality and reproducibility of flow cytometry in the lung: an official American Thoracic Society workshop report. *Am J Respir Cell Mol Biol* 2019;61:150–161.
- Hisert KB, Liles WC, Manicone AM. A flow cytometric method for isolating cystic fibrosis airway macrophages from expectorated sputum. *Am J Respir Cell Mol Biol* 2019;61:42–50.
- Sly PD, Gangell CL, Chen L, Ware RS, Ranganathan S, Mott LS, et al.; AREST CF Investigators. Risk factors for bronchiectasis in children with cystic fibrosis. *N Engl J Med* 2013;368:1963–1970.
- Yu YR, Hotten DF, Malakhau Y, Volker E, Ghio AJ, Noble PW, et al. Flow cytometric analysis of myeloid cells in human blood, bronchoalveolar lavage, and lung tissues. *Am J Respir Cell Mol Biol* 2016;54:13–24.
- Barry SM, Condez A, Johnson MA, Janossy G. Determination of bronchoalveolar lavage leukocyte populations by flow cytometry in patients investigated for respiratory disease. *Cytometry* 2002;50: 291–297.
- Maestre-Battle D, Pena OM, Hirota JA, Gunawan E, Rider CF, Sutherland D, et al. Novel flow cytometry approach to identify bronchial epithelial cells from healthy human airways. *Sci Rep* 2017; 7:42214.
- Hodge SJ, Hodge GL, Holmes M, Reynolds PN. Flow cytometric characterization of cell populations in bronchoalveolar lavage and bronchial brushings from patients with chronic obstructive pulmonary disease. *Cytometry B Clin Cytom* 2004;61:27–34.
- Gustafson MP, Lin Y, Maas ML, Van Keulen VP, Johnston PB, Peikert T, et al. A method for identification and analysis of non-overlapping myeloid immunophenotypes in humans. *PLoS One* 2015;10: e0121546.
- Peterson-Carmichael SL, Harris WT, Goel R, Noah TL, Johnson R, Leigh MW, et al. Association of lower airway inflammation with physiologic findings in young children with cystic fibrosis. *Pediatr Pulmonol* 2009;44:503–511.
- Marguet C, Jouen-Boedes F, Dean TP, Warner JO. Bronchoalveolar cell profiles in children with asthma, infantile wheeze, chronic cough, or cystic fibrosis. *Am J Respir Crit Care Med* 1999;159: 1533–1540.
- Ratjen F, Rietschel E, Griese M, Ballmann M, Kleinau I, Döring G, et al. Fractional analysis of bronchoalveolar lavage fluid cytology in cystic fibrosis patients with normal lung function: bronchoalveolar lavage for the evaluation of anti-inflammatory treatment (BEAT) study group. *Eur Respir J* 2000;15:141–145.
- Tricas L, Echeverría A, Blanco MA, Menéndez M, Belda J. Flow cytometry counting of bronchoalveolar lavage leukocytes with a new profile of monoclonal antibodies combination. *Cytometry B Clin Cytom* 2012;82:61–66.
- Reinhardt N, Chen CI, Loppow D, Schink T, Kleinau I, Jöres RA, et al. Cellular profiles of induced sputum in children with stable cystic fibrosis: comparison with BAL. *Eur Respir J* 2003;22: 497–502.
- Brennan S, Sly PD, Gangell CL, Sturges N, Winfield K, Wikstrom M, et al.; AREST CF. Alveolar macrophages and CC chemokines are increased in children with cystic fibrosis. *Eur Respir J* 2009;34:655–661.
- Armstrong DS, Grimwood K, Carlin JB, Carzino R, Gutiérrez JP, Hull J, et al. Lower airway inflammation in infants and young children with cystic fibrosis. *Am J Respir Crit Care Med* 1997;156: 1197–1204.
- Linnane BM, Hall GL, Nolan G, Brennan S, Stick SM, Sly PD, et al.; AREST-CF. Lung function in infants with cystic fibrosis diagnosed by newborn screening. *Am J Respir Crit Care Med* 2008;178: 1238–1244.
- Alexis NE, Muhlebach MS, Peden DB, Noah TL. Attenuation of host defense function of lung phagocytes in young cystic fibrosis patients. *J Cyst Fibros* 2006;5:17–25.
- Marcos V, Latzin P, Hector A, Sonanini S, Hoffmann F, Lacher M, et al. Expression, regulation and clinical significance of soluble and membrane CD14 receptors in pediatric inflammatory lung diseases. *Respir Res* 2010;11:32.
- Hodge G, Upham JW, Chang AB, Baines KJ, Yerkovich ST, Pizzutto SJ, et al. Increased peripheral blood pro-inflammatory/cytotoxic lymphocytes in children with bronchiectasis. *PLoS One* 2015;10: e0133695.
- Garratt LW, Wright AK, Ranganathan SC, Grigg J, Sly PD; behalf of AREST CF. Small macrophages are present in early childhood respiratory disease. *J Cyst Fibros* 2012;11:201–208.
- Hartl D, Koller B, Mehlhorn AT, Reinhardt D, Nicolai T, Schendel DJ, et al. Quantitative and functional impairment of pulmonary CD4+CD25hi regulatory T cells in pediatric asthma. *J Allergy Clin Immunol* 2007;119:1258–1266.
- Freeman CM, Crutlington S, Stolberg VR, Brown JP, Sonstein J, Alexis NE, et al. Design of a multi-center immunophenotyping analysis of peripheral blood, sputum and bronchoalveolar lavage fluid in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *J Transl Med* 2015;13:19.
- Laurens van der Maaten GH. Visualizing data using t-sne. *J Mach Learn Res* 2008;9:2579–2605.
- Chester C, Maecker HT. Algorithmic tools for mining high-dimensional cytometry data. *J Immunol* 2015;195:773–779.
- Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017;9:75.
- Bank H. Granulocyte preservation circa 1980. *Cryobiology* 1980;17: 187–197.
- Sputtek A, Kühn P, Rowe AW. Cryopreservation of erythrocytes, thrombocytes, and lymphocytes. *Transfus Med Hemother* 2007;34: 262–267.
- Swiebeda D, Guo Y, Sagawa S, Thwaites RS, Nadel S, Openshaw PJM, et al. Omip-062: a 14-color, 16-antibody panel for immunophenotyping human innate lymphoid, myeloid and t cells in small volumes of whole blood and pediatric airway samples. *Cytometry A* 2019;95: 1231–1235.
- Glanville N, Message SD, Walton RP, Pearson RM, Parker HL, Laza-Stanca V, et al. $\gamma\delta$ T cells suppress inflammation and disease during rhinovirus-induced asthma exacerbations. *Mucosal Immunol* 2013;6: 1091–1100.

4.3 SUPPLEMENTARY MATERIAL

The published manuscript contains supplementary material published as an online supplement. The following is the supplementary data as published and available online (<https://doi.org/10.1165/rcmb.2019-0453MA>).

Single Cell Flow Cytometry Profiling of Bronchoalveolar Lavage in Children – Supplementary Material

1. Cryopreservation Protocol
2. Preparation of Samples for Flow Cytometry
3. Supplementary Table 1. BD FACSAria-Fusion configuration.
4. Supplementary Table 2. Antibody cocktail used to phenotype BAL cells in the standard panel
5. Supplementary Table 3. Antibody cocktail used to phenotype BAL cells in augmented panel
6. Supplementary Figure 1. Preliminary gating strategies for both panels
7. Supplementary Figure 2. PI staining against all parameters
8. Supplementary Table 4. Patient Demographics for the standard panel
9. Supplementary Table 5. Fluorescence-activated cell sorting characteristics
10. Supplementary Table 6. Patient Demographics for the deep immune profiling panel

Cryopreservation Protocol

After collection, BAL samples were kept on ice for processing within 1 hour. Samples were centrifuged at 300 x g for 10 mins at 4°C. Cell-free BAL supernatant was then collected and stored at -80°C. The cell pellet was resuspended in 10mL of media (RPMI supplemented with 2% fetal calf serum (FCS)), filtered through a 70µm filter and centrifuged at 300 x g for 10mins at 4°C. Supernatant was discarded and the cell pellet resuspended in media of for cell counting. The sample was then centrifuged at 300 x g for 10mins and resuspended in equal parts media and freezing solution (FCS with 15% dimethyl sulfoxide (DMSO)) at $5-10 \times 10^6$ cells per mL. The freeze mix was added to resuspended cells drop by drop on ice. The samples were transferred to cryovials, and then cooled at -1°C per minute to -80 °C overnight before being transferred to liquid nitrogen for storage.

Preparation of Samples for Flow Cytometry

Cryopreserved BAL cells were thawed in 10mL media (RPMI supplemented with 10% heat-inactivated FCS) with 25U/mL benzonase at 37°C. BAL cells were centrifuged at 300xg for 10 minutes and washed twice in media before viability count. Cell pellets were washed once in FACS buffer (2% FCS, 2mM EDTA in PBS) before being resuspended in 50µL FACS buffer containing human Fc-block for antibody staining. 50µL of antibody cocktails made up at 2X were added 1:1 to the resuspended cells and incubated on ice for 30 minutes. Cells were washed and resuspended in 300µL FACS buffer for acquisition. Propidium iodide (PI) was added as a viability marker prior to sample acquisition. Sorted cell populations were resuspended in RLT (Qiagen) and β-mercaptoethanol (β-ME) and stored at -80°C for future analysis. For all experiments, compensation was performed at the time of sample acquisition using compensation beads (BD Biosciences) and results analysed using FlowJo Version 10.6 software (FlowJO LLC).

Supplementary Table 1. BD FACSAria-Fusion configuration.

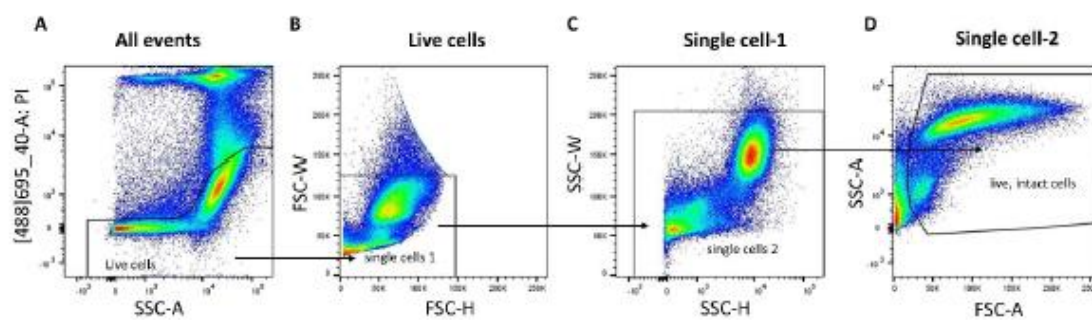
Detector	Laser	Dichroic	Filter	Detector Name
C	488		488/10	SSC
B	488	502LP	530/30	<488>530/30
A	488	655LP	695/40	<488>695/40
F	405		450/50	<405>450/50
E	405	505LP	525/50	<405>525/50
D	405	595LP	610/20	<405>610/20
C	405	635LP	660/20	<405>660/20
B	405	690LP	710/50	<405>710/50
A	405	750LP	780/60	<405>780/60
D	355		379/28	<355>379/28
C	355	450LP	515/30	<355>515/30
B	355	690LP	740/35	<355>740/35
A	355	770LP	820/60	<355>820/60
C	640	630LP	670/30	<640>670/30
B	640	690LP	730/45	<640>730/45
A	640	755LP	780/60	<640>780/60
C	561	550LP	586/15	<561>586/15
B	561	595LP	610/20	<561>610/20
A	561	750LP	780/60	<561>780/60

Supplementary Table 2. Antibody cocktail used to phenotype BAL cells in the standard panel

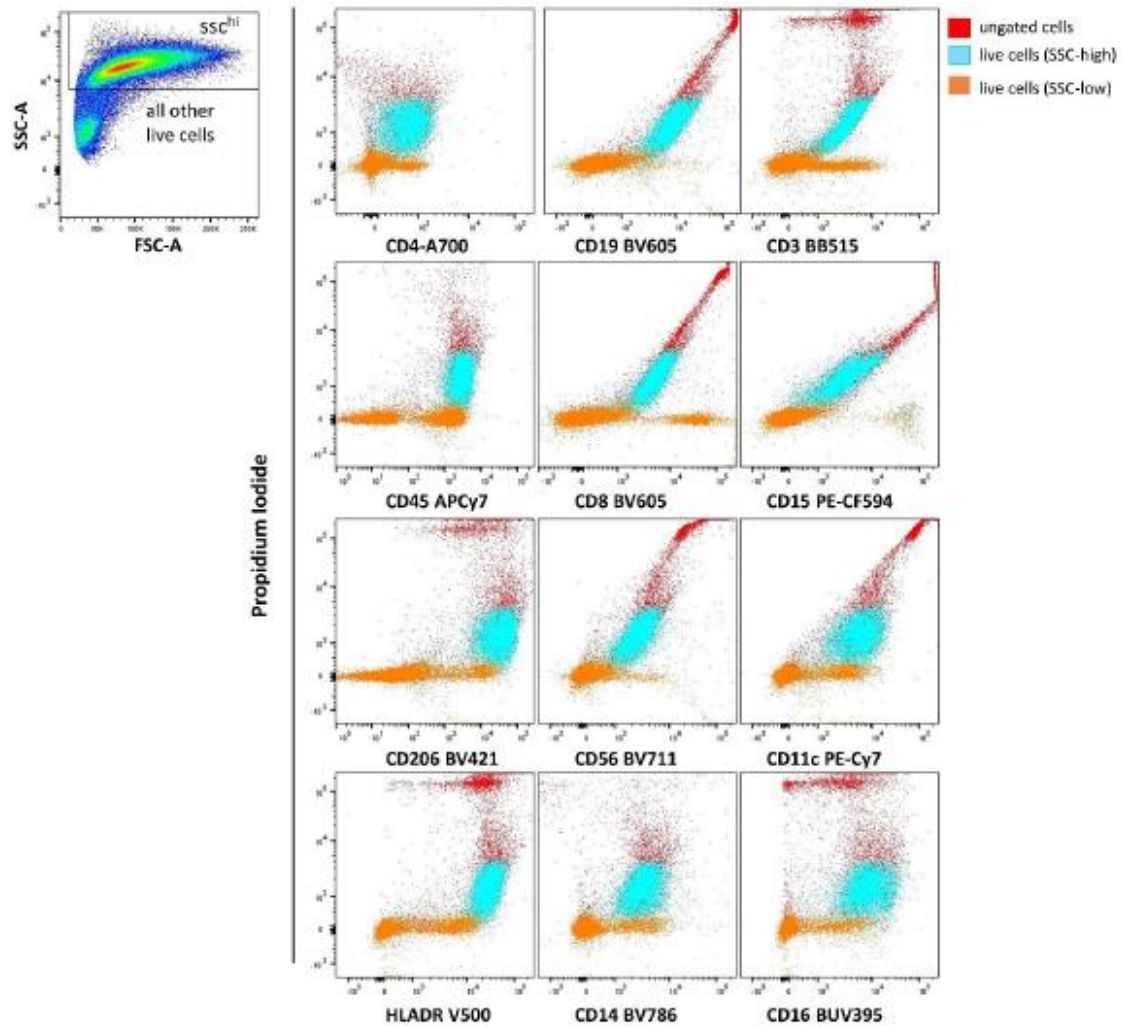
Surface Marker	Flourophore	Clone	Dilution (2X)
CD45	APC-CY7	2D1	1:50
EpCAM	BB515	EBA-1	1:20
CD23	PE-Cy7	M-L233	1:20
CD15	PE-CF594	W6D3	1:50
CD206	BV421	19.2	1:50

Supplementary Table 3. Antibody cocktail used to phenotype BAL cells in augmented panel

Surface Marker	Flourophore	Clone	Dilution (2X)
CD45	APC-CY7	2D1	1:50
CD15	PE-CF594	W6D3	1:50
CD206	BV421	19.2	1:100
CD3	BB515	UCHT1	1:50
CD19	BV605	SJ25C1	1:50
CD4	A700	RPA-T4	1:50
CD8	BV650	RPA-T8	1:50
HLADR	V500	G46-6	1:100
CD14	BV786	M5E2	1:25
CD11c	PE-Cy7	B-ly6	1:50
CD16	BUV395	3G8	1:200
CD56	BV711	NCAM16.2	1:50



Supplementary Figure 1. Preliminary gating strategies for both panels. (A) Firstly, live cells were selected using SSC-A vs PI. (B-C) Within the live cell gate, single cells were identified using FSC-H vs W, followed by SSC-H vs W. (D) Finally, debris was removed, and live, single intact cells were selected. These cells then proceeded to downstream gating as described in the main manuscript.



Supplementary Figure 2. PI staining against all parameters. Overlay of ungated cells (red), live cells with high SSC (predominately highly auto-fluorescent macrophages, blue), and all other live cells (SSC-low, orange).

Supplementary Table 4. Patient Demographics for the standard panel

Patient Number	Gender	Age (months)	Weight (kg)	Volume Inserted (ml)	Volume Obtained (ml)	Return (%)	Viability (proportion of single cells)
1	M	14	10	10	4.5	45.0	76.4
2	M	23	12	12	11	91.7	83.1
3	M	27	14	14	8	57.1	75.1
4	F	36	14	14	8.5	60.7	69.9
5	M	23	10	10	4.5	45.0	73.0
6	F	49	17	17	12	70.6	77.7
7	M	36	16	16	13	81.3	90.7
8	F	61	20	20	11	55.0	80.1
9	F	23	12	12	7	58.3	80.3
10	F	48	15	15	10	66.7	74.0
11	F	60	16	16	13.5	84.4	76.1
12	F	59	18	18	12.5	69.4	72.0
13	M	23	11	11	6.5	59.1	83.5
14	F	59	18	18	14	77.8	73.9

Supplementary Table 5. Fluorescence-activated cell sorting characteristics

Patient Number	Volume Obtained (ml)	Macrophage Number	Macrophage Number/ml	Granulocyte Number	Granulocyte Cells/ml	Lymphocyte Number	Lymphocyte Number/ml	Alveolar Epithelial Cell Number	Alveolar Epithelial Cell Number/ml
1	4.5	142,940	31764.4	129,127	28694.9	101,784	22618.7	10,917	2426.0
2	11	1,761,623	160147.5	28,242	2567.5	115,932	10539.3	20,314	1846.7
3	8	697,170	87146.3	161,229	20153.6	45,337	5667.1	9,115	1139.4
4	8.5	161,154	18959.3	66,623	7838.0	21,930	2580.0	16,676	1961.9
5	4.5	529,279	117617.6	168,925	37538.9	27,248	6055.1	13,301	2955.8
6	12	216,793	18066.1	94,022	7835.2	56,976	4748.0	2,223	185.3
7	13	424,782	32675.5	12,769	982.2	2,902	223.2	2,446	188.2
8	11	46,318	4210.7	5,890	535.5	3,154	286.7	5,038	458.0
9	7	447,015	63859.3	183,418	26202.6	72,426	10346.6	8,795	1256.4
10	10	582,463	58246.3	96,229	9622.9	39,461	3946.1	5,015	501.5
11	13.5	627,993	46518.0	42,973	3183.2	190,370	14101.5	14,402	1066.8
12	12.5	1,747,553	139804.2	83,658	6692.6	51,954	4156.3	37,723	3017.8
13	6.5	485,884	74751.4	71,405	10985.4	36,824	5665.2	9,600	1476.9
14	14	2,069,697	147835.5	80,338	5738.4	158,974	11355.3	7,516	536.9

Supplementary Table 6. Patient Demographics for the deep immune profiling panel (note, not all of the BAL return was used for flow cytometry experiments)

Patient Number	Gender	Age (months)	Weight (kg)	Volume Inserted (ml)	Volume Obtained (ml)	BAL Return (%)	Viability (proportion of single cells)
1	M	36	16	32	21	65.6	76.9
2	M	36	15	30	24	80.0	71.7
3	F	73	18	36	29	80.6	75.6
4	M	59	18	36	32	88.9	77.6
5	F	11	10.5	21	13.5	64.3	61.8
6	F	47	18	36	21	58.3	71.8
7	M	23	13.5	27	16	59.3	84.3

4.4 RESEARCH IN CONTEXT

This project achieved the aim of developing a robust methodology for FACS of paediatric BAL samples. In addition, a cryopreservation method for BAL samples, including analysis of its effect on cell composition, was developed. A broader flow cytometry panel used to phenotype rare immune cells in BAL samples was also established, further contributing to the understanding of the immune cell composition of paediatric BAL fluid.

This project informed the subsequent work of this thesis by providing a method for the isolation of relevant cell types from BAL samples, in turn enabling the development of a reference panel of DNA methylation signatures. The additional development of a cryopreservation method will facilitate the use of BAL specimens in research. These tissue-specific samples are extremely valuable from a research point of view; however, they are often collected at unpredictable times because of the nature of clinical practice, limiting their use in research. The ability to cryopreserve BAL samples and maintain viable cells will drastically improve the ease with which these clinically collected samples can be used in research. The potential value of cryopreserving BAL samples is highlighted by the AREST CF biobank, which contains over 2,000 BAL specimens collected during clinical practice over the past 23 years.¹⁹⁵ Unfortunately, these samples were stored in a manner inappropriate for the maintenance of cell viability, thus were not suitable for laboratory techniques such as flow cytometry and RNA sequencing. Had these samples been cryopreserved, they would have been an even more valuable resource for downstream technological application. Based on the methods developed in this project, the AREST CF biobank has now implemented cryopreservation of all BAL samples into their protocol. The development of a broader flow cytometry panel to phenotype rare immune cells in BAL samples will enable a better characterisation of the local immune response in paediatric respiratory disease. With respect to CF, it will enable the evaluation of the association between cell composition and disease severity, potentially informing predictive biomarkers. The potential utility of this broader flow cytometry panel is further discussed in Chapter 8 (Conclusions and Future Directions).

CHAPTER 5: DNA Methylation Reference Panel

5.1 INTRODUCTION

In any EWAS it is vital to account for cellular heterogeneity within samples. This is because individual cell types have unique DNA methylation profiles. Without accounting for this, differences in DNA methylation patterns observed between samples may reflect differences in cell composition rather than a true difference in DNA methylation. The importance of accounting for cell type composition is a key recommendation in guidelines on the optimal design of EWAS.¹⁸⁷ Determining the cell composition of a sample has added benefits because the associations between cell type proportions and outcomes of interest can be assessed.

There are several methods for managing issues related to cellular heterogeneity in EWAS. One option is to use a purified single cell type, thus obviating the need for adjustment.¹⁸⁷ This approach is most appropriate for situations in which an individual cell type is known to be directly related to the outcome of interest. However, given the relevance of both immune and epithelial cells in CF lung disease, it would be advantageous to study multiple cell types. Broadly, there are two commonly used approaches to adjust for cellular heterogeneity in samples containing multiple cell types: reference-based or reference-free approaches.¹⁹⁶

Reference-based approaches were developed first and involve adjusting for cell composition using a set of reference epigenomes for the cell types in the sample.¹⁹⁴ This process requires (1) knowledge of the cell types in the sample, (2) purification of cell types, (3) methylation profiling of purified cell types and (4) comparison of the DNA methylation pattern of each cell type to identify a set of differentially methylated positions that act as unique methylation signatures for each cell type. These signatures can then be used to estimate the cell composition of a sample containing multiple cell types. In the sample, it is assumed that the observed DNA methylation pattern is the weighted linear sum of the cell type methylation signatures. The model then identifies the best weighting for each cell type to explain the observed sample profile.^{187 194} This approach was first published by Houseman et al.¹⁹⁴ It was initially described for whole blood samples¹⁹⁴ and has since been applied to other tissues, including cord blood,¹⁹⁷ brain tissue¹⁹⁸ and breast tissue.¹⁹⁹ Reference-based cell type adjustment is a well-established and validated method that enables adjustment at the individual sample level.¹⁹⁶ Its main limitation is that it

requires knowledge of the cell types in a sample as well as the genome-wide methylation profile for each cell type. This is not an issue for commonly used samples such as blood, which have publicly available methylation profiles for their constituent cell types. However, less commonly studied samples such as those obtained from BAL have not been as extensively characterised. Another consideration is the age of subjects involved in the development of the reference epigenomes. DNA methylation patterns are known to change with age²⁰⁰ and can be more pronounced in certain cell types.²⁰¹ Thus, depending on the degree to which the methylomes of constituent cell types change with age, a reference epigenome derived from adult subjects may not be ideal for estimating the cell composition of biospecimens collected from infants and vice versa.

An alternative approach to cell type deconvolution is a reference-free method. The major advantage of a reference-free approach is that no prior knowledge of the cell types in a sample and the reference epigenomes of these cells is needed, making it appealing for uncommon sample types. Several reference-free methods based on different statistical models are available. In general, these methods rely much more heavily on assumptions that inform the choice of model, which, if incorrect, can result in poor performance. For example, methods such as EWASher²⁰² and ReFACTor²⁰³ both assume that cell type is one of the largest drivers of variability in the sample. While this is true for most sample types, situations for which this is not the case may result in false negative results.¹⁹⁶ Another disadvantage of reference-free approaches is that the majority do not report the estimated cell type proportions of individual samples, which are of interest in CF.¹⁹⁶

The primary aims of this thesis relate to using EWAS to identify DNA methylation profiles in BAL specimens from children with CF that can predict the development of bronchiectasis in later life. Thus, it is important use suitable methods to account for the cell composition of paediatric BAL samples.

The cellular composition of BAL fluid is well characterised, with predominant cell types being alveolar macrophages, granulocytes, lymphocytes and alveolar epithelial cells (AECs).²⁰⁴ As such BAL samples the cells of interest to the CF lung disease and represent a tissue specific sample. One disadvantage is that relatively few airway epithelial cells are found in BAL, and these cells are affected by CF lung disease. However, no samples routinely collected in clinical practice sample these. Sampling methods which do predominantly obtain epithelial cells include bronchial brushings and tissue biopsies, however neither sample obtains immune cells which is a significant omission given the

crucial role of inflammation in CF lung disease. In addition, tissue biopsies are much more invasive than bronchoscopy and BAL. As outlined in Chapter 4, lymphocyte subsets can be identified in BAL; however, given that CF is characterised by neutrophilic inflammation,⁸⁹ the relevance of these rare populations to CF lung disease may be limited. Reference epigenomes do not exist for the aforementioned cell types, specifically none that are sourced from paediatric BAL. While a reference-free approach could be applied, in this situation a reference-based approach has several advantages. First, it allows for a robust estimation of the proportions of the different cell types in each sample. Given the role of neutrophilic inflammation in CF lung disease, the cell composition of BAL fluid, specifically the neutrophil proportion, may be a predictor of future outcomes. In cross-sectional studies, an elevated neutrophil count in BAL fluid has been associated with lower airway infection^{188 191} and structural lung disease measured via CT scan²⁰⁵; however, no longitudinal analysis has been performed. Second, it will enable the comparison of methylation with the outcome of interest (i.e. the presence or absence of structural lung disease), taking cell type composition into account. Finally, as an assumption-free approach, there will be less chance of choosing a statistical model that is inappropriate for the samples.

The purpose of this project was to develop reference epigenomes for the cell types present in BAL specimens, which can then inform a reference-based adjustment of cell composition in subsequent EWAS.

5.2 METHODS

5.2.1 Ethics

The BAL samples used in this study were obtained as part of the AREST CF study, which has obtained ethics approval (Human Research Ethics Committee [HREC] No. 25054). The parents of all participants gave informed consent for their children's participation in the study.

5.2.2 Bronchoalveolar lavage collection and processing

Flexible bronchoscopy was conducted by a respiratory physician under general anaesthesia. BAL was performed by instilling sterile 0.9% normal saline in aliquots of 1 ml/kg (maximum 20 ml). The aliquots used in this study were the second wash in the same location, increasing the chances of sampling alveolar macrophages.^{193 206} Samples

were kept on ice and cryopreserved within 1 hour of collection using the methods developed in Chapter 4.²⁰⁴

5.2.3 Purification of individual cell types

Cryopreserved samples were thawed using the methods developed in Chapter 4. FACS was used to isolate alveolar macrophages, granulocytes, lymphocytes and AECs. Purified samples were stored in Buffer RLT (Qiagen, Venlo, Netherlands) with β -mercaptoethanol (Gibco, New York, USA) at -80°C until DNA extraction.

5.2.4 DNA extraction

A variety of methods were trialled for DNA extraction. These included the following commercially available kits: Allprep DNA/RNA Mini Kit, Allprep DNA/RNA Micro Kit, QIAamp DNA Mini Kit and QIAamp DNA Micro Kit (all supplied by Qiagen, Venlo, Netherlands) as well as phenol/chloroform extraction. DNA quantity and quality were assessed using a Qubit[®] fluorometer and NanoDrop[™] spectrophotometer, respectively. The best yield of DNA was obtained from the QIAamp DNA Micro Kit, which was subsequently used for DNA extraction. The yield from an individual BAL was found to be insufficient for downstream whole-genome methylation analysis; thus, cell populations purified from several individuals were pooled together. Once extracted, DNA was stored at -30°C until further analysis.

5.2.5 DNA methylation analysis

Genome-wide methylation profiling was performed using the Infinium MethylationEPIC array (EPIC array) (Illumina, San Diego, USA). Arrays were prepared and processed at GenomeScan (Leiden, Netherlands) and Erasmus Medical Centre (Rotterdam, Netherlands). For arrays prepared at GenomeScan, an initial quality control step involving assessment of DNA quality and quantity was performed. A second quality control step using quantitative polymerase chain reaction (qPCR) and a melting curve analysis to assess the quality of bisulphite conversion was also performed. For arrays prepared at Erasmus Medical Centre, quality control involved the assessment of DNA quality, including gel electrophoresis, and quantity.

5.2.6 Data processing and statistical analysis

EPIC array data were analysed using the R programming language²⁰⁷ according to best practice for methylation array analysis.²⁰⁸

5.2.6.1 Quality control

After raw fluorescence array data were imported, extensive quality control was performed. First, as a measure of the quality of array data, the detection p -value was calculated for each probe in each sample using the *minfi*²⁰⁹ Bioconductor²¹⁰ package. The detection p -value represents the significance of the total DNA methylation signal for each probe relative to the background signal level, which was estimated using negative control probes. A detection p -value of < 0.01 was indicative of a robust signal. Poor-quality samples have high mean detection p -values across their probes; thus, any sample with a mean detection p -value > 0.01 was excluded from subsequent analysis. Data were then normalised using subset quantile normalisation (SQN)²¹¹ as implemented in *minfi* to minimise unwanted variations within and between samples. Poor-performing probes, defined as having a detection p -value of > 0.1 in one or more samples, were excluded because their signals were considered unreliable. In addition, probes known to have an SNP at the CpG site and cross-reactive probes that have been shown to map to multiple places in the genome were also excluded.²¹²

5.2.6.2 Data exploration

The purpose of initial data exploration was to thoroughly assess the methylation data derived from cell-specific samples and confirm that, as expected, arrays from individual cell types had similar methylation profiles, and each cell type had a unique methylation profile. The data were required to satisfy these criteria to be suitable for further analysis.

To assist with this, reference panel methylation data were explored in conjunction with data from the subsequent EWAS experiments presented in this thesis, enabling the visualisation of cell-specific data in relation to data from subject samples containing multiple cell types. Data exploration also confirmed that the quality control steps had the intended effect. Multidimensional scaling (MDS) plots were used to visualise variations in the data, with specific assessment of the contribution of variables such as sex, sample type, including cell-specific samples (alveolar macrophages, granulocytes, lymphocytes and AECs) and subject samples (no bronchiectasis, bronchiectasis, raw BAL fluid—disease status unknown), and sample batch (GenomeScan v. Erasmus Medical Centre).

5.2.6.3 Identification of cell-specific DNA methylation profiles

A modified version of the *estimateCellCounts2* function from the FlowSorted.Blood.EPIC package²¹³ was used to identify DNA methylation profiles unique to each cell type and estimate the composition of samples. This function implements the Houseman et al.¹⁹⁴ cell type deconvolution algorithm in conjunction with a custom panel of reference epigenomes to estimate the cell composition of a sample. Prior to cell proportion estimation, the original code of the *estimateCellCounts2* function was modified to allow for the removal of probes excluded in the quality control process.

Reference epigenomes were defined using EPIC array methylation profiles of flow-sorted cells from paediatric BAL specimens. By default, the algorithm chooses the 100 probes with the most statistically significant differential methylation for each cell type compared with all other cell types. These probes can be selected in one of two ways: 1) if the *probeselect* parameter is set to *both*, the 100 probes will consist of the 50 most hypermethylated probes and the 50 most hypomethylated probes for each cell type; 2) if the *probeselect* parameter is set to *any*, then the 100 probes will be selected purely based on their statistical significance, regardless of the direction of the effect. The other important parameter is *processMethod*, which determines how data will be normalised. Although there are several options, SQN²¹¹ was selected because it was the same normalisation method used in the initial data quality control.

5.2.7 Validation

The performance of the BAL reference panel for estimating cell composition was validated in several ways.

5.2.7.1 Publicly available blood dataset

A publicly available EPIC array dataset of artificially created cell mixtures of known proportions of multiple leukocyte cell types was used to validate the utility of the BAL reference epigenomes in estimating the true proportions of constituent cell types.²¹³ The cell types present were T cells (both CD4 and CD8), B cells, natural killer cells, monocytes and neutrophils. By summing the T, B and natural killer cell proportions, this dataset enabled validation of both the lymphocyte and granulocyte reference epigenomes. However, the latter was considered less accurate because it contained both neutrophils and eosinophils, and the validation dataset only reports the proportion of neutrophils.

Concordance between the known artificial cell mixture composition and methylation-based estimates was assessed using Bland–Altman plots and Pearson’s correlations.

5.2.7.2 Bronchoalveolar lavage validation

The custom BAL reference panel was also validated against flow cytometry–derived cell type proportions of BAL specimens. BAL specimens were collected in the same manner as described earlier. Prior to processing, 1 mL of raw lavage was collected and stored at –80°C. The remaining BAL fluid was cryopreserved as described earlier. The raw BAL fluid was then thawed and its DNA extracted and profiled using EPIC arrays. The cryopreserved fraction was then thawed, and its cell composition was determined via flow cytometry using the methods described in Chapter 4.²⁰⁴ Cell type proportions were estimated from the EPIC array methylation data for the same samples using the custom BAL reference panel, as previously described. The cell composition estimates from the EPIC array data were then compared with the cell composition determined via flow cytometry. Concordance between the flow cytometry–based cell composition and methylation-based estimates were assessed via Bland–Altman plots and calculation of Pearson’s correlations.

5.3 RESULTS

5.3.1 Bronchoalveolar lavage collection

Seventeen BAL samples obtained from children with CF were used to develop the reference panel. The median (range) age of the children was 36 months (14–70 months), and 11/17 (64.7%) were female. One child was of South Asian ethnicity, and all other children were of European ethnicity. On average, 6 mL of BAL fluid was used. The median (range) cell composition of samples was 63.7% (5.2–95.9%) alveolar macrophages, 23.7% (2.9–86.1%) granulocytes, 8.1% (0.7–6.5%) lymphocytes and 1.5% (0.2–8.3%) AECs.

5.3.2 DNA extraction and methylation profiling

The purified cell populations were pooled to allow for the subsequent extraction of sufficient DNA for EPIC array profiling. This resulted in five macrophage pools, four granulocyte pools, three lymphocyte pools and two AEC pools. Of the five macrophage pools, three included cells from four subjects, one included cells from three subjects, and one included cells from two subjects. Of the four granulocyte pools, three consisted of

cells from four subjects, and one consisted of cells from five subjects. Of the three lymphocyte pools, two comprised cells from six subjects, and one comprised cells from five subjects. Of the two AEC pools, one contained the cells from 12 subjects, and the other contained the cells from five subjects. DNA was successfully extracted for all pools and profiled using EPIC arrays.

5.3.3 Initial quality control and data exploration

The mean detection p -values for all pooled flow-sorted cell samples were less than 0.01; hence, all samples were retained for further analysis. Following normalisation, visualisation of the data via MDS plots showed that, as expected, samples clustered together by cell type (Figure 5.1). In addition, the batch in which each sample was assayed (GenomeScan v. Erasmus Medical Centre) was not a significant source of variation in the first two principal components.

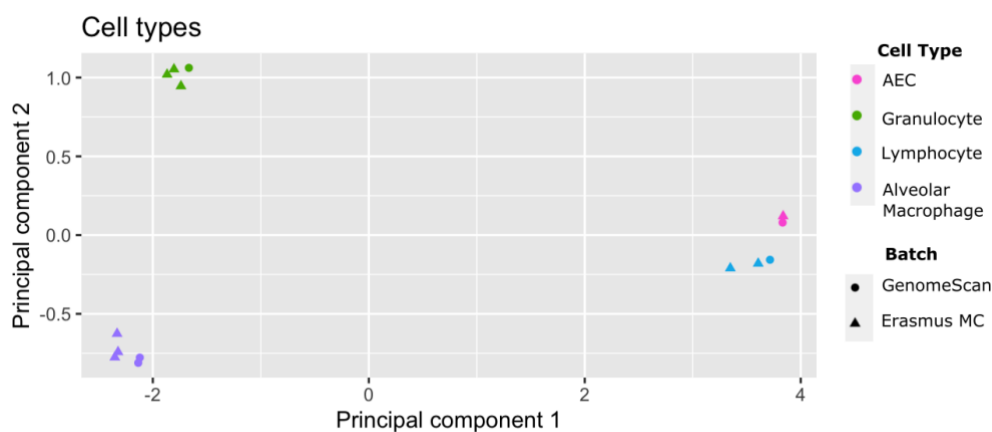


Figure 5.1. Initial multidimensional scaling plot, showing samples clustered together by cell type with no significant batch effect.

Poor-performing probes, multimapping probes and probes with SNPs commonly located at the CpG were then removed. The data were again visualised, including samples from the subsequent EWAS described in this thesis, to assess the relationship between cell-specific arrays and arrays from samples containing multiple cell types. The first principal component of variation was driven by sex. Because the cell-specific data were derived from samples pooled from both males and females, they were located between the two groups (Figure 5.2).

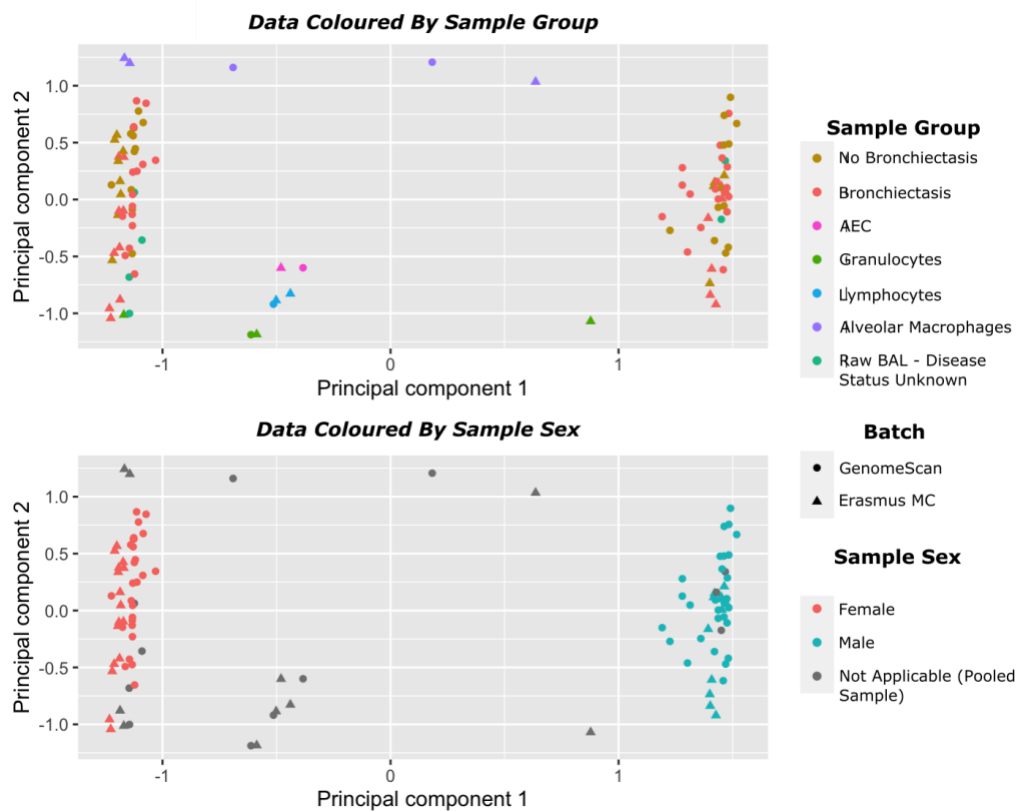


Figure 5.2. Multidimensional scaling plot post initial probe filtering, showing sex is responsible for the variation in the first component.

Probes that mapped to the sex chromosomes were then filtered out, leaving 695,764 probes for downstream analysis. Subsequent exploration of the data confirmed that the sex effect in the first principal component had disappeared (see Figure 5.3). Following probe filtering, variations in the first two principal components were driven by differences in cell type proportions between the samples, with the flow-sorted samples sitting at the extremes of the plot. The second principal component captured the differences between flow-sorted cell types as well as a weak batch effect, while the first appeared to be driven by the differences between macrophages and granulocytes, lymphocytes and AECs.

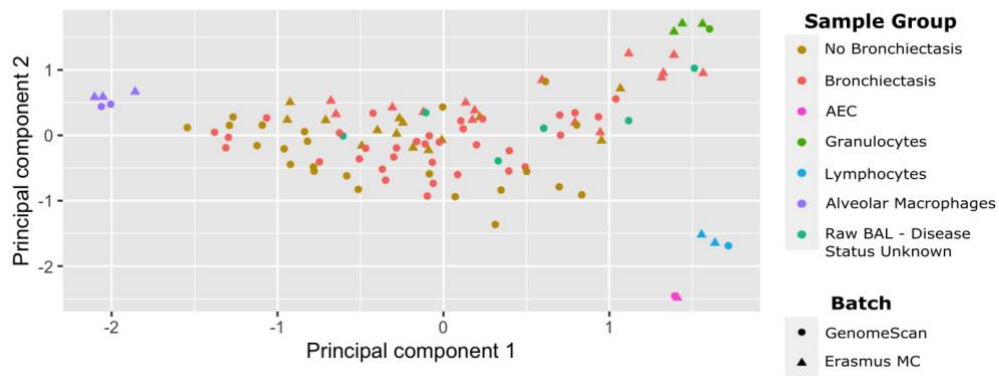


Figure 5.3. Multidimensional scaling plot after filtering out probes that mapped to sex chromosomes. The first and second principal components were now driven by cell-specific probes. The first principal component was driven by the difference between alveolar macrophages and other cell-specific samples. The second principal component was driven by the difference between granulocytes and AECs/lymphocytes as well as a weak batch effect.

Prior to filtering out probes related to sex chromosomes, sex did not appear to be responsible for a significant amount of variation in the cell-specific samples. This was confirmed via visualisation of the MDS plots of only the cell-specific samples before and after filtering out the probes on sex chromosomes (see Figure 5.4). These MDS plots were identical in appearance.

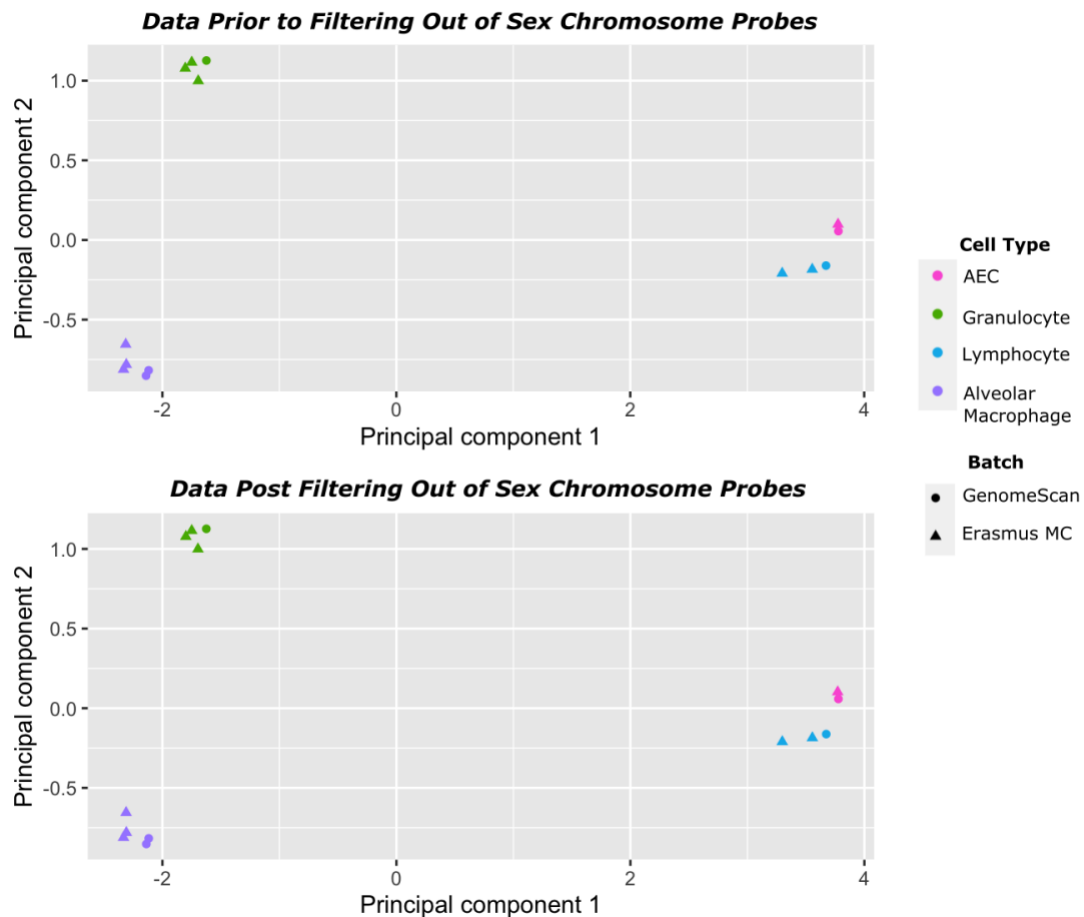


Figure 5.4. Multidimensional scaling plots limited to cell-specific samples. The inclusion or exclusion of probes related to sex chromosomes did not affect the plots, indicating that sex did not contribute to the top two principal components of variation between these samples, and variation was attributable to differences in methylation patterns related to cell type.

Further visualisation of the data using MDS plots once again demonstrated strong clustering of the flow-sorted cell type samples, and there was no evidence of the batch effect until the fourth principal component (Figure 5.5). This indicates that differences between cell types were a significantly greater source of variation. Given this, the data were deemed suitable for further analysis.

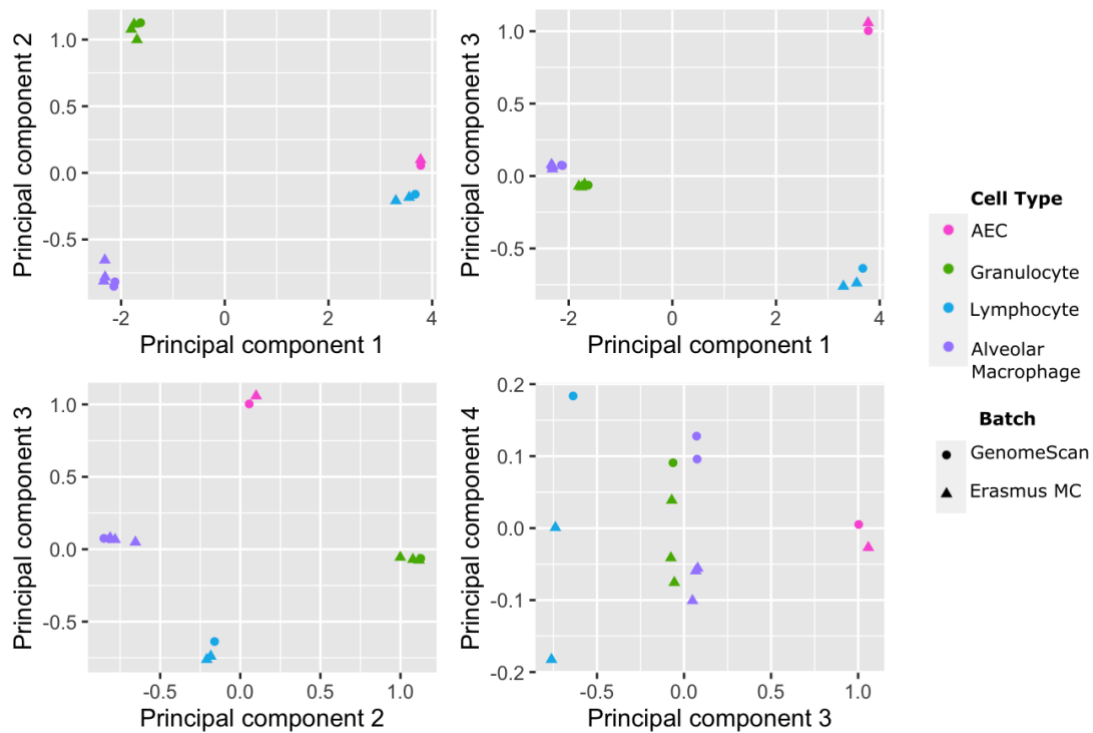


Figure 5.5. Multidimensional scaling plots showing the first four principal components after filtering out poor-performing probes. The plots show that samples clustered together by cell type, and a batch effect was only apparent in the fourth principal component.

5.3.4 Identification of cell-specific DNA methylation profiles

The *estimateCellCounts2* function was used to identify DNA methylation profiles unique to each cell type. SQN normalisation²¹¹ was selected to ensure consistency with the initial quality control steps. Cell composition was estimated with the *probeSelect* parameter set to *both* and repeated with it set to *any*, and the results were compared. The probes removed during quality control were flagged for removal prior to cell type proportion estimation. Figure 5.6 shows an example of the code that was used.

```

cellEstBoth <- estimateCellCounts2Mod(rgSet = patientSamps,
  compositeCellType = "Lavage",
  processMethod = "preprocessQuantile",
  probeSelect = "both",
  cellTypes = unique(targets$Sample_Group[
    cells]),
  referencePlatform =
    "IlluminaHumanMethylationEPIC",
  referenceset = "lavageRef",
  IDOLOptimizedCpGs = NULL,
  returnAll = TRUE,
  meanPlot = FALSE,
  keepProbes = rownames(mValsNoXY))

cellEstBoth$counts

```

Figure 5.6. Example code for the *estimateCellCounts2* function. *patientsSamps* represents the data from which the function estimated the cell composition; *lavageRef* was previously defined as the EPIC array data related to the cell-specific pools.

The *estimateCellCounts2* function was able to identify unique DNA methylation patterns for each cell type using either of the *probeSelect* options. To visualise this, a heatmap showing the cell type–discriminating probes was created (see Figure 5.7), showing a clear separation between the methylation profile of each cell type.

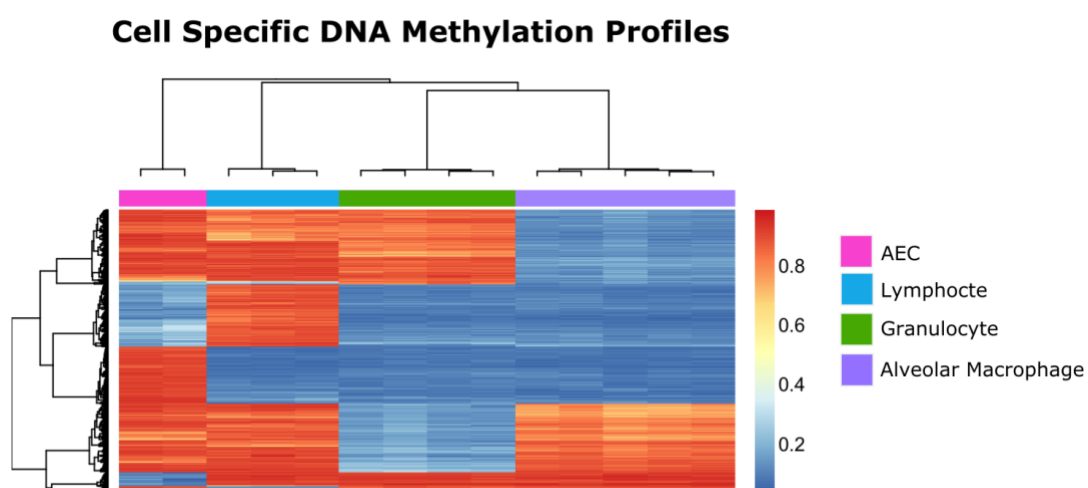


Figure 5.7. Heatmap demonstrating unique DNA methylation profiles for each individual cell type. The probes that were most able to discriminate between the different cell types were selected using the *estimateCellCounts2* function.

5.3.5 Validation

5.3.5.1 Accuracy of cell proportion estimates for artificial cell mixtures

The first validation was performed using a publicly available dataset²¹³ of 12 artificially created cell mixtures. Created mixtures had known proportions of several leukocyte populations (see Figure 5.8), and DNA methylation of the mixtures was determined using the EPIC array. Because the true proportions of constituent cell types were known, this dataset was useful for validating reference epigenomes of relevant cell types. However, only the BAL lymphocyte and granulocyte reference epigenomes could be validated because the artificial cell mixtures did not include alveolar macrophages or AECs.

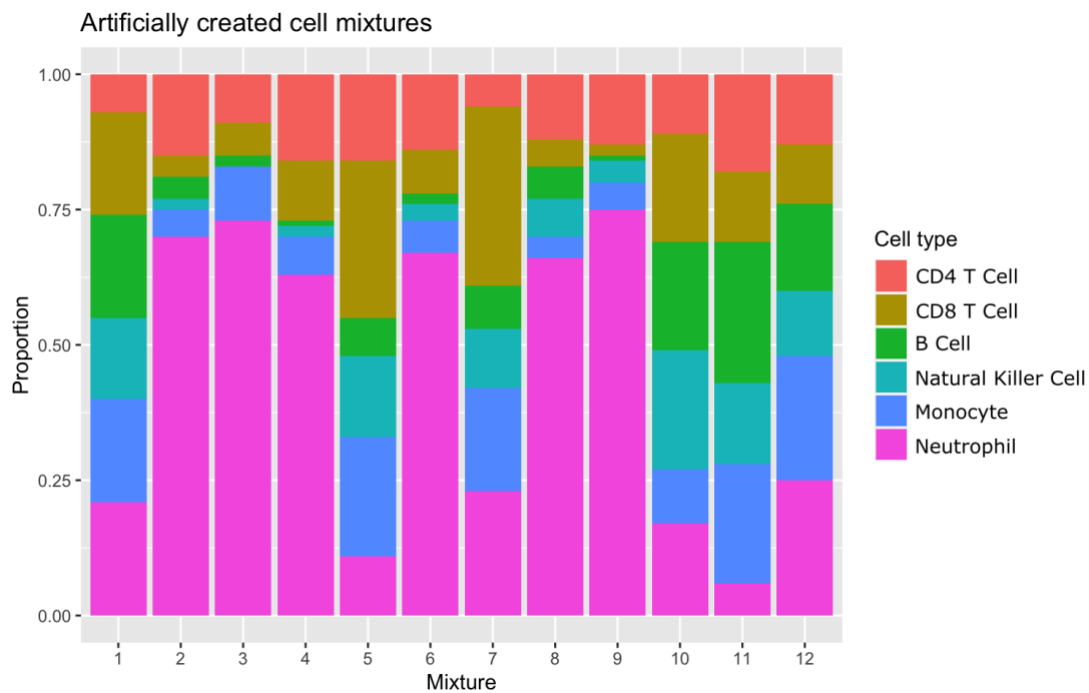


Figure 5.8. Artificial cell mixtures. There were 12 different artificial cell mixtures, all with varying but known compositions of six different cell types.²¹³

The first step was to sum the known proportions of the lymphocyte subsets to enable validation of the BAL-derived lymphocyte reference epigenome (see Figure 5.9). Using the available EPIC array data, the cell proportions of the artificial mixtures were then estimated using the reference epigenomes derived in this study, in conjunction with the Houseman et al.¹⁹⁴ method, on the available EPIC array data. The results are summarised in Figures 5.9 and 5.10. For cell types with available reference epigenomes (lymphocytes and granulocytes), there was a tight correlation between the known cell proportion and the methylation-based cell proportion estimate. For some samples in which more than 50% of the cells were lymphocytes (1, 5, 7, 10, 11, 12), the estimated proportion of

granulocytes was slightly overestimated. As expected, the sum of the estimated cell proportions was slightly above or below 1.¹⁹⁴ The methylation-based estimates included a proportion of alveolar macrophages, even though these were not contained in the artificial cell mixture. Overall, Pearson’s correlation between the true proportions and methylation-based estimates for both lymphocytes and granulocytes using either the *any* or *both probeSelect* option was > 0.998 , indicating excellent performance.

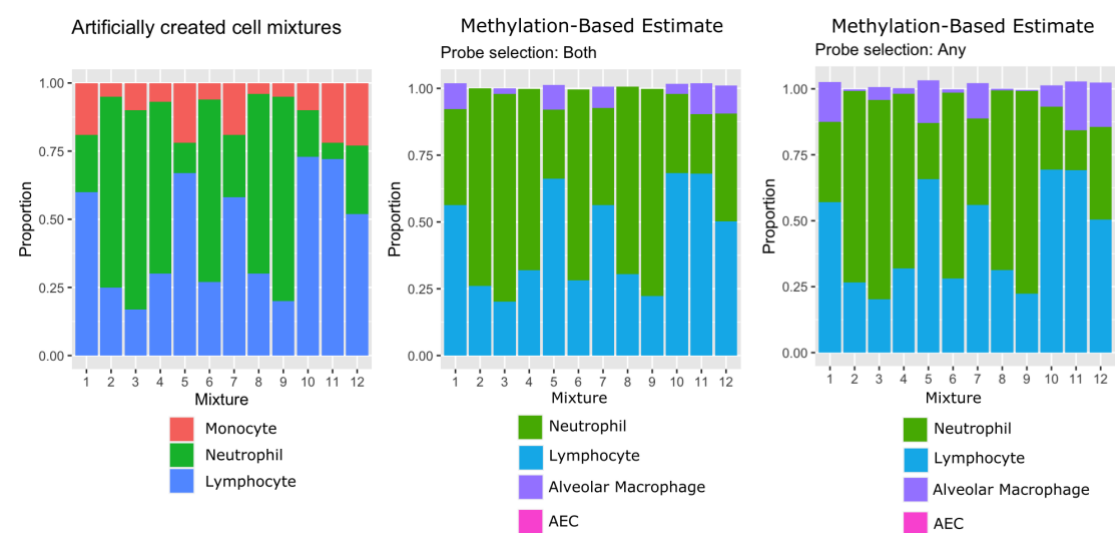


Figure 5.9. Comparison of known proportions in artificial cell mixtures and methylation-based estimated cell proportions. In artificial cell mixtures, T cells, B cells and natural killer cells are combined into a lymphocyte population. For methylation-based estimated cell proportions, both probe options were selected.

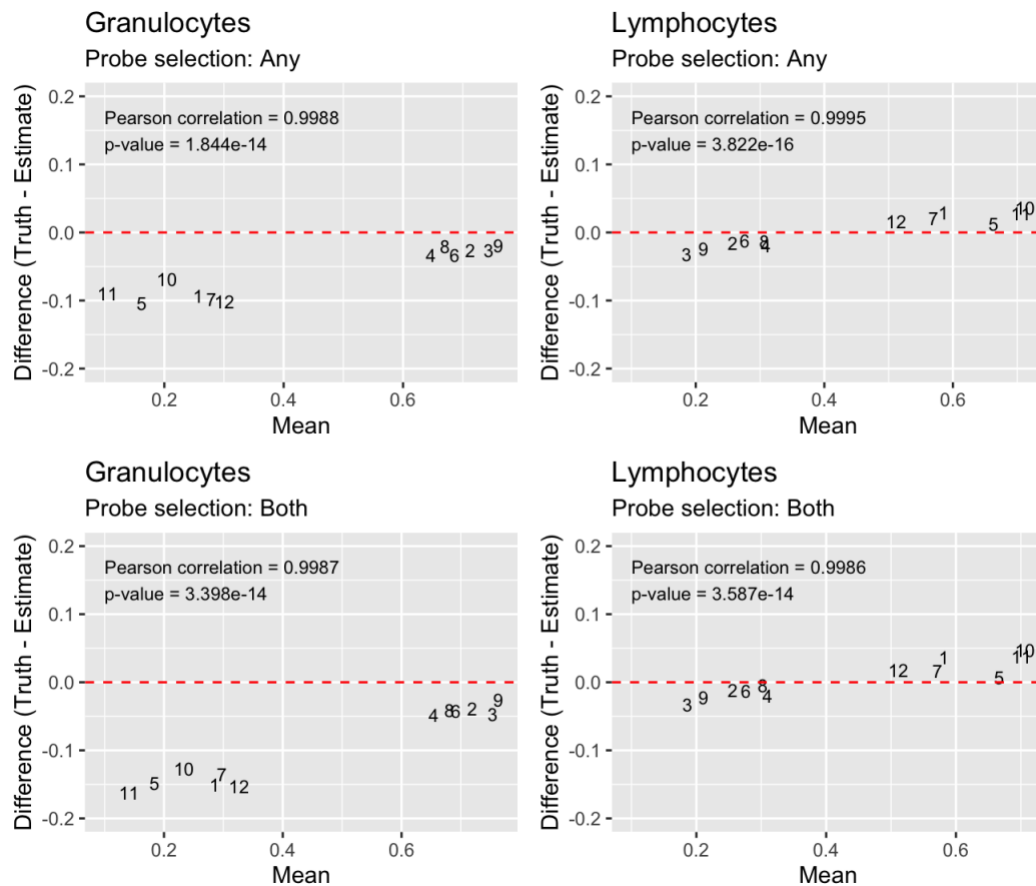


Figure 5.10. Bland–Altman plots comparing known artificial cell mixture proportions (‘Truth’) and methylation-based estimated cell proportions (‘Estimate’). The data point numbers represent the cell mixture to which the data pertain. The estimated lymphocyte proportion is extremely accurate, whereas there is a slight overestimation of granulocytes in samples with a low granulocyte proportion.

5.3.5.2 Comparison of flow cytometry and reference-based DNA methylation estimation of cell composition of bronchoalveolar lavage samples

Six BAL samples were obtained to compare cell type proportions using flow cytometry and methylation-based estimates generated using the Houseman et al.¹⁹⁴ method in conjunction with the BAL reference epigenomes. Five BAL samples were obtained from children with CF, and one was obtained from a control child who underwent BAL for investigation of suppurative lung disease. Four of the six children (66.7%) were female. The median (range) age of the children was 57 months (23–89 months), which is slightly older than those used to derive the reference panel (median age = 36 months, range = 14–70 months). Five children were of European ethnicity, and one was of South Asian ethnicity.

DNA was successfully extracted from raw BAL samples in all cases, and EPIC arrays were performed. All samples passed initial quality control. The cell composition of all samples was determined using flow cytometry on cryopreserved BAL fractions (see Table 5.1).

For flow cytometry data, the proportion of cells identified ranged between 46.2% and 89.0%. The majority of the unidentified cells were CD45-negative and EpCAM-negative, indicating that they were most likely red blood cells. Because red blood cells do not contain DNA, they do not contribute a DNA methylation signature. Consequently, analysis was conducted with both inclusion and exclusion (referred to as ‘scaled’) of the unknown cells. The cell composition for most samples was as expected; however, two samples (CF1, CF5) had a CD45-negative and EpCAM-negative fraction in excess of 30%, which was higher than expected and may have been related to a traumatic blood-stained BAL specimen.

Table 5.1. Cell composition of bronchoalveolar lavage samples determined by flow cytometry

	Samples					
	CF1	CF2	CF3	CF4	CF5	Control
Cell proportions as a proportion of total live cells (%)						
Alveolar macrophages	2.4	46.1	42.5	50.2	31.3	34.8
Lymphocytes	2.5	7.4	5.3	5.0	3.6	13.9
Granulocytes	39.7	26.9	15.8	13.6	14.5	14.6
Alveolar epithelial cells	1.5	2.4	2.3	0.4	0.1	4.7
CD45/EpCAM-negative	48.5	15.3	27.7	28.0	38.5	27.3
Cell proportions scaled to total 100% (%)						
Macrophages	5.2	55.7	64.6	72.5	63.2	51.1
Lymphocytes	5.5	8.9	8.0	7.2	7.2	20.5
Granulocytes	86.1	32.5	23.9	19.7	29.4	21.5
Alveolar epithelial cells	3.2	2.9	3.4	0.6	0.2	6.9

Note: The scaled proportions were calculated because the CD45- and EpCAM-negative cells were most likely red blood cells, which do not contribute to DNA methylation data.

Comparison of the estimated cell proportions using methylation data and flow cytometry cell proportions is shown in Figure 5.11. The AEC and lymphocyte methylation-based estimates correlated well with the flow cytometry data. The alveolar macrophage and granulocyte reference epigenomes were more variable, and in certain subjects (CF1, CF2, CF3 and control) there was a significant discrepancy between flow cytometry and

methylation-based cell composition. Pearson's correlation for four of the six subjects (CF1, CF3, CF4, CF5) was > 0.88 ; however, for CF it was 0.7205, and for the control subject it was -0.1298 .

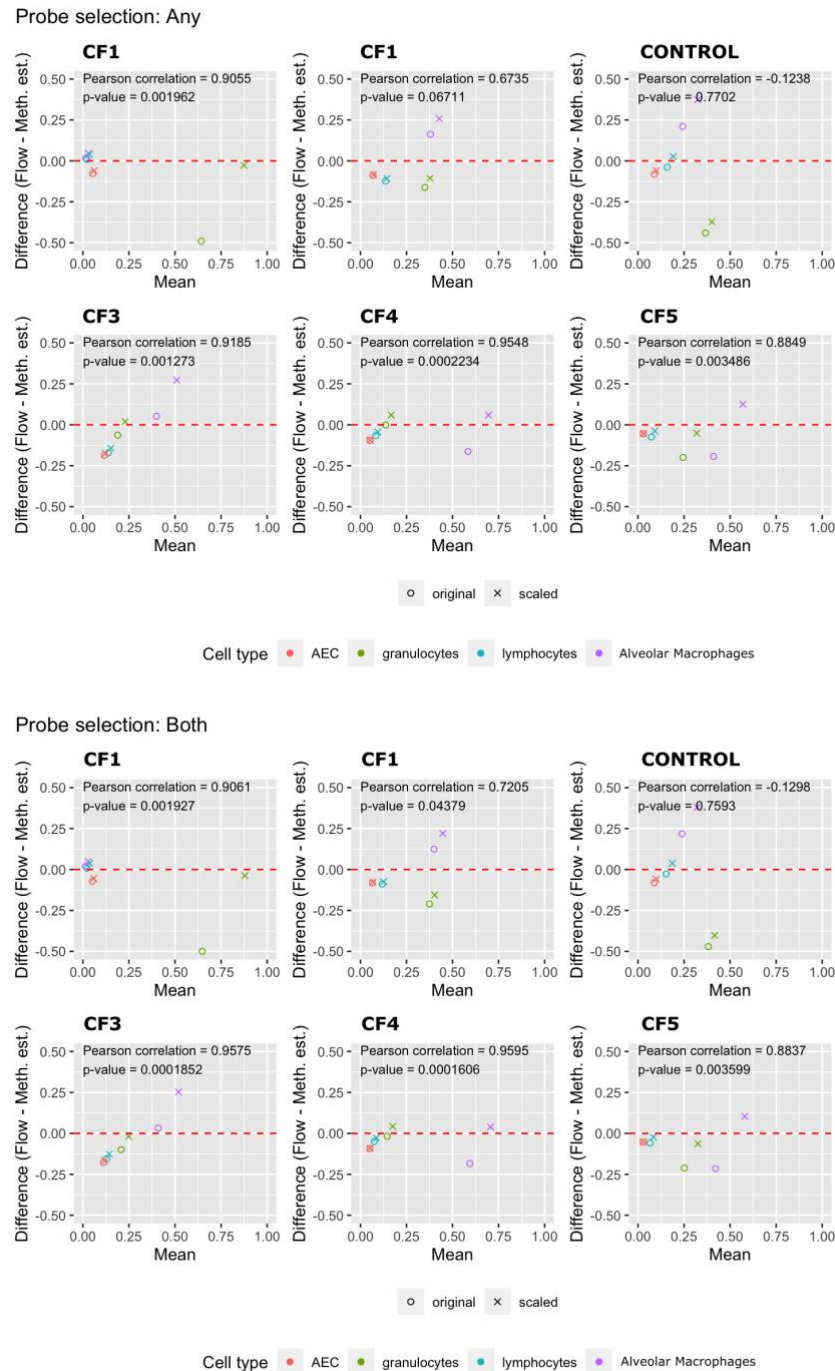


Figure 5.11. Correlation of cell composition determined via flow cytometry ('Flow') and methylation-based estimates ('Meth. est.'). 'Original' is the proportion of cells in the total live cell population, and 'scaled' is the proportion of cells when limited to the four cell types of interest. Correlations between four of the six subjects (CF1, CF3, CF4,

CF5) were strong. Lymphocyte and AEC correlations were tight in all samples, while granulocyte and alveolar macrophage correlations were more variable.

5.4 DISCUSSION

This project succeeded in its aim to develop a set of reference epigenomes that may be used to adjust cell composition in an EWAS study utilising BAL samples from children with CF. The methylation-based cell type proportion estimates generated using the Houseman et al.¹⁸⁹ method in conjunction with BAL reference epigenomes were then validated using two different datasets.

The first validation occurred in a publicly available dataset of artificial cell mixtures derived from adult blood. In this validation, the lymphocyte and granulocyte reference epigenomes identified in this study performed extremely well, with Pearson's correlations of > 0.998 for both reference epigenomes. In a subset of samples, granulocyte proportion was slightly overestimated. This may have been attributable to the fact that the reference epigenome contained eosinophils as well as neutrophils, thus was not a direct reflection of the artificial cell mixture. Further, in their original study, Houseman et al.¹⁹⁴ observed that granulocytes had the most significant discrepancy between the true and estimated proportion, suggesting that the minor discrepancy observed here was unlikely to be solely driven by the BAL-derived reference epigenome. Although alveolar macrophage cells were not present in the artificial cell mixture, their proportion was also estimated. This was likely attributable to the similarity in DNA methylation patterns of circulating monocytes (which *were* present in the artificial cell mixture) and alveolar macrophages. Some alveolar macrophages consist of circulating monocytes that migrate to the lungs and become a resident population.^{214 215} Also, the sum of the estimated cell proportions was slightly above or below 1. This is in keeping with the Houseman et al.¹⁹⁴ study, which estimated a total cell composition of 98.8–102.7% of ground truth data.

The second validation was undertaken using paediatric BAL samples. The performance of the reference epigenomes was not as robust as in the first validation; however, this may relate to issues with the design of the validation experiment. Two of the six subjects had poor correlations, which mainly appeared to result from a discrepancy between the granulocyte and alveolar macrophage estimate. The poor performance of the granulocyte estimate may have been related to previously described difficulties with the Houseman et al.¹⁹⁴ method of determining an accurate granulocyte proportion. It may also have been attributable to the effect of cryopreservation on the BAL fraction for which cell type

proportions were estimated using flow cytometry. In Chapter 4, cryopreservation was shown to reduce the proportion of granulocytes. However, the BAL fraction profiled using EPIC arrays was not cryopreserved, thus retaining a higher proportion of granulocytes. The methylation-based estimate was consistently higher than that of flow cytometry. Therefore, the discrepancy between granulocyte proportions seen in flow cytometry and methylation-based estimates may be best explained by the effect of cryopreservation rather than by poor accuracy of the methylation-based estimate. The variable size in discrepancy could be explained by the fact that cryopreservation primarily reduces the CD16⁺ granulocyte fraction. Thus, samples that originally had a large CD16⁺ granulocyte proportion were likely to have a larger discrepancy than those with a smaller CD16⁺ granulocyte proportion. In Chapter 4, the range of CD16⁺ granulocyte proportions observed in BAL samples was 0.09–9%, indicating substantial variation.

The cause of variation in methylation-based estimates of macrophages is less clear. This may have been a consequence of the previously described issues with the granulocyte estimate. The Houseman et al.¹⁹⁴ method estimates a proportion of cells close to 1; thus, if one cell proportion is inaccurately estimated, there will be a reciprocal effect on other cell proportions. In this case, the consistent overestimation of granulocytes (when compared with flow cytometry data) may have also resulted in the observed underestimation of the alveolar macrophage proportion. Recently, single-cell transcriptomic analysis of adult BAL samples has revealed 13 alveolar macrophage subpopulations.²¹⁶ The composition of a subject's alveolar macrophage pool is most likely related to disease severity, and each of the alveolar macrophage subpopulations will have a unique epigenetic profile. Thus, the discrepancy between flow cytometry and methylation-based estimates of the proportion of alveolar macrophages seen in some subjects (i.e. CF2 and CF3) may have been attributable to differences in the composition of their alveolar macrophage subtypes relative to the samples used to generate the reference epigenome. The relatively large discrepancy between flow cytometry and methylation-based estimates of alveolar macrophage proportions for the control subject may have resulted from altered alveolar macrophage DNA methylation profiles between control and CF subjects.²¹⁷

Given the effect of ageing on DNA methylation,²⁰⁰ the older ages of the validation group may have also affected validation. This is especially relevant to the control subject, who was aged 89 months, 19 months older than any subject included in the reference epigenomes.

This work has several strengths. The use of BAL in an EWAS study is appealing because it enables the simultaneous assessment of the methylation of local immune and epithelial cells, both of which are relevant to CF. Alternative biospecimens such as blood or bronchial brushings would have omitted cell types of interest. The panel of reference epigenomes enabled deconvolution of these multiple cell types using a well-established reference-based deconvolution method. The reference panel was derived from children with CF under 6 years of age, meaning it will be appropriate for EWAS in this patient population. The purification of individual cell populations using FACS was done using previously validated high-quality methods.²⁰⁴ The DNA methylation analysis, in particular the methods used for identifying cell-specific probes, is well established and widely used.^{194 208 213} Finally, the validation of two separate datasets allowed for a detailed assessment of the performance of the reference epigenomes.

The limitations of this work primarily relate to the samples used to derive the reference epigenomes. Given that samples were derived from children with CF under 6 years of age, they may not be appropriate for use in EWAS in different populations. As mentioned, DNA methylation profiles change with age. In a seminal study of monozygotic twins, Fraga et al.²⁰⁰ showed that twins in early life had near-identical methylation profiles, but in later life they differed greatly, illustrating that DNA methylation patterns change with age. Additionally, Wikenius et al.²¹⁸ studied 214 infants and performed genome-wide methylation profiling using the 450K array at 6 and 52 weeks of age, identifying a change in DNA methylation profiles in the first year of life, with statistically significant changes in the methylation of 42 genes. Tan et al.²¹⁹ used the 450K array to perform genome-wide DNA methylation profiling in a cohort of elderly patients at age 73 years and 10 years later at age 82 years. They identified 2,284 CpG sites whose DNA methylation levels had changed over the 10-year period. In addition to age, disease-specific changes in DNA methylation may also limit the broader use of these reference epigenomes. Previous studies have shown differences in alveolar macrophage DNA methylation patterns between adult controls and subjects with CF.²¹⁷ However, given that the primary aim of this project was to develop reference epigenomes for use in studies of children with CF, these limitations were unavoidable. In addition, differences in DNA methylation between cell types are likely to be much greater than those related to age or disease; hence, it is possible that the reference epigenomes identified in this project will be suitable for broader use. This is supported by the robust performance of the lymphocyte and granulocyte epigenomes in the data derived from healthy adult subjects. The poorer

performance of the reference epigenomes when used in conjunction with the Houseman et al.¹⁸⁹ method to estimate cell proportions in the BAL sample from the control subject was more concerning. However, this was likely to be related more to methodological issues with the BAL-based validation, the other major limitation of this project. As previously discussed, the effect of cryopreservation may have altered the observed granulocyte fraction, with a magnitude that could vary between samples.²⁰⁴ In addition, flow cytometry data captured a proportion of undefined cells, which were predominantly CD45- and EpCAM-negative and were likely to be red blood cells. As such, they would not contribute to the DNA methylation signature. However, there is a possibility that the undefined cells represent another unknown cell type. Given these two issues, the cell proportions defined by flow cytometry data derived from cryopreserved samples may not be completely accurate, which then causes difficulty when interpreting discrepancies in cell proportions estimated from methylation data. An alternate approach would have been to create an artificial cell mixture from FACS-purified cell populations, as done by Salas et al.²¹³ This would have provided a more accurate ‘ground truth’ knowledge of cell composition for evaluation of the reference epigenomes because it would avoid the effect of cryopreservation, and there would be no undefined cells in the cell mixture.

Another option to potentially improve the methylation-based estimates of cell composition would be to use the BAL reference epigenomes developed in this project in conjunction with a different reference-based method. For example, MethylResolver²²⁰ is a recently published reference-based method that could be used instead of the Houseman et al. method. The major advantage of MethylResolver is that it can estimate proportions relative to unknown cell fractions. This would be particularly relevant to the BAL validation experiment conducted in which flow cytometry estimated an unknown cell fraction ranging between 15.3% and 48.5% of cells. This may offer improved estimates compared with the Houseman et al. method.

The panel of BAL reference epigenomes has the potential to significantly improve the conduct of EWAS in respiratory research. To date, multiple EWAS have been undertaken on respiratory diseases; however, they have rarely used tissue-specific samples. In EWAS Atlas,²²¹ an online repository of published EWAS studies, there are 19 published EWAS in non-cancer lung diseases such as asthma, CF, chronic obstructive pulmonary disease, acute respiratory distress syndrome and granulomatous lung disease. Blood is the most commonly used biological sample (8/19 studies), followed by nasal and airway epithelial cells (6/19). However, these samples do not contain immune cells relevant to the lung

diseases being investigated.²²² Three studies have utilised BAL samples with varying approaches to cell composition adjustment. One investigated BAL macrophages only, using a reference-free approach to adjust for alveolar macrophage subtype composition.²¹⁷ Another used microscopy-derived lymphocyte cell count to adjust the analysis.²²³ This approach is limited by the performance of microscopy, which is considered less accurate than flow cytometry.^{224 225} The final study²²⁶ used the publicly available eFORGE tool²²⁷ to adjust for cell composition. This tool was derived using a number of biological specimens, including peripheral blood and lung tissue; however, resident lung macrophages (the most common cell type in BAL specimens) were not involved in its development. Thus, the validity of using eFORGE for BAL samples is unknown. By providing a panel of reference epigenomes for commonly found cell types in BAL specimens, this project reduces the barriers for using this sample type in EWAS. It will be important to validate our reference epigenomes prior to use in other diseases and age groups.

An alternative to the reference-based approach to BAL fluid cell composition adjustment in EWAS is the reference-free approach. This approach was used in a study that estimated the presence of two alveolar macrophage subtypes in control and CF subjects.²¹⁷ Although the reference-free approach could not precisely estimate the cell composition of each sample, at the group level it showed that one subtype was predominant in healthy controls and the other more common in CF. To validate this finding, the authors used microscopy to examine the BAL specimens from the two groups. In healthy controls, they found that > 95% of alveolar macrophages had a traditional appearance. In contrast, subjects with CF had large subpopulations of alveolar macrophages that appeared different in size and shape from traditional alveolar macrophages. The authors hypothesised that these abnormal alveolar macrophages were the second subtype identified by the reference-free methylation estimation. Even at the group level, the proportions of alveolar macrophage subtypes, as determined by the two different methods, were not directly compared. Examination of the figures included in the manuscript show the reference-free methylation-based estimate of traditional macrophages in healthy controls was less than the > 95% reported on microscopy, potentially indicating poor correlation between microscopy and reference-free methylation estimates of cell composition. To date, there has been no direct comparison of reference-free and reference-based approaches for BAL.

In summary, despite its limitations, which are primarily related to the method of validation used for BAL samples, this project has defined a novel set of reference epigenomes that will inform reference-based cell composition adjustment of BAL samples in EWAS examining early-life CF lung disease.

5.5 RESEARCH IN CONTEXT

This project achieved its primary aim of developing a set of reference epigenomes, which will inform the analyses in the subsequent EWAS conducted as part of this thesis. The ability to robustly adjust for cell composition will ensure that the EWAS in this thesis are conducted in keeping with best practice recommendations for EWAS methodology.¹⁸⁷

The impact of this project should go beyond the context of this thesis. BAL specimens are often collected in the evaluation of lung diseases, including interstitial lung disease,²²⁸ lung cancer,²²⁹ severe asthma²³⁰ and lung transplantation.²⁰⁶ This means that BAL could be used in EWAS investigating these conditions. The reference epigenomes described in this study may remove a significant barrier to their use in EWAS. Prior to their use in other conditions, they would need to be validated, ideally using artificial cell mixtures derived from BAL samples in these conditions. If this validation occurs, the reference epigenomes described in this study could be widely used in the investigation of DNA methylation in a variety of respiratory diseases.

CHAPTER 6: Epigenome-Wide Association Study of Cystic Fibrosis Lung Disease Severity at 9 Years of Age

6.1 INTRODUCTION

As discussed, biomarkers that predict CF lung disease would be extremely informative for people with CF and their families as well as clinicians, researchers and health authorities. To date, most of the work in this area has focused on non-*CFTR* genetic modifiers.²³¹ One area that has received little attention is epigenetic biomarkers such as DNA methylation. This is despite the knowledge that both CF lung disease severity and DNA methylation are determined by the interaction between genotype and environmental exposures.

This chapter describes an EWAS that aimed to identify predictive biomarkers of CF lung disease severity. There are several important considerations when designing such a study. Michels et al.¹⁸⁷ published guidelines on the optimal design of an EWAS (see Figure 6.1), which highlight that the starting point for any EWAS is a clear research question. The next step is to identify an appropriate study population and, importantly, obtain tissue samples relevant to the research question. After measuring genome-wide methylation, several different analysis techniques may be used; for example, site-specific (also known as probe-wise), regional or cluster-based approaches. Not mentioned in the figure but emphasised in the guidelines is the need to account for the cell composition of the samples in the analysis. The guidelines also mention the importance of the validation of any significant findings.

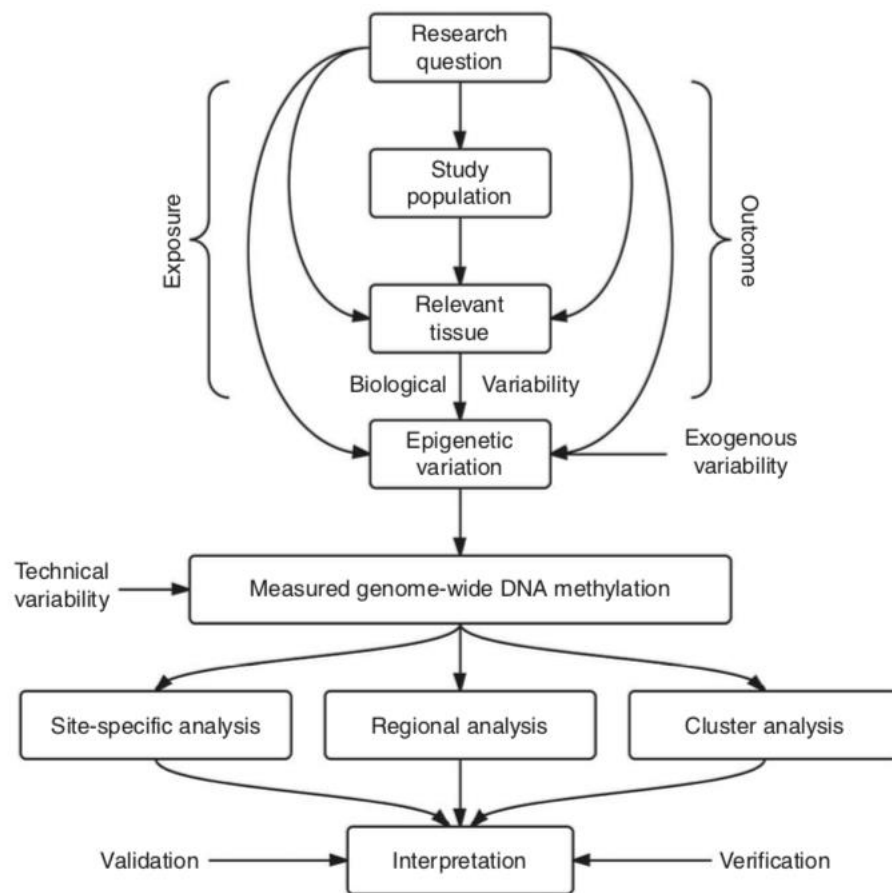


Figure 6.1. Optimal design of epigenome-wide association studies.¹⁸⁷

The starting point for designing an EWAS is a research question. The EWAS presented in this chapter aimed to address Aim 1 and Hypothesis 1 of this thesis:

Aim 1: To identify DNA methylation profiles associated with normal lung structure at 9 years of age in children with CF.

Hypothesis 1: *In children with CF, specific features of DNA methylation profiles of BAL fluid at 6 years of age are predictive of bronchiectasis outcomes at 9 years of age.*

The decision to use CT to assess for the presence or absence of bronchiectasis at 9 years of age reflects the most clinically relevant outcome. As discussed in Chapter 1, CT assessment of lung structure is the most sensitive marker of CF lung disease. Given that the majority of children with CF develop bronchiectasis before the age of 5 years, the absence of bronchiectasis at 9 years of age represents an ‘extreme’ phenotype.²³² Such phenotypes are valuable when identifying determinants of disease severity.^{233 234}

In CF, registry data are highly informative. Traditionally, measures of lung disease severity were collected from around 6 years of age, which is typically when children can reliably perform spirometry. Identifying a biomarker that could be assessed at this age would then enable the tracking of the progression of lung disease using already collected registry data. This could be in addition to assessments of lung structure using CT, which are more sensitive but not performed as widely. This would allow for a much broader population for the validation of any identified biomarkers.

One of the key steps in designing an EWAS is to select relevant tissue. For a study of CF lung disease severity, this should be tissue from the lungs, including immune cells, given the key role of inflammation in the pathophysiology of CF lung disease. BAL represents the ideal method for collecting such tissue because it obtains the cells of interest and is already frequently conducted in clinical practice.²⁰⁴ As discussed in chapter five, a relative disadvantage is the low proportion of epithelial cells, however biospecimens which have a larger proportion of these cells are not collected in routine clinical practice, are more invasive in nature, and omit immune cells which are of relevance to CF lung disease.

Another key aspect of EWAS design is choosing the study population. Specifically, Michels et al.¹⁸⁷ highlight the importance of controlling for confounding factors that may affect the outcome of interest or DNA methylation. Relevant factors that need to be controlled in the proposed EWAS include those that may affect CF lung disease severity (e.g. pancreatic status or *CFTR* genotype) and DNA methylation (sex and age).

The final EWAS design aspect highlighted by Michels et al.,¹⁸⁷ which affects the choice of both biospecimens and analysis, is controlling for cell composition in the sample. The panel of reference epigenomes identified in Chapter 5 will allow for deconvolution of BAL cell composition in the present EWAS.

The EWAS described in this chapter involved the identification of two groups of subjects matched for all relevant factors but who differed based on the presence or absence of bronchiectasis on CT scan at 9 years of age. The genome-wide DNA methylation patterns of BAL specimens collected at 6 years of age were then assessed to determine whether they could predict the presence or absence of bronchiectasis at 9 years of age. The study was designed to adhere to best practice guidelines for EWAS design and was well positioned to identify DNA methylation-based biomarkers that can predict CF lung disease severity.

6.2 METHODS

6.2.1 Ethics

The BAL samples used in this study were obtained as part of the AREST CF study, which has received appropriate ethics committee approval (HREC #25054). The parents of all participants provided informed consent for their children's participation in the study.

6.2.2 Subjects

Given that the majority of people with CF develop structural lung disease in their preschool years, those at 9 years of age without bronchiectasis are in the minority. As a result, these children were viewed as 'cases' and identified first before being matched to those with bronchiectasis. All Melbourne-based subjects in the AREST CF study who had undergone a CT scan at 9 years of age were reviewed for inclusion.

6.2.3 Computed tomography classification

Volumetric CT scans were obtained from subjects, with images acquired at peak inspiration and end expiration. The radiology reports of CT scans obtained from Melbourne-based children at 9 years of age were reviewed. All reports that described the absence of bronchiectasis, borderline bronchiectasis, minimal bronchiectasis or mild bronchiectasis were chosen for further review by three clinicians trained in assessing CT scans in CF. Reviewers classified CT scans into those showing bronchiectasis and those showing no bronchiectasis. One clinician reviewed the scans twice to allow for calculation of both inter- and intra-observer reliability. Only cases for which all reviewers agreed that there was no bronchiectasis were chosen for further analysis as cases. Inter- and intra-observer reliability were determined using publicly available software for calculating fixed marginal kappa measures of agreement.²³⁵

6.2.4 Subject matching

The AREST CF biobank was checked for the availability of BAL samples obtained at 6 years of age from subjects identified as having no bronchiectasis. For subjects with available samples, relevant demographic data, including sex, date of birth, *CFTR* genotype, pancreatic status, age at 6-year-old BAL and age at 9-year-old CT, were collected. A list of Melbourne-based AREST CF subjects who had undergone both a CT scan at 9 years of age and a BAL at 6 years of age (available in the biobank) was then

extracted. Subjects without bronchiectasis were then manually matched to subjects with bronchiectasis based on the following criteria: sex, age, *CFTR* genotype, pancreatic status, age at 6-year-old BAL and age at 9-year-old CT. A paired rather than a group approach to matching was undertaken.

6.2.5 DNA extraction and genome-wide methylation analysis

DNA was extracted from stored raw BAL aliquots using the Qiagen DNA Blood Mini Kit. DNA quantity and quality were assessed using a Qubit® fluorometer and NanoDrop™ spectrophotometer, respectively. Once extracted, the DNA was stored at –30°C until further analysis.

Genome-wide DNA methylation analysis was performed using the EPIC array, which measures methylation at 850,000 CpG sites throughout the genome. This was performed at GenomeScan (Leiden, Netherlands). Initial quality control involving repeat assessment of DNA quality and quantity was carried out. A second quality control step, which used qPCR and a melting curve analysis to assess the quality of bisulphite conversion, was also performed.

6.2.6 DNA methylation analysis

DNA methylation analysis was conducted in the R programming language²⁰⁷ following best practice for analysis of such datasets.²⁰⁸ The analysis pipeline is summarised in Figure 6.2.

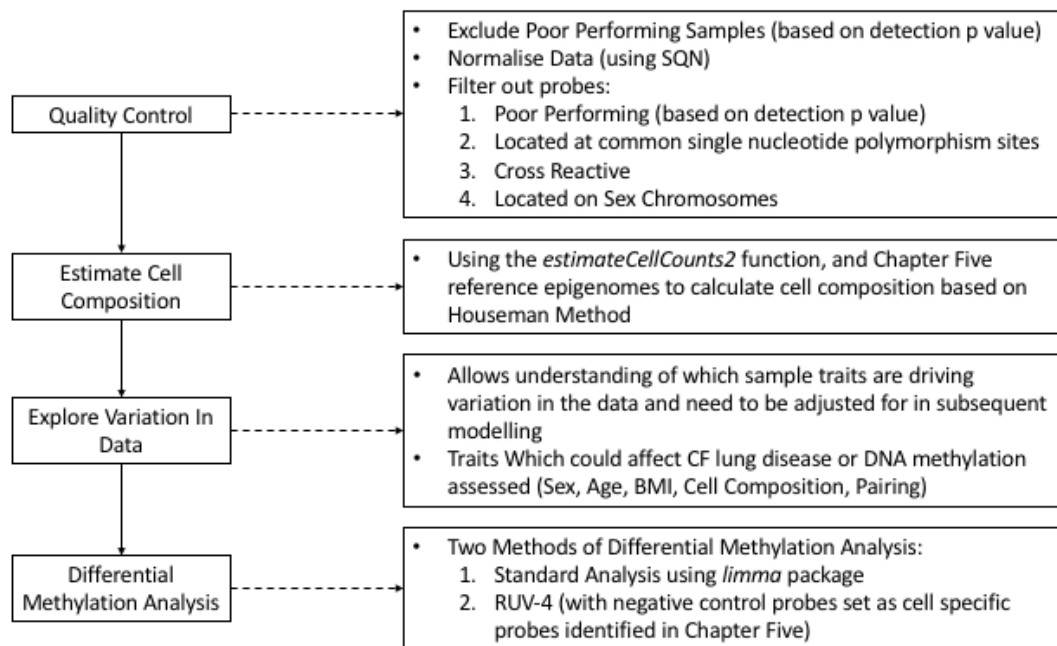


Figure 6.2. Summary of the analysis pipeline. The boxes on the left show the four broad steps, and those on the right outline what was involved in each step.

The initial quality control steps were the same as described in Chapter 5. First, the detection p -value was calculated for each CpG probe in each sample as a measure of the quality of the array data. Any sample with a mean detection p -value > 0.01 was defined as poor quality and excluded from further analysis. Data were then normalised using SQN,²¹¹ as implemented in the *minfi*²⁰⁹ Bioconductor²¹⁰ package. SQN minimises unwanted variations within and between samples. Normalisation is recognised as a routine and essential step in the analysis of high-throughput technologies because it reduces variations related to systematic issues such as sample preparation and array processing, thus improving the comparability of data in downstream analysis.²³⁶ Initial data exploration was performed using MDS plots. This process involves colouring the data points by variables of interest (such as sex or bronchiectasis v. no bronchiectasis). This can be helpful because it can help inform the downstream analysis. For example, if an unwanted source of variation in the data is identified, it can be incorporated in subsequent linear modelling. Next, poor-performing probes—defined as those with a detection p -value of > 0.01 in one or more samples—were excluded. CpG probes located on sex chromosomes that were known to have SNPs at the CpG site or cross-reactive probes that were known to map to multiple places in the genome were also removed.²¹²

Following quality control, the data were again visualised using MDS plots. Reassessment of the data was important because exclusion of probes may have altered the appearance

of the data in the MDS plots. Similar to the previous study, understanding sources of variation in the data will inform subsequent analyses.

The cell composition of the samples was then determined using the reference-based methods developed in Chapter 5. The relative contribution of each of the known sample traits (sex, age at BAL, estimated cell composition, BMI, pairing, bronchiectasis v. no bronchiectasis) was correlated with the first 20 principal components to identify the traits that may be contributing to variation and need to be controlled for.

Selected traits (cell composition, BMI and pairing) were then included in the model for differential methylation analysis. Differential methylation analysis was performed using the *limma* package. A CpG probe-wise linear model, including selected traits, was fitted to determine the differences in methylation between subjects with and without bronchiectasis. Differentially methylated probes (DMPs) were identified using empirical Bayes moderated t-tests,²³⁷ and robust empirical Bayes shrinkage of probe-wise variances was incorporated to protect against hypervariable probes.²³⁸ *p*-values were adjusted for multiple testing using the Benjamini–Hochberg method.²³⁹ The threshold for significant DMPs was a false discovery rate (FDR) of 0.05.

In addition to standard differential methylation analysis, an analysis using the ‘remove unwanted variation (RUV), 4-step’ (RUV-4) method was performed.²⁴⁰ A similar strategy was described by Jaffe et al.,²⁴¹ who found that RUV could be utilised in differential methylation analysis to account for cell type composition. This approach requires the identification of negative control probes that can capture variations associated with confounding factors but are not associated with the factor of interest.²⁴² In this case, the negative control probes are designated as the probes associated with cell type, as defined using the reference panel identified in Chapter 5. The 576F negative control probes were defined as a combination of cell type–discriminating probes from the *any* and *both* sets into a single non-redundant set of cell type–discriminating probes (see Figure 6.3). This analysis did not take sample pairing into account. The same analysis was repeated with the data restricted to only probes that mapped to the validated non-*CFTR* gene modifiers identified in Chapter 2.

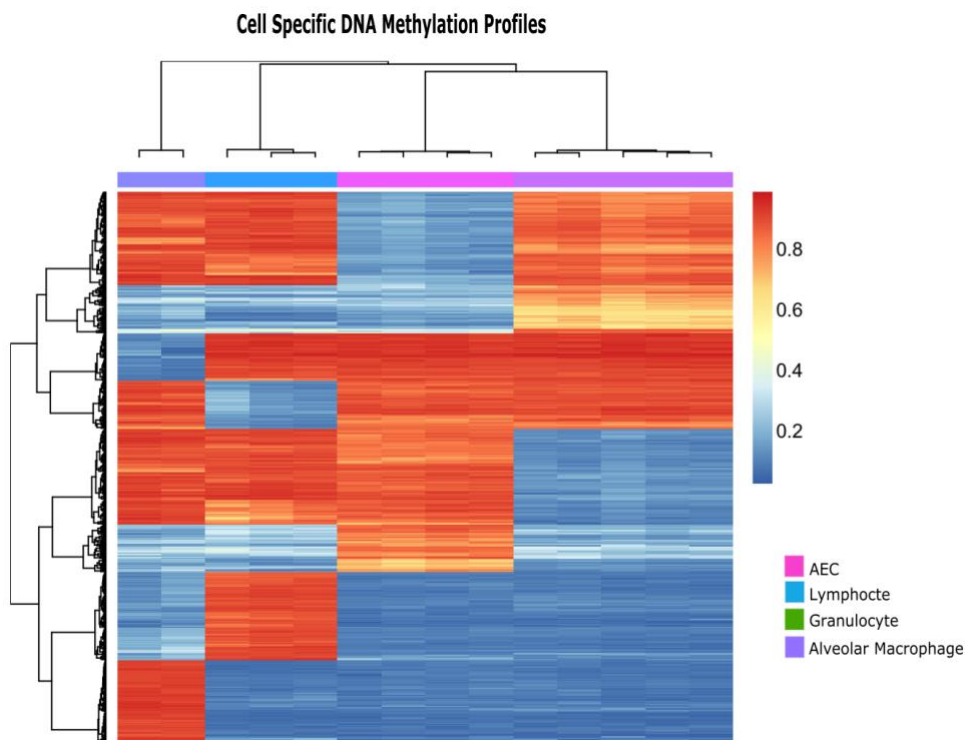


Figure 6.3. Heatmap demonstrating cell-specific probes (identified in Chapter 5) and used as negative control probes in the RUV-4 analysis.

6.3 RESULTS

6.3.1 Subject selection

Initially, 60 CT scans were reviewed for inclusion, and of these, 34 (56.7%) were chosen for complete review. The results of three reviewers' assessment are shown in Table 6.1. Intra-observer agreement for Reviewer 1 was 91.2%, with a fixed marginal kappa value of 0.74 (95% confidence interval [CI] [0.25, 1.00]). Inter-observer agreement was 77.5%, with a fixed marginal kappa value of 0.48 (95% CI [0.30, 0.67]). There were 16 CT scans for which all reviewers agreed there was no bronchiectasis.

Table 6.1. Results of computed tomography review

AREST ID	Reviewer 1 (first review)	Reviewer 1 (second review)	Reviewer 2	Reviewer 3
M1C024	Bronchiectasis	Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C024	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C003	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C022	No bronchiectasis	No bronchiectasis	No bronchiectasis	Bronchiectasis
M1C072	No bronchiectasis	No bronchiectasis	No bronchiectasis	Bronchiectasis
M1C016	No bronchiectasis	Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C021	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C006	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C012	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C023	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C013	No bronchiectasis	No bronchiectasis	Bronchiectasis	Bronchiectasis
M1C026	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C035	Bronchiectasis	No Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C029	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C032	Bronchiectasis	No Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C036	Bronchiectasis	Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C040	Bronchiectasis	Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C038	Bronchiectasis	Bronchiectasis	Bronchiectasis	Bronchiectasis
M1C057	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C045	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C041	Bronchiectasis	Bronchiectasis	Bronchiectasis	No Bronchiectasis
M1C044	No bronchiectasis	No bronchiectasis	Bronchiectasis	Bronchiectasis
M1C046	No bronchiectasis	No bronchiectasis	Bronchiectasis	Bronchiectasis
M1C051	No bronchiectasis	No bronchiectasis	Bronchiectasis	No bronchiectasis
M1C048	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C054	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C055	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C052	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C053	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis
M1C068	Bronchiectasis	Bronchiectasis	No Bronchiectasis	No Bronchiectasis
M1C080	No bronchiectasis	No bronchiectasis	Bronchiectasis	No bronchiectasis
M1C144	No bronchiectasis	No bronchiectasis	Bronchiectasis	No bronchiectasis
M1C065	No bronchiectasis	No bronchiectasis	No bronchiectasis	Bronchiectasis
M1C076	No bronchiectasis	No bronchiectasis	No bronchiectasis	No bronchiectasis

Note: Scans for which there was agreement of no bronchiectasis are highlighted in blue.

6.3.2 Subject matching

Of the 16 subjects with no bronchiectasis on CT at 9 years of age, 14 had undergone BAL at 6 years of age, with samples available in the AREST CF biobank. These 14 subjects were then paired with subjects with bronchiectasis based on all relevant criteria (see Table 6.2). All pairs were matched exactly for sex (eight female pairs and six male pairs), and all subjects had pancreatic insufficiency. Ten of the pairs consisted of two subjects who were Phe508del homozygous. Of the remaining four pairs, two pairs comprised subjects who were Phe508del/Class I compound heterozygotes, and one pair comprised subjects who were Phe508del/unknown compound heterozygotes. In the fourth pair, the subject with no bronchiectasis was Phe508del/unknown, and the subject with bronchiectasis was a Phe508del/Class III compound heterozygote. The mean difference (*SD*) in age at BAL of the pairs was 0.14 years (± 0.15 years). The mean difference (*SD*) in age at the 9-year-old CT scan was 0.89 years (± 0.93 years). Subjects with no bronchiectasis were slightly older at the time of BAL (6.00 ± 0.16 years v. 5.98 ± 0.10 years) and CT scan (9.6 ± 0.91 years v. 9.1 ± 0.56 years). They were also a marginally older cohort (age at time of analysis: 11.6 ± 1.18 years v. 11.5 ± 1.18 years).

Table 6.2: Subject characteristics

Pair	ID	Status	Sex	CFTR Genotype 1	CFTR Genotype 2	CFTR Mutation Class 1	CFTR Mutation Class 2	Pancreatic Status	Age at Matching	Age at BAL	Age at CT
1	M1C057	No Bronchiectasis	Female	Phe508del	Gln414GlnfsX1	II	I	PI	10.82	6.02	9.16
	M1C037	Bronchiectasis	Female	Phe508del	P.Ile864SerfsX28	II	I	PI	11.27	6.12	9.06
2	M1C023	No Bronchiectasis	Female	Phe508del	1717-1G>A	II	I	PI	13.37	6.05	11.93
	M1C008	Bronchiectasis	Female	Phe508del	Gly542X	II	I	PI	12.79	5.89	8.86
3	M1C054	No Bronchiectasis	Female	Phe508del	c.1006insG	II	Not Classified	PI	11.43	6.05	8.99
	M1C089	Bronchiectasis	Female	Phe508del	Unknown	II	Unknown	PI	10.85	6.05	10.43
4	M1C055	No Bronchiectasis	Female	Phe508del	c.1006insG	II	Not Classified	PI	11.43	6.32	8.99
	M1C032	Bronchiectasis	Female	Phe508del	Gly551Asp	II	III	PI	11.32	6.02	9.12
5	M1C006	No Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	12.68	5.93	10.65
	M1C014	Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	12.32	5.99	8.36
6	M1C012	No Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	12.55	6.07	9.12
	M1C050	Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	10.42	5.78	9.23
7	M1C052	No Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	10.10	5.55	9.39
	M1C062	Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	9.54	6.08	8.07
8	M1C026	No Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	11.57	6.04	9.37
	M1C061	Bronchiectasis	Female	Phe508del	Phe508del	II	II	PI	9.59	5.98	8.82
9	M1C053	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	10.10	5.99	8.96
	M1C041	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	10.76	5.89	9.10
10	M1C048	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	10.29	5.99	9.14
	M1C025	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	11.60	6.04	9.21
11	M1C045	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	10.78	6.02	8.86
	M1C018	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	11.64	5.82	9.73
12	M1C029	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	11.42	5.97	9.08
	M1C005	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	12.62	6.06	9.32
13	M1C021	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	12.79	6.00	9.59
	M1C004	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	12.71	5.96	8.95
14	M1C076	No Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	13.55	6.03	10.84
	M1C078	Bronchiectasis	Male	Phe508del	Phe508del	II	II	PI	13.29	6.04	9.36

Note: Subjects were matched as pairs based on relevant criteria: sex, *CFTR* genotype and class, pancreatic status (PI = pancreatic insufficient) and age.

6.3.3 DNA extraction

DNA was extracted from all samples and assessed for its suitability for subsequent DNA methylation analysis. All samples passed initial quality control, with sufficient quality

and quantity to proceed to further analysis. All samples also passed quality control assessment post bisulphite conversion. The details of quality control results for each sample are presented in Table 6.3.

Table 6.3. Initial quality control of DNA samples

Samples details				Entry QC					Conversion QC K2				
Well	Index	GS Code	Customer name	Conc [ng/μl] Entry QC	Volume	Total amount	Total amount passed	Visual inspection Gel	Passed QC	Ct	Tm1	Tm2	Passed qPCR QC
A01	1	103516-001-001	GS Human Brain Control Sample	42.3	11.8	500	N/A	N/A	N/A	26.1	74.8	0.0	N/A
B01	2	103516-001-002	A1	18.9	12	202	Passed	Passed	Passed	28.6	73.3	0.0	Passed
C01	3	103516-001-003	52H	61.2	19	1163	Passed	Passed	Passed	27.8	72.6	0.0	Passed
D01	4	103516-001-004	61G	30.2	144	4353	Passed	Passed	Passed	28.2	72.7	0.0	Passed
E01	5	103516-001-005	26G	87.7	144	12628	Passed	Passed	Passed	29.0	72.3	0.0	Passed
F01	6	103516-001-006	14G	63.8	144	9194	Passed	Passed	Passed	28.2	72.8	0.0	Passed
G01	7	103516-001-007	29G	32.2	144	4630	Passed	Passed	Passed	27.9	72.6	79.5	Passed
H01	8	103516-001-008	M1C005F	135.0	143	19305	Passed	Passed	Passed	26.8	73.2	0.0	Passed
A02	9	103516-001-009	54F12	50.4	144	7255	Passed	Passed	Passed	27.6	73.0	0.0	Passed
B02	10	103516-001-010	89C	19.8	19	376	Passed	Passed	Passed	27.8	72.8	0.0	Passed
C02	11	103516-001-011	55F12	23.6	19	448	Passed	Passed	Passed	27.5	72.5	0.0	Passed
D02	12	103516-001-012	32G	56.8	143	8122	Passed	Passed	Passed	28.5	72.3	0.0	Passed
E02	13	103516-001-013	12F11	22.0	144	3168	Passed	Passed	Passed	28.6	72.5	0.0	Passed
F02	14	103516-001-014	50G	71.5	144	10302	Passed	Passed	Passed	28.0	72.3	0.0	Passed
G02	15	103516-001-015	06F	33.1	144	4763	Passed	Passed	Passed	28.0	72.9	0.0	Passed
H02	16	103516-001-016	62H	19.7	144	2836	Passed	Passed	Passed	28.3	72.9	0.0	Passed
A03	17	103516-001-017	23E10	41.9	144	6037	Passed	Passed	Passed	27.7	73.2	0.0	Passed
B03	18	103516-001-018	37H	36.3	144	5228	Passed	Passed	Passed	27.3	72.7	0.0	Passed
C03	19	103516-001-019	53F	53.8	144	7754	Passed	Passed	Passed	28.6	72.2	0.0	Passed
D03	20	103516-001-020	41G	35.8	144	5162	Passed	Passed	Passed	28.3	72.3	0.0	Passed
E03	21	103516-001-021	48I	15.7	144	2260	Passed	Passed	Passed	28.7	72.4	0.0	Passed
F03	22	103516-001-022	25G12	44.2	144	6369	Passed	Passed	Passed	28.8	72.5	0.0	Passed
G03	23	103516-001-023	45G	13.8	144	1994	Passed	Passed	Passed	28.6	72.4	0.0	Passed
H03	24	103516-001-024	18H12	47.7	144	6868	Passed	Passed	Passed	28.3	72.9	0.0	Passed
A04	25	103516-001-025	21E	32.2	144	4630	Passed	Passed	Passed	27.3	73.4	0.0	Passed
B04	26	103516-001-026	04G11	71.2	143	10182	Passed	Passed	Passed	27.5	72.9	0.0	Passed
C04	27	103516-001-027	76B10	47.7	144	6868	Passed	Passed	Passed	28.0	72.6	0.0	Passed
D04	28	103516-001-028	78B10	22.8	19	433	Passed	Passed	Passed	29.5	72.1	0.0	Passed
E04	29	103516-001-029	57G	21.2	144	3046	Passed	Passed	Passed	29.1	72.3	0.0	Passed
F04	30	103516-001-030	08F	50.8	144	7311	Passed	Passed	Passed	29.2	72.4	0.0	Passed
G04	31	103516-001-031	008G	80.0	144	11520	Passed	Passed	Passed	29.0	72.6	0.0	Passed

Note: Initial quality control ('Entry QC') measured DNA concentration to determine DNA quantity. DNA quality was also determined via gel electrophoresis. Quality control post bisulphite conversion was then undertaken ('Conversion QC K2') using quantitative polymerase chain reaction (qPCR) (Ct) and a melting curve analysis (Tm1 and Tm2). Samples passed or failed either step based on predetermined thresholds set by GenomeScan, which performed the EPIC array.

6.3.4 DNA methylation analysis

6.3.4.1 Quality control

Genome-wide DNA methylation analysis was measured using the EPIC array. The mean detection p -value was < 0.01 for all samples; hence, all samples were deemed suitable for further analysis. Data were normalised using SQN.²¹¹ When data were visualised using MDS plots, there was no clustering by disease status but clear clustering by sex (see Figure 6.4). Poor-performing probes, defined as probes with a detection p -value > 0.01 in one or more sample, probes located on sex chromosomes, cross-reactive probes and probes located at a CpG site, which is a common site for SNPs, were all excluded.²¹² This left 695,764 probes for downstream analysis. When the data were visualised at this stage, the effect of sex had been removed. There was still no clear clustering of samples by disease status (see Figure 6.5). Two samples (3 and 4 from the 'no bronchiectasis' group) appeared to drive the variation in the second principal component.

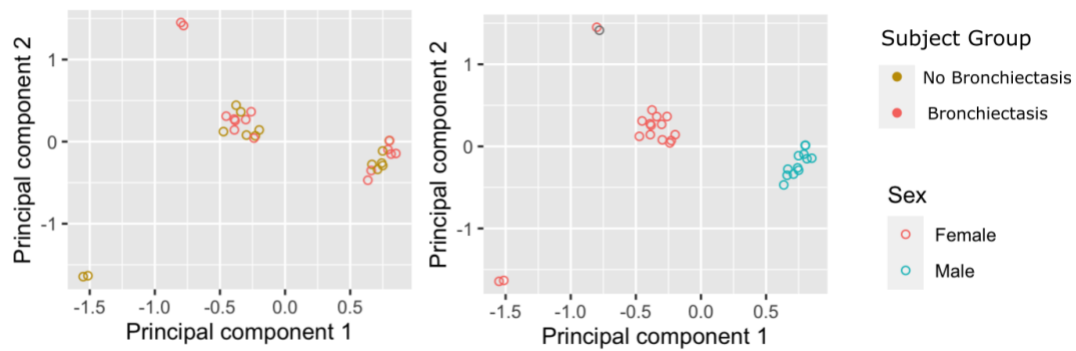


Figure 6.4. Initial data exploration showing that sex was responsible for the variation seen in the first principal component.

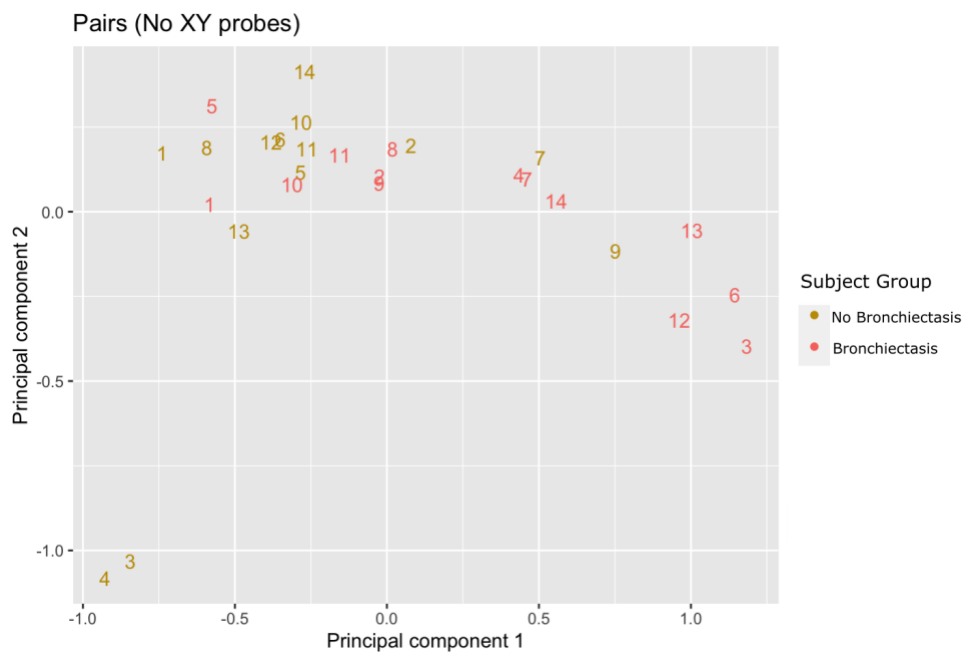


Figure 6.5. Data exploration following quality control. Data points are numbered by their pair number and coloured by disease status. There was no separation by disease status, although Subjects 3 and 4 in the ‘no bronchiectasis’ group appeared to drive the variation in the second principal component.

6.3.4.2 Estimate cell composition

Using the data post quality control and the resources developed in Chapter 5, the cell composition of each sample was estimated. The *estimateCellCounts2* function in R provides two options for probe selection: *any*, which selects the 100 probes in the reference epigenomes with the most statistically significant differential methylation between cell types, regardless of direction of effect, or *both*, which chooses the 50 most

hypermethylated probes and 50 most hypomethylated probes. Both methods were used to estimate the cell composition of the samples (see Figure 6.6). Differences in cell composition between the bronchiectasis and no bronchiectasis groups were then assessed using linear models.²⁴³ Both methods of probe selection resulted in a significantly lower proportion of granulocytes in the no bronchiectasis group, with a corresponding significant increase in macrophages (see Figure 6.7).

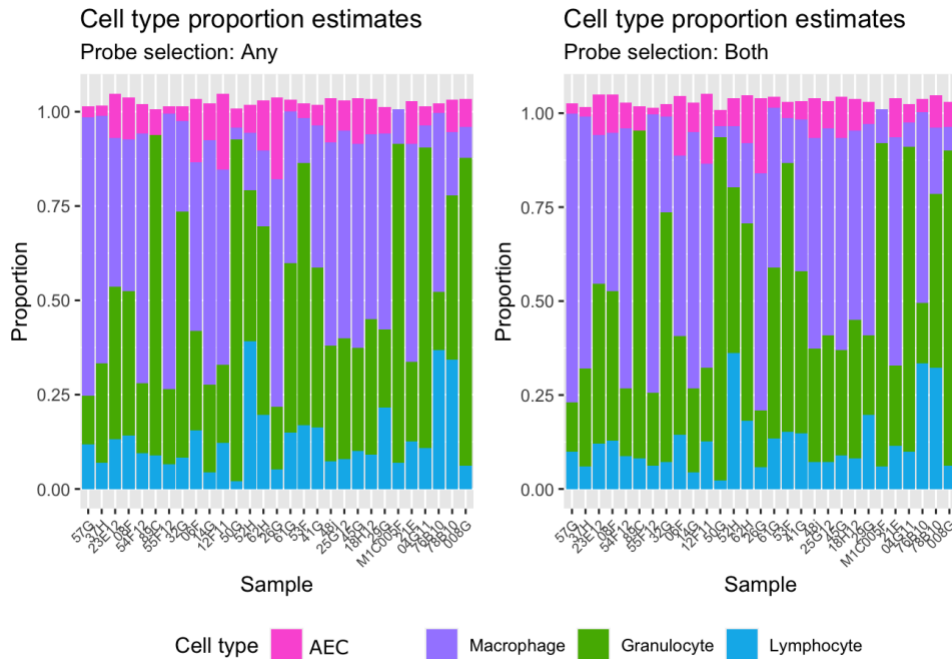


Figure 6.6. Estimated cell proportions of subject samples. Both probe selection options are shown.

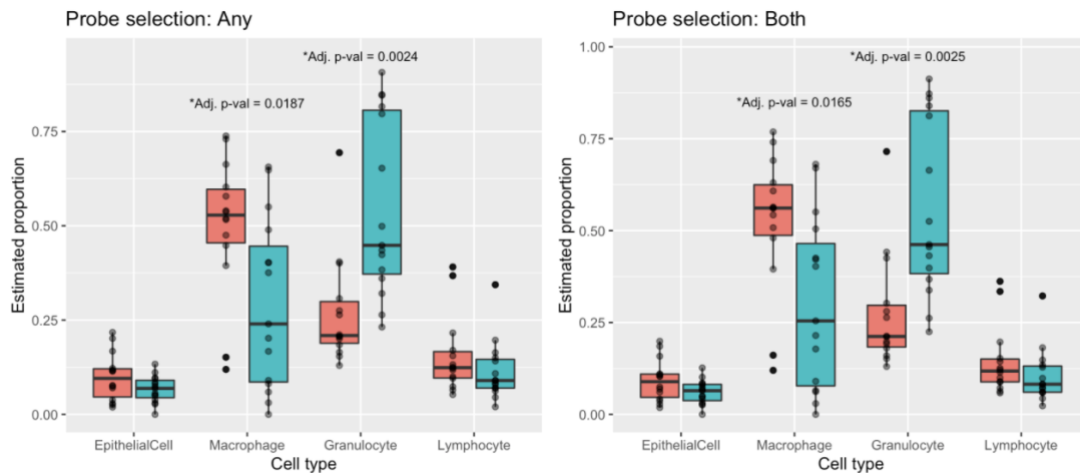


Figure 6.7. Comparison of cell composition of the no bronchiectasis and bronchiectasis groups. There was a significant increase in the granulocyte proportion in the bronchiectasis group and a corresponding increase in macrophages in the no bronchiectasis group.

6.3.4.3 Exploring variations in the data

The aim of data exploration was to understand the sample traits that may be contributing to variations in the data to inform which traits should be included in the differential methylation analysis model. The scree plot in Figure 6.8 demonstrates that most of the variation in the data was captured by the first five principal components. Correlation between known sample traits (pairing, disease status, age at BAL, sex, cell composition and BMI) and the top 20 principal components was then assessed (Figure 6.9). This shows that cell composition most strongly contributed to the first three principal components. Pairing, disease status and age at BAL contribute weakly to the top principal components. Sex and BMI do not appear to contribute to the top five principal components, which capture the majority of the variation in the data, as shown in Figure 6.8.

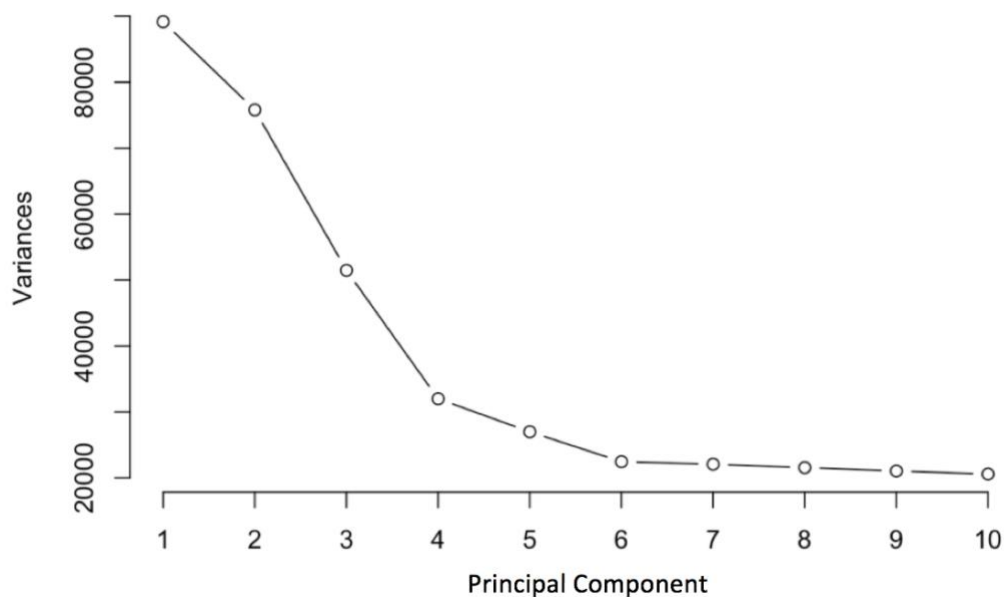


Figure 6.8. Scree plot demonstrating that most of the variation in the data was captured by the first five principal components.

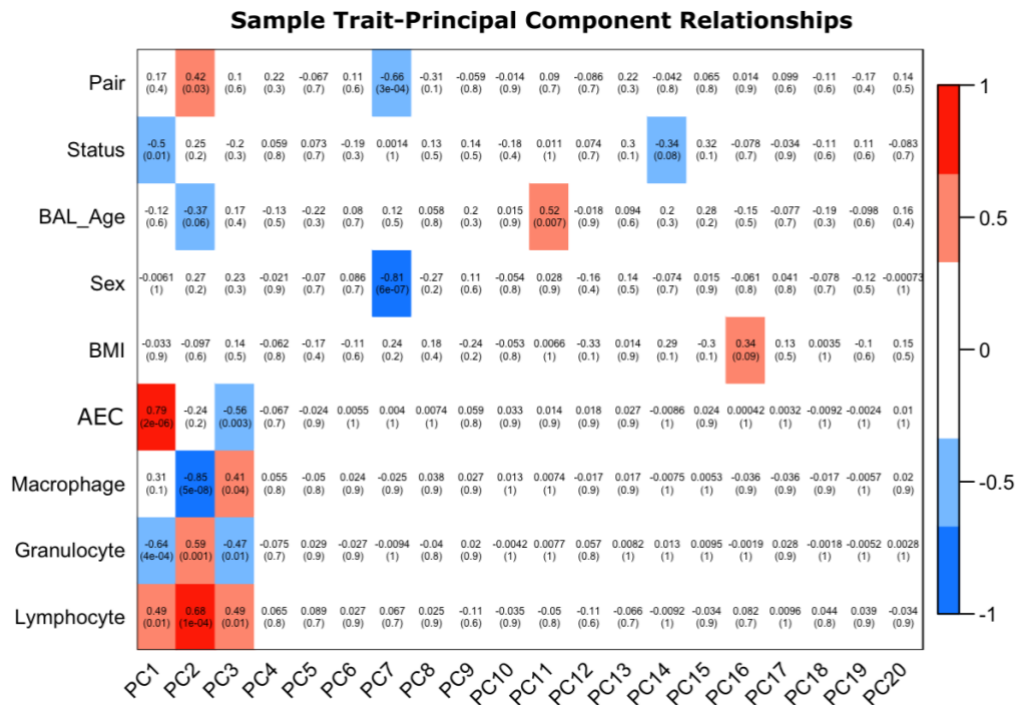


Figure 6.9. Correlation between sample traits and top 20 principal components. The strongest correlations between sample traits and principal components were between cell composition and the first three principal components. Pairing, disease status and BAL age contributed weakly to the top principal components.

To further examine the variations related to cell composition, the sample composition was plotted for each pair (Figure 6.10). Some pairs (1, 2, 5, 7, 10 and 11) had highly similar cell composition across all cell types, whereas others (3, 4, 6, 8, 9, 12, 13 and 14) were divergent, especially for macrophages and granulocytes.

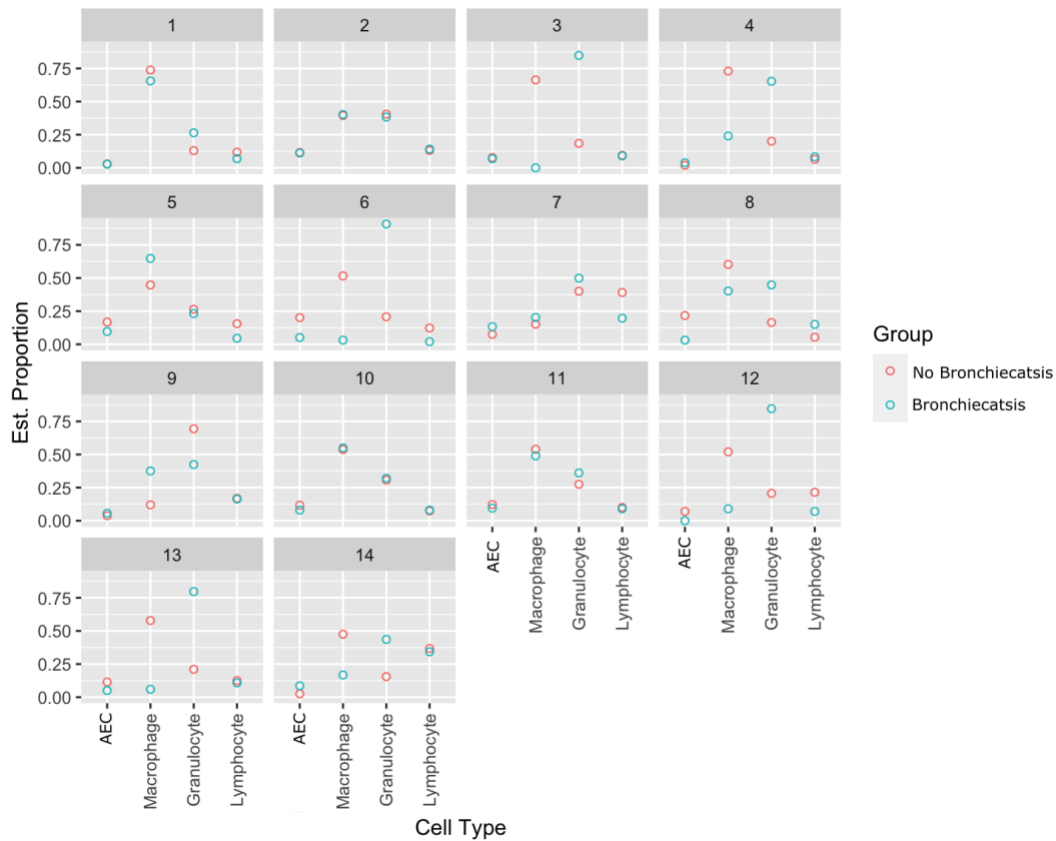


Figure 6.10. Cell composition of samples for each pair.

The MDS plots used to visualise the contribution of cell composition to the top principal components suggest that the proportions of macrophages and granulocytes drove the first principal component (Figure 6.11). With respect to macrophages, there is a clear gradient from samples with a high proportion of macrophages to those with a low proportion across the first principal component. The granulocyte plot shows an inverse relationship.

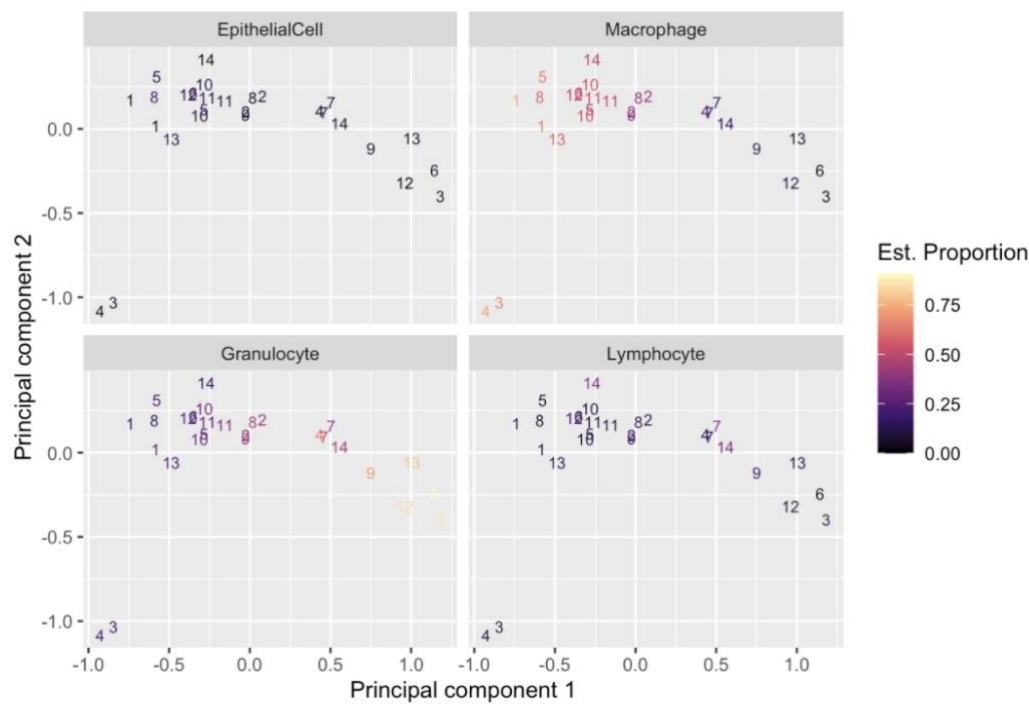


Figure 6.11. MDS plots demonstrating the effect of cell composition on the top principal components. Samples are coloured by the relative proportion of the respective cell type in each plot. The macrophage plot shows a clear gradient between samples with a high and low proportion of macrophages, driving the variation in the first principal component. The inverse relationship is seen in the granulocyte plot. This confirms that these cell types drive the variation seen in the first principal component.

BMI was assessed because it is known to independently affect DNA methylation. When comparing BMI within pairs, some pairs had similar BMIs, while others were divergent (Figure 6.12). When the distribution of BMI was compared for those with and without bronchiectasis, there was less variation in the bronchiectasis group (Figure 6.13); however, when assessed using the Wilcoxon signed-rank test, the difference was not statistically significant ($p = 0.2208$). To determine whether BMI should be included in the model for differential methylation analysis, the distribution of R^2 values for models with and without BMI included was assessed across all CpGs. The superior model was expected to have a greater proportion of CpGs, with an R^2 value greater than 10%. When the two models were compared, the model that included BMI marginally outperformed the one without (56.5% v. 55.9%). This was further assessed via a likelihood ratio test between the two models for each CpG. This demonstrated that including BMI significantly improved the model for 63,307 (8.90%) of all CpGs. Although BMI only appeared to contribute a relatively small amount to the variation in this data, given that it

did significantly improve the model for just under 10% of CpGs, it was retained in the model.

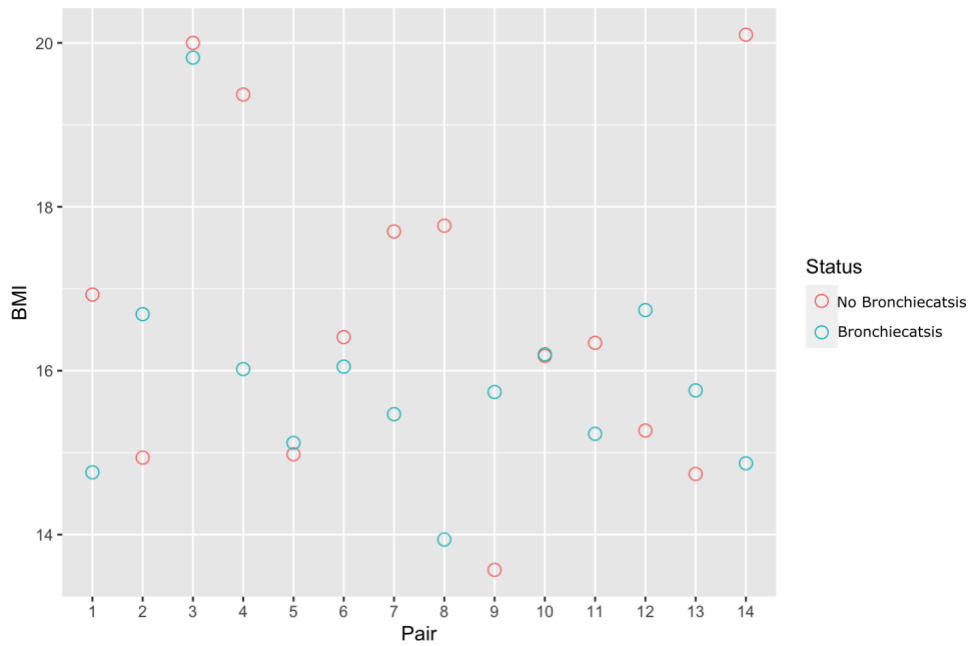


Figure 6.12. Comparison of BMI within pairs. BMI was similar in some pairs and different in others.

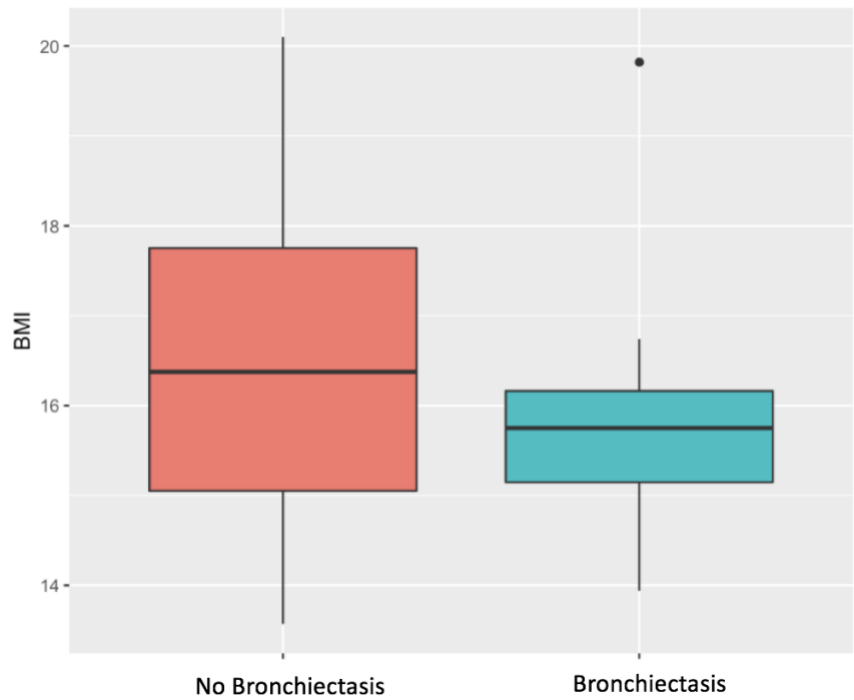


Figure 6.13. BMI distribution of subjects grouped by disease status. The boxes indicate the median and interquartile ranges. Median BMI was slightly lower in the bronchiectasis group, while the no bronchiectasis group had a much wider spread of BMI.

Based on the results of data exploration, the following traits were included in subsequent differential methylation analysis models: cell composition, BMI and pairing.

6.3.4.4 Differential methylation analysis

A CpG probe-wise linear model was fitted to determine the differences in methylation between subjects with and without bronchiectasis. The model included cell composition, BMI and pairing. There were no statistically significant DMPs when subjects with and without bronchiectasis were compared using an FDR of < 0.05 . The top-ranked probes from this analysis are listed in Table 6.4 (see Chapter 6 Appendix).

Examining the data using singular value decomposition (an alternative approach to MDS or principal component analysis), the samples appeared to cluster by disease status (bronchiectasis v. no bronchiectasis) (Figure 6.14). However, when the singular value decomposition was restricted to only the predefined negative control probes, whose methylation patterns should largely reflect cell type differences, the samples once again clustered by disease status (Figure 6.15). This suggests that the observed clustering is reflective of significant differences in cell composition between the two groups, as previously described, rather than true methylation differences between cases and controls.

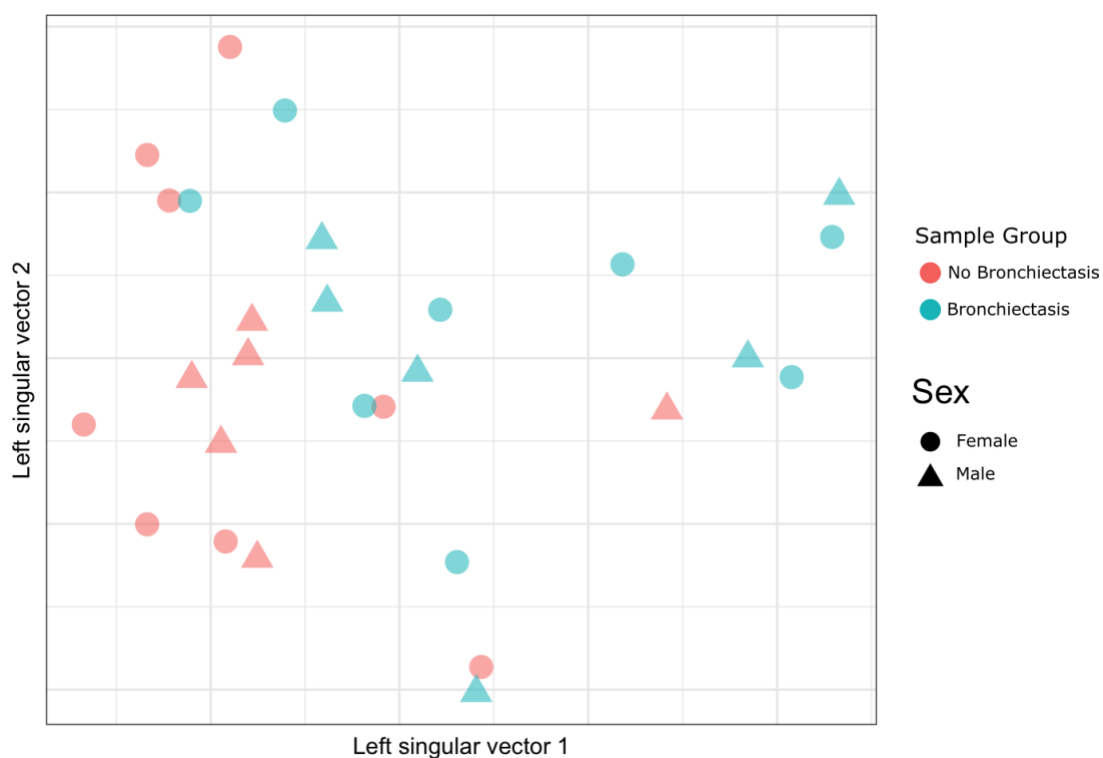


Figure 6.14. Data re-exploration prior to RUV-4 analysis. Samples clustered by disease status but not by sex.

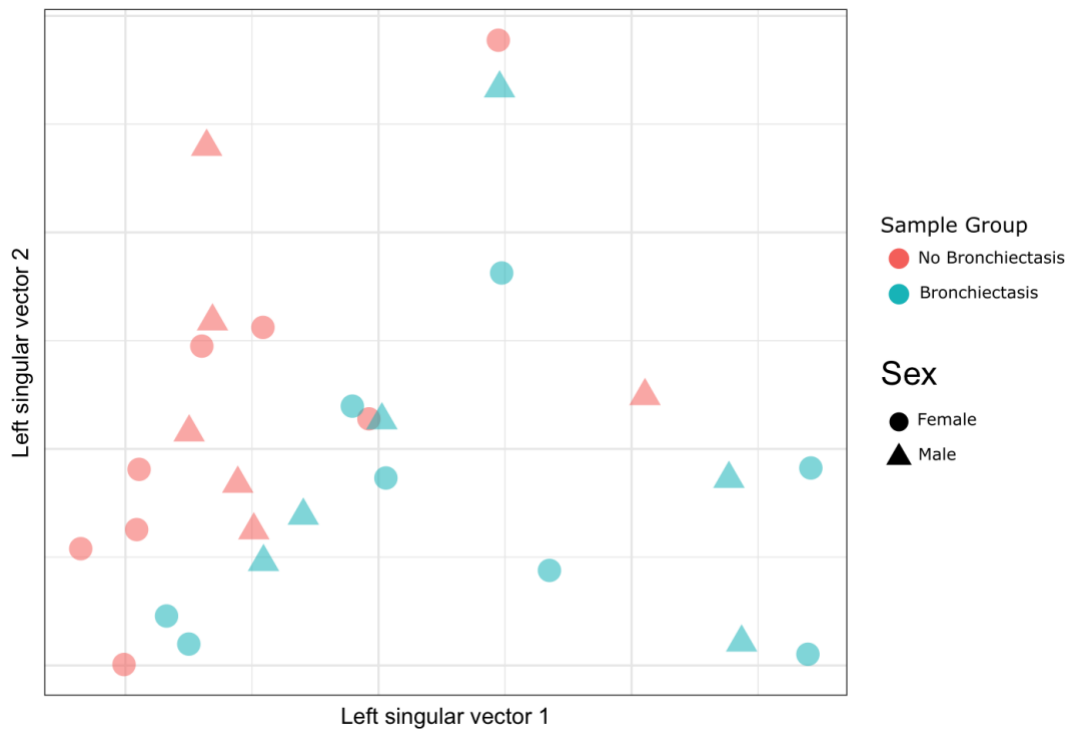


Figure 6.15. Data re-exploration prior to RUV-4 analysis but limited to only negative control probes, which are related to cell type. Again, samples clustered by disease status, not by sex, confirming the different cell compositions of the two groups.

Consequently, unadjusted RUV-4 analysis resulted in 22,081 significant DMPs between the two groups. Because this analysis did not take paired samples into account, these DMPs likely reflect the differences in cell composition observed between the two groups. Further, there were 155 negative control probes found to be differentially methylated between cases and controls. These were excluded as negative control probes for the subsequent adjusted RUV-4 analysis because they were correlated with disease status, indicating that they were inappropriate for use as negative controls (which should not be associated with the outcome of interest).

When the RUV-4 analysis was performed, including adjustment for cell composition based on the negative control probes, there were no significant DMPs at an FDR of < 0.1 . The top-ranked probes from this analysis are listed in Table 6.5 (see Chapter 6 Appendix). This remained true when the analysis was repeated but was limited to probes that mapped to the validated non-*CFTR* gene modifiers identified in Chapter 2 (see Table 6.6 in Chapter 6 Appendix).

6.4 DISCUSSION

This study did not achieve its primary aim of identifying a DNA methylation profile for BAL at 6 years of age that predicts the severity of CF lung disease at 9 years of age. This is despite the strengths of the study, which included robust subject phenotyping, the pairing of subjects with and without bronchiectasis and a sound and thorough analysis. The major limitations of the study were related to its power and the effect of the skewed cell composition between the bronchiectasis and no bronchiectasis groups. Despite not achieving its primary aim, the study will inform the ongoing search for predictive biomarkers of CF lung disease severity by highlighting the importance of choosing samples that are likely to have similar cell compositions, powering the study adequately and evaluating subjects in early life.

The disease status classification undertaken in this experiment was robust. Each CT scan was reviewed four times by three different reviewers. The choice of a dichotomised outcome as opposed to a composite score reflects clinical practice. While multiple CF-specific CT scores exist, including the Brody⁹⁷ and Perth–Rotterdam annotated grid morphometric analysis (PRAGMA)²⁴⁴ scores, these have not been used in clinical practice to date. When CT scans are performed clinically, the presence or absence of bronchiectasis is one of the outcomes assessed and used to inform clinical practice. The intra-observer agreement in this study was 91.2%. This is similar to that reported by others. The original description of the Brody score reported an intra-observer agreement of 84%, and the PRAGMA score had an intra-observer agreement of 92.8% for bronchiectasis assessment.^{97 244} The inter-observer agreement of 77.5% was less than the that reported in other studies, which has ranged from 78% to 90.7%.⁴⁰ The reduced agreement between the three reviewers would have biased scans towards being classified as having bronchiectasis present. The limitation associated with this classification method is the limited number of subjects being classified as having no bronchiectasis, which may have reduced the power of the study. However, this also reflects a strength of the study because the no bronchiectasis classification was robust.

Following classification, 16 of 60 (26.7%) scans were identified as being bronchiectasis free. Few studies have assessed the prevalence of bronchiectasis at 9 years of age. In a group of 16 children with a median (range) age of 9.3 years (2.8–11.9 years), Marchant et al.⁹⁸ found that 100% had bronchiectasis. In a larger study of 117 subjects with a mean age of 12.2 years (range = 0.25–32 years), Helbich et al.⁹⁹ found that 19.7% were

bronchiectasis free. Gustafsson et al.⁵⁸ studied 44 children aged between 5 and 19 years with a mean age of 12 years and found that 41% were bronchiectasis free. Given that no studies have directly assessed the prevalence of bronchiectasis in 9-year-old children with CF, it is difficult to directly compare the prevalence found in this study with previous works, but it does fall within the range of 0–41% of bronchiectasis-free subjects reported by others at a similar age. The prevalence of bronchiectasis in this study (73.3%) may also underestimate the true prevalence in both the AREST CF cohort and CF clinics. The denominator used to calculate prevalence was children who had undergone a CT at 9 years of age rather than all subjects with CF aged 9 years. The decision to perform a CT at 9-years-old is clinically determined by the treating clinician. Patients with well-established bronchiectasis or lung function impairment may not have had a CT scan performed at 9 years because it is unlikely to yield new information. Conversely, patients who are free of bronchiectasis up until the age of 9 years are likely to have a CT scan performed because identification of the development of bronchiectasis will influence their clinical management. This introduces a potential bias towards the underestimation of the prevalence of bronchiectasis in this study. Regardless, the fact that 26.7% of subjects were bronchiectasis free at 9 years supports this being an extreme phenotype worthy of further investigation. Previous attempts aimed at identifying the determinants of CF lung disease severity have used the mildest and most severe quartiles to define extreme phenotypes for investigation.²³³

The matching process used in this study ensured that subjects with and without bronchiectasis were similar. All pairs were identical for sex and pancreatic status. The majority of pairs (10/14) had identical *CFTR* genotypes, and 3/14 were identically matched for genotype class. The last pair consisted of a Phe508del/unknown class subject with no bronchiectasis being matched with a Phe508del/Class III subject with bronchiectasis. The primary aim of matching genotypes was to ensure that subjects in the mild or no bronchiectasis groups were not predominantly made up of those with *CFTR* genotypes, which are considered milder in nature. Given that the only discrepancy occurred in one pair in which ‘severe’ bronchiectasis subjects had a class III mutation, this was not the case. The age at 6-year-old BAL was similar, with a mean difference in pairs of 0.14 years (± 0.15 years). There was greater variation in age at the 9-year-old CT scan, with a mean difference within pairs of 0.89 years (± 0.93 years). The mean age of subjects without bronchiectasis was higher (9.6 years ± 0.91 years v. 9.1 years ± 0.56 years), confirming that these subjects were definitely free from

bronchiectasis at 9 years of age and in fact had longer than the bronchiectasis group to develop structural lung disease. The mean difference in the age of pairs was 0.79 years (± 0.64 years). The two groups were similar in age at the time of analysis (11.6 ± 1.18 years v. 11.5 ± 1.18 years), confirming that they were from similar eras and would have received similar care. The subjects were not matched for other factors such as microbiology, history of pulmonary exacerbations and nutritional status. This decision was made for several reasons. While there are data suggesting these variables may be associated with lung disease severity, clinicians are unable to use these factors to predict future lung disease severity.²⁴⁵ None of these factors, or indeed the combination of these factors robustly predicts future lung disease severity. The analysis method used, RUV-4, is also specifically designed to identify known and unknown sources of variation in data.²⁴⁰ Thus during analysis if any confounders are influencing the methylation data it can be identified, even if they are not specified *a priori*. Further, the aim was to identify a biomarker that can be applied to all children with CF and not just those with particular clinical histories (for example patients without a history of *Pseudomonas aeruginosa* infection). Lastly, given the limited number of subjects available, adding additional parameters for matching would have reduced the number of included participants, thus further reducing the power of the study.

The DNA extracted from stored samples and raw DNA methylation data was of high quality, with all samples passing quality control. The analysis followed best practice for DNA methylation analysis²⁰⁸ and included adjustment for the sample traits shown to be contributing to variations in the data. This included two different methods for adjusting for the cell composition of the samples: fitting estimated cell type proportions in the linear models and removing variations related to cell composition using the RUV-4 approach. Despite this, no sites with statistically significant differential methylation were identified.

This EWAS reported a negative finding, and several factors may have contributed to this. The design of the EWAS included subject selection, subject matching, DNA methylation analysis and statistical analysis. As discussed, the methods used for these steps were all robust. However, limitations of this study that may have contributed to type II errors are related to the power of the study and the skewed cell composition of the samples. Alternatively, the negative finding could be a valid result, with 6 years of age being too late to detect differences in methylation between groups.

The sample size of 28 was relatively small. Despite the use of an extreme phenotype of no bronchiectasis at 9 years of age, which served to increase the power, the study was still likely to be underpowered. Using estimates from simulated data,²⁴⁶ the sample size in this EWAS would not have reached 80% power, even with a 30% difference in mean methylation between the two groups. There were issues preventing the direct comparison of the simulated data with those in this EWAS. The simulated data were based on calculations that used a different threshold of significance (a p -value of 1×10^{-6} , which corresponded to an FDR smaller than the 1% used in this EWAS). Moreover, the simulations were based on data derived at nine CpG sites in a uniform cell type, which differed from this EWAS. However, the simulated data do suggest that this EWAS was likely underpowered. When this study was developed, there were no established methods for power calculations to inform EWAS design. A publicly available tool for determining the power of EWAS has recently been developed;²⁴⁷ however, it could not be applied to this study because the tool was not designed for studies using BAL.

Despite adjusting for cell composition using two different methods, the effect of skewed cell composition is a limitation of this study. It is likely that any differential methylation between the two groups was linked to differences in cell composition. Hence, when the effect of cell composition was removed, so too was any differential methylation. This is supported by the findings of the RUV-4 analysis, which, prior to removing variations related to cell composition, identified 22,081 significant DMPs between the two groups. However, the aim of this study was to identify potential biomarkers of lung disease severity that could aid in clinical practice. Thus, any differential methylation, which is merely a surrogate measure of differences in cell composition, would not be useful because it would represent a more convoluted method of measuring a parameter that can already be measured in many clinical laboratories. The difference in cell composition between the two groups, driven by an increased proportion of granulocytes in the bronchiectasis group, was not surprising given the previous works showing the role of neutrophilic inflammation in CF lung disease. For example, the proportion of neutrophils in BAL samples correlated cross-sectionally with structural lung disease severity in a group of 23 children with CF aged 1–3 years.²⁰⁵ Other studies have also shown that markers of neutrophilic inflammation, such as free neutrophil elastase in BAL fluid, are associated with the future development and progression of structural lung disease in CF. The presence of free neutrophil elastase in early life has been correlated with disease severity.⁷⁶ In a seminal study, the detection of free neutrophil elastase at 3 months was

shown to predict bronchiectasis at 1 and 3 years of age.⁸⁹ Other works have also demonstrated that free neutrophil elastase is associated with bronchiectasis progression.⁴⁶
²⁴⁸ More recently, an alternate marker of neutrophilic inflammation, methionine sulfoxide, was also correlated with structural lung disease severity.²⁰⁵ The elevated neutrophil proportion in the bronchiectasis group at 6 years of age is in keeping with the evidence that neutrophilic inflammation in early life is associated with structural lung disease.

Another potential limitation of this study is the age at which the epigenetic profile was assessed. Lung disease in CF is present in infancy and persists and progresses through the early years. By the age of 6 years, when methylation was measured in this EWAS, the majority of children have established structural lung disease. As highlighted by the cohort without bronchiectasis in this EWAS, some children with CF will have normal lung structure at this age. However, it is possible that DNA methylation changes, which would likely precede structural lung damage, had progressed in all children with CF by the age of 6 years. Thus, at this age it may not be possible to separate CF subjects based on their epigenetic profiles or other molecular biomarkers. If this concept is true, it would support the evaluation of DNA methylation-based biomarkers in earlier life. For predictive biomarkers determined at 6 years of age, it may be more appropriate to examine biomarkers derived from structural lung disease change to predict future lung disease severity. This approach was successfully used in a recent study in which the severity of structural lung disease changes at 5–6 years of age was able to predict the severity of lung function impairment over the next decade.²⁴⁹

Although the present study did not identify a DNA methylation profile at 6 years of age that can predict disease status at 9 years of age, it generated several useful outcomes. First, the extremely well-phenotyped cohort in the study provides a platform for the investigation of other potential biomarkers of disease severity. The possibilities for further investigation of this cohort are further discussed in Chapter 8. In addition, the results of this EWAS provided useful lessons for a similar study presented in Chapter 7 of this thesis. In particular, this EWAS highlights the need for increased study power as well as choosing samples from earlier life, which are less likely to differ in cell composition. If future studies incorporate these lessons, the likelihood of finding DNA methylation-based biomarkers of future CF lung disease severity will be maximised.

6.5 APPENDIX

Table 6.4. Top-ranked probes from standard differential methylation analysis using *limma*. (logFC: estimate of the log2-fold-change corresponding to the effect. AveExpr: average log2-expression for the probe over all arrays and channels. t: moderated t-statistic. P.Value: raw p-value. adj.P.Val: adjusted P-value. B: log-odds that the gene is differentially methylated. Chr: chromosome the probe is located on. pos: genomic coordinates of probe based on University of California, Santa Cruz (UCSC) genome browser. UCSC_RefGene_Name: Name of the gene that the probe maps to. UCSC_RefGene_Group: Location in gene that the probe is located.)

	logFC	AveExpr	t	P.Value	adj.P.Val	B	chr	pos	UCSC_RefGene_Name	UCSC_RefGene_Group
cg08086906	-0.6312749	-3.9721143	-5.963707	3.68E-06	0.5787279	-0.1384698	chr16	86588932	MTHFSD;MTHFSD;MTHFSD;MTHFSD;MTHFSD;MTHFSD;MTHFSD;FLJ30679	TSS200;TSS200;TSS200;TSS200;TSS200;TSS200;Body
cg14753740	-1.1841422	-1.8155914	-5.904091	4.26E-06	0.5787279	-0.1864078	chr5	54468940	CDC20B;CDC20B;MIR449C;CDC20B;CDC20B;CDC20B;CDC20B	1stExon;1stExon;TSS1500;5'UTR;1stExon;5'UTR;5'UTR
cg05050482	0.5974218	-3.5470879	5.875938	4.57E-06	0.5787279	-0.2092122	chr6	30710482	FLOT1	TSS200
cg09786634	-1.0748715	-1.321128	-5.836888	5.03E-06	0.5787279	-0.2410202	chr17	78912393	RPTOR;RPTOR	Body;Body
cg09069203	0.4528994	1.8458068	5.829601	5.10E-06	0.5787279	-0.2443242	chr1	214251836		
cg06440065	-0.7147448	-2.4002866	-5.830922	5.11E-06	0.5787279	-0.2458981	chr19	30206617	C19orf12;C19orf12	TSS200;TSS1500
cg23610373	0.5430589	-2.6791507	5.738648	6.42E-06	0.5787279	-0.3219518	chr1	149298587		
cg03641032	-0.4436379	0.8777105	-5.731297	6.51E-06	0.5787279	-0.3254974	chr17	78911768	RPTOR;RPTOR	Body;Body
cg11949518	-1.9388687	0.8334538	-5.681212	7.41E-06	0.5854486	-0.3698704	chr17	78912765	RPTOR;RPTOR	Body;Body
cg24545338	0.6874729	-2.7777173	5.483738	1.21E-05	0.774365	-0.5379973	chr14	31343687	COCH;COCH	TSS200;TSS200

Table 6.5. Top-ranked probes from standard differential methylation analysis using RUV-4. (F.p: F test p-values for testing all of the factors simultaneously. F.p.BH: False Discovery Rate (FDR) adjusted F test p-values. Adjustment via Benjamini Hochberg method. p_X1.Control: P-values for the factor of interest. p.BH_X1.Control: FDR-adjusted p-values. b_X1.Control: The estimated coefficients of the factor(s) of interest. sigma2: Estimates of the features' variances using the empirical Bayes shrinkage method of Smyth. var.b_X1.Control: "Empirical Bayes" estimate of the variance of coefficient estimates. fit.ctf: Vector of negative control genes. Mean: Mean adjusted M-value across all samples. Chr: chromosome the probe is located on. pos: genomic coordinates of probe based on University of California, Santa Cruz (UCSC) genome browser. strand: + is the forward strand and – is the reverse strand. UCSC_RefGene_Name: Name of the gene that the probe maps to. UCSC_RefGene_Group: Location in gene that the probe is located.)

Probe	F.p	F.p.BH	p_X1.Control	p.BH_X1.Control	b_X1.Control	sigma2	var.b_X1.Control	fit.ctf	mean	chr	pos	strand	UCSC_RefGene_Name	UCSC_RefGene_Group
cg09069203	9.16E-07	0.3833844	9.16E-07	0.3833844	-0.4449318	0.02419181	0.005077728	FALSE	1.841098	chr1	214251836	-		
cg14573826	1.64E-06	0.3833844	1.64E-06	0.3833844	0.5069386	0.03370886	0.007075304	FALSE	2.666418	chr14	65409667	+	GPX2;GPX2;GPX2;CHURC1-FNTB;CHURC1-FNTB	TSS200;TSS200;TSS200;Body;Body
cg06440065	1.65E-06	0.3833844	1.65E-06	0.3833844	0.6770298	0.06016182	0.012627636	FALSE	-2.394294	chr19	30206617	-	C19orf12;C19orf12	TSS200;TSS1500
cg24545338	2.92E-06	0.4069494	2.92E-06	0.4069494	-0.6848828	0.06608186	0.013870219	FALSE	-2.77735	chr14	31343687	+	COCH;COCH	TSS200;TSS200
cg13349671	3.39E-06	0.4069494	3.39E-06	0.4069494	0.5228843	0.0392597	0.008240394	FALSE	-3.289752	chr12	120105332	+	PRKAB1	TSS1500
cg14527280	3.82E-06	0.4069494	3.82E-06	0.4069494	-0.7601811	0.08424962	0.017683531	FALSE	-3.090168	chr6	3458119	-	SLC22A23	TSS1500
cg26586287	4.09E-06	0.4069494	4.09E-06	0.4069494	0.4594433	0.03105616	0.006518517	FALSE	3.786366	chr11	94613673	+		
cg25963583	5.01E-06	0.4353941	5.01E-06	0.4353941	-0.6175816	0.0575978	0.012089461	FALSE	1.122894	chr14	65539104	-	MAX;MAX	Body;Body
cg17410093	6.57E-06	0.451546	6.57E-06	0.451546	0.3636232	0.02069317	0.004343382	FALSE	2.447854	chr2	232650620	+	COPS7B;COPS7B;COPS7B;COPS7B;COPS7B;COPS7B	TSS1500;TSS1500;TSS1500;TSS1500;5'UTR;Body
cg12202341	6.62E-06	0.451546	6.62E-06	0.451546	0.3929656	0.02419332	0.005078044	FALSE	3.223121	chr9	131977681	-		

Table 6.6. Top-ranked probes from standard differential methylation analysis using RUV-4, limited to probes that map to validated non-*CFTR* gene modifiers. (F.p: F test p-values for testing all of the factors simultaneously. F.p.BH: False Discovery Rate (FDR) adjusted F test p-values. Adjustment via Benjamini Hochberg method. p_X1.Control: P-values for the factor of interest. p.BH_X1.Control: FDR-adjusted p-values. b_X1.Control: The estimated coefficients of the factor(s) of interest. sigma2: Estimates of the features' variances using the empirical Bayes shrinkage method of Smyth. var.b_X1.Control: "Empirical Bayes" estimate of the variance of coefficient estimates. fit.ctl: Vector of negative control genes. Mean: Mean adjusted M-value across all samples. Chr: chromosome the probe is located on. pos: genomic coordinates of probe based on University of California, Santa Cruz (UCSC) genome browser. strand: + is the forward strand and – is the reverse strand. UCSC_RefGene_Name: Name of the gene that the probe maps to. UCSC_RefGene_Group: Location in gene that the probe is located.)

Probe	F.p	F.p.BH	p_X1.Control	p.BH_X1.Control	b_X1.Control	sigma2	var.b_X1.Control	fit.ctl	mean	chr	pos	strand	UCSC_RefGene_Name	UCSC_RefGene_Group
cg16636767	0.000526803	0.4879727	0.000526803	0.4879727	0.2430327	0.01750993	0.003765788	TRUE	0.5790899	chr11	13694647	+	FAR1	5'UTR
cg11012463	0.001371108	0.4879727	0.001371108	0.4879727	-0.2407009	0.0209218	0.004499564	TRUE	-2.1069535	chr1	33804514	+	PHC2;PHC2	5'UTR;Body
cg21883598	0.001411009	0.4879727	0.001411009	0.4879727	-0.2566826	0.02394248	0.005149211	FALSE	-2.578064	chr15	45404157	+	DUOX2	Body
cg18898017	0.001675752	0.4879727	0.001675752	0.4879727	0.2128519	0.01710562	0.003678834	FALSE	2.3128086	chr17	39681660	+	KRT19	Body
cg10593292	0.002018555	0.4879727	0.002018555	0.4879727	0.2826402	0.03146888	0.006767883	FALSE	1.5974498	chr3	195539133	+	MUC4;MUC4;MUC4;MUC4;MUC4	5'UTR;1stExon;5'UTR;1stExon;1stExon;5'UTR
cg14645244	0.00224991	0.4879727	0.00224991	0.4879727	-0.6061063	0.14841761	0.031919564	FALSE	0.4023847	chr6	32552205	-	HLA-DRB1	Body
cg24767336	0.002294029	0.4879727	0.002294029	0.4879727	0.5765023	0.1348878	0.029009763	FALSE	-2.7719695	chr19	41860095	-	TGFB1	TSS1500
cg09330885	0.005099071	0.748601	0.005099071	0.748601	0.6123106	0.18592748	0.039986657	FALSE	1.0235882	chr3	195536485	-	MUC4;MUC4;MUC4	Body;Body;Body
cg24622817	0.005290914	0.748601	0.005290914	0.748601	-0.1963949	0.01931851	0.004154752	TRUE	-2.8003555	chr22	40859826	-	MKL1;MKL1;MKL1;MKL1	TSS1500;5'UTR;5'UTR;5'UTR
cg09152047	0.005408098	0.748601	0.005408098	0.748601	0.3522835	0.06252729	0.013447486	FALSE	-3.6501088	chr1	110282835	+	GSTM3;GSTM3	Body;Body

CHAPTER 7: Epigenome-Wide Association Study of Cystic Fibrosis Lung Disease Severity at 5 Years of Age

7.1 INTRODUCTION

This EWAS aimed to address Aim 2 and Hypothesis 2 of this thesis:

Aim 2: To identify early-life DNA methylation profiles associated with normal lung structure at 5 years of age in children with CF.

Hypothesis 2: In children with CF, specific features of DNA methylation profiles of BAL fluid at 1 year of age are predictive of bronchiectasis outcomes at 5 years of age.

Two groups of matched children with CF, who differed solely based on the presence or absence of bronchiectasis at 5 years of age, were compared. The genome-wide methylation patterns of BAL specimens collected at 1 year of age were then assessed to determine whether they could predict outcomes at 5 years of age.

As with the previous EWAS, this study was designed to adhere to published best practice guidelines.¹⁸⁷ Tissue-specific samples were used for DNA methylation profiling, and previously developed methods to account for cell composition were used in the analysis. In addition, to ensure robustness, the gold standard measure of lung disease severity was used in the classification of disease status. The outcome of interest, bronchiectasis status at 5 years of age, is of high clinical significance in CF. In CF registry data, 5 years of age is used as the starting point for measuring lung disease severity, primarily because this is the age at which spirometry is commonly performed.²⁵⁰ As a result of improvements in care, the vast majority of children now have normal spirometry at this age.²⁵⁰ However, the majority of children at 5 years of age have bronchiectasis.^{52 95} Thus, being able to predict the presence or absence of this pathology at 5 years of age has the potential to influence how CF clinicians manage patients through their infancy and preschool years. This is often a challenging time for clinicians because of the inability of patients to report symptoms, the insensitivity of physical examinations, the inability to obtain expectorated lower airway samples and the need for general anaesthesia to obtain volume-controlled CT scans. Any biomarkers measurable in infancy that can accurately predict outcomes at

5 years of age would aid clinicians in managing patients, overcoming many current challenges.

This EWAS has several strengths over the study previously described. First, by studying subjects at earlier time points, this study was better powered because more children in the cohort had completed the necessary time points. In addition, the numbers in this EWAS were increased via the inclusion of AREST CF subjects from Perth as well as Melbourne. Finally, rather than taking a paired approach to matching the two groups, matching at the group level was undertaken. This meant that groups did not have to have the exact same numbers of subjects, allowing for the inclusion of more subjects in the bronchiectasis (a common disease status at 5 years of age) group.

BAL performed at 1 year of age is less likely to show major cellular differences between subjects with and without bronchiectasis, meaning that any differences between groups are less likely to be lost when accounting for cell composition. For example, compared with older subjects, 1-year-old subjects with CF have had fewer pulmonary exacerbations, fewer antibiotic courses and fewer environmental exposures, all of which may contribute to altered cell composition. Assessment of DNA methylation at 1 year of age will still enable the measurement of the effect of important determinants of methylation profiles, including genotype and antenatal exposures.

Assessing a biomarker at 1 year of age will be of benefit to clinicians because it will allow for early intervention prior to the onset of symptoms, minimising structural and functional lung impairment. The preschool years are a period of rapid lung growth and development, with early-life impairment resulting in reduced attainment of peak lung function and ongoing impairment in adult life.⁸⁴ Thus, any intervention that minimises pathology during the preschool years will not only reduce early-life morbidity but also have lifelong benefits.

7.2 METHODS

7.2.1 Ethics

The BAL samples and clinical data used in this study were obtained as part of the AREST CF study. The study was approved by the Royal Children's Hospital Ethics Committee (HREC #25054). The parents of all participants provided informed consent for their children's participation.

7.2.2 Subjects

The children included in this study were all subjects from the AREST CF cohort, who had BAL specimens collected (and frozen) at 1 year of age and a CT scan at 5 years of age. Subjects from both Melbourne and Perth were included.

7.2.3 Computed tomography classification

Volumetric CT scans were obtained from subjects under general anaesthesia. As per AREST CF protocols, peak inspiratory images were obtained at an airway pressure of 25 cm H₂O, and end expiratory images were obtained at 0 cm H₂O.²⁵¹ All CT scans were scored using the PRAGMA method²⁴⁴ by scorers who had received appropriate training. The PRAGMA method was specifically designed for use on CT scans of preschool children with CF. It involves the selection of 10 CT slices, equally spaced between the apex and base of the lungs. A square grid is then overlaid on each slice. To adjust for lung size, the size of the grid squares is set to 1/20th of the lung width at the carina. The scorer then reviews each grid and applies a hierarchical scoring system from highest to lowest: bronchiectasis, mucus plugging, bronchial wall thickening, atelectasis and normal. The same method is applied to the expiratory images, and the scorer classifies each cell as either having trapped air or healthy. Disease scores are then calculated based on the summation of these findings. Total disease is determined by dividing the number of cells showing bronchiectasis, mucus plugging or bronchial wall thickening by the total number of annotated cells. The bronchiectasis sub-score is calculated by dividing the number of cells annotated with bronchiectasis by the total number of annotated cells. The intra-observer reliability of this score was established by the same assessor repeatedly scoring the same 30 scans from children with CF, 1 month apart. The intra-observer reliability for the bronchiectasis sub-score was 0.928 (95% CI [0.854, 0.965]), with a mean difference between rescores of 0.32 ($p = 0.005$). Inter-observer reliability was determined by two separate observers scoring the same 30 scans. The bronchiectasis sub-score was 0.907 (95% CI [0.815, 0.955]), with a mean difference between assessors of 0.38 ($p = 0.006$).

Subjects were classified into dichotomised groups according to bronchiectasis status (present or absent). To do so, the bronchiectasis sub-score was used—subjects with a score of 0 were classified as having no bronchiectasis, and all others were classified as having bronchiectasis.

7.2.4 Subject matching

Relevant demographic factors were extracted from the AREST CF database for the included subjects: sex, date of birth, *CFTR* genotype, age at 1 year-old BAL, age at 5-year-old CT scan and CF centre. For continuous variables, unpaired *t*-tests were used to assess for significant differences between the two groups. For categorical variables, Fisher's exact *t*-test was used. For categorical data where one variable had more than two groups, Pearson's chi-squared test was used. Statistical tests were performed using GraphPad Prism version 8.4.3.

7.2.5 DNA extraction and genome-wide methylation analysis

DNA was extracted from stored raw BAL aliquots using the Qiagen DNA Blood Mini Kit. DNA quantity and quality were assessed using a Qubit[®] fluorometer and NanoDrop[™] spectrophotometer, respectively. Once extracted, DNA was stored at -30°C .

Genome-wide DNA methylation analysis was performed using the EPIC array at Erasmus Medical Centre (Rotterdam, Netherlands). Prior to methylation measurement, DNA quality and quantity were again assessed, including by gel electrophoresis.

7.2.6 DNA methylation analysis

DNA methylation analysis was conducted using the R programming language²⁰⁷ following best practice for analysis of such datasets.²⁰⁸ The analysis pipeline is summarised in Figure 7.1.

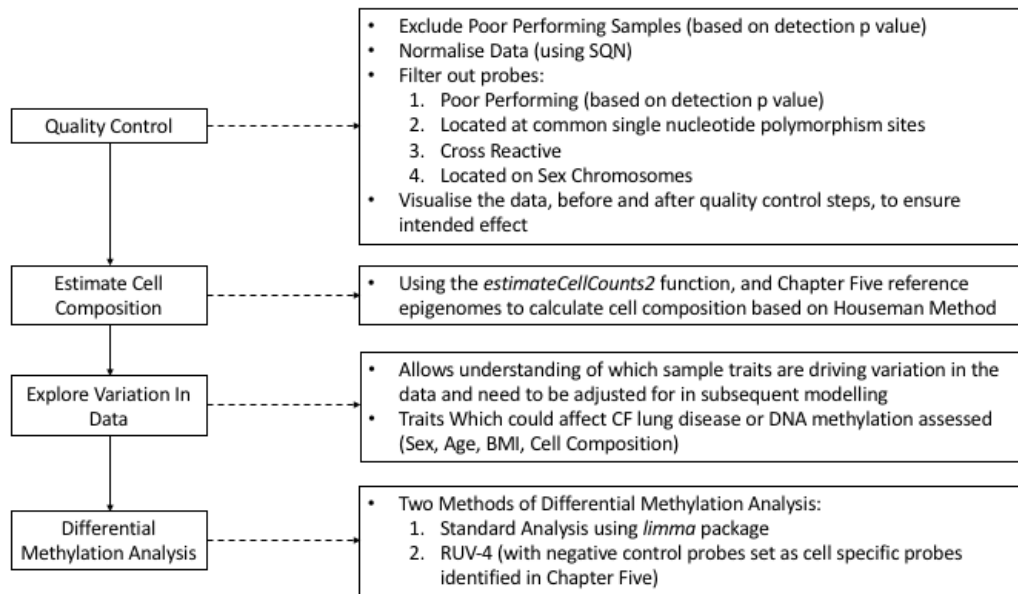


Figure 7.1. Analysis pipeline. The boxes on the left show the four broad steps, and those on the right outline what was involved in each step.

The initial quality control steps were similar to those described Chapters 5 and 6. First, the mean detection p -value was calculated for each sample as a measure of the quality of the array data. Any sample with a mean detection p -value > 0.01 was defined as poor quality and excluded from further analysis. Data were then normalised using the SQN²¹¹ function as implemented in the *minfi*²⁰⁹ package. To assess sources of variation that may inform downstream analysis, data exploration was performed using MDS plots. Next, poor-performing probes—those with a detection p -value > 0.01 in one or more samples—were excluded. Probes located on sex chromosomes, those with SNPs at the CpG site of interest and cross-reactive probes known to map to multiple places in the genome were also removed.²¹²

Following quality control, data were visualised using MDS plots to assess the effect of the quality control steps and identify sources of variation that may need to be accounted for in subsequent linear modelling. The cell composition of the samples was then determined using the reference-based methods developed in Chapter 5. The known sample traits (sex, age at BAL, cell composition, BMI and bronchiectasis vs. no bronchiectasis) were correlated with the first 20 principal components to identify the traits that may be contributing to variation and need to be controlled for.

Differential methylation analysis was performed using the *limma*²⁴³ package. A CpG probe-wise linear model, including estimated cell type proportions for AECs, macrophages, granulocytes and lymphocytes, was fitted to identify differences in

methylation between subjects with and without bronchiectasis. DMPs were identified using empirical Bayes moderated t-tests.²³⁷ Robust empirical Bayes shrinkage of the probe-wise variances was used to protect against hypervariable probes.²³⁸ The Benjamini–Hochberg procedure²³⁹ was used to adjust p -values for multiple testing. The threshold for significant DMPs was an FDR of 0.1.

As described in Chapter 6, a differential methylation analysis using the RUV-4 method²⁴⁰²⁵² was also performed. Rather than explicitly including confounding variables in the linear model, the RUV-4 approach relies on estimating unwanted variations in the data using negative control probes. These negative control probes should capture the variations associated with confounding factors but should not be affected by the factor of interest.²⁴² In this case, similar to the approach used by Jaffe and Irizarry,²⁴¹ the negative control probes were the probes associated with cell type, selected using the reference panel and methods described in Chapter 6. The negative control probes were defined as 200 cell-specific probes per cell type (800 in total). They were identified using *minfi*'s *pickCompProbes* function with probe select set at *any*, meaning that they were the 200 most statistically significant DMPs for each cell type, regardless of the direction of differential methylation (Figure 7.2). The threshold for significant DMPs for the RUV-4 analysis was an FDR of 0.1.

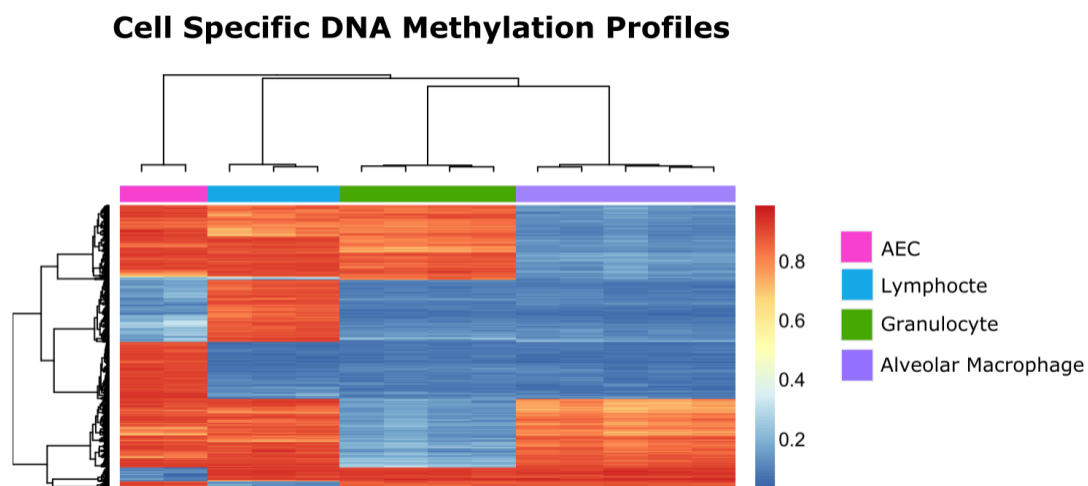


Figure 7.2. Heatmap showing the methylation levels of cell-specific probes used as negative control probes in RUV analysis. There is clear separation of each cell type.

7.2.7 Performance of significant differentially methylated probes as a diagnostic test

Significant DMPs were evaluated for their diagnostic utility using area under the receiver operating characteristics curve (AUC), sensitivity, specificity, positive predictive value

(PPV) and negative predictive value (NPV) for certain thresholds. This analysis was performed using GraphPad Prism version 8.4.3.

7.2.8 Assessment of the functional impact of differentially methylated probes

Publicly available datasets were used to assess the potential functional impact of DMPs. This involved the genomic location of specific DMPs using published genome annotation.²⁵³ If located in proximity to a gene, expression profiles were then assessed using two datasets: Descartes Human Gene Expression During Development atlas²⁵⁴ and the National Institutes of Health Genotype–Tissue Expression project.²⁵⁵ The Gene Ontology Resource²⁵⁶ was used to determine relevant biological processes. The University of California, Santa Cruz (UCSC) Genome Browser was used to assess the location of the DMP in the gene as well as the location of any surrounding CpG islands.²⁵⁷

Methylation patterns around the DMPs of interest were assessed by plotting the mean methylation of each probe located within 10 kb of either side of the DMP for both groups. The Ensembl Regulatory Build database was used to determine whether the DMP was located in a known gene regulatory area in human embryonic stem cells, lung fibroblasts or CD14 monocytes.^{258 259} Further data regarding other epigenetic markers in the region were obtained from the Descartes Human Chromatin Accessibility During Development atlas²⁶⁰ and the Encyclopedia of DNA Elements (ENCODE)²⁶¹ and Blueprint²⁶² projects. The Descartes Human Chromatin Accessibility During Development atlas includes data derived from the assay for transposase-accessible chromatin using sequencing (ATAC-seq) on foetal lung tissue. The data shown has peaks at areas of chromatin accessibility and uses the Cicero co-accessibility score²⁶³ to predict chromatin hubs. These sites are thought to affect gene expression. The ENCODE and Blueprint databases were used to obtain data on epigenetic markers in adult lung tissue, neutrophils and monocytes. The epigenetic markers included H3K4me1 (peaks at enhancer regions), H3K4me3 (peaks in promoter regions), H3K427ac (peaks in both enhancer and promoter regions), CCCTC-binding factor or CTCF (indicates a modification of 3D chromatin structure) and ATAC-seq (peaks indicate accessible chromatin).

7.3 RESULTS

7.3.1 Subject selection

Fifty-eight subjects met the inclusion criteria. Of these, 23 (39.7%) were classified as having no bronchiectasis and 35 (60.3%) as having bronchiectasis. There were no significant differences in the demographic factors between the two groups (Table 7.1). When the *CFTR* genotypes of the two groups were compared (Table 7.2), the bronchiectasis group had a greater proportion of subjects with two class II mutations (54.3% v. 43.5%) and one class II combined with either a class III or IV mutation (22.9% v. 17.4%). The no bronchiectasis group had a higher proportion of class II and class I compound heterozygotes (26.1% v. 14.3%) as well as subjects who did not have both mutations classified (13.0% v. 8.6%). Pearson's chi-squared test was used to test for a statistically significant difference in *CFTR* genotype distribution between the two groups. However, because this calculation requires a minimum frequency of five in each group,²⁶⁴ subjects with class II and either class III or class IV mutations, class II and an unknown mutation group or two unclassified mutations were combined. There was no significant difference found ($p = 0.510$).

Table 7.1. Subject demographics in the no bronchiectasis and bronchiectasis groups.

	No bronchiectasis ($n = 23$)	Bronchiectasis ($n = 35$)	Statistical test of difference
Sex (% female)	45.7%	52.2%	0.789 [^]
Cystic fibrosis centre (% Melbourne)	40.0%	39.1%	>0.99 [^]
Age at 1-year-old BAL	1.12	1.15	0.50 [#]
Age at 5-year-old CT scan	5.16	5.23	0.57 [#]
Age at matching	9.30	9.57	0.57 [#]

Note: BAL: bronchoalveolar lavage; CT: computed tomography; [^] Fisher's exact t -test; [#] unpaired t -tests.

Table 7.2. *CFTR* genotypes in the no bronchiectasis and bronchiectasis groups.

<i>CFTR</i> genotype by class	No bronchiectasis ($n = 23$)	Bronchiectasis ($n = 35$)
II/I	26.1%	14.3%
II/II	43.5%	54.3%
II/III or IV	17.4%	22.9%
II/?	13.0%	5.7%
??/?	0%	2.9%

Note: When the II/III or IV, II/? and ??/? classes were combined in the Pearson's chi-squared test, there were no significant differences between the groups ($p = 0.510$).

7.3.2 DNA methylation analysis

7.3.2.1 Quality control

Following DNA methylation analysis, two samples had a mean detection p -value of > 0.01 and, as per the predetermined analysis plan, were excluded from further analysis. Both samples were from the bronchiectasis group, reducing the total sample size to 56, 33 (58.9%) of whom had bronchiectasis. The effect of these exclusions on subject demographics is summarised in Tables 7.3 and 7.4. Pearson's chi-squared test was again used to test for a statistically significant difference in *CFTR* genotype distribution between the two groups. Groups were combined in the same manner as earlier, and again there was no significant difference ($p = 0.500$).

Table 7.3. Subject demographics in the no bronchiectasis and bronchiectasis groups after exclusion of poor-quality samples.

	No bronchiectasis ($n = 23$)	Bronchiectasis ($n = 33$)	Statistical test of difference
Sex (% female)	45.5%	52.2%	0.786 [^]
Cystic fibrosis centre (% Melbourne)	39.4%	39.1%	>0.999 [^]
Age at 1-year-old BAL	1.12	1.16	0.37 [#]
Age at 5-year-old CT scan	5.16	5.23	0.58 [#]
Age at matching	9.30	9.61	0.51 [#]

Note: BAL: bronchoalveolar lavage; CT: computed tomography; [^] Fisher's exact t -test; [#] unpaired t -tests.

Table 7.4. *CFTR* genotypes in the no bronchiectasis and bronchiectasis groups after exclusion of poor-quality samples.

<i>CFTR</i> genotype by class	No bronchiectasis ($n = 23$)	Bronchiectasis ($n = 35$)
II/I	26.1%	15.2%
II/II	43.5%	57.6%
II/III or IV	17.4%	24.2%
II/?	13.0%	3.03%
?/?	0%	0.0%

Note: When the II/III or IV, II/? and ?/? classes were combined in the Pearson's chi-squared test, there were no significant differences between the groups ($p = 0.500$).

Data were normalised using SQN. When data were visualised using MDS plots, there was no clustering by disease status but a clear clustering by sex (Figure 7.3.) Poor-performing probes, defined as probes with a detection p -value of > 0.01 in one or more samples, probes located on sex chromosomes, cross-reactive probes and probes with a common

SNP at the CpG site, were all excluded. This left 695,764 probes for downstream analysis. When the data were visualised at this stage, the effect of sex had been removed. There was still no apparent clustering of samples by disease status (Figure 7.4).

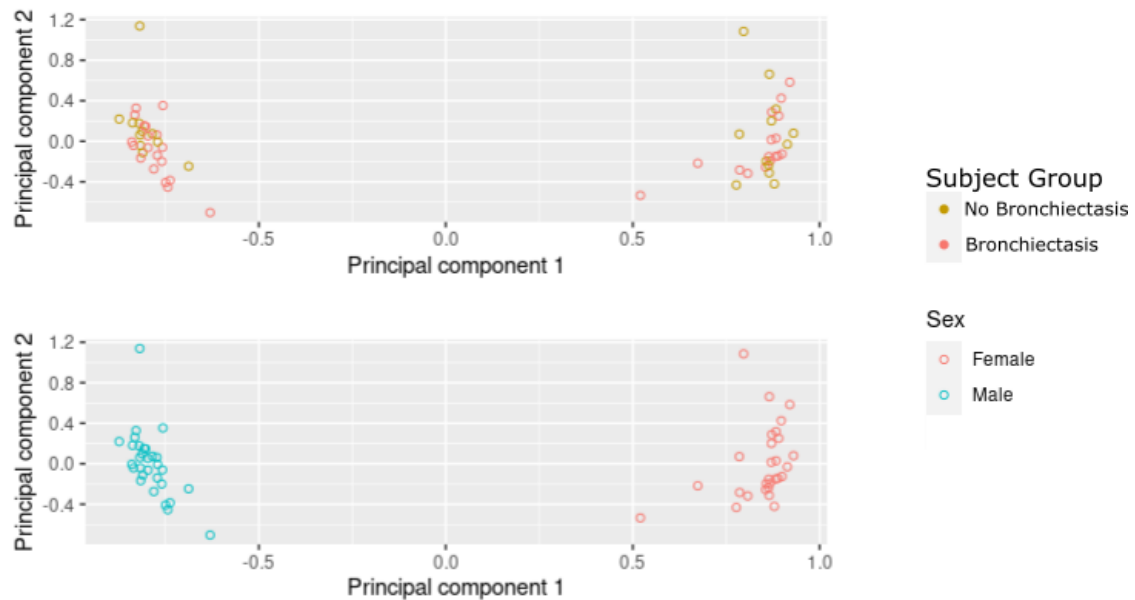


Figure 7.3. Initial data exploration showing that sex was responsible for the variation seen in the first principal component.

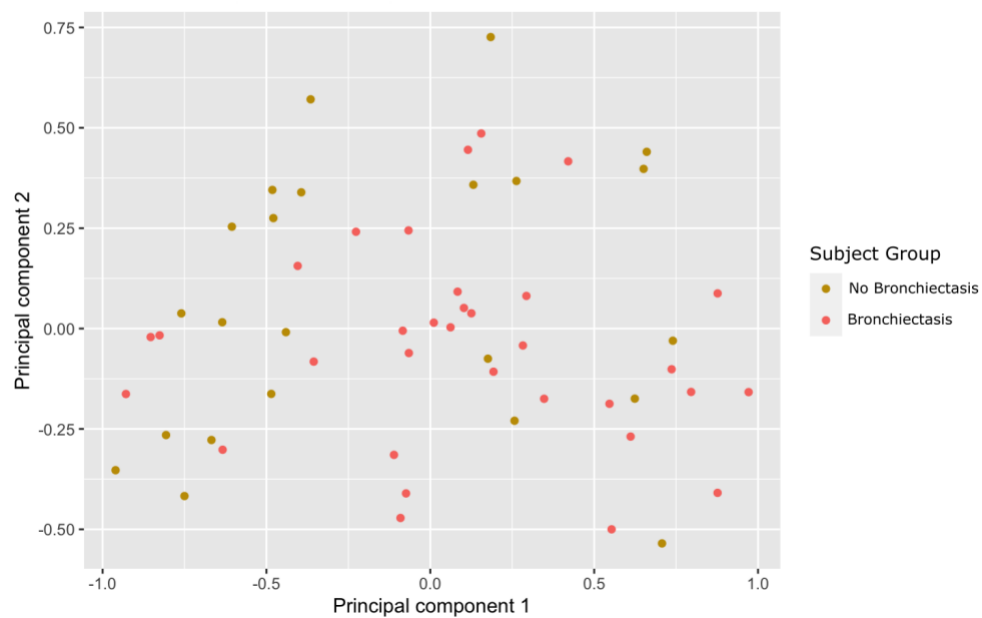


Figure 7.4. Data exploration after quality control, showing no clustering by disease status.

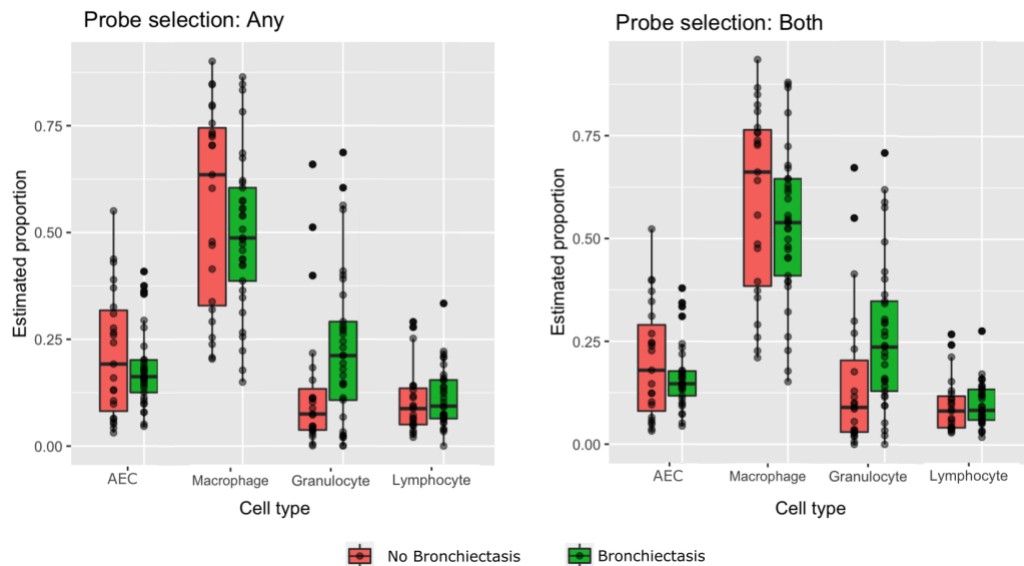


Figure 7.6. Comparison of cell composition in the bronchiectasis and no bronchiectasis groups. There are no significant differences using either probe selection method.

7.3.2.3 Exploring variations in the methylation data

Data were examined to identify the traits that should be included in the differential methylation analysis. The scree plot in Figure 7.7 shows that most of the variation in the data was captured in the first three to five principal components. Correlations between known sample traits (disease status, age at BAL, sex, cell composition and BMI) and the top 20 principal components were then assessed (Figure 7.8). This showed that cell composition most strongly contributed to the first three principal components. Age at BAL collection, sex and BMI did not significantly contribute to the top 20 principal components of variation.

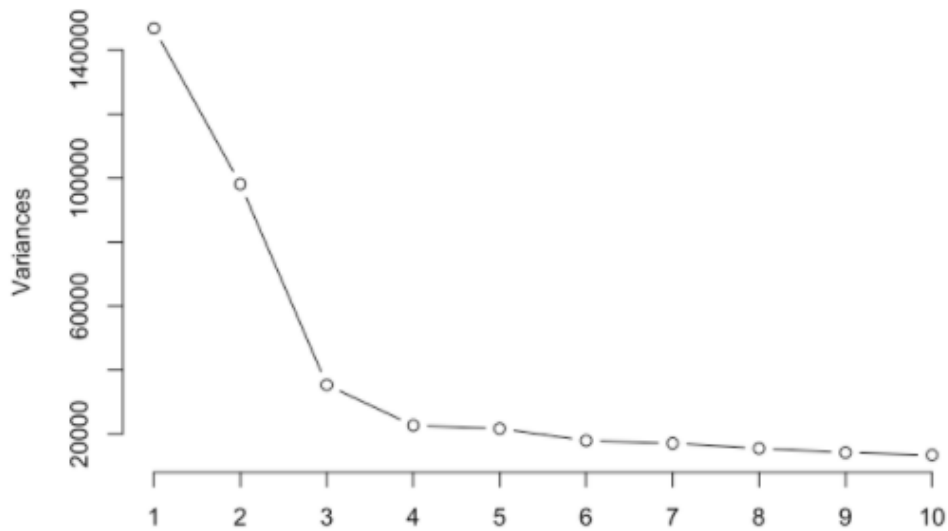


Figure 7.7. Scree plot demonstrating that most of the variation in the data was captured by the first five principal components.

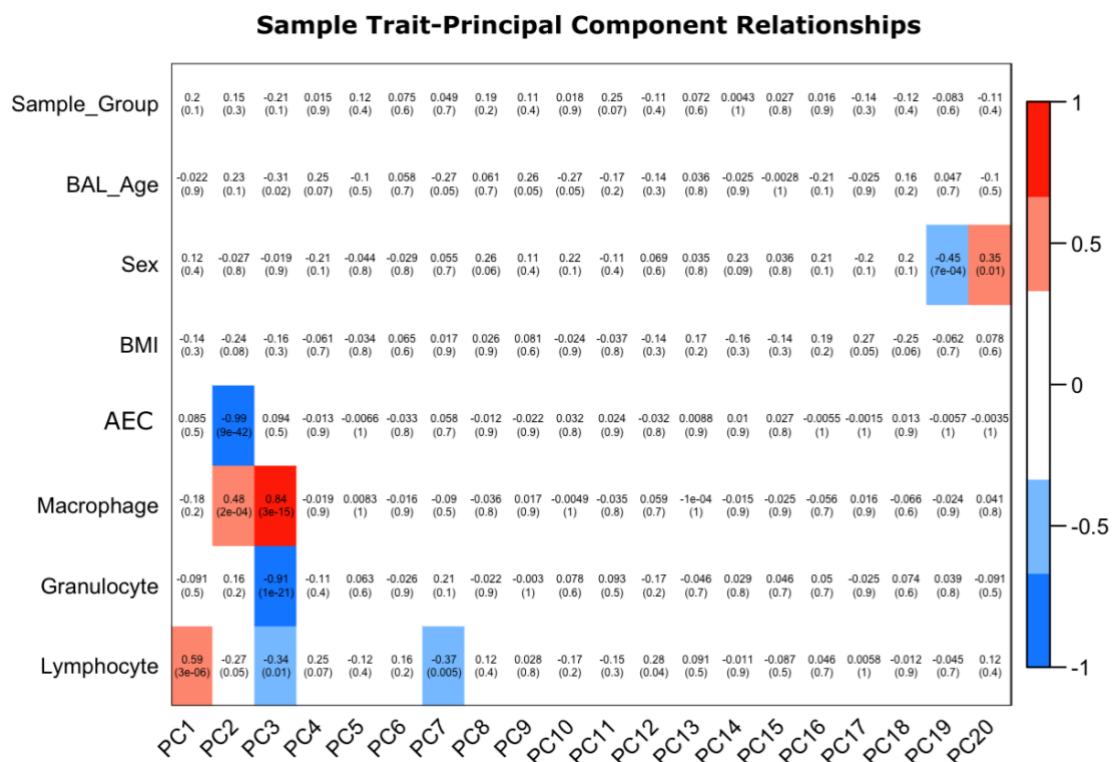


Figure 7.8. Correlation between sample traits and top 20 principal components. The strongest correlations between sample traits and principal components were between cell composition and the first three principal components.

The contribution of cell composition to variations in the data was further visualised using MDS plots (Figure 7.9). Similar to previous analyses (Chapter 6), the effect of macrophages and granulocytes was driving the variation in the first principal component.

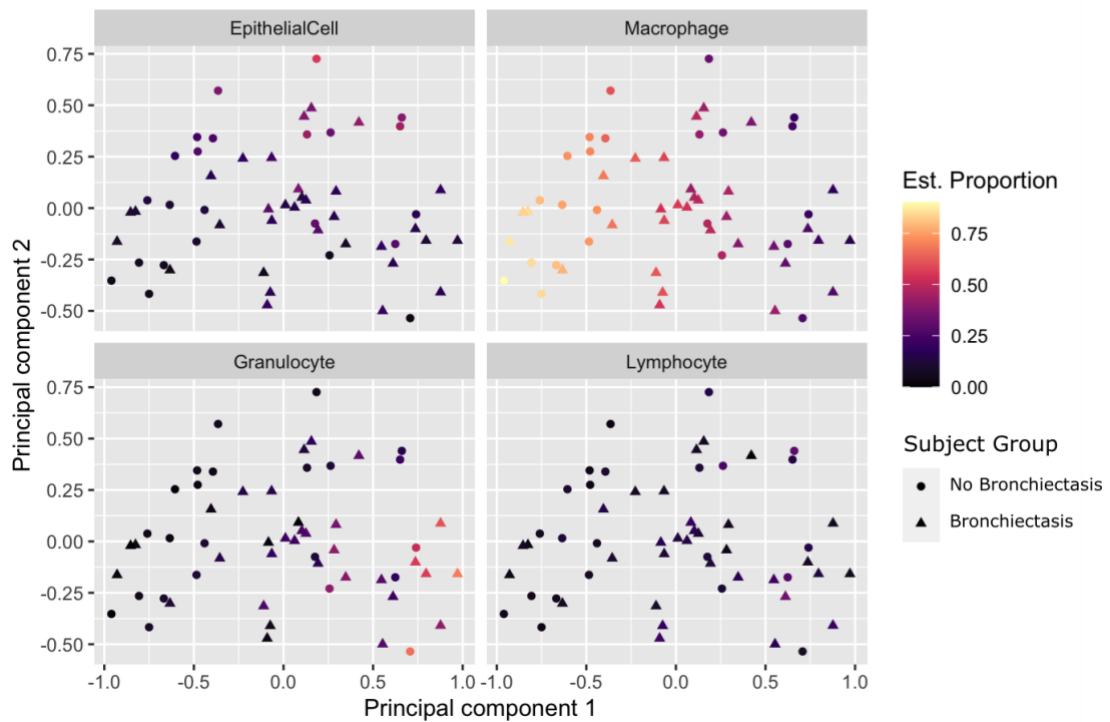


Figure 7.9. MDS plots demonstrating the effect of cell composition on the top principal components. The samples are coloured by the relative proportion of the respective cell type in each plot. The macrophage plot shows a clear gradient between samples with a high and low proportion of macrophages, driving the variation in the first principal component. The inverse relationship is seen in the granulocyte plot. This confirms that these cell types were driving the variation seen in the first principal component.

7.3.2.4 Identification of differentially methylated probes

A probe-wise linear model was fitted to determine the differences in methylation between subjects with and without bronchiectasis. The model included estimated cell type proportions to adjust for cell composition but not for other known sample traits, which did not appear to contribute to the principal components driving the variations in the data. There were no statistically significant DMPs when subjects with and without bronchiectasis were compared at an FDR of < 0.1 (see Table 7.11 in Chapter 7 Appendix).

A differential methylation analysis using RUV-4 was then performed. As a point of reference, an RUV-4 analysis without adjustment for unwanted variation did not identify any statistically significant DMPs. The RUV-4 analysis was then repeated, including adjustment for $k = 3$ components of variation that should primarily be related to cell composition. This analysis generated seven significant DMPs at an FDR of < 0.01 (Table 7.5 and Figure 7.10).

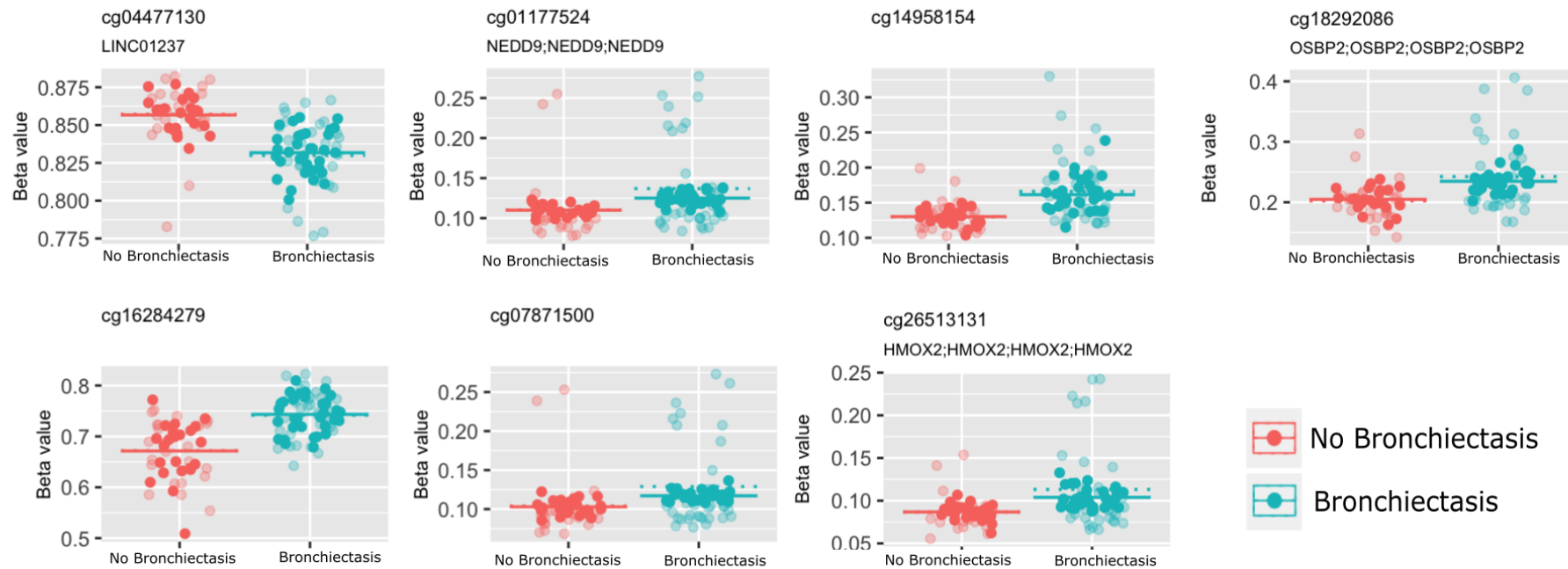


Figure 7.10. Summary of significant differentially methylated probes. Probes that map to known genes have the gene listed under the probe number. The opaque dots represent the results prior to adjustment for cell composition. Solid dots represent post-RUV-4 adjustment data, showing that adjustment narrowed the spread of methylation values.

Table 7.5. Significant differentially methylated probes identified using RUV-4 analysis.

Probe	Mean Methylation Bronchiectasis Group	Mean Methylation No Bronchiectasis Group	False discovery rate	False discovery rate adjusted for multiple comparisons*
cg04477130	0.832	0.857	4.60E-08	0.0320
cg01177524	0.125	0.110	1.85E-07	0.0645
cg14958154	0.161	0.130	4.96E-07	0.0842
cg18292086	0.235	0.205	6.38E-07	0.0842
cg16284279	0.743	0.671	6.73E-07	0.0842
cg07871500	0.117	0.103	7.26E-07	0.0842
cg26513131	0.104	0.087	9.69E-07	0.0963

Note: RUV-4: remove unwanted variation, 4-step; *Using the Benjamini–Hochberg procedure.²³⁹

The effect of the RUV-4 analysis was visualised using MDS plots, revealing that macrophages and granulocyte proportions were no longer driving variations in the first principal component (Figure 7.11). To further assess the effect of the RUV-4 adjustment, *p*-value histograms were plotted for both the adjusted and unadjusted analyses (Figure 7.12). A ‘good’ *p*-value distribution is expected to have a peak for low values and show a uniform distribution for higher *p*-values.²⁶⁵ The *p*-value histogram of the unadjusted analysis did not follow this pattern; rather, it showed only a gradual decrease in incidence as *p*-value increased. However, in keeping with the identification of significant DMPs,²⁶⁵ the adjusted analysis histogram had a sharp peak at the lowest *p*-value, followed by a uniformly lower frequency for all other *p*-values, suggesting a marked improvement over the unadjusted analysis.

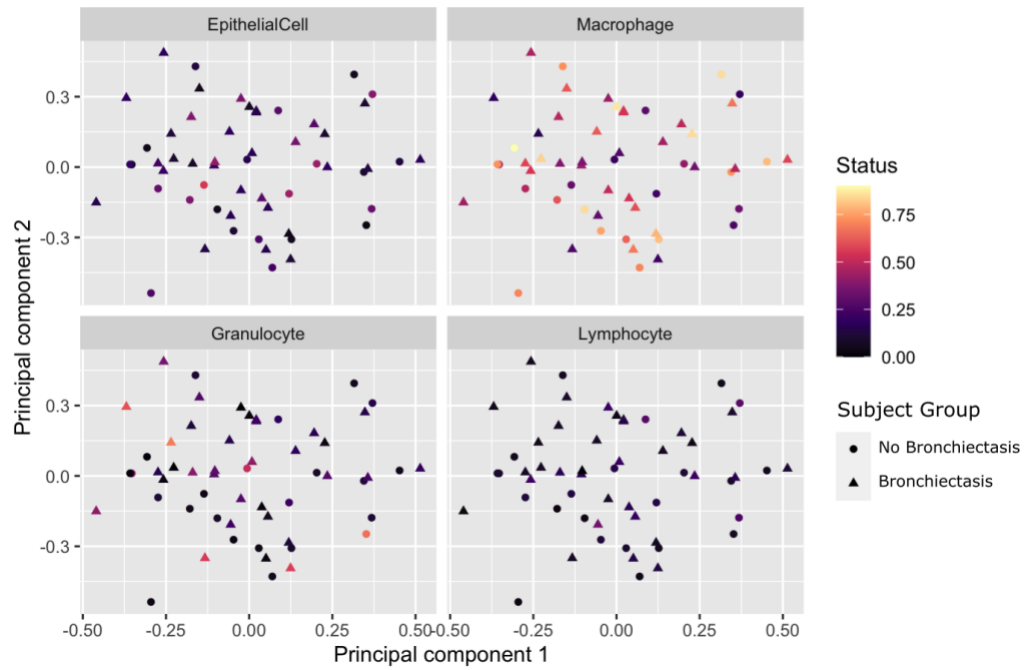


Figure 7.11. MDS plots post RUV-4 adjustment, demonstrating the effect of cell composition on the top principal components. Samples are coloured by the relative proportion of the respective cell type in each plot. Compared with Figure 7.9, as expected, cell composition no longer drives the variation in the first two principal components.

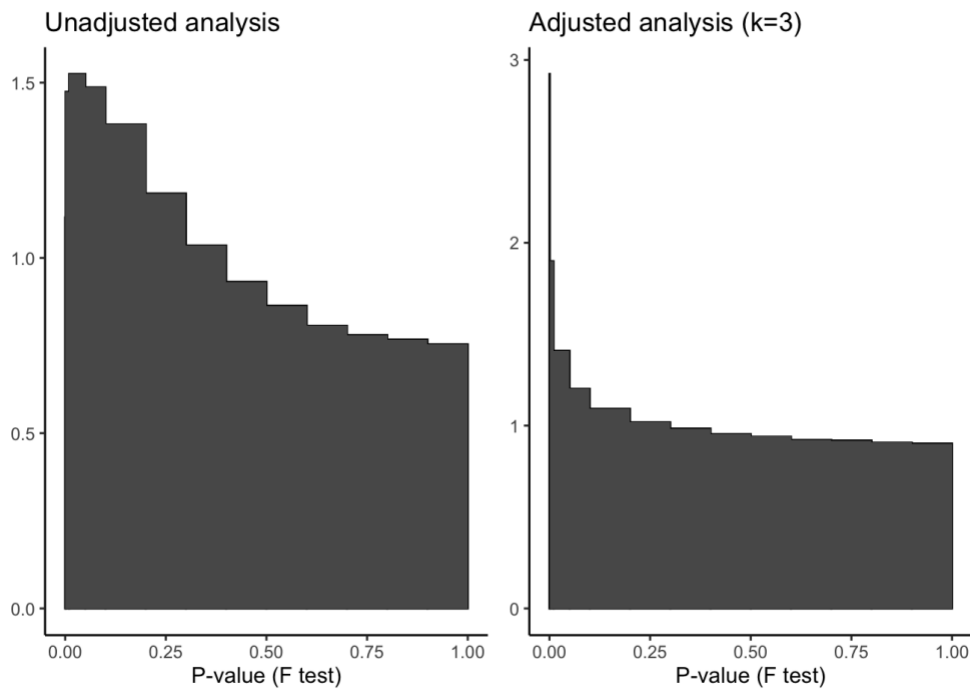


Figure 7.12. *p*-value histogram of both the unadjusted and adjusted RUV analysis. The x-axis represents the different *p*-values, and the y-axis is the frequency with which these values occur. The unadjusted analysis, which shows only a gradual decrease in frequency as the *p*-values increase, is in keeping with a ‘poor’ distribution of *p*-values (a distribution resulting from chance). The adjusted analysis, which has a sharp peak at the lowest *p*-values followed by a uniform lower frequency of values for all other *p*-values, suggests a ‘good’ *p*-value distribution (because there is a higher frequency of *p*-values below the threshold of significance). This is the expected profile for an analysis with significant findings.

7.3.2.5 Performance of significant differentially methylated probes as a diagnostic test

The AUC for the seven DMPs ranged between 0.863 ($SE = 0.0490$, 95% CI [0.767, 0.959], $p < 0.0001$) for cg16284279 to 0.918 ($SE = 0.0343$, 95% CI [0.851, 0.985], $p < 0.0001$) for cg04477130 (Table 7.6). The receiver operating characteristic curves are plotted in Figure 7.13.

Table 7.6. AUC analysis for each of the seven significant differentially methylated probes and composite scores

	AUC	SE	95% CI	p-value
Individual DMPs				
cg04477130	0.918	0.0343	0.851–0.985	< 0.0001
cg01177524	0.916	0.0357	0.846–0.986	< 0.0001
cg14958154	0.905	0.0398	0.827–0.983	< 0.0001
cg18292086	0.897	0.0412	0.817–0.978	< 0.0001
cg26513131	0.893	0.0429	0.802–0.977	< 0.0001
cg07871500	0.883	0.0478	0.789–0.976	< 0.0001
cg16284279	0.863	0.0490	0.767–0.959	< 0.0001
Composite scores				
All DMPs added	0.959	0.0273	0.906–1.000	< 0.0001
All DMPs multiplied	0.987	0.0138	0.960–1.000	< 0.0001
Hypomethylated DMPs added	0.976	0.0187	0.940–1.000	< 0.0001
Hypomethylated DMPs multiplied	0.990	0.0113	0.967–1.000	< 0.0001

Note: AUC: area under the receiver operating characteristic curve; DMP: differentially methylated probe.

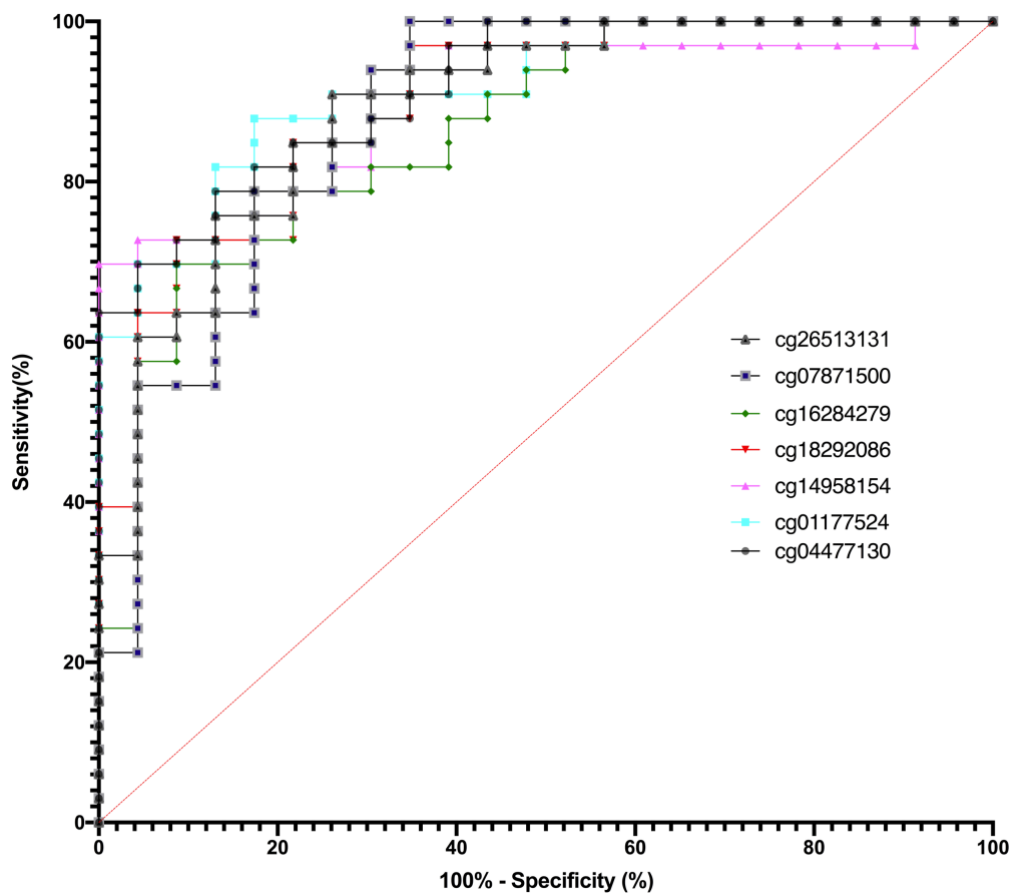


Figure 7.13. Receiver operator curves for significant differentially methylated positions.

To assess whether combining scores would improve diagnostic performance, a series of composite scores were created (Table 7.6 and Figure 7.14). First, the methylation levels at the seven DMPs were added, resulting in an AUC of 0.959 ($SE = 0.273$, 95% CI [0.906, 1.00], $p < 0.0001$). When the methylation levels of the DMPs were multiplied, there was a further improvement in AUC to 0.987 ($SE = 0.0138$, 95% CI [0.960, 1.000], $p < 0.0001$). Because six of the seven significant DMPs showed lower levels of methylation in the no bronchiectasis group, the effect of limiting the composite scores to positions with a consistent direction of effect was explored. For these analyses, cg04477130 was excluded. The limited composite score yielded improved performance when compared with the original composite score: 0.976 ($SE = 0.0187$, 95% CI [0.940, 1.000], $p < 0.0001$) after addition, and 0.990 ($SE = 0.0113$, 95% CI [0.967, 1.000], $p < 0.0001$) after multiplication.

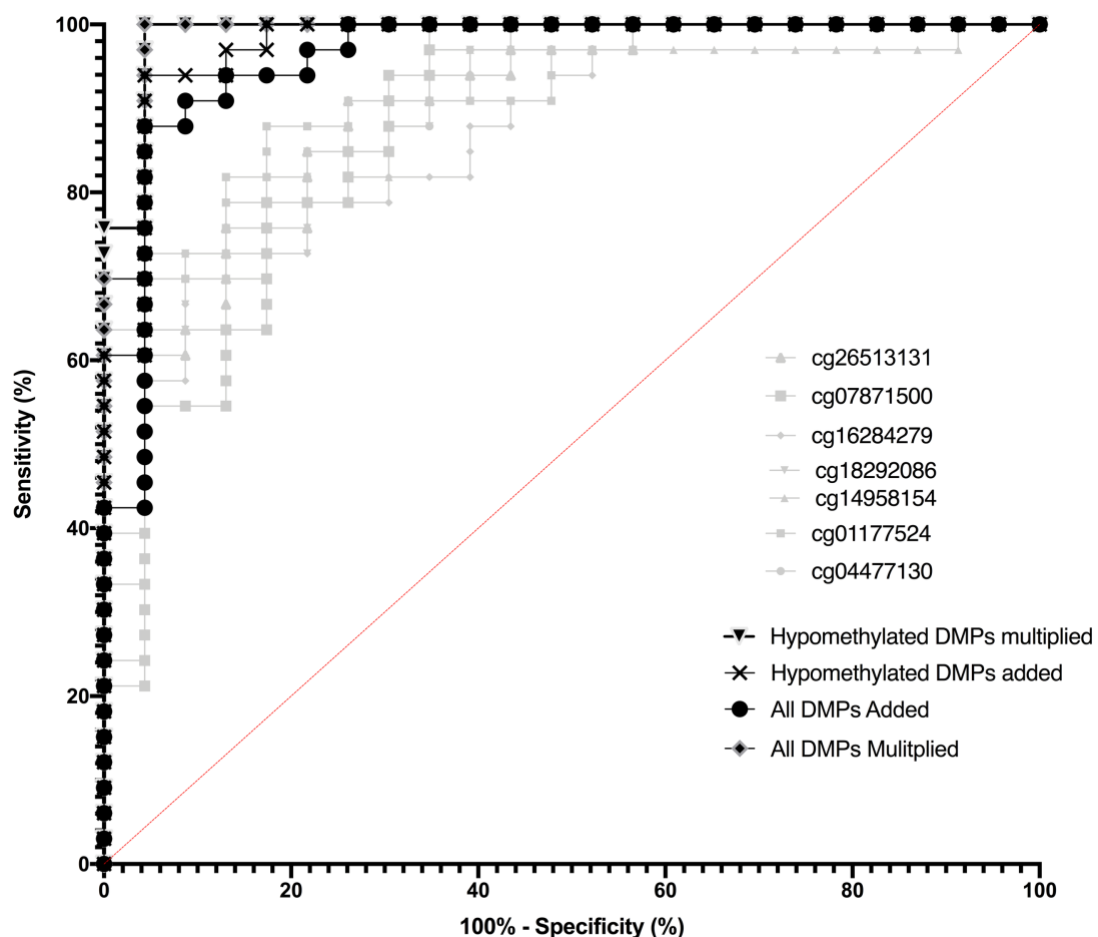


Figure 7.14. Receiver operator curves for the composite scores. The individual differentially methylated probe curves are shown in light grey for comparison.

The diagnostic performance of the leading DMP and composite score was further explored by calculating the sensitivity, specificity, PPV and NPV for selected thresholds.

For cg04477130 across the entire cohort, methylation levels ranged from 0.80 to 0.88. The median (range) level in the no bronchiectasis group was 0.86 (0.83–0.88) compared with 0.83 (0.80–0.86) in the bronchiectasis group. When a threshold of 0.83 or above was used to identify subjects who will remain bronchiectasis free, sensitivity was 100%, specificity was 30.3%, PPV was 50.0%, and NPV was 100%. While this represents an excellent performance if the test is negative (i.e. predicts the child will have bronchiectasis), the low PPV and specificity indicate a high false positive rate (Table 7.7). If this cut-off were used in the current dataset, it would predict that 46 children will remain bronchiectasis free, which would be incorrect for 23 (50.0%) of them. If a cut-off of 0.86 or above were used to identify subjects who will not develop bronchiectasis, sensitivity would be 56.5%, specificity would be 97.0%, PPV would be 92.9%, and NPV would be 76.2%. With this threshold, 13 of 14 (92.9%) subjects predicted to remain bronchiectasis free would do so (Table 7.8). Ten (23.8%) of the 42 children predicted to develop bronchiectasis would indeed remain bronchiectasis free.

Table 7.7. Categorisation of results if methylation level of 0.83 or above at cg04477130 in 1-year-old BAL is used to predict 5-year-old CT outcome

	CT: No bronchiectasis	CT: Bronchiectasis	
DNA methylation: No bronchiectasis	23	23	46
DNA methylation: Bronchiectasis	0	10	10
	23	33	56

Note: BAL: bronchoalveolar lavage; CT: computed tomography.

Table 7.8. Categorisation of results if methylation level of 0.86 or above at cg04477130 in 1-year-old BAL is used to predict 5-year-old CT outcome

	CT: No bronchiectasis	CT: Bronchiectasis	
DNA methylation: No bronchiectasis	13	1	14
DNA methylation: Bronchiectasis	10	32	42
	23	33	56

Note: BAL: bronchoalveolar lavage; CT: computed tomography.

For the composite score, which was calculated by multiplying the six DMPs with a consistent direction of effect, the range in scores was 0.57×10^{-5} to 7.2×10^{-5} . The median (range) of the no bronchiectasis group was 1.7×10^{-5} (0.57×10^{-5} to 3.5×10^{-5}), which was lower than that of the bronchiectasis group 4.0×10^{-5} (2.6×10^{-5} to 7.2×10^{-5}). If a cut-off of 2.5×10^{-5} or lower was used to identify subjects who would remain bronchiectasis free, sensitivity was 100%, specificity was 95.7%, PPV was 100%, and

NPV was 97.1%. Using the current data, all subjects predicted to remain bronchiectasis free would do so (Table 7.9). In subjects predicted to develop bronchiectasis, only 1 of 34 (2.9%) would remain bronchiectasis free. Thus, the test would correctly predict the outcome of 55 (98.2%) of the 56 subjects.

Table 7.9. Categorisation of results if composite methylation score in 1-year-old BAL (determined by multiplying six CpGs with uniform direction of effect) is 2.5×10^{-5} or below and is used to predict 5-year-old CT outcome

	CT: No bronchiectasis	CT: Bronchiectasis	
DNA methylation: No bronchiectasis	22	0	22
DNA methylation: Bronchiectasis	1	33	34
	23	33	56

Note: BAL: bronchoalveolar lavage; CT: computed tomography.

7.3.2.6 Assessment of functional effect of differentially methylated probes

Five of the seven DMP-associated CpG sites mapped to known genes, four of which are protein-coding genes and one of which is a long non-coding RNA. A summary of the terms and definitions used in annotation of the functional effect of DMPs is shown in Table 7.10.

Table 7.10. Summary of epigenetic markers and regulatory region definitions

Epigenetic markers	
H3K27AC	Marker of promoter and enhancer regions ^{266 267}
H3K4me1	Marker of enhancer region ^{267 268}
H3K4me3	Marker of promoter region ^{267 268}
ATAC-seq	Measure of chromatin accessibility ^{267 268}
CTCF	Measure of CTCF-binding sites ²⁶⁸
Regulatory regions	
Promoter	Regions at the 5' end of genes where transcription factors and RNA polymerase bind to initiate transcription ²⁶⁹ Characterised by open chromatin, ^{267 268 270} H3K4me3 ^{267 268 270} and H3K27ac ^{266 267}
Enhancer	Regions that bind transcription factors and interact with promoters to stimulate transcription of distant genes ²⁶⁹ Characterised by open chromatin, ^{267 268 270} H3K4me ^{267 268 270} and H3K27ac ^{266 267}
CTCF-only site	Several different activities that either promote or suppress gene expression. Activities include functioning as an insulator (blocking the actions of cis-located enhancers/promoters) or altering the structure of three-dimensional chromatin loops to either bring promoters and enhancers closer together

(supporting expression) or positioning them further apart (expression less likely)^{268 271}

Characterised by open chromatin, CTCF signal and absence of H3K4me3 and H3K27ac signals²⁶⁸

The probe cg26513131 maps to the heme oxygenase 2 (*HMOX2*) gene. *HMOX2* is expressed in many tissues, including respiratory epithelium and immune cells (see Figures 7.15 and 7.16 in Chapter 7 Appendix). Gene ontology reveals it is involved in the biological processes of CF lung disease, including neutrophil degranulation, response to hypoxia and iron ion homeostasis. The CpG maps to intron 2 of *HMOX2* (Figure 7.17 in Chapter 7 Appendix). There are no CpG islands in this region. Surrounding CpG probes show subtle signs of differential methylation (Figure 7.18 in Chapter 7 Appendix). Ensembl annotates the region as a CTCF site in all three cell lines (Figure 7.17 in Chapter 7 Appendix). This annotation is supported by other data that show in the same area a site of accessible chromatin in respiratory tissue (Figure 7.19 in Chapter 7 Appendix) and CTCF peaks with no other histone marks in adult lung tissue and neutrophils (Figure 7.17 in Chapter 7 Appendix). In neutrophils, there may be an *HMOX2* enhancer immediately upstream, as evidenced by H3K27ac and H3K4me signals (Figure 7.17 in Chapter 7 Appendix); however, whether the CTCF site acts as an insulator or chromatin loop anchor at this site is unknown.

The probe cg18292086 maps to the oxysterol-binding protein 2 (*OSBP2*) gene. *OSBP2* is expressed in immune cells and at a low level in lung tissue (Figures 7.20 and 7.21 in Chapter 7 Appendix). Gene ontology predicts that the primary biological function of *OSBP2* is sterol transportation, which is the directed movement of sterols (steroids with one or more hydroxyl groups and a hydrocarbon side chain) into, out of or within a cell or between cells. The associated CpG maps to the intronic region of the gene (Figure 7.22 in Chapter 7 Appendix). The probe immediately downstream of cg18292086 shows some differential methylation in the same direction (Figure 7.23 in Chapter 7 Appendix). Ensembl annotates the area as a promoter flanking region and a CTCF-binding site in embryonic stem cells, but only a CTCF site in lung fibroblasts and monocytes (Figure 7.22 in Chapter 7 Appendix). The annotation as a CTCF site in lung tissue is supported by it being an area of chromatin accessibility in multiple cell types in foetal lung tissue (Figure 7.24 in Chapter 7 Appendix) and a CTCF peak with no histone marks in adult lung tissue (Figure 7.22 in Chapter 7 Appendix). In immune cells, the pattern of epigenetic markers is less clear (Figure 7.22 in Chapter 7 Appendix). In neutrophils, there

is a distal H3K4me3 peak in keeping with a promoter, whereas in monocytes, H3K27ac and H3K9me1 peaks suggest an enhancer region.

The probe cg07871500 maps to two genes: lymphocyte cytosolic protein 1 (*LCP1*) and leucine-rich repeat-containing 63 (*LRRC63*). *LCP1* is primarily expressed in immune cells, with minimal expression in respiratory epithelial cells (Figures 7.25 and 7.26 in Chapter 7 Appendix). It is involved in a number of biological processes relevant to CF lung disease, including T cell activation in immune responses, cell migration, regulation of intracellular protein transport, IL-12 mediated signalling and regulation of podosome assembly in macrophages. *LRRC63* is not expressed in lung or immune cells (Figure 7.27 in Chapter 7 Appendix) and its function is unknown. The surrounding CpGs show limited differential methylation (Figure 7.28 in Chapter 7 Appendix). Ensembl annotates this area as a promoter region for both *LCP1* and *LRRC63* in monocytes only (Figure 7.29 in Chapter 7 Appendix). In contrast to Ensembl, other data support a promoter region in lung tissues with accessible chromatin (Figure 7.30 in Chapter 7 Appendix) and an area of H3K4me3 (Figure 7.29 in Chapter 7 Appendix). There is no clear pattern of histone marks in neutrophils or monocytes (Figure 7.29 in Chapter 7 Appendix).

The probe cg01177524 maps to the neural precursor cell-expressed developmentally downregulated protein 9 (*NEDD9*) gene. *NEDD9* is expressed highly in lung tissue (Figures 7.31 and 7.32 in Chapter 7 Appendix). In an atlas of foetal lung tissue gene expression, *NEDD9* is most highly expressed in vascular endothelial cells, bronchiolar cells and AECs (Figure 7.31 in Chapter 7 Appendix). There is also some expression in myeloid and lymphoid cells. Gene ontology reveals it is involved in biological processes such as cytoskeleton organisation, cell adhesion, cell migration, integrin-mediated signalling and positive regulation of tyrosine kinase activity. There are multiple *NEDD9* transcripts—the CpG maps to an intronic area in some and to an exon in others (Figure 7.33 in Chapter 7 Appendix). There are CpGs in the area showing similar levels of differential methylation (Figure 7.34 in Chapter 7 Appendix). In Ensembl, the area is defined as an active promoter flanking region in embryonic stem cells, monocytes and lung fibroblasts (Figure 7.33 in Chapter 7 Appendix). This annotation is supported by other data sources. The area corresponds to accessible chromatin in lung tissue (Figure 7.35 in Chapter 7 Appendix) and monocytes (Figure 7.33 in Chapter 7 Appendix) and is also a site of epigenetic markers associated with promoter regions (defined in Table 7.10) in lung tissue, neutrophils and monocytes (Figure 7.33 in Chapter 7 Appendix).

The probe cg04477130 maps to the long intergenic non-protein coding RNA 1237 (*LINC01237*) gene. *LINC01237* is expressed at low levels in lung tissue and immune cells (Figures 7.36 and 7.37 in Chapter 7 Appendix). As its name suggests, *LINC01237* is a long non-coding RNA that could work in concert with other epigenetic markers to regulate the gene expression of other cells. Two CpG islands are located downstream of cg0441730 (Figure 7.38 in Chapter 7 Appendix). Most CpG probes in the area were hypermethylated in the no bronchiectasis group, with some probes showing a similar magnitude of differential methylation to the DMP (Figure 7.39 in Chapter 7 Appendix). In Ensembl, the area is an unannotated open chromatin region in embryonic stem cells but has no regulatory features in the other cell types (Figure 7.38 in Chapter 7 Appendix). Unannotated open chromatin regions are measured open chromatin regions that cannot be explained by genome segmentation, which is the process whereby data from multiple assays are integrated into a single annotation of the genome.²⁷² Data remain unannotated when the findings are inconsistent with the information derived from other assays. This annotation is supported by other data sources. In foetal lung tissue and adult monocytes, there are small ATAC-seq peaks supporting accessible chromatin but no clear pattern of other epigenetic markers to allow further classification (Figures 7.38 and 7.40 in Chapter 7 Appendix).

The probe cg16284279 does not map to a gene. There are no genes located within 10 kb of either side of the probe (Figure 7.41 in Chapter 7 Appendix). There are also no surrounding CpG probes or CpG islands. However, there are multiple CpG dinucleotides within 10 kb either side of the probe that are not covered by the EPIC array; hence, DNA methylation at these sites remains unknown. There are other epigenetic markers in the area. The area corresponds to promoter flanking regions in embryonic stem cells and lung fibroblasts (Figure 7.41 in Chapter 7 Appendix). In other data, the area does correspond to accessible chromatin, but there is no clear pattern of other epigenetic markers (Figures 7.41 and 7.42 in Chapter 7 Appendix).

The probe cg14958154 also does not map to a gene. Members of the homeobox D (*HOXD*) family of genes (*HOXD11*, *HOXD12* and *HOXD13*) were located within 10 kb of either side of the probe (Figure 7.43 in Chapter 7 Appendix). These *HOXD* genes were not expressed in lung tissue or immune cells. Gene ontology shows that these genes are associated with organogenesis in utero as well as regulation of RNA polymerase II activity. There were several CpG islands both upstream and downstream of cg14958154 (Figure 7.43 in Chapter 7 Appendix), with few probes showing evidence of differential

methylation (Figure 7.44 in Chapter 7 Appendix). Ensembl annotates the area as an active promoter in embryonic stem cells but no other cell types (Figure 7.43 in Chapter 7 Appendix). In other data, there is no clear pattern of epigenetic markers (Figures 7.43 and 7.45 in Chapter 7 Appendix).

7.4 DISCUSSION

This project achieved its primary aim of identifying a DNA methylation pattern at 1 year of age that predicts the presence or absence of structural lung disease at 5 years of age. The seven significant DMPs identified had good predictive ability when assessed using AUC analyses, and this was further improved by the use of composite scores. The seven DMPs also mapped to several genes that may be of interest as potential modifiers of lung disease severity, thus could represent therapeutic targets. Despite the positive findings, much work is needed, including replication in a larger external cohort, before these DNA methylation-based biomarkers can be used in clinical practice.

This EWAS addressed the issues identified with the EWAS described in Chapter 6. The current study included 56 subjects compared with 28 in the first EWAS, increasing the power of the study. The cell composition of the samples used in this EWAS was also less skewed than in the prior EWAS, which may have made it easier to adjust for this variable without eliminating any true differential methylation between the samples.

Similar to the first EWAS, robust methods were used to determine outcomes, biospecimen choice, cell composition adjustment and statistical analysis. The outcome—the presence or absence of bronchiectasis at 5 years of age—represents a crucial clinically relevant outcome. It shows whether children will leave infancy and the preschool years with normal lung structure or irreversible lung damage. Research has shown that any structural lung disease detected in early life will persist and progress^{46 47 249}; therefore, to realise the goal of people with CF living a life unaffected by the disease, clinicians must aim for subjects to have no structural lung disease at the age of 5 years. The methods used for the classification of bronchiectasis status differed from those in the first EWAS. Rather than three clinicians reviewing the scans with a dichotomised response of bronchiectasis being present or absent, the bronchiectasis sub-score of the PRAGMA score was used. When the two different methods were compared, intra-observer reliability was similar (92.8% for PRAGMA v. 91.2% for the three-reviewer method); however, inter-observer reliability was better for PRAGMA (90.7% v. 77.5%).²⁴⁴ The prevalence of bronchiectasis in this cohort was 60.3%, which is similar to previously reported data.

In a small study of 23 children with a median age of 4.8 years, Stiglbauer et al.⁹⁵ found that 65% of children with CF had bronchiectasis. In a large study of 153 children with a mean age of 5 years, Wainwright et al.⁵² found that 57% had bronchiectasis.

Once again, this EWAS leveraged the methods developed earlier in this thesis, using tissue-specific BAL samples. The importance of this is reinforced by the fact that several of the significant CpGs mapped to genes expressed in lung tissue such as *NEDD9* and *HMOX2*. Had blood rather than BAL samples been used, the effects of the contribution of respiratory epithelium and alveolar macrophages would have been lost. The RUV-4 analysis used in this EWAS performed well, highlighting the value of the reference epigenomes developed in Chapter 5. As shown in Figure 7.11, the RUV-4 analysis was able to remove the effect of cell composition and enabled the identification of significant DMPs.

The predictive ability of the seven DMPs is promising. The lowest AUC for the DMPs was 0.863, and the highest was 0.918. In the review presented in Chapter 3, only one study used AUC analysis: Meng et al.²⁷³ identified a single DMP associated with progression of idiopathic scoliosis and found it had a similarly high AUC of 0.827 (95% CI [0.780, 0.876]). AUC performance may be categorised either as acceptable (0.7–0.8), excellent (0.8–0.9) and outstanding (> 0.9)²⁷⁴ or as low accuracy (0.5–0.7), moderate accuracy (0.7–0.9) and high accuracy (> 0.9).²⁷⁵ Using either of these classifications, the performance of the seven DMPs was excellent and was further improved by the calculation of composite scores. Given the complex nature of CF lung disease, it is unsurprising that the expression of several genes may interact to determine disease severity. In other diseases such as ischaemic heart disease, type II diabetes and inflammatory bowel disease, polygenic risk scores have been calculated.²⁷⁶ These scores involve determining the risk of a disease by combining the results of the analysis of mutations at multiple genes of interest. The usual starting point for creating such a score is a genome-wide analysis, followed by the selection of genes of interest.²⁷⁷ Novel methods have been developed to analyse whole genome data rather than identifying a subset of genes for analysis.²⁷⁸ Studies have shown they can yield an AUC of up to 0.87 in type 1 diabetes mellitus²⁷⁹ and 0.82–0.86 in Alzheimer’s disease.²⁸⁰ While these scores are not used in routine clinical practice, preliminary studies show that they can improve care for patients identified as high risk. In a Finnish study, providing patients with a risk score for ischaemic heart disease led to increased rates of smoking cessation and weight loss.²⁸¹ The composite scores calculated in this EWAS were an attempt to create a poly-

epigenetic risk score for CF lung disease severity. By combining the results of the significant DMPs, the performance of the scores was improved. The best-performing composite score had an AUC of 0.990, which corresponds to excellent predictive ability. In a recently published study of DNA methylation–based biomarkers of chronic kidney disease, a poly-epigenetic score derived from eight DMPs was found to have a similar AUC of 0.976.²⁸² In addition, when a specific threshold (2.5×10^{-5}) was chosen, it clearly separated subjects who would or would not go on to develop structural lung disease. Using this threshold, only one of 56 subjects (1.8%) was misclassified, with the score predicting the subject would develop bronchiectasis when in fact they would remain bronchiectasis free. Currently, it is presumed that most children will develop bronchiectasis by the age of 5 years. However, as studies demonstrate, only around 60% of children will have structural lung disease at this age,^{52 95} suggesting a misclassification of 40%.

Significant DMPs are associated with genes whose activity may be involved in the pathophysiology of CF lung disease, warranting further evaluation as potential therapeutic targets.

The cg26513131 probe mapped to *HMOX2*, a direct relative of *HMOX1*, which is one of the few validated gene modifiers of CF lung disease severity.²³¹ Gene ontology has revealed that it is implicated in multiple relevant pathways, and given that biomarkers of neutrophilic inflammation have been shown to be associated with cross-sectional and future lung disease severity, its role in neutrophil degranulation in particular may be of interest.^{89 205 283} For example, in a cross-sectional study, children with a neutrophil phenotype characterised by increased degranulation had more severe structural lung disease.²⁸³ If *HMOX2* expression modified the immune response such that there was less neutrophil degranulation, it would likely result in milder lung disease. The exact functional effect of differential methylation at this CpG is unknown. The DMP is located at a likely CTCF site, which can have multiple actions on gene expression. CTCF sites can act as insulators, blocking the interaction between enhancers and promoters, thus reducing gene expression. DNA methylation can influence this process by altering CTCF-binding sites. For example, a CTCF site acts as an insulator for the gene *IGF2*. When CTCF binds, it blocks the action of an enhancer, and *IGF2* is not expressed.²⁷¹ However, if DNA is methylated at the CTCF site, then CTCF can no longer bind, the enhancer can interact with the *IGF2* promoter, and *IGF2* is expressed. As a result it is possible that

differential methylation at cg26513131 may in turn affect CTCF binding, resulting in altered *HMOX2* expression.

The probe cg18292086 aligned to *OSPB2*, which has previously been implicated in protein folding diseases, including CF. One of the key roles of *OSPB2*, as catalogued by gene ontology, is in the movement of sterols through a cell. In cells expressing dysfunctional CFTR resulting from a class II mutation (such as Phe508del), CFTR misfolding in the endoplasmic reticulum leads to the accumulation of unesterified cholesterol in the cell.²⁸⁴ *OSPB2* has been implicated in modifying the ability of cells to process accumulated cholesterol via vesicle-associated membrane proteins.²⁸⁴ Given that all subjects in this EWAS had at least one class II mutation, this pathway is of definite relevance. Patients with a class II mutation who are able to process misfolded CFTR in a manner that results in less cellular damage would likely have milder lung disease. The functional impact of one probe in the intronic region on methylation is unknown. The area also appears to be a CTCF site. The bronchiectasis-free group had lower levels of methylation at this site, which would favour increased gene expression; however, how this would interact with the CTCF site at this locus is unknown.

The probe cg07871500 mapped to *LCP1* and *LRRC63*. Of these two genes, *LCP1* is known to be involved in pathways relevant to CF lung disease pathophysiology. *LCP1* is involved in multiple aspects of the immune response, including macrophage and T cell responses. Moreover, *LCP1* is expressed in higher levels in the bronchial epithelium of people with CF compared with healthy controls.²⁸⁵ Proteomic analysis has also revealed increased levels of *LCP1* in the sputum of adults with CF compared with controls.²⁸⁶ Given this, subjects with CF with lower levels of *LCP1* expression may have milder lung disease. However, the findings of this EWAS do not support this hypothesis. The CpG site was located within a likely *LCP1* promoter region, and the group with milder lung disease had lower levels of methylation at this site, which would support higher gene expression. However, given the multiple other epigenetic markers in the area, it is difficult to determine the functional effect of methylation at one probe.

The probe cg01177524 mapped to *NEDD9*, which was highlighted as a potential gene of interest in a large genome-wide association and linkage study investigating CF lung disease.²⁸⁷ This is because a locus containing Cas scaffolding protein family member 4 (*CASS4*), a relative of *NEDD9*, and aurora kinase A (*AURKA*), a gene with which *NEDD9* directly interacts, was found to be linked to lung disease severity. *NEDD9* is involved in

multiple pathways relevant to CF lung disease, including the development of cilia and cell migration, which could be implicated in the proinflammatory response characteristic of CF. The area corresponds to a likely promoter region in multiple cell types. The lower level of methylation seen in the bronchiectasis-free group could potentially lead to increased gene expression.

The remaining three probes mapped to either a long non-coding RNA gene or intergenic areas. Hence, these may represent descriptive biomarkers and are less likely to be directly involved in CF lung disease pathophysiology. However, non-coding RNA could influence CF lung disease via epigenetic regulation of other genes.

The primary limitation of this EWAS was its size and the low likelihood that its significant findings can be replicated in an external cohort. The total number of subjects in the study ($n = 56$) was relatively small and would need repeated external validation in larger cohorts before these significant DMPs could be used as a clinical biomarker. In addition, the subjects were not matched for factors such as microbiological infection, nutritional status, or history of pulmonary exacerbations. This decision was made due to the same reasons described in chapter six (page 181). One barrier to successful replication in external cohorts may be the magnitude of differential methylation at the DMPs. The mean absolute level of differential methylation for the DMPs ranged from 1.41–7.19%, and five of the seven probes had an absolute differential methylation of less than 3%. The relatively small magnitude of differential methylation means validation in external cohorts is less likely. In the current EWAS, the DMPs had excellent predictive power when assessed using AUC analysis because the spread of values in either group was small. A larger study may demonstrate a larger spread in methylation values, and given the small magnitude of differential methylation, this may result in the sites no longer being DMPs and reduced predictive performance when assessed using AUC.

There are other potential barriers to the successful replication of this project. This study used the bronchiectasis sub-score of the PRAGMA method to classify disease status. This method of CT assessment is not widely used and is relatively labour intensive. Automated methods of PRAGMA scoring are being developed, which would be advantageous. In the interim, validation cohorts may need to use a method of disease classification closer to that used in the initial EWAS, which would be a more widely accessible method. Another potential barrier to validation experiments will be the availability of patient cohorts who have undergone the requisite collection of BAL samples and CT scans. AREST CF has

led the way in early-life CF cohorts; thus, the biological samples and corresponding patient phenotyping is unrivalled. Other cohorts such as the Study of Host Immunity and Early Lung Disease in Cystic Fibrosis (SHIELD CF) based in Ireland²⁸⁸ have collected similar information and may be suitable for validation experiments.

7.5 CONCLUSIONS

This EWAS achieved its primary aim of identifying a DNA methylation profile of BAL specimens at 1 year of age that predicted the presence or absence of bronchiectasis at 5 years of age. The DMPs performed well as predictive biomarkers, especially when they were combined into a composite score. The next step will be to validate these biomarkers in external cohorts. If replication is successful, they could become clinically valid biomarkers that may improve the care delivered to children with CF.

7.6 APPENDIX

Table 7.11. Top-ranked probes from standard differential methylation analysis using *limma*.(Chr: chromosome the probe is located on. pos: genomic coordinates of probe based on University of California, Santa Cruz (UCSC) genome browser. UCSC_RefGene_Name: Name of the gene that the probe maps to. UCSC_RefGene_Group: Location in gene that the probe is located. logFC: estimate of the log2-fold-change corresponding to the effect. AveExpr: average log2-expression for the probe over all arrays and channels. t: moderated t-statistic. P.Value: raw p-value. adj.P.Val: adjusted P-value. B: log-odds that the gene is differentially methylated.)

	chr	pos	UCSC_RefGene_Name	UCSC_RefGene_Group	logFC	AveExpr	t	P.Value	adj.P.Val	B
cg19905587	chr2	85554135	TGOLN2	Body	0.33866823	-2.0741802	5.40839294	1.41E-06	0.54796807	2.41425682
cg16284279	chr6	77743952			-0.5154492	1.33983971	-5.3391475	1.82E-06	0.54796807	2.26778238
cg12145739	chr14	24912198	SDR39U1;SDR39U1;SDR39U1;SDR39U1;LOC101927045	TSS200;TSS200;TSS200;TSS200;Body	-0.4703872	-1.4517297	-5.2250314	2.75E-06	0.54796807	2.02795848
cg12250896	chr10	134001112	DPYSL4	Body	0.49039759	-3.6186413	5.14518726	3.67E-06	0.54796807	1.86035608
cg12041429	chr17	7585875	TP53;TP53;TP53;TP53	5'UTR;5'UTR;5'UTR;5'UTR	0.3743635	2.99263785	5.06812216	4.84E-06	0.54796807	1.69880067
cg09850105	chr8	145111516	OPLAH	Body	0.23871927	1.64305787	5.03924322	5.35E-06	0.54796807	1.64017591
cg07506458	chr19	38899594	FAM98C	3'UTR	0.33799462	3.1544482	5.00272508	6.11E-06	0.54796807	1.56190919
cg04477130	chr2	242986565	LINC01237	Body	0.31542747	2.4223238	4.9939165	6.30E-06	0.54796807	1.54397284
cg19283778	chr6	96917485			-0.3340654	1.4895364	-4.9209439	8.17E-06	0.63174012	1.39103744
cg15556709	chr7	43739176	C7orf44	5'UTR	0.41899494	2.64100389	4.84158385	1.08E-05	0.66423316	1.22561696

Figures Related to Annotation of cg26513131/*HMOX2*

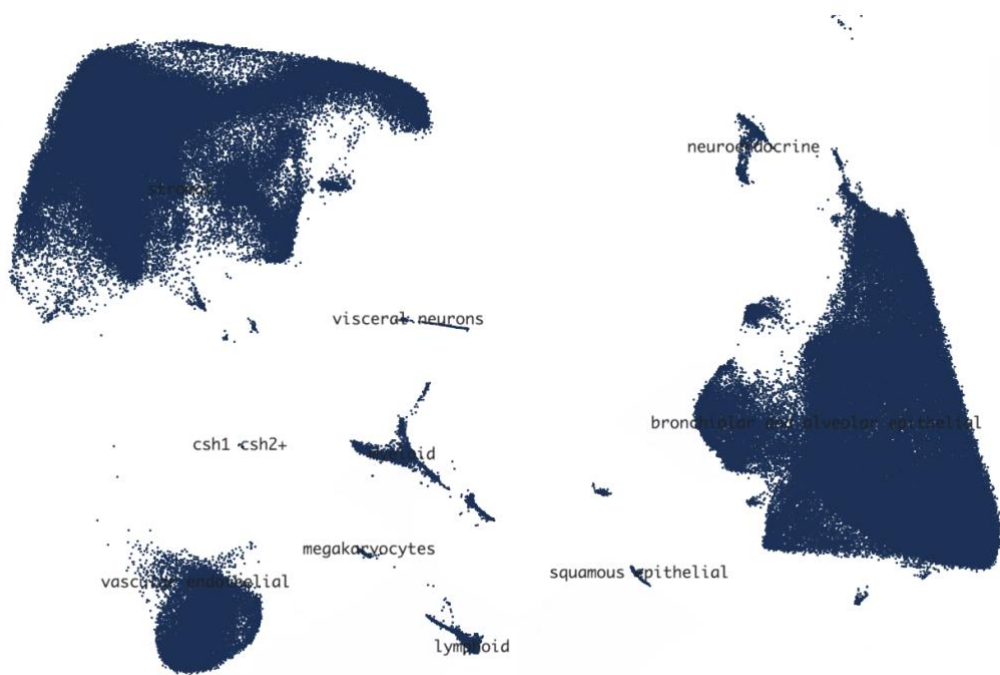


Figure 7.15. Gene expression of *HMOX2* from human foetal lung tissue from Descartes Human Gene Expression During Development. Blue indicates high expression. *HMOX2* is expressed highly in all lung cells.

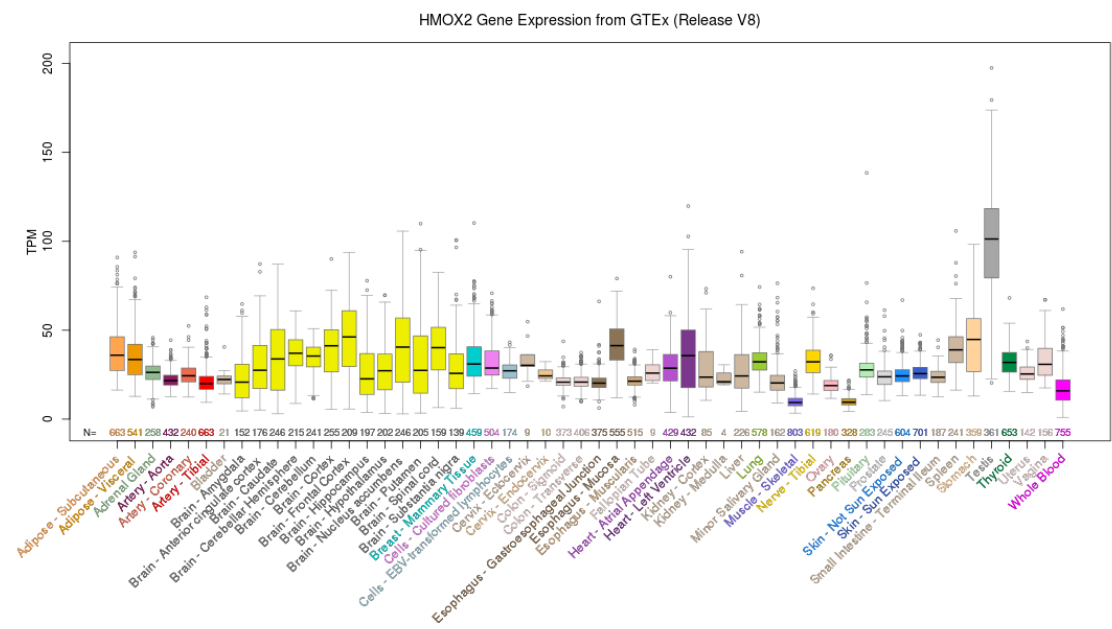


Figure 7.16. Gene expression of *HMOX2* using the National Institutes of Health Genotype–Tissue Expression Project. *HMOX2* is expressed in lung tissue relatively highly, and also expressed in blood.

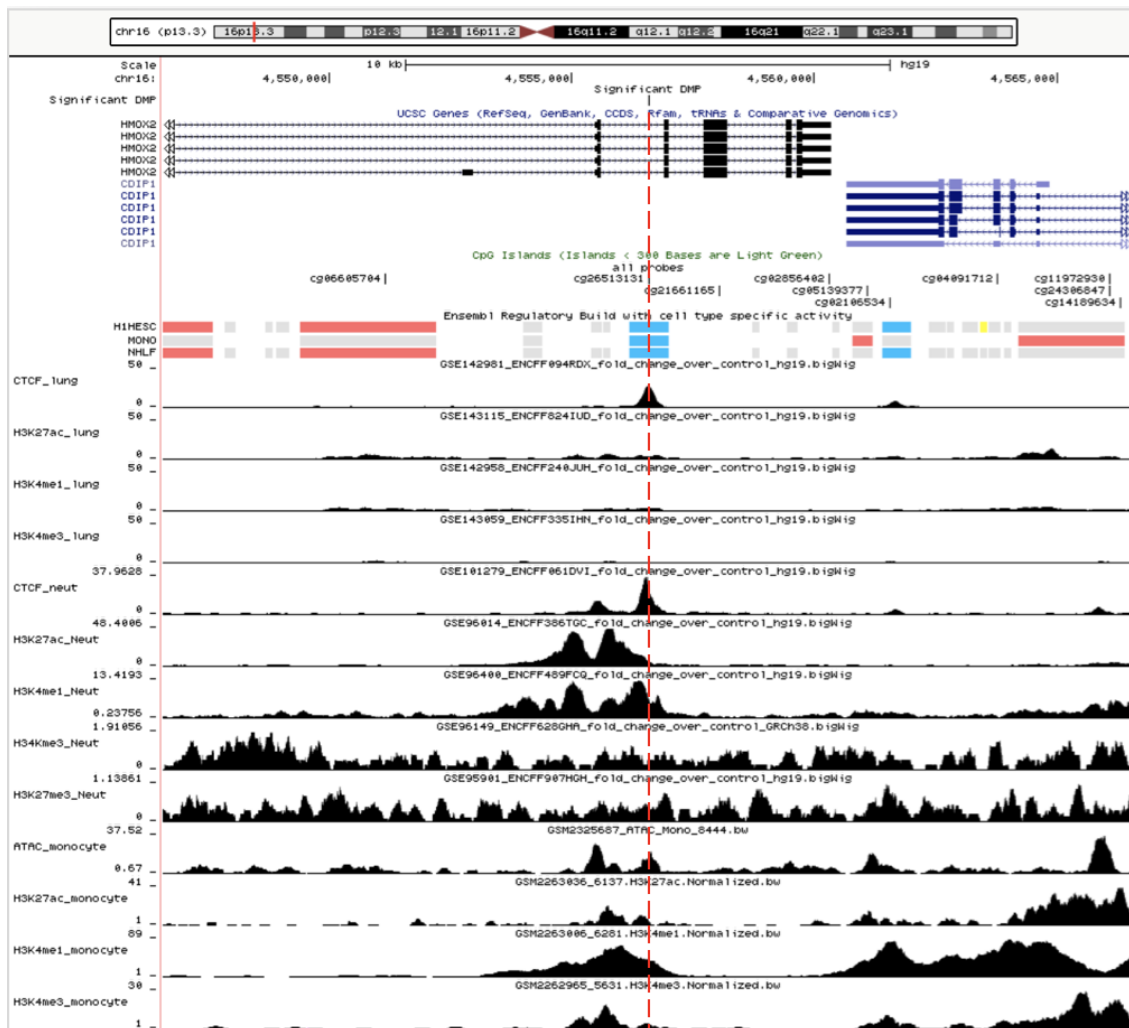


Figure 7.17: Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensemble Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. The DMP coincides with a CTCF peak in lung tissue, consistent with a CTCF site. In neutrophils, there is also a CTCF peak, with a likely proximal enhancer region.

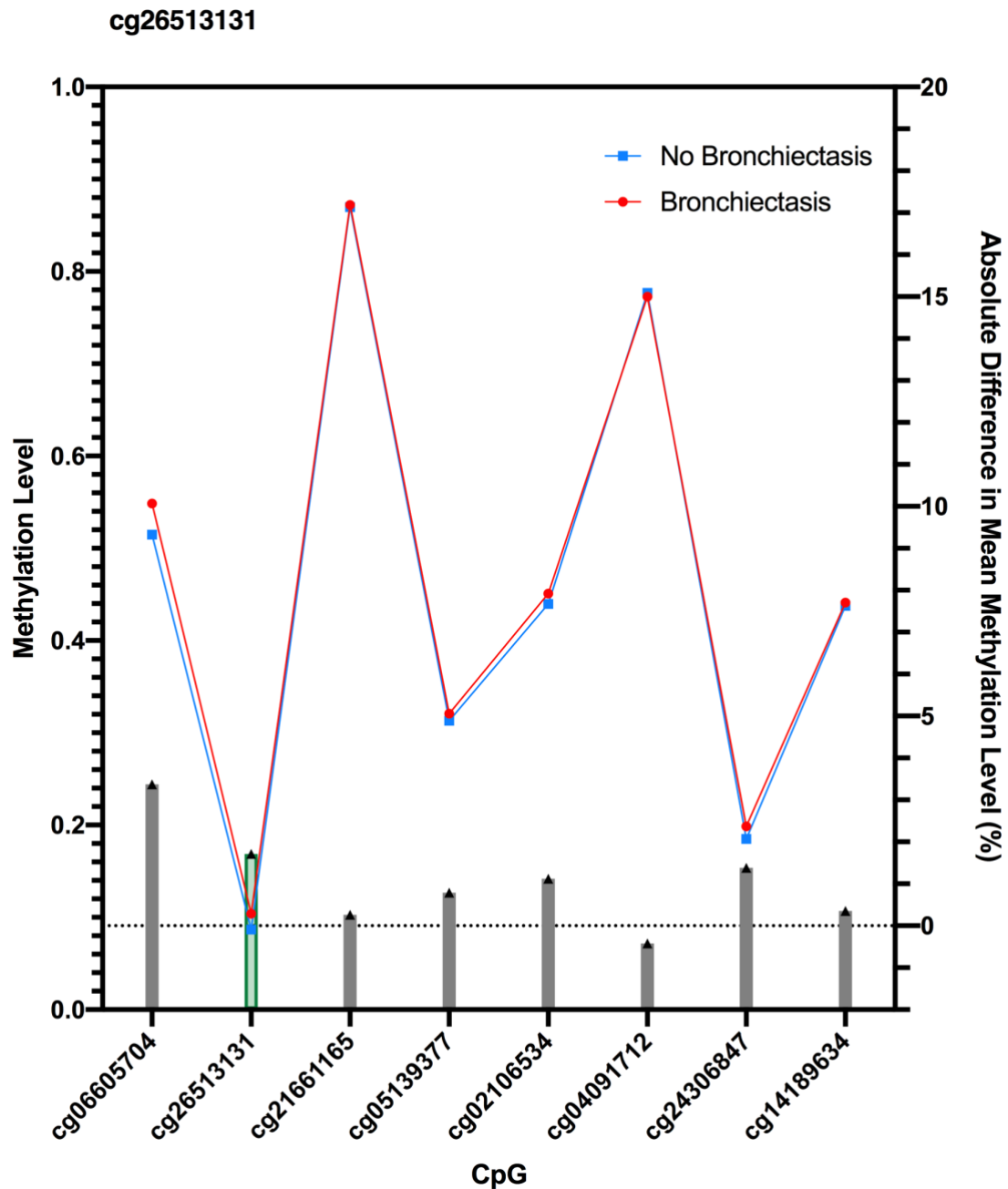


Figure 7.18. A graph of the methylation of probes located 10 kb either side of the significant DMP. (highlighted in green). The left y-axis corresponds to the data shown with lines and is the mean methylation of either group at the respective CpGs. The right y-axis is a bar chart of the difference in mean methylation between the bronchiectasis and no bronchiectasis groups. No difference in methylation (i.e. 0) is highlighted by a horizontal dashed line.



Figure 7.19. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe at an open chromatin site in many cell types and chromatin hubs in bronchoalveolar epithelial cells, ciliated epithelial cells, lymphoid cells, myeloid cells and vascular endothelial cells.

Figures Related to Annotation of cg18292086/OSPB2

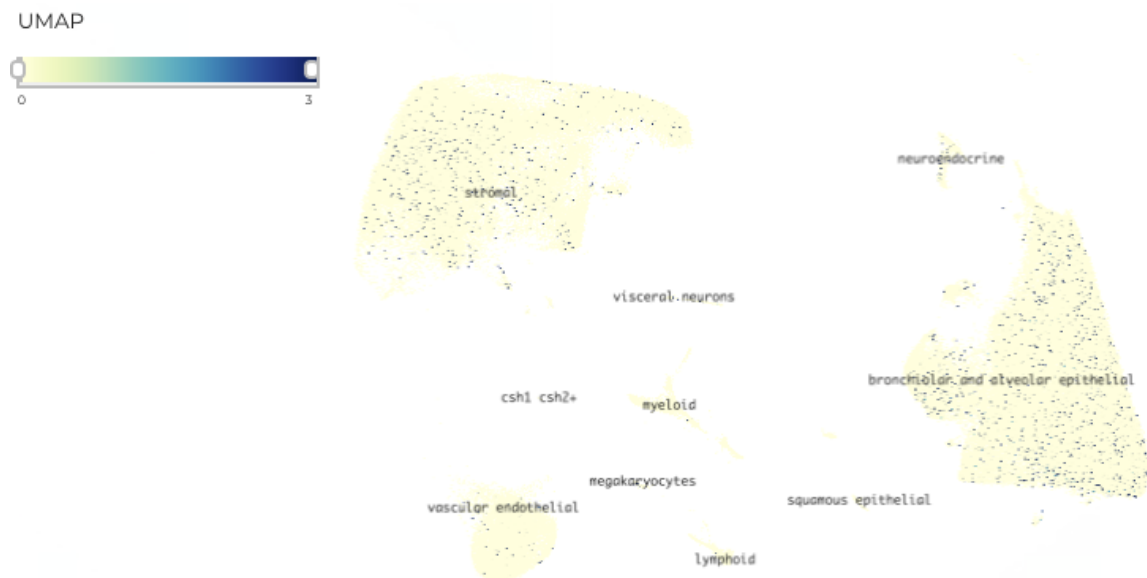


Figure 7.20. Gene expression of *OSBP2* from human foetal lung tissue from Descartes Human Gene Expression During Development. There is low level of *OSBP2* expression in multiple cells.

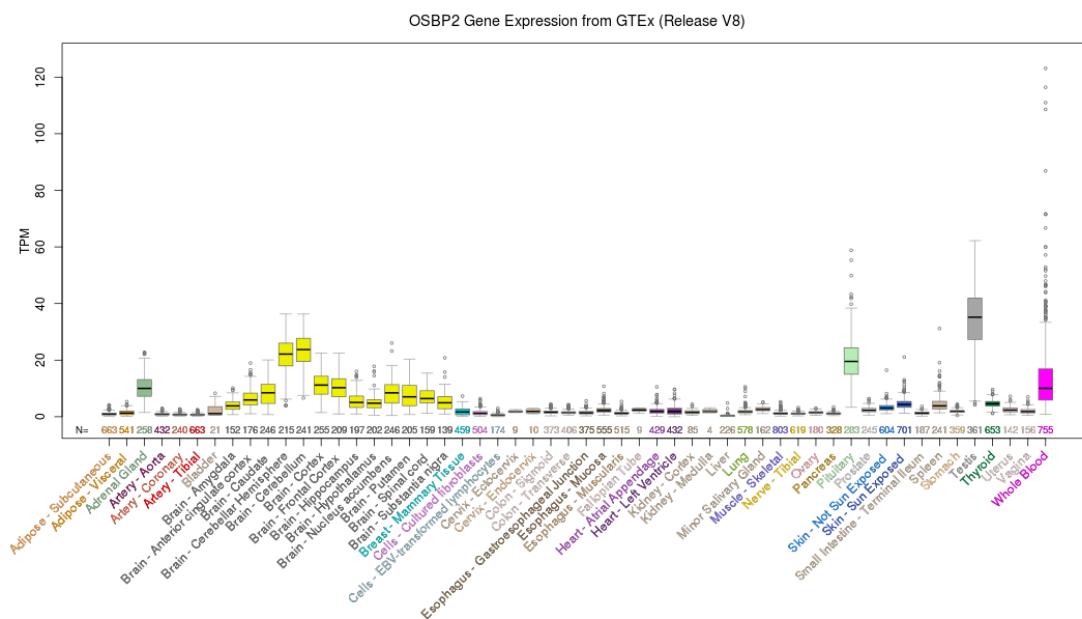


Figure 7.21. Gene expression of *OSBP2* using the National Institutes of Health Genotype–Tissue Expression Project. *OSBP2* is expressed at low levels in the lung and at relatively higher levels in blood.

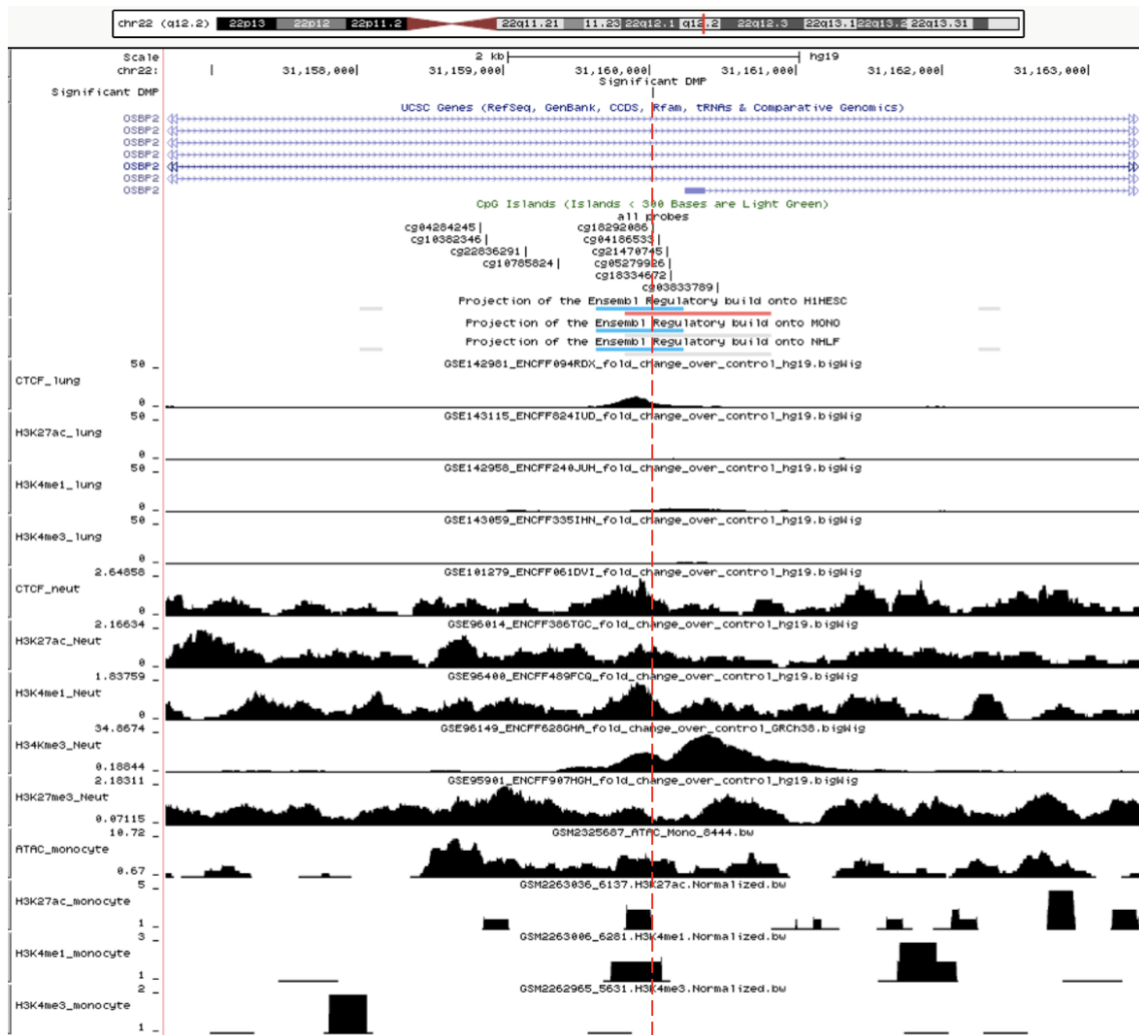


Figure 7.22: Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensembl Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. In lung tissue, the DMP corresponds with a likely CTCF site. There is no clear pattern for neutrophils, but there is a distal H3K4me3 peak consistent with a promoter. In monocytes, H3K27ac and H3K4me1 peaks suggest an enhancer region.

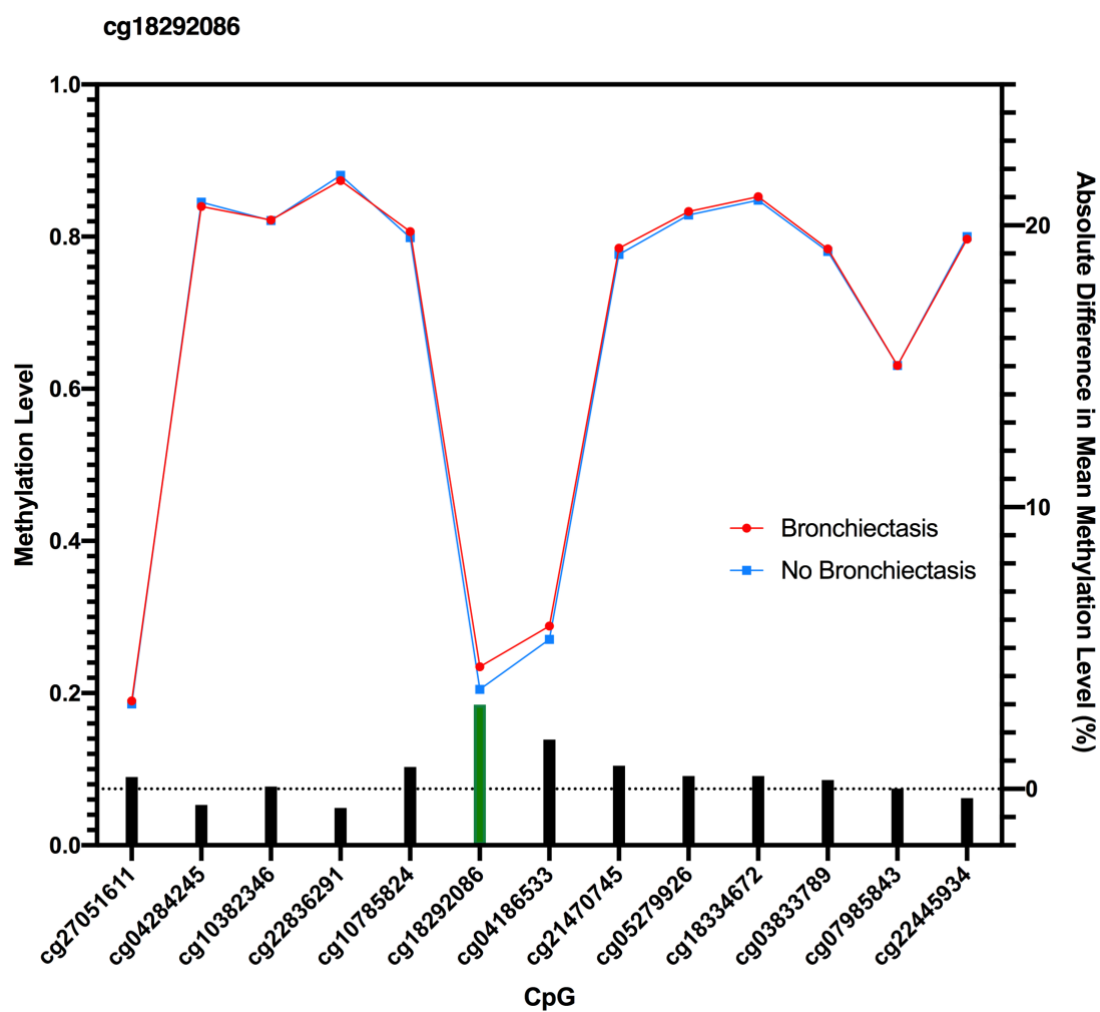


Figure 7.23. A graph of the methylation of probes located 10 kb either side of the significant DMP (highlighted in green). The left y-axis corresponds to the data shown by lines and is the mean methylation of either group at the respective CpGs. The right y-axis is a bar chart of the difference in mean methylation between the bronchiectasis and no bronchiectasis groups. No difference in methylation (i.e. 0) is highlighted by a horizontal dashed line.



Figure 7.24. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). The DMP corresponds to open chromatin in bronchoalveolar and ciliated epithelial cells as well as vascular endothelial cells. Chromatin hubs are found in ciliated epithelial cells and vascular endothelial cells.

Figures Related to Annotation of cg07871500/*LCP1*, *LRRC63*

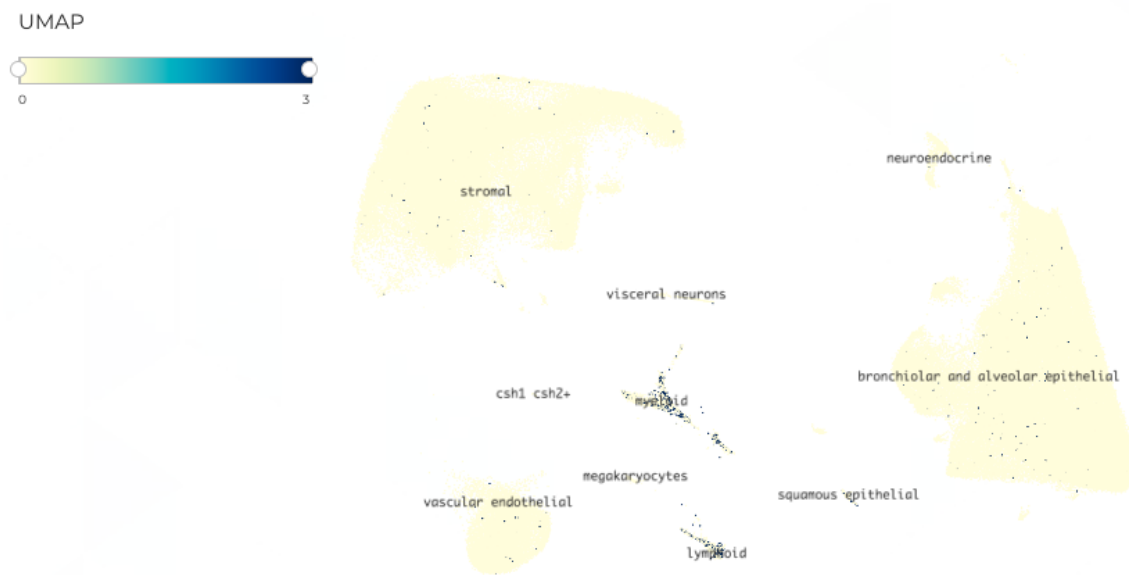


Figure 7.25. Gene expression of *LCP1* from human foetal lung tissue from Descartes Human Gene Expression During Development. *LCP1* is expressed primarily in myeloid and lymphoid cells.

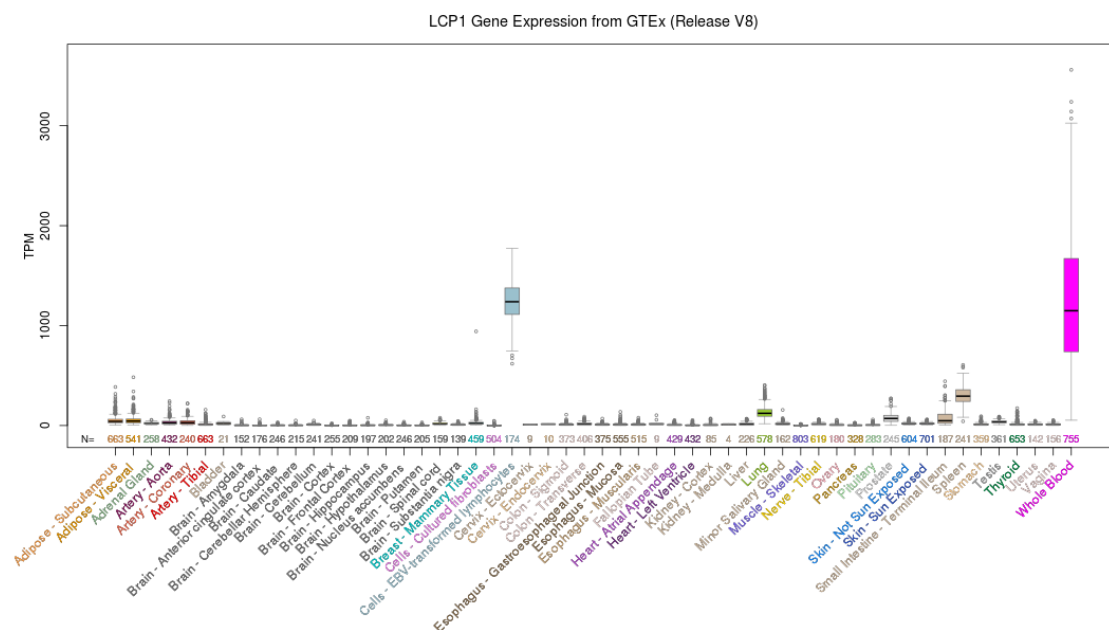


Figure 7.26. Gene expression of *LCP1* using the National Institutes of Health Genotype–Tissue Expression Project. *LCP1* is highly expressed in blood and also expressed at low levels in lung tissue.

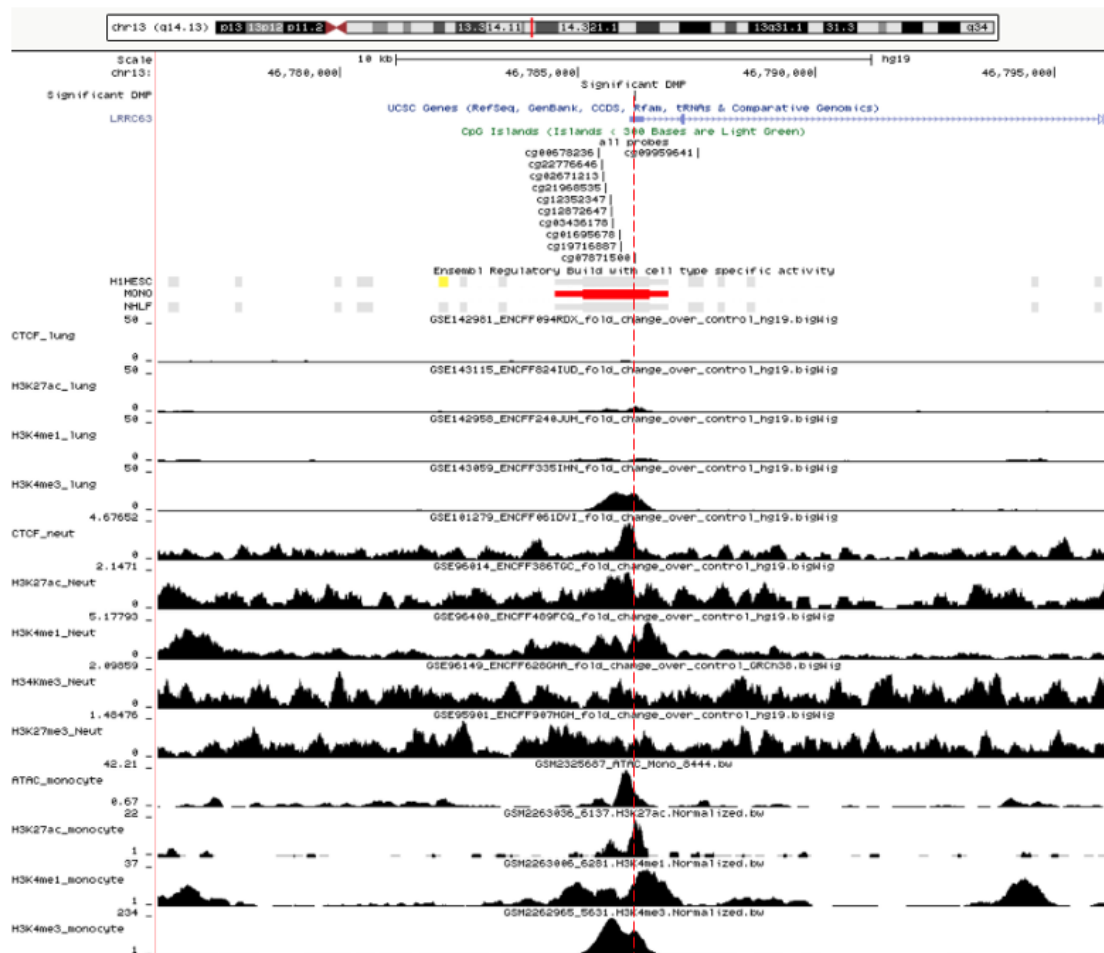


Figure 7.29. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensembl Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. In lung tissue, the area is a likely promoter region with an H3K4me3 peak.

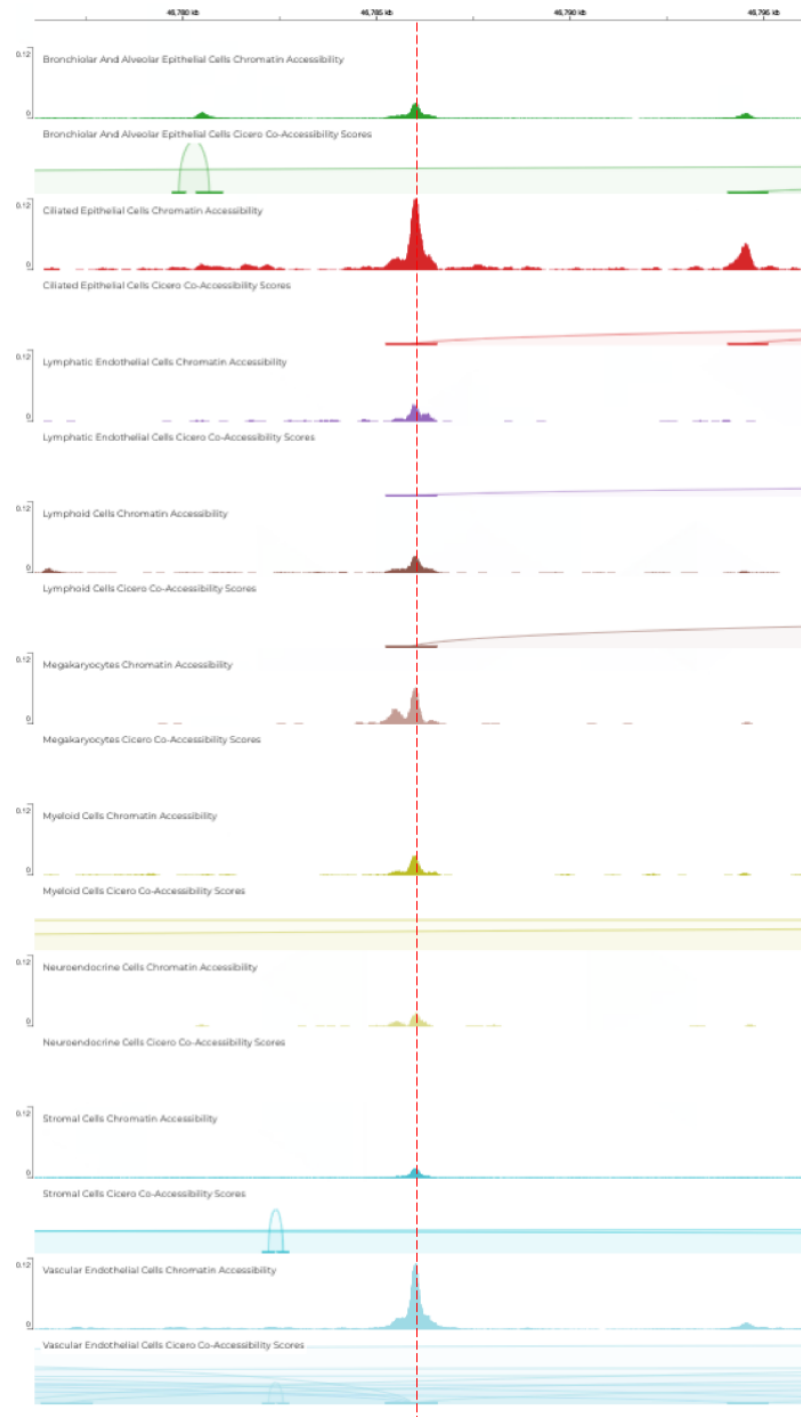


Figure 7.30. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe, which corresponds to open chromatin and predicted chromatin hubs in ciliated epithelial cells and vascular endothelial cells.

Figures Related to Annotation of cg01177524/*NEDD9*

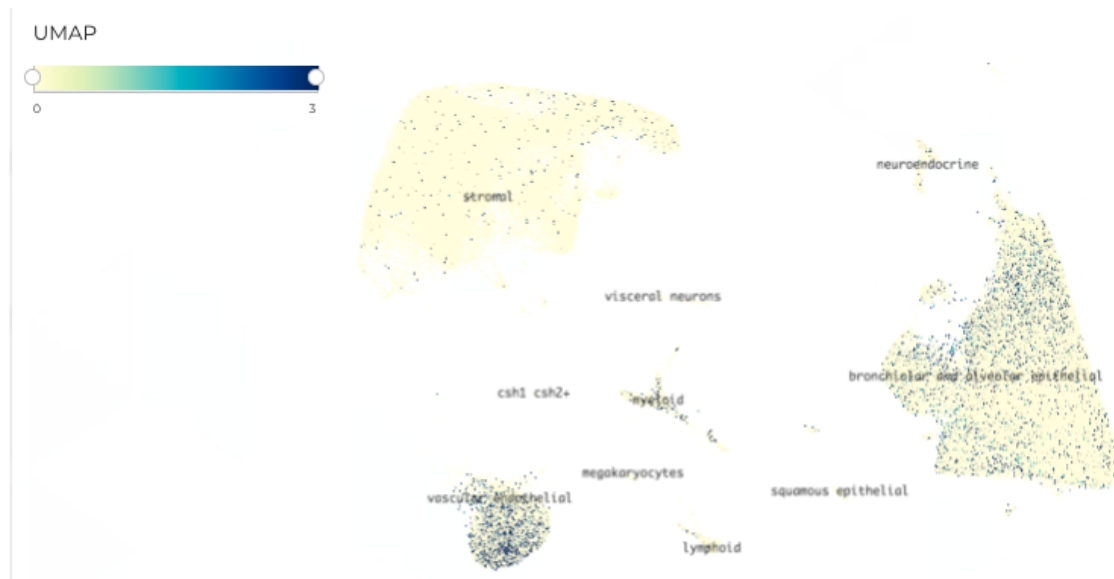


Figure 7.31. Gene expression of *NEDD9* from human foetal lung tissue from Descartes Human Gene Expression During Development. *NEDD9* is highly expressed in vascular endothelial cells.

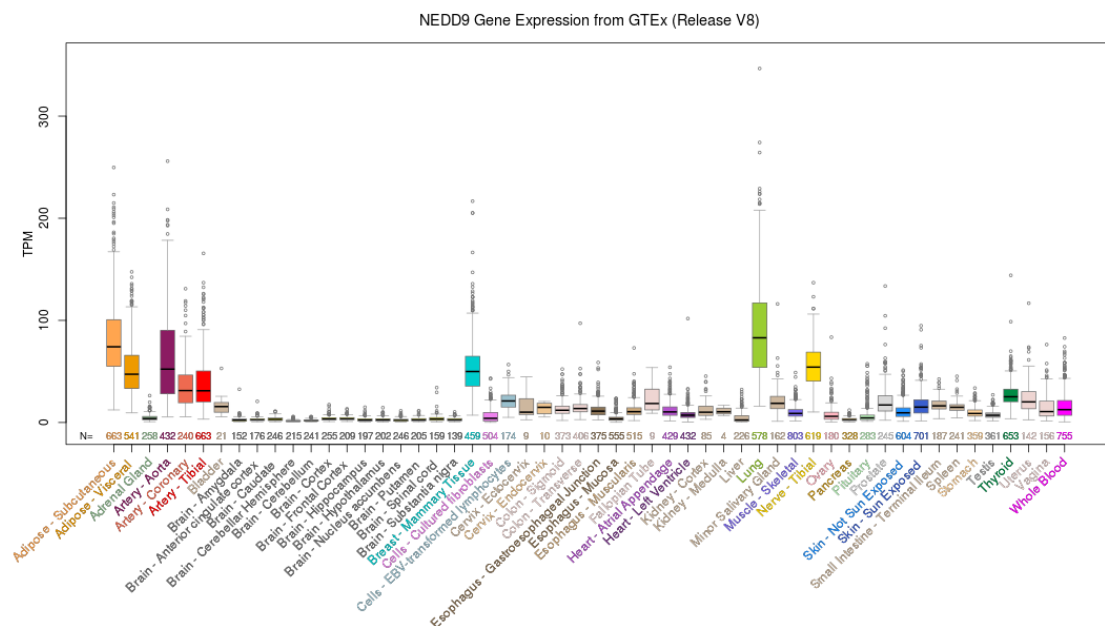


Figure 7.32. Gene expression of *NEDD9* using the National Institutes of Health Genotype–Tissue Expression Project. *NEDD9* is highly expressed in the lung and expressed at low levels in blood.

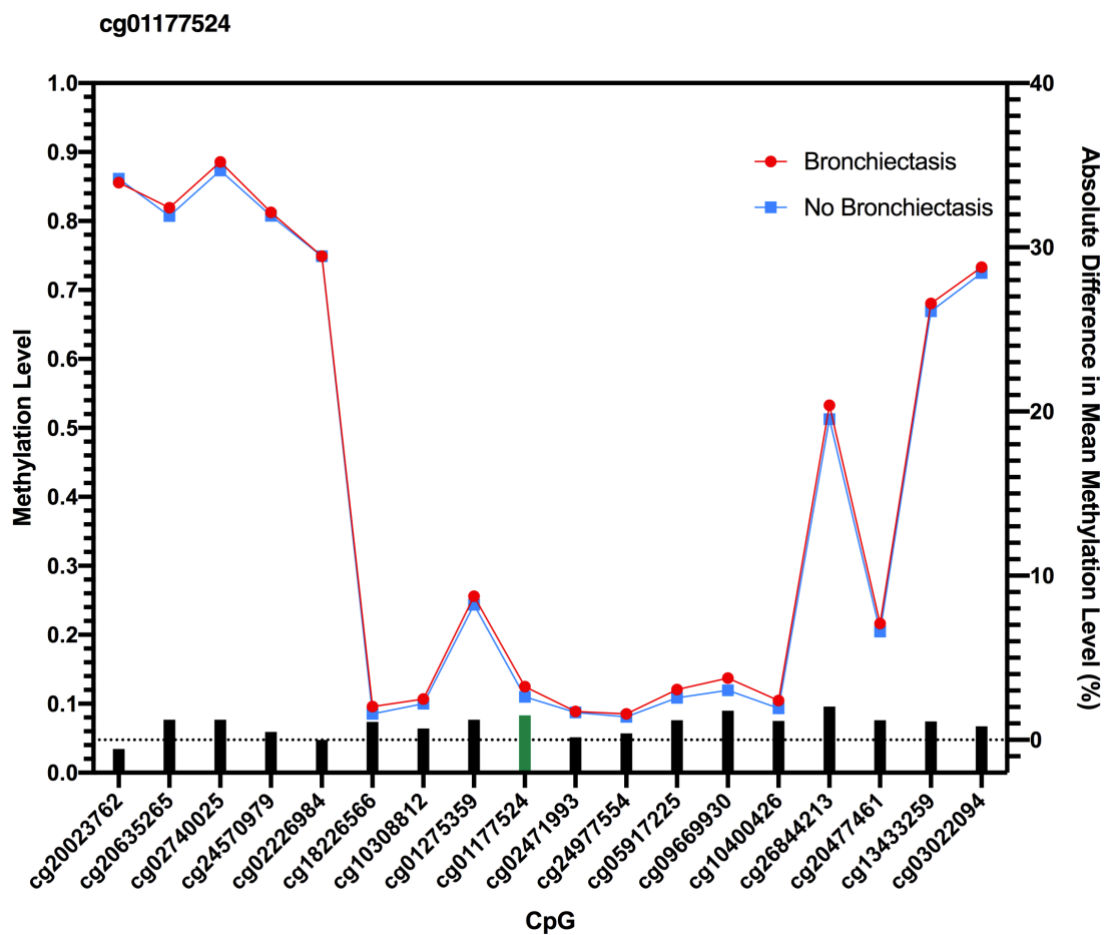


Figure 7.34. A graph of the methylation of probes located 10 kb either side of the significant DMP (highlighted in green). The left y-axis corresponds to the data shown by lines and is the mean methylation of either group at the respective CpGs. The right y-axis is a bar chart of the difference in mean methylation between the bronchiectasis and no bronchiectasis groups. No difference in methylation (i.e. 0) is indicated by a horizontal dashed line.

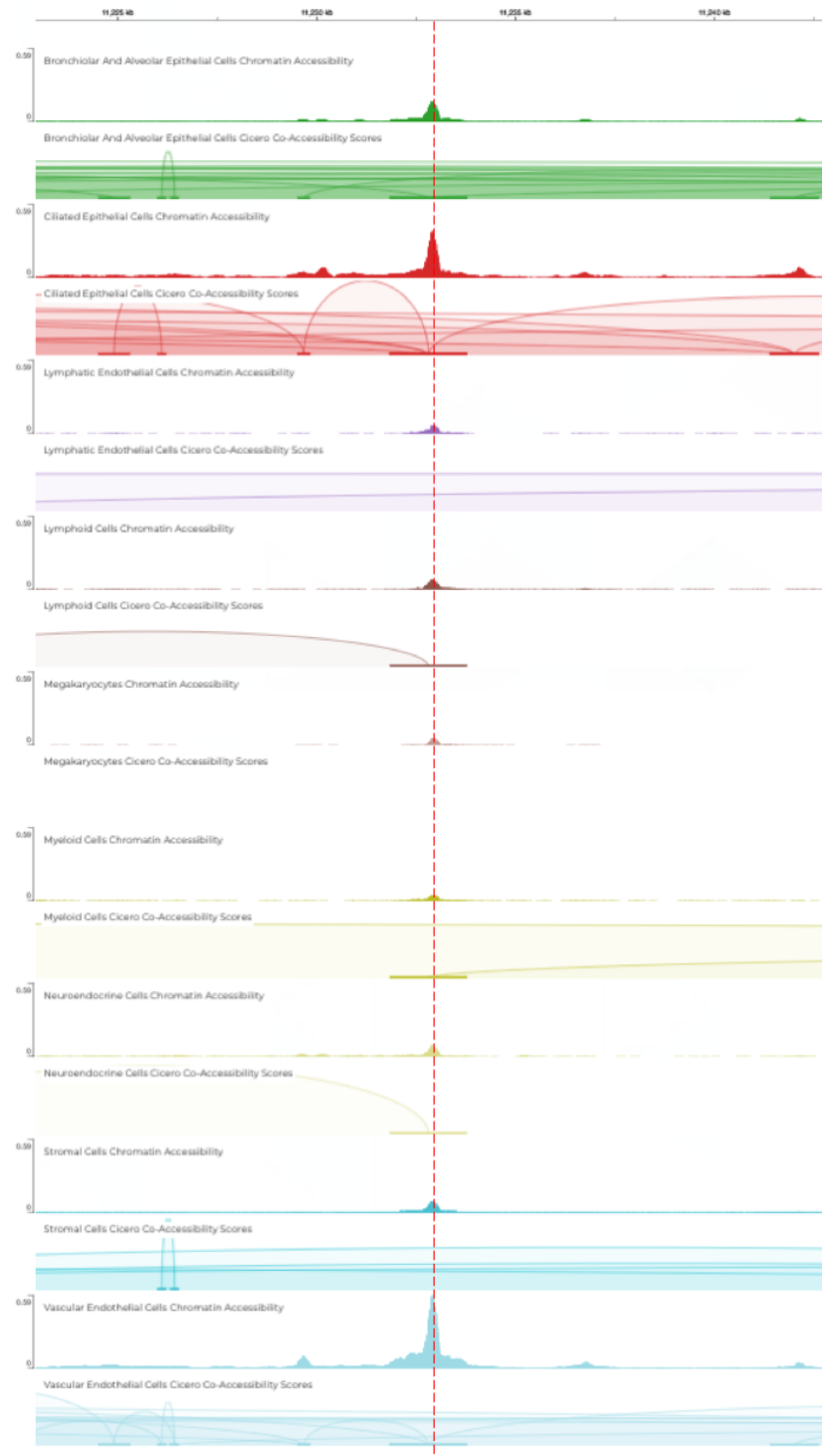


Figure 7.35. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe, which corresponds to open chromatin and predicted chromatin hubs in several cell types.

Figures Related to Annotation of cg04477130/*LINC01237*

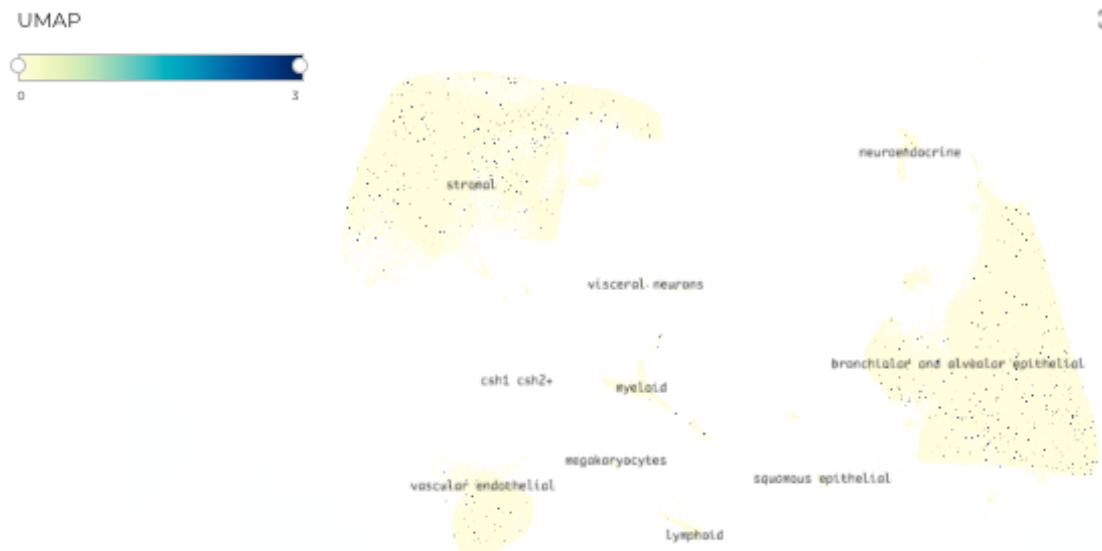


Figure 7.36. Gene expression of *LINC01237* from human foetal lung tissue from Descartes Human Gene Expression During Development. There is minimal *LINC01237* expression.

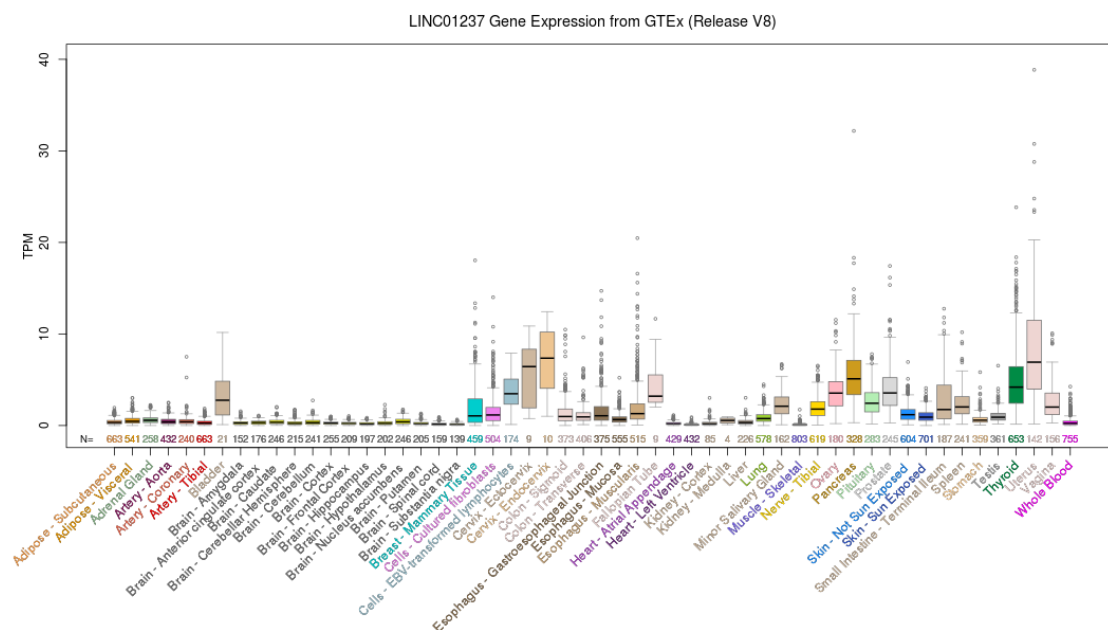


Figure 7.37. Gene expression of *LINC01237* using the National Institutes of Health Genotype–Tissue Expression Project. *LINC01237* is expressed at very low levels in lung tissue.

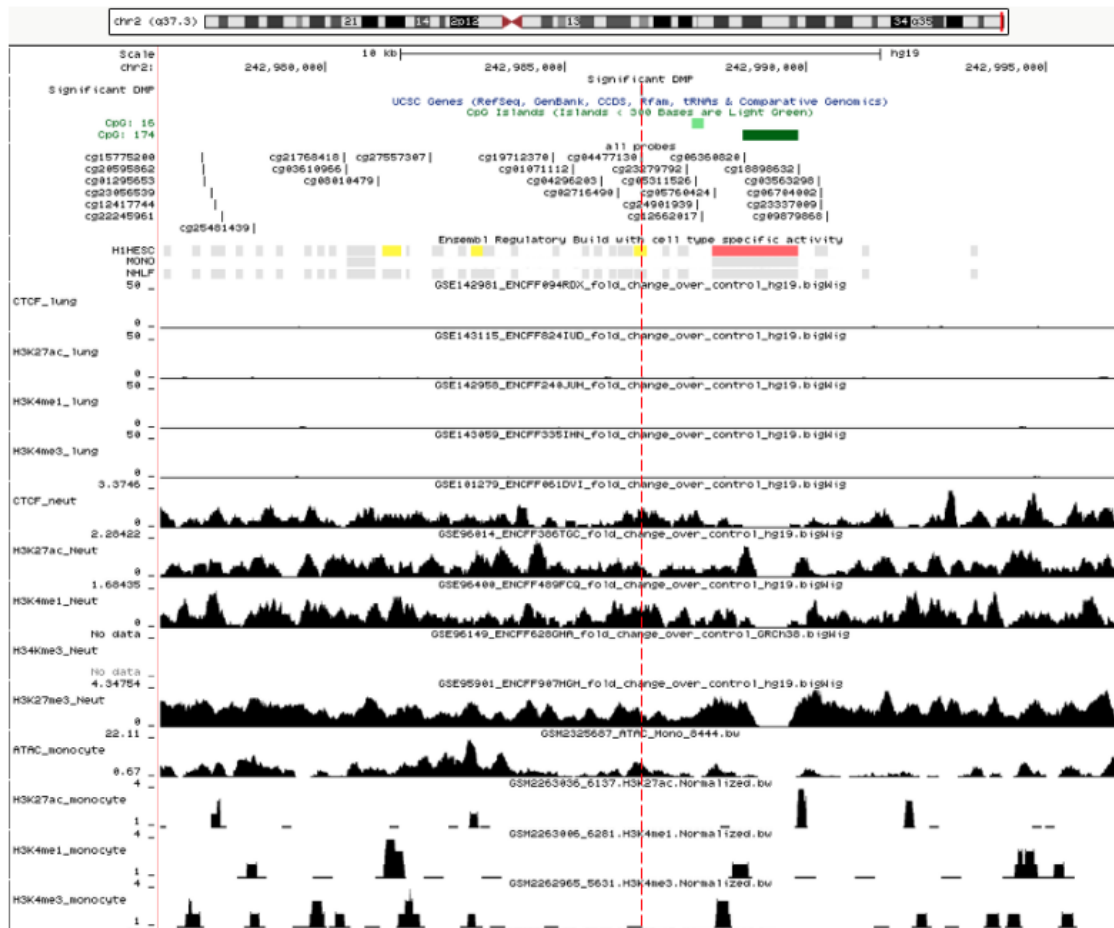


Figure 7.38. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensembl Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. There is no clear pattern of epigenetic markers.

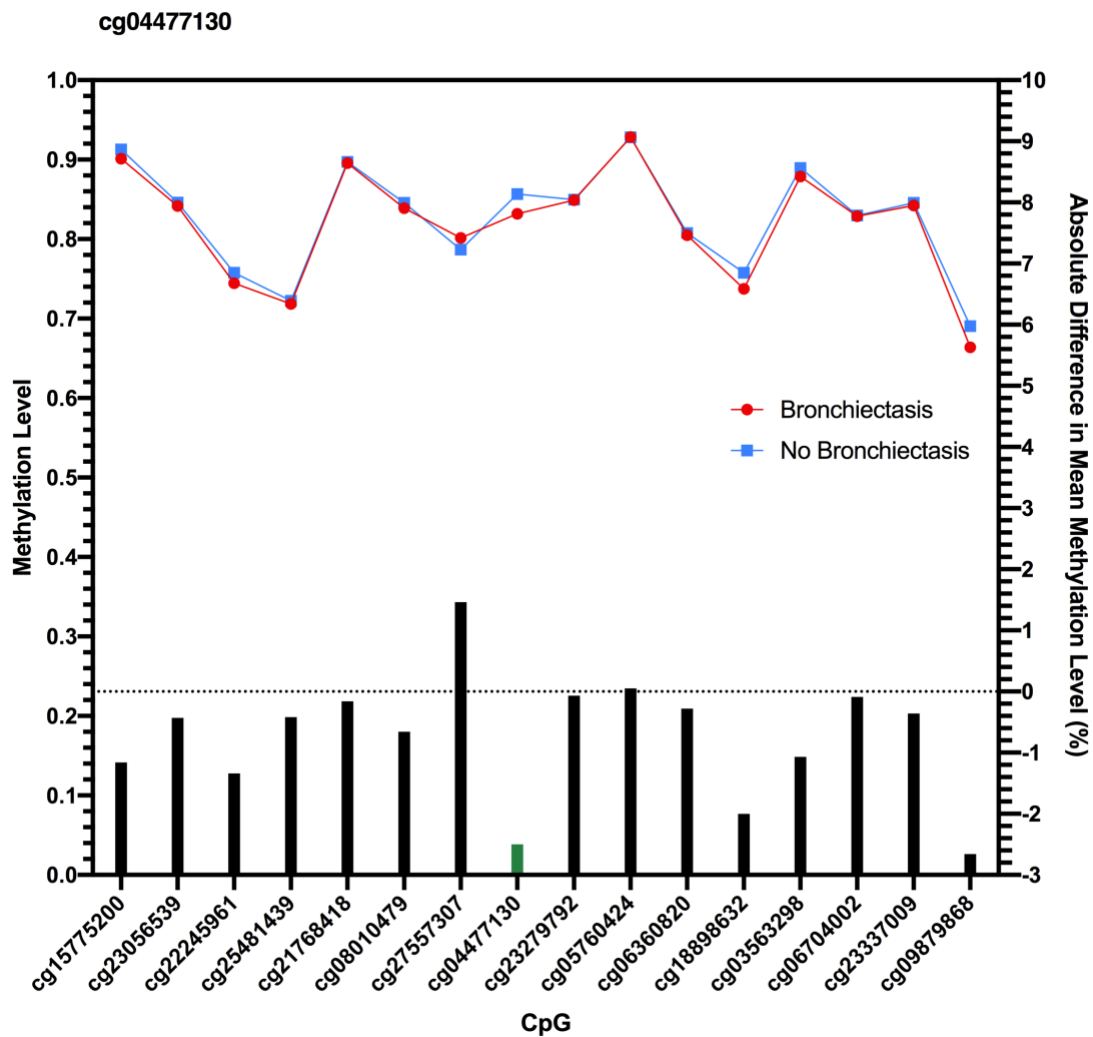


Figure 7.39. A graph of the methylation of probes located 10 kb either side of the significant DMP (highlighted in green). The left y-axis corresponds to the data shown by lines and is the mean methylation of either group at the respective CpGs. The right y-axis is a bar chart of the difference in mean methylation between the bronchiectasis and no bronchiectasis groups. No difference in methylation (i.e. 0) is indicated by a horizontal dashed line.



Figure 7.40. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). There are no predicted chromatin hubs that correspond with the DMP.

Figures Related to Annotation of cg16284279

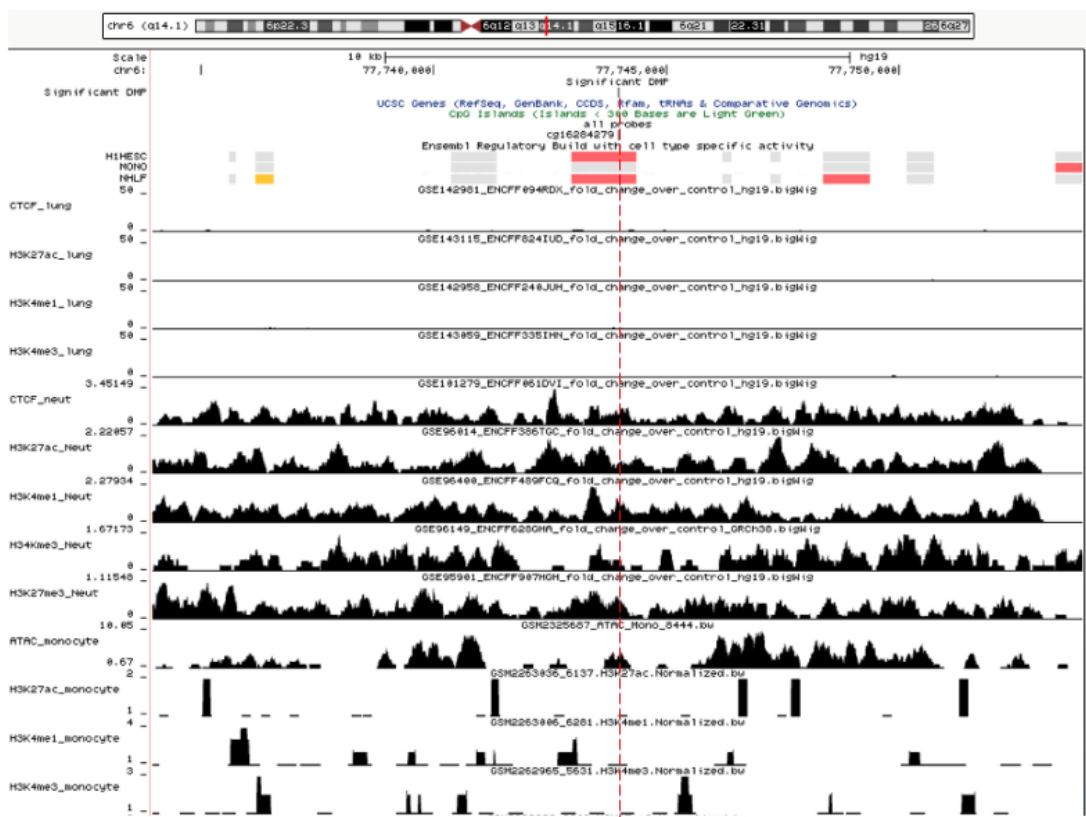


Figure 7.41. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensembl Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. There is no clear pattern of epigenetic markers.

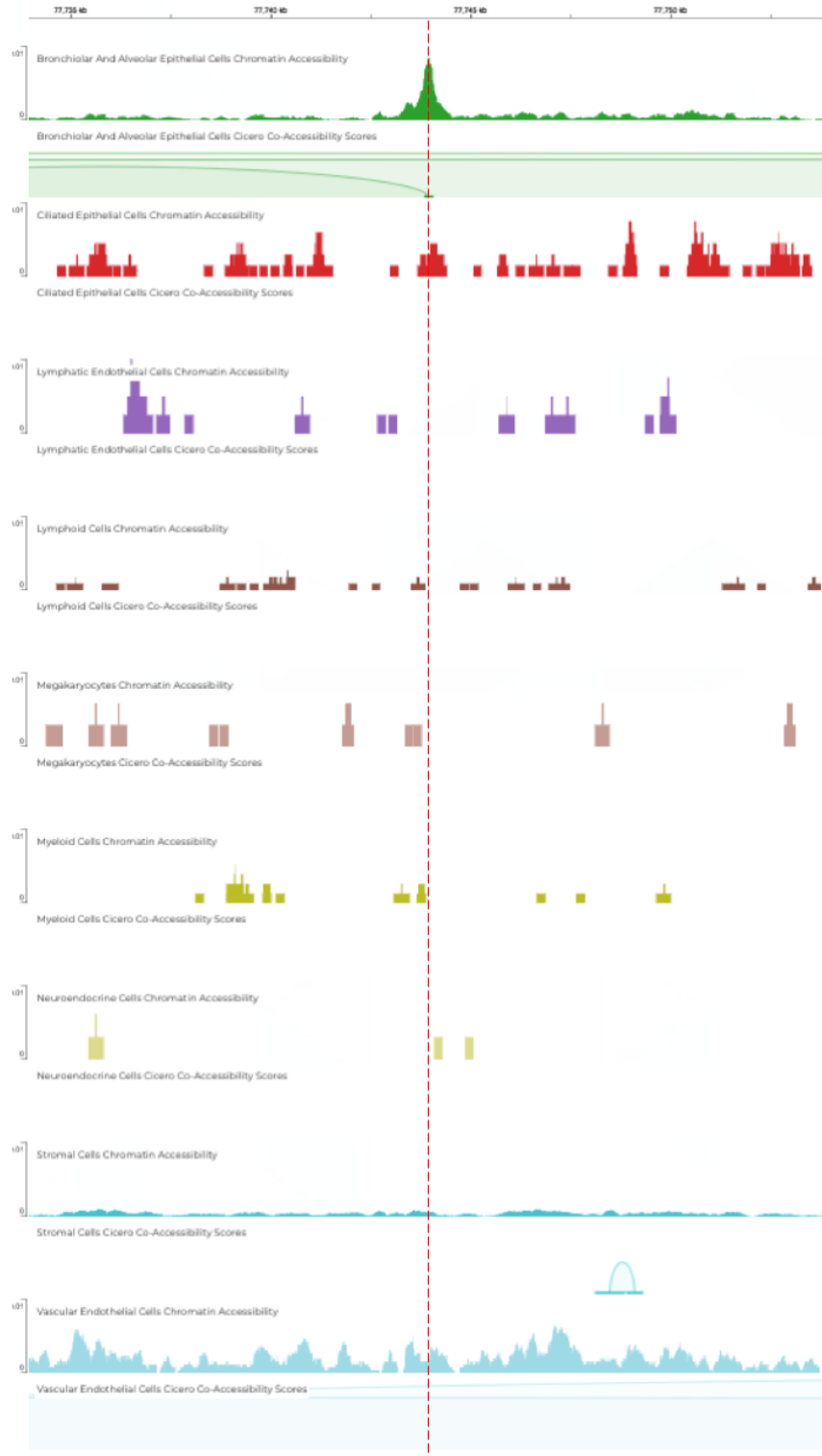


Figure 7.42. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP), which corresponds with open chromatin and a chromatin hub in bronchoalveolar epithelial cells.

Figures Related to Annotation of cg14958154

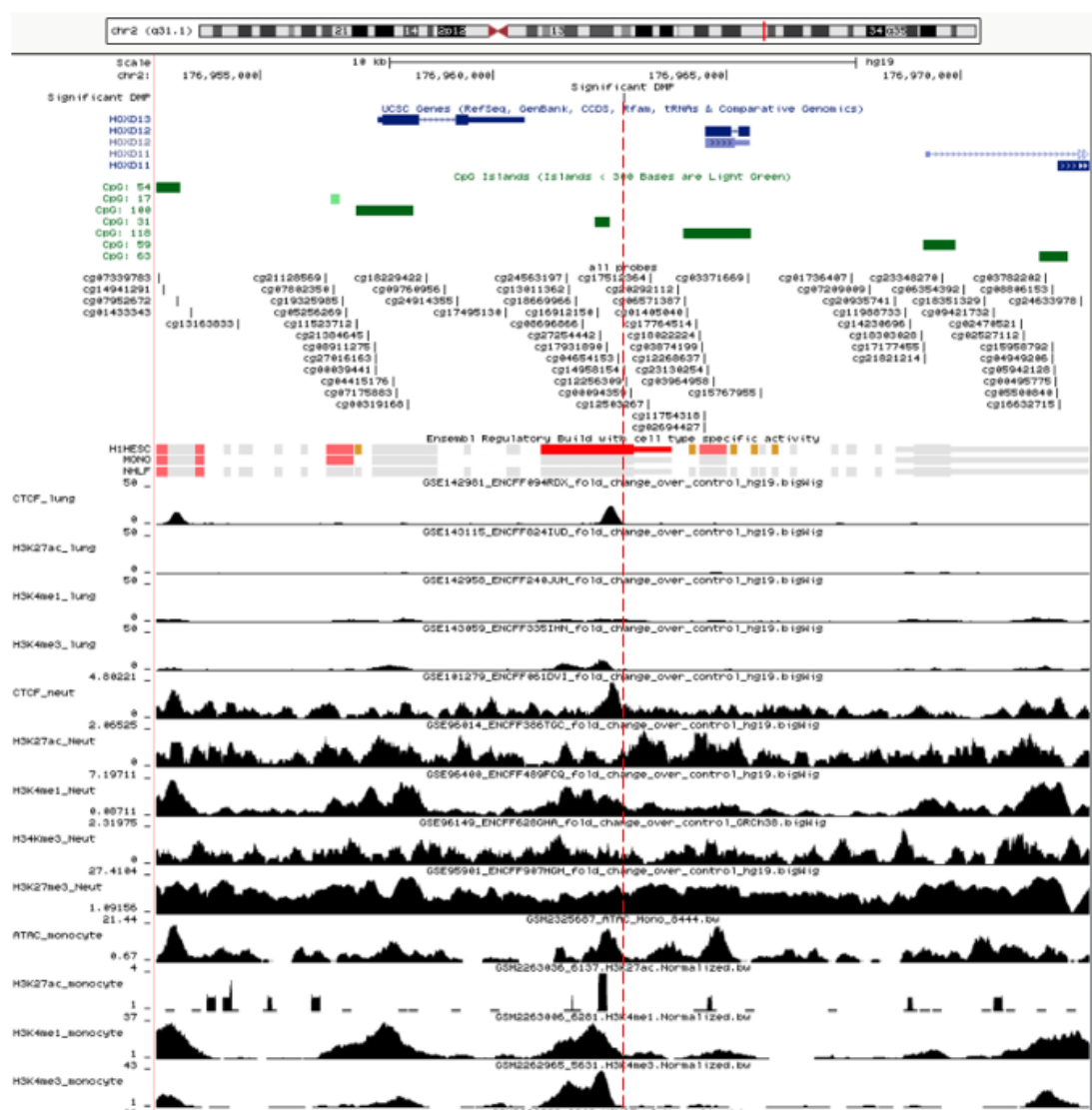


Figure 7.43. Data from UCSC Genome Browser, with data from Ensembl Regulatory Build, ENCODE and Blueprint added. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP). Colour legend for Ensembl Regulatory Build: blue: active CTCF-binding site; orange: predicted active enhancer; light red: predicted active promoter flanking region; bright red: predicted active promoter; gold: unannotated active transcription factor binding sites; yellow: unannotated active open chromatin regions; light grey: inactive region. There is no clear pattern of epigenetic markers.

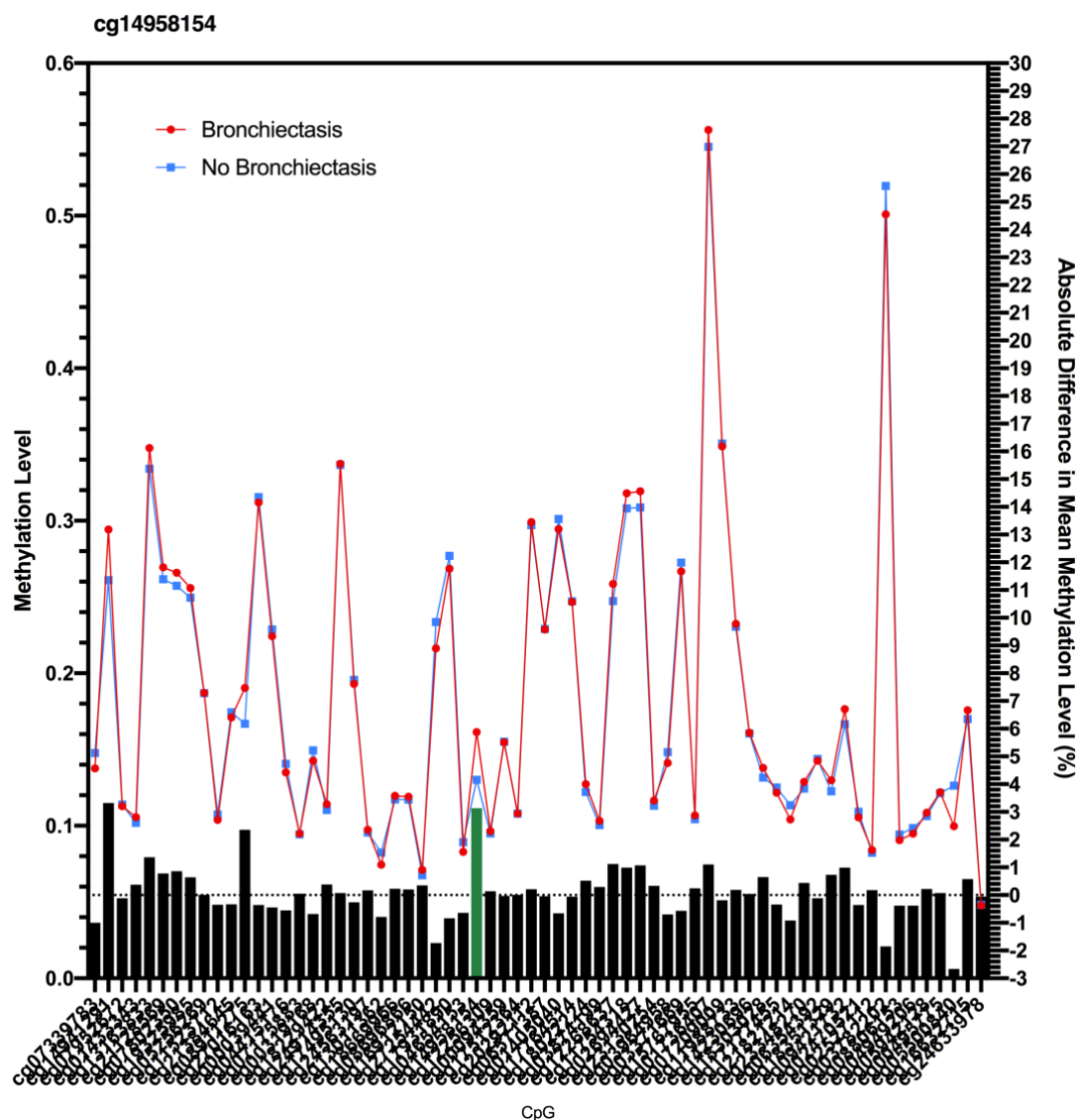


Figure 7.44. A graph of the methylation of probes located 10 kb either side of the significant DMP (highlighted in green). The left y-axis corresponds to the data shown by lines and is the mean methylation of either group at the respective CpGs. The right y-axis is a bar chart of the difference in mean methylation between the bronchiectasis and no bronchiectasis groups. No difference in methylation (i.e. 0) is indicated by a horizontal dashed line.

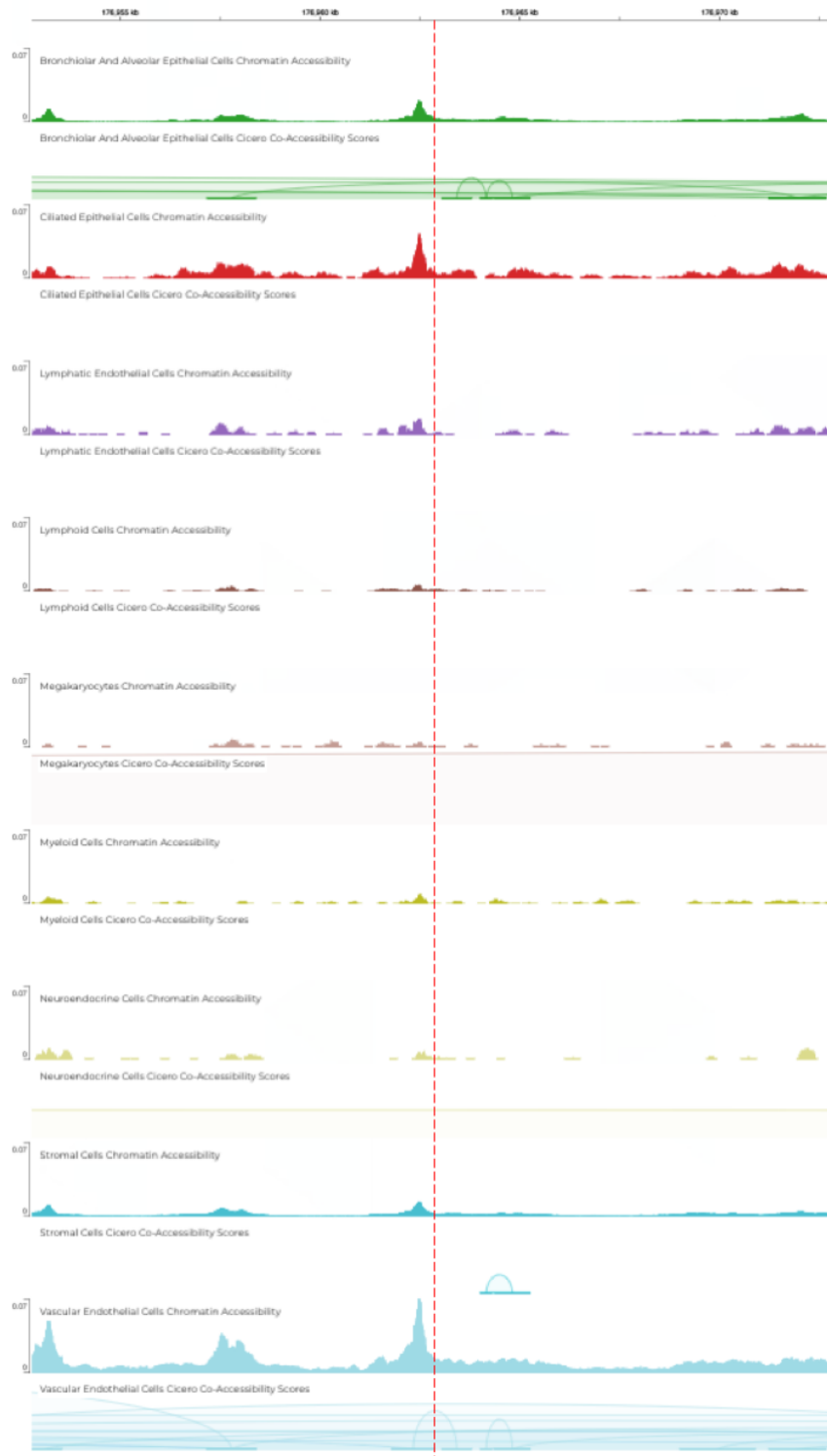


Figure 7.45. ATAC-seq data from foetal lung tissue. The peaks correspond to chromatin accessibility. In the Cicero co-accessibility scores, chromatin hubs are indicated by blocks, with linking lines indicating interaction. The height of the line indicates the strength of the interaction. The vertical dashed red line indicates the position of the significant differentially methylated probe (DMP), which does not directly align with any predicted chromatin hubs.

CHAPTER 8: Conclusions and Future Directions

8.1 OVERVIEW OF THE THESIS

Lung disease is the major source of morbidity and mortality in CF. Chapter 1 reviewed the burden of lung disease in CF, including that it arises in early life and is best measured by the assessment of lung structure using CT. The chapter also described the value of predictive biomarkers of future lung disease severity, highlighting that no such biomarkers exist. The field of epigenetics was then reviewed, with a focus on DNA methylation and an argument for why DNA methylation-based biomarkers may be relevant to CF lung disease.

Before embarking on the experiments that sought to fulfil the primary aims and objectives of the thesis, several background works were completed. To better understand previous efforts to identify predictive biomarkers, the field of non-*CFTR* genetic modifiers of lung disease was systematically reviewed in Chapter 2. Chapter 3 presented a narrative review of DNA methylation biomarkers of future health outcomes in paediatric diseases. The key findings of both reviews informed the subsequent experiments, highlighting the need for sensitive and clinically relevant disease outcome measures, appropriate biological specimens and appropriate statistical testing.

Two necessary and novel methodological projects were then described in Chapters 4 and 5. In Chapter 4, methods for cryopreserving and purifying BAL specimens into individual cell populations were developed. These methods were then used in Chapter 5 to develop a panel of reference epigenomes for each cell population found in BAL samples. The ability of these reference epigenomes to accurately estimate the cell composition of BAL samples was then validated using two different datasets.

With Chapters 2 and 3 informing the optimal study design and Chapter 5 providing reference epigenomes, the necessary preparations had been undertaken for the studies that would address the main aims and hypotheses of this thesis. These projects were described in Chapters 6 and 7. While the project described in Chapter 6 was unable to identify predictive biomarkers, the experiments conducted in Chapter 7 identified an early-life epigenetic profile that predicts future lung disease severity.

8.2 REVIEW OF SIGNIFICANT FINDINGS

The most significant findings of this thesis include the DMPs identified in Chapter 7 and the development of several novel methods, including the cryopreservation of BAL specimens, flow cytometry–based immune cell profiling and DNA methylation reference epigenomes for the cell types found in BAL samples.

The study identified seven significant DMPs in BAL samples at 1 year of age that were associated with disease severity at 5 years of age. When these DMPs were assessed using tests of diagnostic performance, they performed well, with an AUC ranging between 0.863 and 0.918. The diagnostic performance of the DMPs was then further improved by calculating composite scores, with the best-performing composite score having an AUC of 0.990. Several of the DMPs also mapped to genes that are relevant to the pathophysiology of CF lung disease, including *HMOX2*, *OSBP2*, *LCP1* and *NEDD9*. Annotation of the DMP locations using publicly available databases showed they were frequently located at sites of other epigenetic regulation such as CTCF or promoter sites. Thus, differential methylation at these positions may interact with other epigenetic markers to alter the expression of genes relevant to CF lung disease pathophysiology.

Several robust methods were developed in this thesis. A protocol for the cryopreservation of BAL cells was developed and validated. It was demonstrated that the cryopreservation and thawing process maintained the viability of 61.8–90.7% of cells. Further, cryopreservation did not alter the proportion of most cell types in the BAL specimens. The only cell type affected was CD16⁺ granulocytes, which were reduced, as seen in previous studies regarding the effect of cryopreservation. Flow cytometry protocols for phenotyping the most common cell types in BAL specimens as well as rarer immune cell types were developed. FACS methods for flow cytometry were also developed, enabling the purification of the major cell types in BAL samples for downstream analysis. DNA methylation–based reference epigenomes for the four major cell types in paediatric BAL samples were identified and validated, showing sound performance. The reference epigenomes for granulocytes and lymphocytes were also validated in artificial cell mixtures derived from adult blood and were shown to have excellent performance.

The significance of the identification of the DMPs is that they provide a target for further evaluation as predictive biomarkers of future CF lung disease severity. As discussed in Chapter 1, there are currently no biomarkers used in clinical practice. While extensive pipelines²⁸⁹ exist for other aspects of CF care, including CFTR-directed, anti-

inflammatory and anti-infective therapies, predictive biomarkers have received less attention. The seven DMPs represent novel targets that are worthy of validation studies in external cohorts. In addition, given that several of the DMPs may have functional effects on genes implicated in the pathophysiology of CF lung disease, they may also represent novel targets for therapeutic intervention. For example, one of the DMPs mapped to *HMOX2*, a gene involved in neutrophil degranulation. Therapeutic interventions that attempt to mimic the pattern of *HMOX2* gene expression could be developed and their resultant effect on neutrophil degranulation observed in subjects without bronchiectasis. Limiting the potential utility of the DMPs as a therapeutic target is the fact they were not significant DMP's in the first EWAS study. Hence they may only be involved in the pathophysiology of CF lung disease in infancy and hence only be a therapeutic target in this limited, but still important, age group.

The methodological developments are significant in their ability to facilitate research in CF and other paediatric lung diseases using tissue-specific cells obtained via BAL. Importantly, BAL is a safe procedure that is currently used in clinical practice. The ability to preserve viable cells from BAL and then use flow cytometry to phenotype cell composition as well as purify populations of interest for downstream analysis is a significant advance. The reference epigenomes will allow for further research regarding DNA methylation in these diseases because they can be used to deconvolute the cell composition of DNA methylation data derived from BAL samples and the multiple cell types they contain.

Some limitations of the work presented in this thesis should be acknowledged. The DMPs were demonstrated in only one early-life population, thus must be validated in an external cohort. The importance of validating significant findings has been highlighted in guidelines on the design of EWAS.¹⁸⁷ Several factors may reduce the chances of successful replication. For example, the small degree of differential methylation at the DMPs may make replication in an external cohort less likely. The mean absolute level of differential methylation for the DMPs ranged from 1.41–7.19%, and five of the seven probes had an absolute differential methylation of less than 3%. The significant differential methylation observed in the current study was attributable to a tight spread of methylation values; however, in a larger external study, if the spread of values is broader, it is unlikely that the DMPs will be successfully replicated. In addition, there was no differential methylation at these positions in the older cohort of children, who had DNA methylation assessed at 6 years of age with respect to the presence or absence of

bronchiectasis at 9 years of age. This indicates that differential methylation may only occur in early life and that children with CF may have uniform methylation at these sites by 6 years of age. This would be a significant limitation because it would mean that the DMPs are only suitable for use in early life. There are several limitations to the EWAS that was undertaken in the older cohort. In addition to the possibility that CF lung disease progresses at a molecular level so that all subjects have a uniform DNA methylation pattern by the age of 6 years, study power was another major limitation. The initial EWAS only included 28 subjects, and despite the study design attempting to leverage the increase in power associated with an extreme phenotype (in this case normal lung structure at 9 years of age), it was most likely underpowered.

The primary limitation of the methodological work relates to the reference epigenomes. Because the reference epigenomes were derived from children with CF, and DNA methylation is affected by age and disease, they may not be suitable for broader use. These concerns are reduced by the excellent performance of the granulocyte and lymphocyte epigenomes when used to estimate the cell composition of samples derived from healthy adults. The validation of the reference epigenomes in estimating the cell composition of BAL samples was also limited by the methods used in the validation experiments. Flow cytometry performed on cryopreserved samples determined the ground truth cell composition against which the methylation-based estimates were compared. However, given the effect of cryopreservation on CD16⁺ granulocytes, the accuracy of flow cytometry estimates is limited, making it difficult to truly assess the performance of the reference epigenomes.

The projects that did and did not identify DMPs both had several strengths. Both studies were designed in keeping with best practice guidelines for EWAS; in particular, they each had a clear research question with a clinically relevant outcome, used tissue-specific samples and accounted for the cell composition of BAL specimens. The outcome—presence or absence of bronchiectasis—was assessed using robust methods in both studies. Although different methods were used, they both had high intra- and inter-observer reliability. A further strength was the excellent performance of the identified DMPs when assessed using AUC analysis. Thus, if replicated, they will be able to accurately predict future lung disease severity and have high clinical utility. For example, the best composite DMP score had an AUC of 0.990, and using a threshold score of 2.5×10^{-5} , the future disease status of only one child of 56 was predicted incorrectly.

The strengths of the developed methods relate to their careful development and validation. The cryopreservation protocol was demonstrated to maintain the viability of an average of 78% of cells. As discussed, the effect of cryopreservation on cell composition was also carefully assessed, and it was shown that only CD16⁺ granulocytes were affected. Similarly, the flow cytometry panels were validated using t-distributed stochastic neighbour embedding, a form of unsupervised analysis, confirming the results identified with a traditional manual gating approach. The reference epigenomes were validated using two datasets. The excellent performance of the two reference epigenomes, which were tested on artificial cell mixtures derived from healthy adult blood samples, is a strength of the study because it demonstrates that the epigenomes are likely to perform robustly in cohorts outside of children with CF.

There has been limited work examining DNA methylation in CF, and no studies have examined the role of DNA methylation as a predictive biomarker. Hence, the identification of seven DMPs is a novel finding. Magalhães et al.¹⁸⁶ examined DNA methylation in a cohort of 32 adult subjects with CF. The study had several important differences compared with the studies presented in this thesis, including the age of subjects (adults v. children), outcome of interest (FEV₁ v. bronchiectasis), biospecimen used (nasal epithelial cells v. BAL), DNA methylation analysis (approximately 450,000 CpG sites v. > 850,000 CpG sites) and study design (cross-sectional v. longitudinal). When the authors divided the group into extreme phenotypes based on FEV₁, they identified 187 DMPs related to 116 genes. They further assessed the correlation between DNA methylation and FEV₁ across the entire cohort rather than just those with extreme phenotypes and identified 50 DMPs. None of these DMPs overlapped with the seven DMPs identified in this thesis, which may reflect differences in study design and, in particular, the fact that the seven DMPs are only valid in early life.

The identified DMPs are a significant step towards identifying predictive biomarkers of early-life CF lung disease severity. As discussed in Chapter 1, if the DMPs are successfully validated, this will have broad implications for the CF community. Families of infants with CF will have more information regarding the likely trajectory for their children and be able to make better-informed decisions regarding available treatments. For clinicians, predictive biomarkers could inform the management of children during the preschool years, which are particularly challenging because of the lack of routinely assessed objective markers of disease severity and variability in disease progression. For CF centres and health authorities, predictive biomarkers could inform resource allocation,

with those predicted to have more severe disease receiving more intensive care. For researchers, there are several implications. The genes to which the DMPs are mapped could be investigated as therapeutic targets, in particular interventions that mimic the epigenetic profiles associated with normal lung structure. When assessing the effect of other new interventions such as CFTR-directed therapies, the predictive biomarkers could be assessed for whether they predict therapeutic response. Also, given that CFTR-directed therapies may reduce the need for other treatment, studies of CF treatments that can be safely stopped will need to be undertaken. Predictive biomarkers could be used to help researchers identify children predicted to have normal lung structure at 5 years of age, which would be the ideal group to assess initially in trials of therapeutic cessation.

The developed methods could potentially be used in a wide range of studies investigating pulmonary diseases. Several new ‘omic’ technologies (discussed in more detail below), which can provide profound insights into molecular mechanisms using small numbers of cells, have been developed. These technologies are suitable for BAL samples; however, viable cells are often needed. Thus, the cryopreservation methods developed will enable BAL samples collected as part of clinical care to be stored for further investigation. In addition, cells often need to be purified into subpopulations for further analysis, and the flow cytometry methods developed in this regard will also be highly useful. The value of the cryopreservation and flow cytometry methods is highlighted by the AREST CF cohort implementing them as routine practices when biobanking BAL samples. A portion of BAL fluid is cryopreserved, and a portion of the sample is profiled using the augmented flow cytometry panel developed in Chapter 4 to provide in-depth immune phenotype information.

Several lessons were learned during the development of this thesis that would have informed a better study design had they been previously known. First, the effect of cryopreservation on CD16⁺ granulocytes likely caused significant issues with validation of the reference epigenomes when extremely accurate data regarding cell composition was needed. A better study design would have been to use flow cytometry to purify cell types and create artificial BAL cell mixtures, similar to those created for adult blood. This would have resulted in a more accurate determination of sample cell composition against which the methylation-based estimate could be compared. The other lesson was regarding the age groups studied. The first EWAS assessed DNA methylation at 6 years of age. However, as discussed, by this age methylation patterns may be uniform in subjects with

CF. A better study design for both EWAS may have been to assess DNA methylation in early life.

8.3 FUTURE DIRECTIONS

8.3.1 Validation of predictive biomarkers

As discussed, validation of DMPs as predictive biomarkers in an external cohort is the crucial next step if they are to progress towards being used in clinical practice. As demonstrated in the systematic review of non-*CFTR* genetic modifiers of lung disease severity in Chapter 2, predictive biomarkers often fail replication in a different cohort. A pragmatic starting point would be to assess whether the findings could be replicated in a cohort comprising different CF centres. If the DMPs are valid biomarkers, they should show a similar performance in external cohorts. A relative barrier to this is that few external cohorts exist that collect the required samples and outcome measures to validate the DMPs. However, SHIELD CF in Ireland represents one suitable cohort because subjects undergo similar monitoring to that of children in AREST CF.²⁸⁸ For example, BAL is performed at 1 year of age and CT is performed at 5 years of age, so it would be possible to use this cohort to validate the DMPs.

If the findings were successfully replicated in a cohort such as SHIELD CF, then a prospective trial could be justified. This trial may be conducted to not only validate the DMPs and their ability to predict structural lung disease but, if commenced in conjunction with the introduction of elexacaftor/tezacaftor/ivacaftor in the preschool age group, assess whether the DMPs remain valid biomarkers after the initiation of this treatment and their ability to predict therapeutic response. As discussed in more detail subsequently in this chapter, elexacaftor/tezacaftor/ivacaftor therapy will significantly alter the burden of CF lung disease, and it will be important that any further evaluation of predictive biomarkers considers the effect of this therapy.

In addition, a prospective trial could assess whether the DMPs remain valid biomarkers in biospecimens obtained in a less-invasive manner. If predictive biomarkers could be validated in a less-invasive biospecimen, it would significantly aid the uptake into clinical practice. While the obvious advantage of BAL is that it obtains immune cells (including resident populations in the lung) as well as respiratory epithelium, the major downside is the need for general anaesthesia to obtain samples. It is possible that similar methylation patterns may be seen in other sample types such as blood or nasal epithelial cells that can

be obtained in a less-invasive manner. For CpG sites at which DNA methylation is primarily determined by underlying genotype, DNA methylation observed will be uniform across all cell types. Conversely, CpG sites for which environmental factors are involved in determining DNA methylation, the pattern of DNA methylation may vary between cell types. For the seven DMPs, the role of genotype and environment in determining methylation is currently unknown. Another approach would be to isolate circulating cell-free DNA from respiratory epithelium in the blood and assess its methylation patterns to determine whether they correlate with the methylation of BAL specimens. Although DNA methylation analysis of cell-free DNA is a relatively novel technique, the origin of cell-free DNA can be determined by its methylation pattern, suggesting that this type of DNA could be studied as a biomarker.²⁹⁰ This method enables the identification of cell-free DNA from lung tissue, and it would have to be refined to specifically identify lung epithelial cells. Cell-free DNA has been studied as a biomarker of lung cancer;²⁹¹ however, it has not been explored in other lung diseases.

8.3.2 Approach to novel therapeutic targets

Several of the DMPs mapped to genes that may be implicated in the pathophysiology of CF lung disease, thus could be investigated as potential therapeutic targets. A prime example of this is *HMOX2*. Gene ontology suggests that *HMOX2* is involved in neutrophil degranulation, and it is a relative of *HMOX1*, one of the few validated genetic modifiers of CF lung disease. Gene ontology also indicates it is involved in other pathways that may be relevant to CF lung disease, such as response to hypoxia and iron ion homeostasis. The first steps to identifying whether the differential methylation seen in Chapter 7 could be of relevance as a therapeutic target is to assess whether it has a functional outcome. In other words, does the differential methylation seen at cg26513131 affect gene expression of *HMOX2*? The presence of a likely CTCF site at this locus supports the region as being involved in *HMOX2* expression. DNA methylation and CTCF sites are known to interact to determine gene expression.²⁷¹ In addition, *HMOX2* is expressed in the relevant cell types so may be involved in CF lung disease pathophysiology. However, the small difference in methylation levels between the two groups and the lack of differential methylation of surrounding positions make it less likely that the differences observed in the EWAS will result in functional outcomes.

To assess whether *HMOX2* methylation is associated with lung disease severity, future experiments could involve simultaneously assessing *HMOX2* methylation, gene

expression and protein levels. *HMOX2* methylation could be assessed at the probe of interest (cg26513131) but also at other relevant CpGs, which would provide a broader picture of the effect of *HMOX2* methylation. Both gene expression and DNA methylation could be assessed at the sample or single-cell level. The advantage of sample-level assessment is that well-established methods for DNA methylation analysis (such as those used in this thesis) and RNA expression exist and are less expensive.²⁹² The advantage of single-cell assessment would be assessing exactly which cell population (e.g. macrophages, AECs) is driving any observed differences in methylation and gene expression. This may also be possible with sample-level assessment if BAL samples are purified into individual cell populations. To assess *HMOX2* at the protein level, commercially available antibodies exist and could be utilised to measure protein expression.²⁹³ By combining the information from experiments, the relationship between the differential methylation observed in the EWAS and functional outcomes could be assessed. For example, if lower methylation at cg26513131 has a functional effect, it would be associated with increased gene transcription and increased measurable protein. If altered *HMOX2* methylation was associated with a functional impact on *HMOX2* protein levels, further studies could assess the relationship with disease severity. In-vitro and animal models could also be used to assess the effect of blocking or enhancing the actions of *HMOX2* therapeutically.

The same experimental methods could be used to study the other genes identified in Chapter 7 such as *OSPB2*, *LCP1* and *NEDD9* and whether they may be viable therapeutic targets that should be explored further.

8.3.3 Multiomic approach to predictive biomarkers

The thesis has highlighted the potential of DNA methylation-based biomarkers of future CF lung disease severity, and this chapter has outlined the steps required to validate these biomarkers. However, as discussed, the biomarkers may not be successfully replicated or have poorer diagnostic performance in an external cohort. Thus, other potential biomarkers should be explored in parallel with the validation of those identified in this thesis.

One approach to identifying potential biomarkers is to consider various omic approaches. Broadly speaking, omics refers to the high-throughput study of a biological process. Multiple omic approaches (Figure 8.1) may have relevance to CF biomarkers.²⁹⁴ The subsequent discussion describes the various omic levels relevant to the ‘host’, however

there are environmental omics approaches (such as microbiomics and exposomics) which may be relevant to CF lung disease which are not discussed.

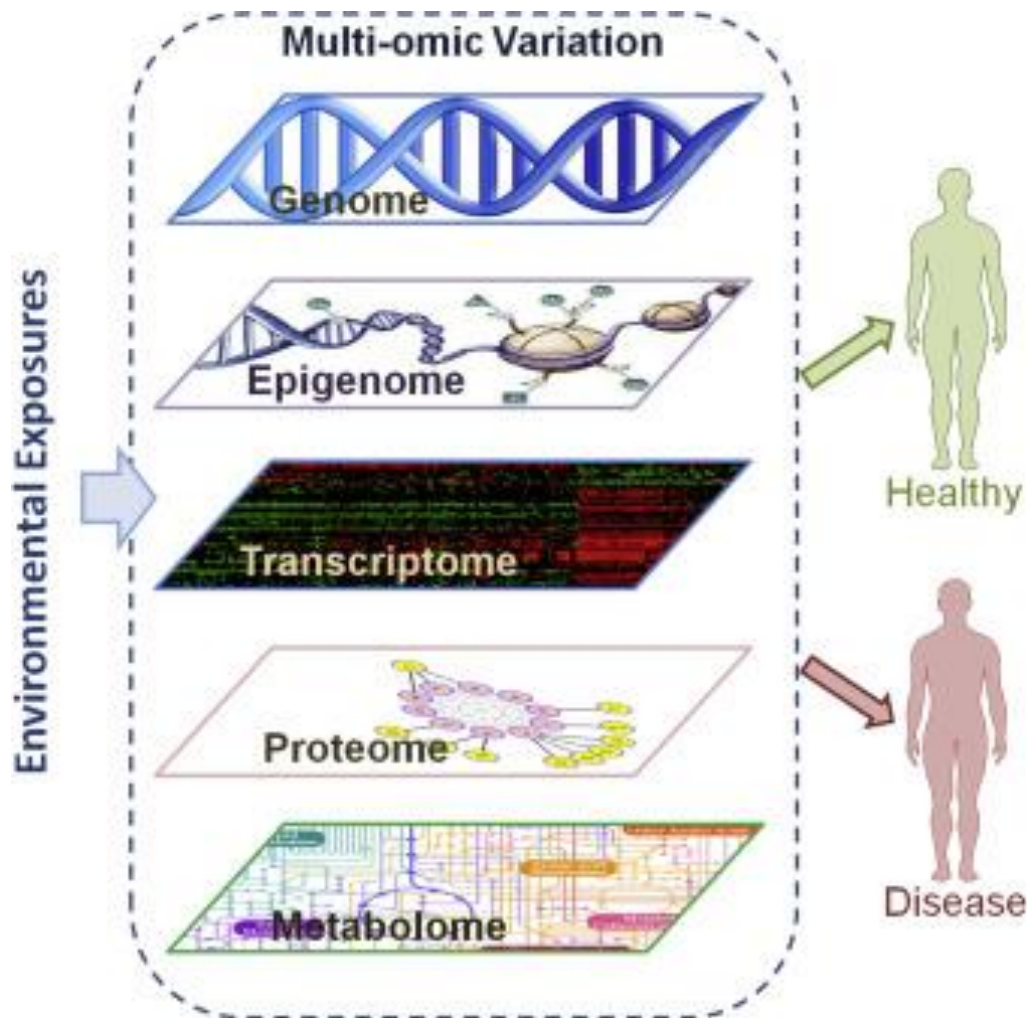


Figure 8.1. Different host omics approaches.²⁹⁵ Each omic layer may be relevant to predictive biomarkers of CF lung disease severity, and a multiomic risk score could combine several omic inputs to improve predictive performance.

8.3.3.1 Genomics

Genomics refers to the study of the DNA sequence of an organism.²⁹⁴ As discussed in Chapter 2, extensive work has attempted to identify non-*CFTR* modifiers genetic biomarkers of lung disease severity. There remains ongoing work in this area, with a recent publication identifying mutations in *NLRP3* and *NLRC4* which were associated with *pseudomonas aeruginosa* infection risk and lung function impairment in children with CF.²⁹⁶ The authors also provided *in vitro* evidence to support that mutation p.Q705K in *NLRP3* is associated by more severe disease by showing that macrophages affected by this mutation display a proinflammatory response characterised by increased secretion of IL-1 β . An approach that may be relevant to CF lung disease and has not yet been studied

is the polygenic risk score. Similar to the composite scores described in Chapter 7, a polygenic risk score involves combining a patient's genotypes at multiple genes to predict the risk of a defined outcome. There are well-established methods for generating polygenic risk scores.²⁷⁷ A major advantage of a polygenic risk score compared with attempting to identify an individual gene that can act as a biomarker is that it incorporates the weighted results of the analysis of multiple genes. Given the broad and complex pathophysiology of CF lung disease, it is likely that multiple genes interact to determine disease severity; hence, a polygenic risk score is more likely to be a valid predictive biomarker than an individual gene. Several studies have attempted to develop polygenic risk scores for the development of diseases such as type 1 diabetes mellitus. In a systematic review, Padilla-Martinez et al.²⁹⁷ identified four studies that have incorporated data from 30–67 different genes to develop a polygenic risk score for type 1 diabetes mellitus, with an AUC of 0.75–0.93. Data from previous genome-wide association studies of CF lung disease severity identified in Chapter 2 could potentially be used as a discovery cohort for identifying a polygenic risk score, and then a new cohort recruited for validation. One limitation with this approach is that FEV₁ has been used as the outcome measure in prior genome-wide association studies, and data have primarily been generated from older subjects. The mean age of subjects involved in genome-wide association studies of CF lung disease severity to date is 19.5 years.²⁹⁸ As a result, any polygenic risk scores developed using these data may not be applicable to early-life CF. Another approach that could focus on early life and use the most sensitive measures of lung disease severity would be to use cohorts such as AREST CF and SHIELD CF for both the discovery and validation of a polygenic risk score.

8.3.3.2 Epigenomics

Epigenomics refers to the study of the various epigenetic mechanisms,²⁹⁴ which were reviewed in Chapter 1. The focus of this thesis has been DNA methylation; however, a number of other epigenetic mechanisms may act as biomarkers of disease severity, including circulating histones, post-translational histone modifications and microRNAs.²⁹⁹ An example of the exploration of these mechanisms as predictive biomarkers is the work of Garcia-Gimenez et al.,²⁹⁹ who explored circulating histones and septic shock outcomes. Using mass spectrometry, the authors quantified circulating histones H3 and H2B in the plasma of subjects with sepsis and healthy controls. Those with sepsis were divided into survivors and non-survivors. The subjects with sepsis had higher levels of circulating H3 and H2B compared with controls, and non-survivors had

higher levels compared with survivors. When AUC analysis was performed, both H3 and H2B demonstrated the ability to predict survival v. non-survival. The AUC for H3 was 0.806 (95% CI [0.554, 1.058], $p = 0.034$), and for H2B it was 0.792 (95% CI [0.562, 1.021], $p = 0.043$). Post-translational histone modifications have mainly been studied as potential biomarkers in oncological conditions. For example, in breast cancer, the chromatin-binding profile of the oestrogen receptor α gene was analysed along with genome-wide H3K4me3 and H3K27me3 profiles.³⁰⁰ The outcome was treatment response to an aromatase inhibitor. The chromatin-binding profile of oestrogen receptor α and an H3K27me3 profile associated with treatment response were identified; however, H3K4me3 was not informative. Diagnostic performance was not assessed. The field of microRNA biomarkers has also concentrated on cancer, primarily on aiding specific cancer diagnosis. For instance, a commercially available product can analyse the profile of eight microRNAs in lung samples from subjects with lung cancer and classify the type of cancer present.³⁰¹ In a validation sample of 410 subjects, the microRNA profile accurately predicted cancer type in 384 (93.7%) of cases. Importantly, techniques have been developed that enable the analysis of epigenetic modifications on samples with a small number of cells, and workflows that enable the simultaneous evaluation of multiple mechanisms are well described.^{299 302} This includes analysis of samples that have been frozen at -80°C and formalin-fixed samples. This means that there are potentially multiple biobanks of stored samples from subjects with CF on which a broad epigenomic analysis could be performed.³⁰³

8.3.3.3 Transcriptomics

Transcriptomics refers to the study of RNA transcripts from DNA in a cell or tissue.²⁹⁴ There are multiple transcriptomic approaches, including the use of microarrays to profile specific RNA transcripts or RNA sequencing (RNA-seq), which is the untargeted profiling of transcripts. Single-cell RNA-seq (scRNA-seq) is a relatively novel method and is particularly powerful for determining the cell composition of a heterogeneous sample.²⁹² A prime example of this is where scRNA-seq was used to identify a novel cell type in respiratory epithelium known as a pulmonary ionocyte.^{304 305} Of relevance to CF, this cell type appears to be the main source of CFTR expression in respiratory epithelium. Data derived from scRNA-seq profiling of BAL samples are limited. A study of adult subjects with chronic obstructive pulmonary disease demonstrated 13 different alveolar macrophage subtypes, which is significantly more than previously thought.²¹⁶ The potential role for scRNA-seq profiles as predictive biomarkers has been explored in a

recent publication regarding subjects with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Liao et al.³⁰⁶ compared the scRNA-seq profile of BAL specimens collected from subjects with mild and severe SARS-CoV-2 infection and controls. They found that severe infection was characterised by an altered cellular profile, with an increased proportion of mononuclear phagocytes such as alveolar macrophages, monocyte-derived macrophages and recruited monocytes. In subjects with severe infection, in addition to an increased mononuclear phagocyte fraction, the relative proportions of the different cell types were also altered, with depletion of tissue-resident macrophages and a relative abundance of recruited monocytes, including subsets that were proinflammatory and profibrotic in nature. In addition to identifying an scRNA-seq profile associated with disease severity, the work also offered some mechanistic insights and highlighted potential therapeutic targets. The recruited monocytes, which were more abundant in subjects with severe disease, expressed *IL-6*, and IL-6 protein levels were elevated when measured in BAL. This highlighted IL-6 as a potential therapeutic target, and trials of anti-IL-6 treatments have been commenced.³⁰⁷ In adults with CF, scRNA-seq has demonstrated that sputum is characterised by an abundance of recruited monocytes³⁰⁸; however, no BAL analysis or comparison of CF disease severity has taken place. Future work using early-life BAL samples collected via the AREST CF cohort and utilising the cryopreservation and flow cytometry methods developed in this thesis will be able to assess whether scRNA-seq can predict CF lung disease severity.

8.3.3.4 Proteomics

The proteome is the complete set of proteins expressed by a cell, tissue or organism.²⁹⁴ In CF lung disease, there are obvious protein level targets for potential biomarkers of lung disease severity. Given the key role of inflammation, inflammatory cytokines are a potential biomarker. Multiple studies have assessed the cross-sectional correlation between BAL cytokines and lung disease severity.^{51 189 309} This includes analysis of IL-8, IL-1 β and IL-1 α , which are elevated in those with more severe disease. Limited work has assessed the predictive ability of inflammatory cytokines. In a seminal paper, the detection of free neutrophil elastase in early-life BAL was associated with the later development of bronchiectasis.⁸⁹ For example, the detection of neutrophil elastase in BAL samples at 3 months of age was associated with seven times higher odds of developing bronchiectasis at 1 year of age and four times higher odds of bronchiectasis at 3 years of age. Tests of diagnostic performance were not undertaken in this study.

There are also potential protein-level biomarkers of disease severity related to mucin composition, a marker of the altered ASL that some propose is the primary driver of CF lung disease pathophysiology. In a recent paper using samples from the AREST CF cohort, Esther et al.³¹⁰ demonstrated that children with CF have elevated total mucins and mucin flakes relative to healthy controls. There was an association between mucin concentration and inflammation but not infection. The deranged mucin profile appears to be an early feature of lung disease in CF because it is present even in children with normal CT scans. Cross-sectional or longitudinal associations between mucin composition and CF lung disease severity have not been explored.

Future work using the AREST CF cohort, in particular stored samples from the cohorts already phenotyped in this thesis, will be able to assess protein-level biomarkers of future lung disease severity.

8.3.3.5 Metabolomics

Metabolomics is the study of metabolites found within a biological sample.²⁹⁴ Several metabolomic assays have been conducted in CF subjects that may have value as predictive biomarkers. Esther et al.³¹¹ developed an assay that calculates sialic acid–urea ratio, demonstrating its ability to measure airway surface hydration. The gold standard measurement of airway surface hydration is the calculation of the percentage of the overall weight of ASL that comprises dry solid content (% solids). The sialic acid–urea metabolomics assay had a strong correlation with % solids ($r^2 = 0.92$, $p < 0.001$) when assessed in the ASL collected from human airway epithelium. The assay was then successfully validated in BAL samples from mice and children with CF. Given the proposed importance of abnormal airway surface hydration in the pathophysiology of CF lung disease, it is possible that this could be a mechanistic biomarker of disease severity. Metabolomic assays have also been used to measure inflammation and have been associated with disease severity in cross-sectional studies. In a study by Chandler et al.,²⁰⁵ an initial untargeted metabolomic analysis identified 22 unique metabolites in the BAL samples of children with CF. The association with structural lung disease severity, measured using the PRAGMA score, was assessed. Of the 22 metabolites, methionine sulfoxide, a product of methionine oxidation by myeloperoxidase-derived oxidants, was associated with disease severity ($p = 5.59 \times 10^{-5}$). Methionine sulfoxide is a product of neutrophilic inflammation. The authors further supported this finding by confirming methionine sulfoxide levels in BAL samples using a targeted assay and then measuring

myeloperoxidase as a protein-level validation of this finding. The targeted assay confirmed the prior associations identified using the untargeted approach (Spearman's $\rho = 0.582$, $p = 0.0029$), and methionine sulfoxide and myeloperoxidase levels were correlated (Spearman's $\rho = 0.803$, $p = 3.9 \times 10^{-6}$). Myeloperoxidase levels were also correlated with disease severity measured by PRAGMA (Spearman's $\rho = 0.706$, $p = 1.7 \times 10^{-4}$).

Similar to protein-level biomarkers, banked biosamples from the subjects in the two cohorts involved in this thesis may be used to investigate metabolomic biomarkers of disease severity.

8.3.3.6 Polyomic risk score

Multiple omic approaches may be used to help identify valid predictive biomarkers of CF lung disease severity. As highlighted by the performance of the composite scores in Chapter 7 and the polygenic risk score concept discussed earlier in this chapter, the performance of predictive biomarkers can often be improved by combining multiple individual biomarkers. Thus, given the complexity of CF lung disease, the best predictive biomarker may be a polyomic risk score that incorporates multiple layers of omic data. Polyomic risk scores have been calculated for diseases other than CF, and methods for their calculation have been described.³¹² In a study of adult subjects with rheumatoid arthritis that attempted to identify a diagnostic biomarker, data from 354 cases and 333 controls were used.³¹³ Subjects had their genotype determined by a microarray, and whole-genome methylation assessed using Illumina 450K Array (the predecessor to the EPIC array used in this thesis). The integration of genotype and methylation data was best able to predict case v. control status, with an AUC of 0.927 as opposed to 0.737 for genotype alone and 0.821 for methylation data alone. In a study of women with breast cancer, the effect of the integration of multiomic data with established clinical risk factors (such as age, race, tumour subtype and pathological stage) was assessed.³¹⁴ Multiomic analysis of tumour tissue was performed and included assessment at the genome (copy number variant), epigenome (genome-wide DNA methylation and microRNA profile) and transcriptome (bulk RNA-seq). The outcome of interest was survival 3 years post diagnosis. The study showed that both DNA methylation and bulk RNA-seq data outperformed clinical risk factors in predicting survival, and that the best-performing model integrated these three inputs with an AUC of 0.753.

In addition to improving the predictive ability of biomarkers, a polyomic approach also has benefits in improving our understanding of CF lung disease pathophysiology. For any biomarker, multiomic data will help to identify whether it is involved in disease mechanisms and thus be a target for new therapeutic approaches. For example, in Chapter 7, if DNA methylation had been simultaneously assessed with RNA-seq, the effect of the altered DNA methylation at the seven DMPs on gene transcription could have been assessed. If altered DNA methylation was also associated with changes at the RNA level, the investigation of the DMPs as therapeutic targets would have been supported. If no change in transcription was identified, then it is likely the DMPs are descriptive biomarkers, and further analysis should be limited to their predictive performance.

8.3.4 Relevance of predictive biomarkers in an era of highly effective modulator therapy

At the commencement of this thesis in March 2017, the only highly effective modulator therapy available and routinely used was ivacaftor. This treatment was suitable for approximately 10% of people with CF and, at that stage in Australia, was only used in patients 6 years of age and older. Since then, tremendous advances in highly effective modulator therapy have been made. With the publication of a series of trials that have demonstrated the safety of ivacaftor in younger ages, ivacaftor is now approved for use in patients 4 months of age and older.³¹⁵⁻³¹⁷ In addition, the advent of elexacaftor/tezacaftor/ivacaftor therapy, which is highly effective for patients with one or more copies of Phe508del, means that 90% of patients will now have an effective CFTR-directed therapy.^{318 319} This change will be truly transformational for patients with CF, and given that the vast majority of patients will have access to a treatment that corrects the primary defect, the potential utility of predictive biomarkers should be reconsidered.

Even with the advances that elexacaftor/tezacaftor/ivacaftor therapy offers, predictive biomarkers will still be relevant to CF care. As discussed, patients may have variable responses to modulator therapy, and the ability to predict therapeutic responses (known as *theratypes*) will be beneficial. As highlighted in Chapter 2, a genetic modifier of the response to ivacaftor therapy has been identified.^{320 321} If a polyomic assessment of treatment response to elexacaftor/tezacaftor/ivacaftor is undertaken, then a polyomic score that predicts the likely response to this treatment could be determined, potentially assisting clinicians with how they use other treatments. In particular, a priority for CF

clinicians and researchers in the coming years will be to identify existing treatments that should be continued alongside modulator therapy and those that may be safely stopped, and a biomarker of modulator therapeutic response could inform these decisions.³²² In addition, there are ongoing trials attempting to identify novel modulators.²⁸⁹ In the future, patients may be eligible for multiple modulators, and biomarkers that predict responses to different modulators may assist clinicians in deciding on the best modulator for each individual patient.

8.4 CONCLUSIONS

This thesis was successful in achieving one of its two stated aims, identifying an early-life DNA methylation profile that can predict lung disease severity at 5 years of age. Additional outputs of the thesis include the development of methods for cryopreservation, flow cytometry and DNA methylation reference epigenomes for BAL specimens. Collectively, the body of work described in this thesis has established a strong foundation for biomarker research in CF, an area that continues to be relevant even in an era of highly effective modulator therapy. Several directions will be pursued, including validation of the identified DMPs and a polyomic approach to identifying biomarkers. It is hoped that this work will identify robust and validated predictive biomarkers of future CF lung disease severity, improving the care delivered to people with CF around the world.

REFERENCES

1. Anderson DH. Cystic fibrosis of the pancreas and its relation to celiac disease. *Am J Dis Child*. 1938 Aug;56(2):344–99.
2. Cystic Fibrosis Australia. Cystic fibrosis in Australia 2014: 17th annual report from the Australian Cystic Fibrosis Data Registry. North Ryde (NSW): Cystic Fibrosis Australia; 2016.
3. Spoonhower KA, Davis PB. Epidemiology of cystic fibrosis. *Clin Chest Med* [Internet]. 2016;37(1):1–8. doi:10.1016/j.ccm.2015.10.002
4. Ong T, Marshall SG, Karczeski BA, Stern DL, Cheng E, Cutting GR. Cystic fibrosis and congenital absence of the vas deferens. In RA Pagon, MP Adam, HH Ardinger HH, et al., editors. *GeneReviews* [Internet]. Seattle (WA): University of Washington; 1993.
5. Zielenski J, Rozmahel R, Bozon D, et al. Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genom* [Internet]. 1991;10(1):214–28.
6. Vankeerberghen A, Cuppens H, Cassiman JJ. The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. *J Cyst Fibros* [Internet]. 2002;1(1):13–29.
7. O’Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet* [Internet]. 2009;373(9678):1891–904. doi:10.1016/s0140-6736(09)60327-5
8. Pittman JE, Ferkol TW. The evolution of cystic fibrosis care. *Chest* [Internet]. 2015;148(2):533–42. doi:10.1378/chest.14-1997
9. Shanthikumar S, Massie J. A review of treatments that improve cystic fibrosis transmembrane conductance regulator function. *Clin Med Insights Ther*. 2017 Jan;9:1–7.
10. Sawicki GS, Sellers DE, Robinson WM. High treatment burden in adults with cystic fibrosis: challenges to disease self-management. *J Cyst Fibros* [Internet]. 2009;8(2):91–6. doi:10.1016/j.jcf.2008.09.007
11. Cystic Fibrosis Foundation Patient Registry. 2015 annual data report. Bethesda (MD): Cystic Fibrosis Foundation; 2016.
12. Cystic Fibrosis Trust. UK cystic fibrosis registry: annual data report 2015. London (United Kingdom): Cystic Fibrosis Trust; 2016
13. Life expectancy at birth, total (years) [Internet]. Washington (DC): The World Bank Group; c2020 [cited 2020 Dec 22]. Available from: <http://data.worldbank.org/indicator/SP.DYN.LE00.IN>

14. Kirkby S, Hayes D, Jr. Pediatric lung transplantation: indications and outcomes. *J Thorac Dis* [Internet]. 2014;6(8):1024–31. doi:10.3978/j.issn.2072-1439.2014.04.27
15. Yusef RD, Christie JD, Edwards LB, et al. The registry of the International Society for Heart and Lung Transplantation: thirtieth adult lung and heart–lung transplant report—2013; focus theme: age. *J Heart Lung Transplant* [Internet]. 2013;32(10):965–78. doi:10.1016/j.healun.2013.08.007
16. Pillarisetti N, Williamson E, Linnane B, et al. Infection, inflammation, and lung function decline in infants with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2011;184(1):75–81. doi:10.1164/rccm.201011-1892OC
17. Regnis JA, Robinson M, Bailey DL, et al. Mucociliary clearance in patients with cystic fibrosis and in normal subjects. *Am J Respir Crit Care Med* [Internet]. 1994;150(1):66–71. doi:10.1164/ajrccm.150.1.8025774
18. Boucher RC. Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. *Ann Rev Medicine* [Internet]. 2007;58:157–70. doi:10.1146/annurev.med.58.071905.105316
19. Matsui H, Grubb BR, Tarran R, et al. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* [Internet]. 1998;95(7):1005–15. doi: 10.1016/s0092-8674(00)81724-9.
20. Hoegger MJ, Fischer AJ, McMenimen JD, et al. Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Sci* [Internet]. 2014;345(6198):818–22. doi:10.1126/science.1255825
21. Quinton PM. Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis. *Lancet* [Internet]. 2008;372(9636):415–7. doi:10.1016/s0140-6736(08)61162-9
22. Zabner J, Smith JJ, Karp PH, Widdicombe JH, Welsh MJ. Loss of CFTR chloride channels alters salt absorption by cystic fibrosis airway epithelia in vitro. *Mol Cell* [Internet]. 1998;2(3):397–403. doi: 10.1016/s1097-2765(00)80284-1
23. Smith JJ, Travis SM, Greenberg EP, Welsh MJ. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* [Internet]. 1996;85(2):229–36. doi: 10.1016/s0092-8674(00)81099-5.
24. Pezzulo AA, Tang XX, Hoegger MJ, et al. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nat* [Internet]. 2012;487(7405):109–13. doi:10.1038/nature11130
25. Schultz A, Puvvadi R, Borisov SM, et al. Airway surface liquid pH is not acidic in children with cystic fibrosis. *Nat Commun* [Internet]. 2017;8(1):1409. doi:10.1038/s41467-017-00532-5

26. Bonfield TL, Konstan MW, Berger M. Altered respiratory epithelial cell cytokine production in cystic fibrosis. *J Allergy Clin Immunol* [Internet]. 1999;104(1):72–8.
27. Tirouvanziam R, de Bentzmann S, Hubeau C, et al. Inflammation and infection in naive human cystic fibrosis airway grafts. *Am J Respir Cell Mol Biol* [Internet]. 2000;23(2):121–7. doi:10.1165/ajrcmb.23.2.4214
28. Khan TZ, Wagener JS, Bost T, Martinez J, Accurso FJ, Riches DW. Early pulmonary inflammation in infants with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 1995;151(4):1075–82. doi:10.1164/ajrccm.151.4.7697234
29. Meyerholz DK, Stoltz DA, Namati E, et al. Loss of cystic fibrosis transmembrane conductance regulator function produces abnormalities in tracheal development in neonatal pigs and young children. *Am J Respir Crit Care Med* [Internet]. 2010;182(10):1251–61. doi:10.1164/rccm.201004-0643OC
30. Stoltz DA, Meyerholz DK, Pezzulo AA, et al. Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth. *Sci Transl Med* [Internet]. 2010;2(29):29–31. doi:10.1126/scitranslmed.3000928
31. Fischer AJ, Singh SB, Adam RJ, et al. Tracheomalacia is associated with lower FEV1 and *Pseudomonas* acquisition in children with CF. *Pediatr Pulmonol* [Internet]. 2014;49(10):960–70. doi:10.1002/ppul.22922
32. Waters V, Ratjen F. Pulmonary exacerbations in children with cystic fibrosis. *Ann Am Thorac Soc* [Internet]. 2015;12(Suppl 2):S200-6. doi:10.1513/AnnalsATS.201502-098AW
33. Flume P, Vandevanter D. Pulmonary exacerbations. In A Bush, D Bilton, M Hodson, editors. *Hodson and Geddes' cystic fibrosis*. 4th ed. Boca Raton (FL): CRC Press; 2016.p. 221-236
34. Quittner AL, Buu A, Messer MA, Modi MA, Watrous AC. Development and validation of the Cystic Fibrosis Questionnaire in the United States: a health-related quality-of-life measure for cystic fibrosis. *Chest* [Internet]. 2005;128(4):2347–54. doi:10.1378/chest.128.4.2347
35. Gee L, Abbott J, Conway SP, Etherington S, Webb A. Development of a disease specific health related quality of life measure for adults and adolescents with cystic fibrosis. *Thorax* [Internet]. 2000;55(11):946–54.
36. Habib AR, Manji J, Wilcox PG, Javer AR, Buxton JA, Quon BS. A systematic review of factors associated with health-related quality of life in adolescents and adults with cystic fibrosis. *Ann Am Thorac Soc* [Internet]. 2015;12(3):420–8. doi:10.1513/AnnalsATS.201408-393OC

37. Bedrossian CW, Greenberg SD, Singer DB, Hansen JJ, Rosenberg HS. The lung in cystic fibrosis: a quantitative study including prevalence of pathologic findings among different age groups. *Hum Pathol* [Internet]. 1976;7(2):195–204.
38. Kang EY, Miller RR, Muller NL. Bronchiectasis: comparison of preoperative thin-section CT and pathologic findings in resected specimens. *Radiol* [Internet]. 1995;195(3):649–54. doi:10.1148/radiology.195.3.7753989
39. de Jong PA, Tiddens HA. Cystic fibrosis specific computed tomography scoring. *Proc Am Thorac Soc* [Internet]. 2007;4(4):338–42. doi:10.1513/pats.200611-175HT
40. Loeve M, Hop WC, de Bruijne M, et al. Chest computed tomography scores are predictive of survival in patients with cystic fibrosis awaiting lung transplantation. *Am J Respir Crit Care Med* [Internet]. 2012;185(10):1096–103. doi:10.1164/rccm.201111-2065OC
41. Brody AS, Sucharew H, Campbell JD, et al. Computed tomography correlates with pulmonary exacerbations in children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2005;172(9):1128–32. doi:10.1164/rccm.200407-989OC
42. Tepper LA, Utens EM, Caudri D, et al. Impact of bronchiectasis and trapped air on quality of life and exacerbations in cystic fibrosis. *Eur Respir J* [Internet]. 2013;42(2):371–9. doi:10.1183/09031936.00137612
43. Loeve M, Gerbrands K, Hop WC, Rosenfeld M, Hartmann IC, Tiddens HA. Bronchiectasis and pulmonary exacerbations in children and young adults with cystic fibrosis. *Chest* [Internet]. 2011;140(1):178–85. doi:10.1378/chest.10-1152
44. Bortoluzzi CF, Volpi S, D’Orazio C, et al. Bronchiectases at early chest computed tomography in children with cystic fibrosis are associated with increased risk of subsequent pulmonary exacerbations and chronic pseudomonas infection. *J Cyst Fibros* [Internet]. 2014;13(5):564–71. doi:10.1016/j.jcf.2014.03.006
45. Mott LS, Park J, Murray CP, et al. Progression of early structural lung disease in young children with cystic fibrosis assessed using CT. *Thorax* [Internet]. 2012;67(6):509–16. doi:10.1136/thoraxjnl-2011-200912
46. de Jong PA, Nakano Y, Hop WC, et al. Changes in airway dimensions on computed tomography scans of children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2005;172(2):218–24. doi:10.1164/rccm.200410-1311OC
47. Nasr SZ, Kuhns LR, Brown RW, Hurwitz ME, Sanders GM, Strouse PJ. Use of computerized tomography and chest x-rays in evaluating efficacy of aerosolized recombinant human DNase in cystic fibrosis patients younger than age 5 years: a preliminary study. *Pediatr Pulmonol* [Internet]. 2001;31(5):377–82.

48. Hisert KB, Heltshe SL, Pope C, et al. Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections. *Am J Respir Crit Care Med* [Internet]. 2017;195(12):1617–28.
49. Ronan NJ, Einarsson GG, Twomey M, et al. CORK study in cystic fibrosis: sustained improvements in ultra-low-dose chest CT scores after CFTR modulation with ivacaftor. *Chest* 2018 Feb;153(2):395–403.
50. Davis SD, Fordham LA, Brody AS, et al. Computed tomography reflects lower airway inflammation and tracks changes in early cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2007;175(9):943–50. doi:10.1164/rccm.200603-343OC
51. Wainwright CE, Vidmar S, Armstrong DS, et al. Effect of bronchoalveolar lavage-directed therapy on *Pseudomonas aeruginosa* infection and structural lung injury in children with cystic fibrosis: a randomized trial. *JAMA* [Internet]. 2011;306(2):163–71. doi:10.1001/jama.2011.954
52. Stick S, Tiddens H, Aurora P, et al. Early intervention studies in infants and preschool children with cystic fibrosis: are we ready? *Eur Respir J* [Internet]. 2013;42(2):527–38. doi:10.1183/09031936.00108212
53. Ramsey KA, Ranganathan S. Interpretation of lung function in infants and young children with cystic fibrosis. *Respir* [Internet]. 2014;19(6):792–9. doi:10.1111/resp.12329
54. Kerem E, Reisman J, Corey M, Canny GJ, Levison H. Prediction of mortality in patients with cystic fibrosis. *NEJM* [Internet]. 1992;326(18):1187–91. doi:10.1056/nejm199204303261804
55. Miller MR, Crapo R, Hankinson J, et al. General considerations for lung function testing. *Eur Respir J* [Internet]. 2005;26(1):153–61. doi:10.1183/09031936.05.00034505
56. Aurora P, Bush A, Gustafsson P, et al. Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2005;171(3):249–56. doi:10.1164/rccm.200407-895OC
57. Gustafsson PM, De Jong PA, Tiddens HA, Lindblad A. Multiple-breath inert gas washout and spirometry versus structural lung disease in cystic fibrosis. *Thorax* [Internet]. 2008;63(2):129–34. doi:10.1136/thx.2007.077784
58. Horsley AR, Gustafsson PM, Macleod KA, et al. Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax* [Internet]. 2008;63(2):135–40. doi:10.1136/thx.2007.082628

59. Marostica PJ, Weist AD, Eigen H, et al. Spirometry in 3- to 6-year-old children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2002;166(1):67–71. doi:10.1164/rccm.200111-056OC
60. Ramsey KA, Ranganathan S, Park J, et al. Early respiratory infection is associated with reduced spirometry in children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2014;190(10):1111–6. doi:10.1164/rccm.201407-1277OC
61. Harrison AN, Regelman WE, Zirbes JM, Milla CE. Longitudinal assessment of lung function from infancy to childhood in patients with cystic fibrosis. *Pediatr Pulmonol* [Internet]. 2009;44(4):330–9. doi:10.1002/ppul.20994
62. Belessis Y, Dixon B, Hawkins G, et al. Early cystic fibrosis lung disease detected by bronchoalveolar lavage and lung clearance index. *Am J Respir Crit Care Med* [Internet]. 2012;185(8):862–73. doi:10.1164/rccm.201109-1631OC
63. Owens CM, Aurora P, Stanojevic S, et al. Lung Clearance Index and HRCT are complementary markers of lung abnormalities in young children with CF. *Thorax* [Internet]. 2011;66(6):481–8. doi:10.1136/thx.2010.150375
64. Vermeulen F, Proesmans M, Boon M, Havermans T, De Boeck K. Lung clearance index predicts pulmonary exacerbations in young patients with cystic fibrosis. *Thorax* [Internet]. 2014;69(1):39–45. doi:10.1136/thoraxjnl-2013-203807
65. Milla CE, Ratjen F, Marigowda G, et al. Lumacaftor/ivacaftor in patients aged 6–11 years with cystic fibrosis and homozygous for F508del-CFTR. *American journal of respiratory and critical care medicine* 2017;195(7):912-20.
66. Ratjen F, Hug C, Marigowda G, et al. Efficacy and safety of lumacaftor and ivacaftor in patients aged 6–11 years with cystic fibrosis homozygous for F508del-CFTR: a randomised, placebo-controlled phase 3 trial. *The Lancet Respiratory medicine* 2017;5(7):557-67.
67. Shaw M, Khan U, Clancy JP, et al. Changes in LCI in F508del/F508del patients treated with lumacaftor/ivacaftor: Results from the prospect study. *Journal of Cystic Fibrosis* 2020
68. Amin R, Subbarao P, Jabar A, et al. Hypertonic saline improves the LCI in paediatric patients with CF with normal lung function. *Thorax* [Internet]. 2010;65(5):379–83. doi:10.1136/thx.2009.125831
69. Robinson PD, Cooper P, Van Asperen P, Fitzgerald D, Selvadurai H. Using index of ventilation to assess response to treatment for acute pulmonary exacerbation in children with cystic fibrosis. *Pediatr Pulmonol* [Internet]. 2009;44(8):733–42. doi:10.1002/ppul.20956

70. Fuchs SI, Gappa M, Eder J, Unsinn KM, Steinkamp G, Ellemunter H. Tracking Lung Clearance Index and chest CT in mild cystic fibrosis lung disease over a period of three years. *Respir Med* [Internet]. 2014;108(6):865–74. doi:10.1016/j.rmed.2014.03.011
71. Stanojevic S, Davis SD, Retsch-Bogart G, et al. Progression of lung disease in preschool patients with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2017;195(9):1216–25. doi:10.1164/rccm.201610-2158OC
72. Ranganathan SC, Hall GL, Sly PD, Stick SM, Douglas TA. Early lung disease in infants and pre-school children with cystic fibrosis: what have we learnt and what should we do about it? *Am J Respir Crit Care Med* [Internet]. 2017;195(12):1567–1575. doi:10.1164/rccm.201606-1107CI
73. Simpson SJ, Mott LS, Esther CR, Jr., Stick SM, Hall GM. Novel end points for clinical trials in young children with cystic fibrosis. *Expert Rev Respir Med* [Internet]. 2013;7(3):231–43. doi:10.1586/ers.13.25
74. Long FR, Williams RS, Castile RG. Structural airway abnormalities in infants and young children with cystic fibrosis. *J Pediatr* [Internet]. 2004;144(2):154–61. doi:10.1016/j.jpeds.2003.09.026
75. Sly PD, Brennan S, Gangell C, et al. Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *Am J Respir Crit Care Med* [Internet]. 2009;180(2):146–52. doi:10.1164/rccm.200901-0069OC
76. Ranganathan SC, Dezateux C, Bush A, et al. Airway function in infants newly diagnosed with cystic fibrosis. *Lancet* [Internet]. 2001;358(9297):1964–5.
77. Mott LS, Graniel KG, Park J, et al. Assessment of early bronchiectasis in young children with cystic fibrosis is dependent on lung volume. *Chest* [Internet]. 2013;144(4):1193–8. doi:10.1378/chest.12-2589
78. de Jong PA, Mayo JR, Golmohammadi K, et al. Estimation of cancer mortality associated with repetitive computed tomography scanning. *Am J Respir Crit Care Med* [Internet]. 2006;173(2):199–203. doi:10.1164/rccm.200505-810OC
79. Ramsey KA, Rosenow T, Turkovic L, et al. Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2016;193(1):60–7. doi:10.1164/rccm.201507-1409OC
80. Hall GL, Logie KM, Parsons F, et al. Air trapping on chest CT is associated with worse ventilation distribution in infants with cystic fibrosis diagnosed following newborn screening. *PloS One* [Internet]. 2011;6(8):e23932. doi:10.1371/journal.pone.0023932

81. Ellemunter H, Fuchs SI, Unsinn KM, et al. Sensitivity of Lung Clearance Index and chest computed tomography in early CF lung disease. *Respir Med* [Internet]. 2010;104(12):1834–42. doi:10.1016/j.rmed.2010.06.010
82. Burrows B, Knudson RJ, Lebowitz MD. The relationship of childhood respiratory illness to adult obstructive airway disease. *Am Rev Respir Dis* [Internet]. 1977;115(5):751–60. doi:10.1164/arrd.1977.115.5.751
83. Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *Lancet Respir Med* [Internet]. 2013;1(9):728–42. doi:10.1016/s2213-2600(13)70118-8
84. Chow CW, Landau LI, Taussig LM. Bronchial mucous glands in the newborn with cystic fibrosis. *Eur J Pediatr* [Internet]. 1982;139(4):240–3.
85. Fakhoury K, Durie PR, Levison H, Canny GJ. Meconium ileus in the absence of cystic fibrosis. *Arch Dis Child* [Internet]. 1992;67(10 Spec No):1204–6.
86. Linnane BM, Hall GL, Nolan G, et al. Lung function in infants with cystic fibrosis diagnosed by newborn screening. *Am J Respir Crit Care Med* [Internet]. 2008;178(12):1238–44. doi:10.1164/rccm.200804-551OC
87. Hoo AF, Thia LP, Nguyen TT, et al. Lung function is abnormal in 3-month-old infants with cystic fibrosis diagnosed by newborn screening. *Thorax* [Internet]. 2012;67(10):874–81. doi:10.1136/thoraxjnl-2012-201747
88. Sly PD, Gangell CL, Chen L, et al. Risk factors for bronchiectasis in children with cystic fibrosis. *NEJM* [Internet]. 2013;368(21):1963–70. doi:10.1056/NEJMoa1301725
89. Ranganathan SC, Stocks J, Dezateux C, et al. The evolution of airway function in early childhood following clinical diagnosis of cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2004;169(8):928–33. doi:10.1164/rccm.200309-1344OC
90. Nguyen TT, Thia LP, Hoo AF, et al. Evolution of lung function during the first year of life in newborn screened cystic fibrosis infants. *Thorax* [Internet]. 2014;69(10):910–7. doi:10.1136/thoraxjnl-2013-204023
91. Davies G, Thia LP, Stocks J, et al. Minimal change in structural, functional and inflammatory markers of lung disease in newborn screened infants with cystic fibrosis at one year. *J Cyst Fibros* [Internet]. 2020;19(6):896–901. doi:10.1016/j.jcf.2020.01.006
92. Long FR, Williams RS, Adler BH, Castile RG. Comparison of quiet breathing and controlled ventilation in the high-resolution CT assessment of airway disease in infants with cystic fibrosis. *Pediatr Radiol* [Internet]. 2005;35(11):1075–80. doi:10.1007/s00247-005-1541-4

93. Mott LS, Park J, Gangell CL, et al. Distribution of early structural lung changes due to cystic fibrosis detected with chest computed tomography. *The Journal of pediatrics* [Internet]. 2013;163(1):243-8.e1-3. doi: 10.1016/j.jpeds.2012.12.042
94. Stiglbauer R, Schurawitzki H, Eichler I, Götz M. High resolution CT in children with cystic fibrosis. *Acta Radiol* [Internet]. 1992;33(6):548–53.
95. Demirkazik FB, Ariyurek OM, Ozcelik U, Göçmen A, Hassanabad HK, Kiper N. High resolution CT in children with cystic fibrosis: correlation with pulmonary functions and radiographic scores. *Eur J Radiol* [Internet]. 2001;37(1):54–9.
96. Brody AS, Klein JS, Molina PL, Quan J, Bean JA, Wilmott RW. High-resolution computed tomography in young patients with cystic fibrosis: distribution of abnormalities and correlation with pulmonary function tests. *J Pediatr* [Internet]. 2004;145(1):32–8. doi:10.1016/j.jpeds.2004.02
97. Marchant JM, Masel JP, Dickinson FL, Masters IB, Chang AB. Application of chest high-resolution computer tomography in young children with cystic fibrosis. *Pediatr Pulmonol* [Internet]. 2001;31(1):24–9.
98. Helbich TH, Heinz-Peer G, Eichler I, et al. Cystic fibrosis: CT assessment of lung involvement in children and adults. *Radiol* [Internet]. 1999;213(2):537–44. doi:10.1148/radiology.213.2.r99nv04537
99. Maffessanti M, Candusso M, Brizzi F, Piovesana F. Cystic fibrosis in children: HRCT findings and distribution of disease. *J Thorac Imaging* [Internet]. 1996;11(1):27–38.
100. Drumm ML, Konstan MW, Schluchter MD, et al. Genetic modifiers of lung disease in cystic fibrosis. *NEJM* [Internet]. 2005;353(14):1443–53. doi:10.1056/NEJMoa051469
101. Whiting P, Al M, Burgers L, et al. Ivacaftor for the treatment of patients with cystic fibrosis and the G551D mutation: a systematic review and cost-effectiveness analysis. *Health Technol Assess* [Internet]. 2014;18(18):1–106. doi:10.3310/hta18180
102. Rowbotham NJ, Smith S, Leighton PA, et al. The top 10 research priorities in cystic fibrosis developed by a partnership between people with CF and healthcare providers. *Thorax* [Internet]. 2018;73(4):388–90. doi:10.1136/thoraxjnl-2017-210473
103. Van Goor F, Hadida S, Grootenhuis PD, et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc Natl Acad Sci U S A* [Internet]. 2009;106(44):18825–30. doi:10.1073/pnas.0904709106
104. Rowe SM, Miller S, Sorscher EJ. Cystic fibrosis. *NEJM* [Internet]. 2005;352(19):1992–2001. doi:10.1056/NEJMra043184

105. Kerem E, Corey M, Kerem BS, et al. The relation between genotype and phenotype in cystic fibrosis—analysis of the most common mutation (delta F508). *NEJM* [Internet]. 1990;323(22):1517–22. doi:10.1056/nejm199011293232203
106. Johansen HK, Nir M, Hoiby N, Koch C, Schwartz M. Severity of cystic fibrosis in patients homozygous and heterozygous for delta F508 mutation. *Lancet* [Internet]. 1991;337(8742):631–4.
107. Santis G, Osborne L, Knight R, Hodson ME, Ramsay M. Genetic influences on pulmonary severity in cystic fibrosis. *Lancet* [Internet]. 1990;335(8684):294.
108. Cystic Fibrosis Genotype–Phenotype Consortium. Correlation between genotype and phenotype in patients with cystic fibrosis. *NEJM* [Internet]. 1993;329(18):1308–13. doi:10.1056/nejm199310283291804
109. Koch C, Cuppens H, Rainisio M, et al. European Epidemiologic Registry of Cystic Fibrosis (ERCF): comparison of major disease manifestations between patients with different classes of mutations. *Pediatr Pulmonol* [Internet]. 2001;31(1):1–12.
110. Santis G, Osborne L, Knight RA, Hodson ME. Independent genetic determinants of pancreatic and pulmonary status in cystic fibrosis. *Lancet* [Internet]. 1990;336(8723):1081–4.
111. Mekus F, Ballmann M, Bronsveld I, Bijman J, Veeze H, Tümmler B. Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin Res* [Internet]. 2000;3(4):277–93.
112. Vanscoy LL, Blackman SM, Collaco JM, et al. Heritability of lung disease severity in cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2007;175(10):1036–43. doi:10.1164/rccm.200608-1164OC
113. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet* [Internet]. 2008;9(4):255–66. doi:10.1038/nrg2322
114. Collaco JM, Blackman SM, McGready J, Naughton KM, Cutting GR. Quantification of the relative contribution of environmental and genetic factors to variation in cystic fibrosis lung function. *J Pediatr* [Internet]. 2010;157(5):802–7.e3. doi:10.1016/j.jpeds.2010.05.018
115. Waddington C. The epigenotype. *Endeavour*. 1942;18–20.
116. Deans C, Maggert KA. What do you mean, “epigenetic”? *Genet*. 2015 Apr 1;199(4):887–96.
117. Haig D. Commentary: the epidemiology of epigenetics. *Int J Epidemiol*. 2012 Feb;41(1):13–16.
118. Nanney DL. Epigenetic control systems. *Proc Natl Acad Sci*. 1958 Jul;44(7):712–17.

119. Holliday R. Epigenetics: an overview. *Dev Genet.* 1994;15(6):453–57.
120. Russo VE, Martienssen RA, Riggs AD. Epigenetic mechanisms of gene regulation. Plainview (NY): Cold Spring Harbor Laboratory Press; 1996.
121. Bird A. Perceptions of epigenetics. *Nat.* 2007 May;447(7143):396–398.
122. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. *NEJM.* 2018 Apr;378(14):1323–34.
123. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nat.* 1997 Sep;389(6648):251–60. doi:10.1038/38444
124. Annunziato A. DNA packaging: nucleosomes and chromatin. *Nat Educ.* 2008;1(1):26.
125. Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet.* 2006 Apr;15(1 Suppl):R17–29. doi:10.1093/hmg/ddl046
126. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* [Internet]. 2009;19(1):92–105. doi:10.1101/gr.082701.108
127. Piatek MJ, Werner A. Endogenous siRNAs: regulators of internal affairs. *Biochem Soc Trans.* 2014 Jan;42(4):1174–9. doi:10.1042/BST20140068
128. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol.* 2007;8(3):209–20. doi: 10.1038/nrm2124
129. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* [Internet]. 2011;12(4):246–58. doi:10.1038/nrm3089
130. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol.* 2018 Jan;19(1):45–58. doi:10.1038/nrm.2017.77
131. Guo M, Schimmel P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol.* 2013 Mar;9(3):145–53. doi:10.1038/nchembio.1158
132. Cech TR, Steitz JA. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell.* 2014 Mar;157(1):77–94. doi:10.1016/j.cell.2014.03.008
133. Horn PJ, Peterson CL. Molecular biology: chromatin higher order folding—wrapping up transcription. *Sci* [Internet]. 2002;297(5588):1824–7. doi:10.1126/science.1074200
134. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Sci* [Internet]. 1975;187(4173):226–32.

135. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* [Internet]. 2007;39(4):457–66. doi:10.1038/ng1990
136. Guo H, Zhu P, Yan L, et al. The DNA methylation landscape of human early embryos. *Nat*. 2014 Jul;511(7511):606–10. doi:10.1038/nature13544
137. Sundrani D, Khot V, Joshi S. DNA Methylation for Prediction of Adverse Pregnancy Outcomes. In: Garcia-Gimenez J, ed. *Epigenetic Biomarkers and Diagnostics*. London: Elsevier 2016.
138. Dhe-Paganon S, Syeda F, Park L. DNA methyl transferase 1: regulatory mechanisms and implications in health and disease. *Int J Biochem Mol Biol* [Internet]. 2011;2(1):58–66.
139. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 2013;14(3):204–20. doi:10.1038/nrg3354
140. Ollikainen M, Smith KR, Joo EJ, et al. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum Mol Genet*. 2010 Nov;19(21):4176–88. doi:10.1093/hmg/ddq336
141. Gordon L, Joo JE, Powell JE, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res*. 2012;22(8):1395–406. doi:10.1101/gr.136598.111
142. Teh AL, Pan H, Chen L, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res*. 2014 Jul;24(7):1064–74. doi:10.1101/gr.171439.113
143. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* [Internet]. 2013;49(2):359–67. doi:10.1016/j.molcel.2012.10.016
144. Gervin K, Hammero M, Akselsen HE, et al. Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome Res* [Internet]. 2011;21(11):1813–21. doi:10.1101/gr.119685.110
145. Bierne H, Hamon M, Cossart P. Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med*. 2012 Dec;2(12):a010272. doi:10.1101/cshperspect.a010272
146. Cole J, Morris P, Dickman MJ, Dockrell DH. The therapeutic potential of epigenetic manipulation during infectious diseases. *Pharmacol Ther* [Internet]. 2016;167:85–99. doi:10.1016/j.pharmthera.2016.07.013

147. Kopp BT, Ortega-García JA, Sadreameli SC, et al. The impact of secondhand smoke exposure on children with cystic fibrosis: a review. *Int J Environ Res Public Health* [Internet]. 2016;13(10):1003. doi:10.3390/ijerph13101003
148. Guerrero-Preston R, Goldman LR, Brebi-Mieville P, et al. Global DNA hypomethylation is associated with in utero exposure to cotinine and perfluorinated alkyl compounds. *Epigenet*. 2010 Aug;5(6):539–46. doi:10.4161/epi.5.6.12378
149. Ivorra C, Fraga MF, Bayon GF, et al. DNA methylation patterns in newborns exposed to tobacco in utero. *J Transl Med*. 2015 Jan;13:25. doi:10.1186/s12967-015-0384-5
150. Joubert BR, Haberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* [Internet]. 2012;120(10):1425–31. doi:10.1289/ehp.1205412
151. Rauschert S, Melton PE, Burdge G, et al. Maternal smoking during pregnancy induces persistent epigenetic changes into adolescence, independent of postnatal smoke exposure and is associated with cardiometabolic risk. *Front Genet* [Internet]. 2019;10:770. doi:10.3389/fgene.2019.00770
152. Joubert BR, Håberg SE, Bell DA, et al. Maternal smoking and DNA methylation in newborns: in utero effect or epigenetic inheritance? *Cancer Epidemiol Biomarkers Prev*. 2014 Jun;23(6):1007–17. doi:10.1158/1055-9965.epi-13-1256
153. Breton CV, Byun HM, Wenten M, Pan F, Yang A, Gilliland FD. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med*. 2009 Sep;180(5):462–7. doi:10.1164/rccm.200901-0135OC
154. Wiklund P, Karhunen V, Richmond RC, et al. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clin Epigenet*. 2019 Dec;11(1):97. doi:10.1186/s13148-019-0683-4
155. Corey M, McLaughlin FJ, Williams M, Levison H. A comparison of survival, growth, and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *J Clin Epidemiol* [Internet]. 1988;41(6):583–91.
156. Voisin S, Almen MS, Moschonis G, Chrousos GP, Manios Y, Schiøth HB. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. *Eur J Hum Genet*. 2015 May;23(5):654–62. doi:10.1038/ejhg.2014.139
157. Figueroa V, Milla C, Parks EJ, Schwarzenberg SJ, Moran A. Abnormal lipid concentrations in cystic fibrosis. *Am J Clin Nutr* [Internet]. 2002;75(6):1005–11. doi:10.1093/ajcn/75.6.1005

158. Woestenenk JW, Schulkes DA, Schipper HS, van der Ent CK, Houwen RJH. Dietary intake and lipid profile in children and adolescents with cystic fibrosis. *J Cyst Fibros*. 2017 May;16(3):410–17. doi:10.1016/j.jcf.2017.02.010
159. Vaughan WJ, Lindgren FT, Whalen JB, Abraham S. Serum lipoprotein concentrations in cystic fibrosis. *Sci* [Internet]. 1978;199(4330):783–6.
160. Wijnands KP, Obermann-Borst SA, Steegers-Theunissen RP. Early life lipid profile and metabolic programming in very young children. *Nutri Metab Cardiovasc Dis* [Internet]. 2015;25(6):608–14. doi:10.1016/j.numecd.2015.02.010
161. de la Rocha C, Perez-Mojica JE, Leon SZ, et al. Associations between whole peripheral blood fatty acids and DNA methylation in humans. *Sci Rep*. 2016 May;6(1):25867. doi:10.1038/srep25867
162. Cassidy L, Morton A, Etherington C, Peckham D. 205 folate deficiency in adult patients at a regional cystic fibrosis (CF) centre. *J Cyst Fibros*. 2014 Jun;13:S99.
163. Chakrabarty B, Kabra SK, Gulati S, et al. Peripheral neuropathy in cystic fibrosis: a prevalence study. *J Cyst Fibros*. 2013 Dec;12(6):754–60. doi:10.1016/j.jcf.2013.01.005
164. Maqbool A, Schall JJ, Mascarenhas MR, Dougherty KA, Stallings VA. Vitamin B(12) status in children with cystic fibrosis and pancreatic insufficiency. *J Pediatr Gastroenterol Nutr*. 2014 Jun;58(6):733–8. doi:10.1097/MPG.0000000000000313
165. Jacob RA, Gretz DM, Taylor PC, et al. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J Nutr* [Internet]. 1998;128(7):1204–12. doi:10.1093/jn/128.7.1204
166. Rampersaud GC, Kauwell GP, Hutson AD, Cerda JJ, Bailey LB. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *Am J Clin Nutr* [Internet]. 2000;72(4):998–1003. doi:10.1093/ajcn/72.4.998
167. Haggarty P, Hoad G, Campbell DM, Horgan GW, Piyathilake C, McNeill G. Folate in pregnancy and imprinted gene and repeat element methylation in the offspring. *Am J Clin Nutr* [Internet]. 2013;97(1):94–9. doi:10.3945/ajcn.112.042572
168. Fryer AA, Nafee TM, Ismail KM, Carroll WD, Emes RD, Farrell WE. LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. *Epigenet* [Internet]. 2009;4(6):394–8.
169. Perng W, Rozek LS, Mora-Plazas M, et al. Micronutrient status and global DNA methylation in school-age children. *Epigenet* [Internet]. 2012;7(10):1133–41. doi:10.4161/epi.21915
170. Quittner AL, Goldbeck L, Abbott J, et al. Prevalence of depression and anxiety in patients with cystic fibrosis and parent caregivers: results of The International Depression

- Epidemiological Study across nine countries. *Thorax*. 2014 Dec;69(12):1090–7. doi:10.1136/thoraxjnl-2014-205983
171. Smith BA, Modi AC, Quittner AL, Wood BL. Depressive symptoms in children with cystic fibrosis and parents and its effects on adherence to airway clearance. *Pediatr Pulmonol*. 2010 Jul;45(8):756–63. doi:10.1002/ppul.21238
 172. Hilliard ME, Eakin MN, Borrelli B, Green A, Riekert KA. Medication beliefs mediate between depressive symptoms and medication adherence in cystic fibrosis. *Health Psychol*. 2015;34(5):496–504.
 173. Havermans T, Colpaert K, Dupont LJ. Quality of life in patients with cystic fibrosis: association with anxiety and depression. *J Cyst Fibros* [Internet]. 2008;7(6):581–4. doi:10.1016/j.jcf.2008.05.010
 174. Riekert KA, Bartlett SJ, Boyle MP, Krishnan JA, Rand CS. The association between depression, lung function, and health-related quality of life among adults with cystic fibrosis. *Chest* [Internet]. 2007;132(1):231–7. doi:10.1378/chest.06-2474
 175. McGowan PO, Sasaki A, D'Alessio AC, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci*. 2009 Mar;12(3):342–8. doi:10.1038/nn.2270
 176. Esposito EA, Jones MJ, Doom JR, MacIsaac JL, Gunnar MR, Kobor MS. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Dev Psychopathol* [Internet]. 2016;28(4pt2):1385–99. doi:10.1017/s0954579416000055
 177. Serpeloni F, Radtke K, de Assis SG, Henning F, Nätt D, Elbert T. Grandmaternal stress during pregnancy and DNA methylation of the third generation: an epigenome-wide association study. *Transl Psychiatry*. 2017 Aug;7(8):e1202. doi:10.1038/tp.2017.153
 178. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016 Dec;17(1):61. doi:10.1186/s13059-016-0926-z
 179. van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016 Sep;7(1):11115. doi:10.1038/ncomms11115
 180. Yousefi M, Karmaus W, Zhang H, Ewart S, Arshad H, Holloway JW. The methylation of the LEPR/LEPROT genotype at the promoter and body regions influence concentrations of leptin in girls and BMI at age 18 years if their mother smoked during pregnancy. *Int J Mol Epidemiol Genet* [Internet]. 2013;4(2):86–100.

181. Klengel T, Mehta D, Anacker C, et al. Allele-specific FKBP5 DNA demethylation mediates gene–childhood trauma interactions. *Nat Neurosci*. 2013 Jan;16(1):33–41. doi:10.1038/nn.3275
182. Robinson WH, Lindstrom TM, Cheung RK, Sokolove J. Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat Rev Rheumatol* [Internet]. 2013;9(5):267–76. doi:10.1038/nrrheum.2013.14
183. Ideozu JE, Zhang X, Pan A, et al. Increased expression of plasma-induced ABCC1 mRNA in cystic fibrosis. *Int J Mol Sci*. 2017 Aug;18(8):1752.
184. Magalhães M, Rivals I, Claustres M, et al. DNA methylation at modifier genes of lung disease severity is altered in cystic fibrosis. *Clin Epigenet*. 2017 Dec;9(1):19.
185. Magalhães M, Tost J, Pineau F, et al. Dynamic changes of DNA methylation and lung disease in cystic fibrosis: lessons from a monogenic disease. *Epigenomics* [Internet]. 2018;10(8):1131–45. doi:10.2217/epi-2018-0005
186. Michels KB, Binder AM, Dedeurwaerder S, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* [Internet]. 2013;10(10):949–55. doi:10.1038/nmeth.2632
187. Brennan S, Sly PD, Gangell CL, et al. Alveolar macrophages and CC chemokines are increased in children with cystic fibrosis. *Eur Respir J* [Internet]. 2009;34(3):655–61. doi:10.1183/09031936.00178508
188. Peterson-Carmichael SL, Harris WT, Goel R, et al. Association of lower airway inflammation with physiologic findings in young children with cystic fibrosis. *Pediatr Pulmonol*. 2009 May;44(5):503–11. doi:10.1002/ppul.21044
189. Reinhardt N, Chen CI, Loppow D, et al. Cellular profiles of induced sputum in children with stable cystic fibrosis: comparison with BAL. *Eur Respir J*. 2003 Sep;22(3):497–502. doi:10.1183/09031936.03.00043603
190. Armstrong DS, Grimwood K, Carlin JB, et al. Lower airway inflammation in infants and young children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 1997;156(4):1197–204. doi:10.1164/ajrccm.156.4.96-11058
191. Marguet C, Jouen-Boedes F, Dean TP, Warner JO. Bronchoalveolar cell profiles in children with asthma, infantile wheeze, chronic cough, or cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 1999;159(5):1533–40. doi:10.1164/ajrccm.159.5.9805028
192. Ratjen F, Rietschel E, Griesse M, et al. Fractional analysis of bronchoalveolar lavage fluid cytology in cystic fibrosis patients with normal lung function. Bronchoalveolar lavage for the evaluation of anti-inflammatory treatment (BEAT) study group. *Eur Respir J* [Internet]. 2000;15(1):141–5. doi:10.1183/09031936.00.15114100

193. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* [Internet]. 2012;13(1):86. doi:10.1186/1471-2105-13-86
194. Garratt LW, Kicic A, Robertson C, Ranganathan S, Sly PD, Stick SM. The AREST CF experience in biobanking—more than just tissues, tubes and time. *J Cyst Fibros* [Internet]. 2017;16(5):622–27. doi:10.1016/j.jcf.2017.08.003
195. Teschendorff AE, Zheng SC. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* [Internet]. 2017;9(5):757–68. doi:10.2217/epi-2016-0153
196. Lin X, Tan JYL, Teh AL, et al. Cell type-specific DNA methylation in neonatal cord tissue and cord blood: a 850K-reference panel and comparison of cell types. *Epigenetics* [Internet]. 2018;13(9):941–58. doi:10.1080/15592294.2018.1522929
197. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* [Internet]. 2013;8(3):290–302. doi:10.4161/epi.23924
198. Teschendorff AE, Gao Y, Jones A, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* [Internet]. 2016;7(1):10478. doi:10.1038/ncomms10478
199. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005 Jul;102(30):10604–09. doi:10.1073/pnas.0500398102
200. Bakulski KM, Feinberg JI, Andrews SV, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* [Internet]. 2016;11(5):354–62. doi:10.1080/15592294.2016.1161875
201. Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods* [Internet]. 2014;11(3):309–11. doi:10.1038/nmeth.2815
202. Rahmani E, Zaitlen N, Baran Y, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat Methods* [Internet]. 2016;13(5):443–5. doi:10.1038/nmeth.3809
203. Shanthikumar S, Burton M, Saffery R, Ranganathan SC, Neeland MR. Single-cell flow cytometry profiling of BAL in children. *Am J Respir Cell Mol Biol* [Internet]. 2020;63(2):152–9. doi:10.1165/rcmb.2019-0453MA

204. Chandler JD, Margaroli C, Horati H, et al. Myeloperoxidase oxidation of methionine associates with early cystic fibrosis lung disease. *Eur Respir J* [Internet]. 2018;52(4):1801118. doi:10.1183/13993003.01118-2018
205. Levy L, Juvet SC, Boonstra K, et al. Sequential broncho-alveolar lavages reflect distinct pulmonary compartments: clinical and research implications in lung transplantation. *Respir Res*. 2018 Dec;19(1):102. doi:10.1186/s12931-018-0786-z
206. R Core Team. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing; 2014. Available from: <http://www.r-project.org/>
207. Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data. *F1000Research* [Internet]. 2016;5:1281. doi:10.12688/f1000research.8839.3
208. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinform* [Internet]. 2017;33(4):558–60. doi:10.1093/bioinformatics/btw691
209. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* [Internet]. 2004;5(10):1–16. doi:10.1186/gb-2004-5-10-r80
210. Touleimat N, Tost J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* [Internet]. 2012;4(3):325–41. doi:10.2217/epi.12.21
211. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* [Internet]. 2016;17(1):208. doi:10.1186/s13059-016-1066-1
212. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* [Internet]. 2018;19(1):64. doi:10.1186/s13059-018-1448-7
213. Blériot C, Chakarov S, Ginhoux F. Determinants of resident tissue macrophage identity and function. *Immun* [Internet]. 2020;52(6):957–70. doi:10.1016/j.immuni.2020.05.014
214. Tan SY, Krasnow MA. Developmental origin of lung macrophage diversity. *Dev* [Internet]. 2016;143(8):1318–27. doi:10.1242/dev.129122

215. Bassler K, Fujii W, Kapellos TS, et al. Alterations of multiple alveolar macrophage states in chronic obstructive pulmonary disease. *bioRxiv* [Internet]. 2020 May. doi:10.1101/2020.05.28.121541
216. Chen Y, Armstrong DA, Salas LA, et al. Genome-wide DNA methylation profiling shows a distinct epigenetic signature associated with lung macrophages in cystic fibrosis. *Clin Epigenet* [Internet]. 2018;10(1):152. doi:10.1186/s13148-018-0580-2
217. Wikenius E, Moe V, Smith L, Heiervang ER, Berglund A. DNA methylation changes in infants between 6 and 52 weeks. *Sci Rep* [Internet]. 2019;9(1):17587. doi:10.1038/s41598-019-54355-z
218. Tan Q, Heijmans BT, Hjelmberg JV, Soerensen M, Christensen K, Christiansen L. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int J Epidemiol* [Internet]. 2016;45(4):1146–58. doi:10.1093/ije/dyw132
219. Arneson D, Yang X, Wang K. MethylResolver—a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. *Commun Biol* [Internet]. 2020;3(1):422. doi:10.1038/s42003-020-01146-2
220. Li M, Zou D, Li Z, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* 2019 Jan;47(D1):D983–8. doi:10.1093/nar/gky1027
221. Braga FAV, Kar G, Berg M, et al. A cellular census of healthy lung and asthmatic airway wall identifies novel cell states in health and disease. *bioRxiv* [Internet]. 2019:527408. doi:10.1101/527408
222. Yang IV, Konigsberg I, MacPhail K, et al. DNA methylation changes in lung immune cells are associated with granulomatous lung disease. *Am J Respir Cell Mol Biol* [Internet]. 2019;60(1):96–105. doi:10.1165/rcmb.2018-0177OC
223. Tricas L, Echeverria A, Blanco MA, et al. Flow cytometry counting of bronchoalveolar lavage leukocytes with a new profile of monoclonal antibodies combination. *Cytometry B Clin Cytom.* 2012 Mar;82(2):61–6. doi:10.1002/cyto.b.20630
224. Hodge SJ, Hodge GL, Holmes M, Reynolds PM. Flow cytometric characterization of cell populations in bronchoalveolar lavage and bronchial brushings from patients with chronic obstructive pulmonary disease. *Cytometry B Clin Cytom.* 2004 Sep;61(1):27–34. doi:10.1002/cyto.b.20020
225. Ringh MV, Hagemann-Jensen M, Needhamsen M, et al. Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells. *EBioMedicine* [Internet]. 2019;46:290–304. doi:10.1016/j.ebiom.2019.07.006

226. Breeze CE, Paul DS, van Dongen J, et al. eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep* [Internet]. 2016;17(8):2137–50. doi:10.1016/j.celrep.2016.10.059
227. Meyer KC, Raghu G, Baughman RP, 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* [Internet]. 2012;185(9):1004–14. doi:10.1164/rccm.201202-0320ST
228. Calabrese F, Lunardi F, Pezzuto F, et al. Are there new biomarkers in tissue and liquid biopsies for the early detection of non-small cell lung cancer? *J Clin Med* [Internet]. 2019;8(3): 414. doi:10.3390/jcm8030414
229. Saglani S, Nicholson AG, Scallan M, et al. Investigation of young children with severe recurrent wheeze: any clinical benefit? *Eur Respir J* [Internet]. 2006;27(1):29–35. doi:10.1183/09031936.06.00030605
230. Shanthikumar S, Neeland MN, Saffery R, Ranganathan S. Gene modifiers of cystic fibrosis lung disease: A systematic review. *Pediatr Pulmonol* [Internet]. 2019;54(9):1356–66. doi:10.1002/ppul.24366
231. Sly PD, Wainwright CE. Diagnosis and early life risk factors for bronchiectasis in cystic fibrosis: a review. *Expert Rev Respir Med* [Internet]. 2016;10(9):1003–10. doi:10.1080/17476348.2016.1204915
232. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of phenotypic extremes identifies CAV2 and TMC6 as interacting modifiers of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *PLoS Genet* [Internet]. 2015;11(6):e1005273. doi:10.1371/journal.pgen.1005273
233. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet* [Internet]. 2012;44(8):886–9. doi:10.1038/ng.2344
234. Randolph JJ. Online kappa calculator [Computer software]; 2016 Available from: <http://justusrandolph.net/kappa/>
235. Wu Z, Aryee MJ. Subset quantile normalization using negative control features. *J Comput Biol* [Internet]. 2010;17(10):1385–95. doi:10.1089/cmb.2010.0049
236. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinform* [Internet]. 2005;21(9):2067–75. doi:10.1093/bioinformatics/bti270

237. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat* [Internet]. 2016;10(2):946–63. doi:10.1214/16-aos920
238. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995 Jan;57(1):289–300.
239. Gagnon-Bartsch JA, Jacob L, Speed TP. Removing unwanted variation from high dimensional data with negative controls. Berkeley (CA): University of California Department of Statistics; 2013. Report No.: 820.
240. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* [Internet]. 2014;15(2):R31. doi:10.1186/gb-2014-15-2-r31
241. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostat* [Internet]. 2012;13(3):539–52. doi:10.1093/biostatistics/kxr034
242. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* [Internet]. 2015;43(7):e47. doi:10.1093/nar/gkv007
243. Rosenow T, Oudraad MC, Murray CP, et al. PRAGMA-CF: a quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2015;191(10):1158–65. doi:10.1164/rccm.201501-0061OC
244. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* [Internet]. 2015;44(4):1429–41. doi:10.1093/ije/dyv041
245. Graw S, Henn R, Thompson JA, Koestler DC. pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC Bioinformatics* [Internet]. 2019;20(1):218. doi:10.1186/s12859-019-2804-7
246. Rosenow T, Mok LC, Turkovic L, et al. The cumulative effect of inflammation and infection on structural lung disease in early cystic fibrosis. *Eur Respir J* [Internet]. 2019;54(1):1801771. doi:10.1183/13993003.01771-2018
247. Turkovic L, Caudri D, Rosenow T, et al. Structural determinants of long-term functional outcomes in young children with cystic fibrosis. *Eur Respir J* [Internet]. 2020;55(5):1900748. doi:10.1183/13993003.00748-2019
248. Cystic Fibrosis Foundation Patient Registry. 2019 annual data report. Bethesda (MD): Cystic Fibrosis Foundation; 2020.

249. Stick SM, Brennan S, Murray C, et al. Bronchiectasis in infants and preschool children diagnosed with cystic fibrosis after newborn screening. *J Pediatr* [Internet]. 2009;155(5):623–8.e1. doi:10.1016/j.jpeds.2009.05.005
250. Maksimovic J, Gagnon-Bartsch JA, Speed TP, Oshlack A. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res* [Internet]. 2015;43(16):e106. doi:10.1093/nar/gkv526
251. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* [Internet]. 2017;45(4):e22. doi:10.1093/nar/gkw967
252. Descartes human gene expression during development [Internet]. Seattle (WA): Brotman Baty Institute; c2020. Available from: <https://descartes.brotmanbaty.org/bbi/human-gene-expression-during-development/>
253. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013 Jun;45(6):580–5.
254. Carbon S, Ireland A, Mungall CJ, et al. AmiGO: online access to ontology and annotation data. *Bioinform* [Internet]. 2009;25(2):288–9. doi:10.1093/bioinformatics/btn615
255. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC known genes. *Bioinform* [Internet]. 2006;22(9):1036–46. doi:10.1093/bioinformatics/btl048
256. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res* [Internet]. 2020;48(D1):D682–8. doi:10.1093/nar/gkz966
257. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. *Genome Biol* [Internet]. 2015;16(1):56. Available from: <http://europepmc.org/abstract/MED/25887522> doi:10.1186/s13059-015-0621-5
258. Descartes human chromatin accessibility during development [Internet]. Seattle (WA): Brotman Baty Institute; c2020. Available from: <https://descartes.brotmanbaty.org/bbi/human-chromatin-during-development/>
259. Davis CA, Hitz BC, Sloan CA, et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* [Internet]. 2018;46(D1):D794–801. doi:10.1093/nar/gkx1081
260. Martens JH, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematol* [Internet]. 2013;98(10):1487–9. doi:10.3324/haematol.2013.094243
261. Pliner HA, Packer JS, McFaline-Figueroa JL, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell* [Internet]. 2018;71(5):858–71.e8. doi:10.1016/j.molcel.2018.06.044

262. Hazra A, Gogtay N. Biostatistics series module 4: comparing groups—categorical variables. *Indian J Dermatol* [Internet]. 2016;61(4):385–92. doi:10.4103/0019-5154.185700
263. Breheny P, Stromberg A, Lambert J. p-value histograms: inference and diagnostics. *High-Throughput* [Internet]. 2018;7(3):23. doi:10.3390/ht7030023
264. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nat* [Internet]. 2015;518(7539):317–30. doi:10.1038/nature14248
265. Kuznetsova T, Prange KHM, Glass CK, de Winther MPJ. Transcriptional and epigenetic regulation of macrophages in atherosclerosis. *Nat Rev Cardiol* [Internet]. 2020;17(4):216–28. doi:10.1038/s41569-019-0265-3
266. Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nat* [Internet]. 2020;583(7818):699–710. doi:10.1038/s41586-020-2493-4
267. Regulatory features: EMBL-EBI [Internet]. Ensembl; c2020. [cited 2020 Dec 22] Available from: https://asia.ensembl.org/info/genome/funcgen/regulatory_features.html
268. Rhie SK, Schreiner S, Witt H, et al. Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. *Sci Adv* [Internet]. 2018;4(12):eaav8550. doi:10.1126/sciadv.aav8550
269. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med* [Internet]. 2015;47(6):e166. doi:10.1038/emm.2015.33
270. Zerbino DR, Johnson N, Juetteman T, et al. Ensembl regulation resources. *Database*. 2016;2016:bav119. doi:10.1093/database/bav119
271. Meng Y, Lin T, Liang S, et al. Value of DNA methylation in predicting curve progression in patients with adolescent idiopathic scoliosis. *EBioMedicine* [Internet]. 2018;36:489–96. doi:10.1016/j.ebiom.2018.09.014
272. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* [Internet]. 2010;5(9):1315–6. doi:10.1097/JTO.0b013e3181ec173d
273. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatr* [Internet]. 2007;96(5):644–7. doi:10.1111/j.1651-2227.2006.00178.x
274. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* [Internet]. 2018;50(9):1219–24. doi:10.1038/s41588-018-0183-z
275. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* [Internet]. 2020;15(9):2759–72. doi:10.1038/s41596-020-0353-1

276. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *GenomeMed* [Internet]. 2020;12(1):44. doi:10.1186/s13073-020-00742-5
277. Onengut-Gumuscu S, Chen WM, Robertson CC, et al. Type 1 diabetes risk in African-ancestry participants and utility of an ancestry-specific genetic risk score. *Diabetes Care* [Internet]. 2019;42(3):406–15. doi:10.2337/dc18-1727
278. Escott-Price V, Shoai M, Pither R, Williams J, Hardy J. Polygenic score prediction captures nearly all common genetic risk for Alzheimer’s disease. *Neurobiol Aging* [Internet]. 2017;49:214.e7–11. doi:10.1016/j.neurobiolaging.2016.07.018
279. European Society of Human Genetics. Individual access to genomic disease risk factors has a beneficial impact on lifestyles. *Science Daily* [Internet]. 2018 Jun 15 [cited 2020 Nov 19]. Available from: www.sciencedaily.com/releases/2018/06/180615185408.htm
280. Lecamwasam A, Novakovic B, Meyer B, Ekinci EI, Dwyer KM, Saffery R. DNA methylation profiling identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. *Nephrol Dial Transplant* [Internet]. 2020. doi:10.1093/ndt/gfaa226
281. Margaroli C, Garratt LW, Horati H, et al. Elastase exocytosis by airway neutrophils is associated with early lung damage in children with cystic fibrosis. *Am J Respir Crit Care Med* [Internet]. 2019;199(7):873–81. doi:10.1164/rccm.201803-0442OC
282. Ernst WL, Shome K, Wu CC, Gong X, Frizzell RA, Aridor M. VAMP-associated proteins (VAP) as receptors that couple cystic fibrosis transmembrane conductance regulator (CFTR) proteostasis with lipid homeostasis. *J Biol Chem* [Internet]. 2016;291(10):5206–20. doi:10.1074/jbc.M115.692749
283. Ogilvie V, Passmore M, Hyndman L, et al. Differential global gene expression in cystic fibrosis nasal and bronchial epithelium. *Genomics* [Internet]. 2011;98(5):327–36. doi:10.1016/j.ygeno.2011.06.008
284. Pattison SH, Gibson DS, Johnston E, et al. Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa*. *Eur Respir J* [Internet]. 2017;50(1):1601569. doi:10.1183/13993003.01569-2016
285. Wright FA, Strug LJ, Doshi VK, et al. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nat Genet* [Internet]. 2011;43(6):539–46. doi:10.1038/ng.838
286. SHIELD CF (the study of host immunity and early lung disease in cystic fibrosis) [Internet]. Dublin (Ireland): National Children’s Research Centre. Available from: <https://www.nationalchildrensresearchcentre.ie/research-innovation/research-areas/immunity-infection/cystic-fibrosis/shield-cf/>

287. Drug development pipeline [Internet]. Bethesda (MD): Cystic Fibrosis Foundation; 2020 [cited 2020 Nov 26]. Available from: <https://www.cff.org/trials/pipeline>
288. Liu X, Ren J, Luo N, et al. Comprehensive DNA methylation analysis of tissue of origin of plasma cell-free DNA by methylated CpG tandem amplification and sequencing (MCTA-Seq). *Clin Epigenet* [Internet]. 2019;11(1):93. doi:10.1186/s13148-019-0689-y
289. Liu M, Oxnard G, Klein E, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020 Sep;31(9):1266–7.
290. Chambers DC, Carew AM, Lukowski SW, Powell GE. Transcriptomics and single-cell RNA-sequencing. *Respirology* [Internet]. 2019;24(1):29–36. doi:10.1111/resp.13412
291. HMOX2 [Internet]. The Human Protein Atlas; c2020 [cited 2020 Nov 26]. Available from: <https://www.proteinatlas.org/ENSG00000103415-HMOX2/antibody>
292. Micheel C, Nass SJ, Omenn GS, et al. Evolution of translational omics: lessons learned and the path forward [Internet]. Washington (DC): National Academies Press; 2012. [cited 2020 Nov 26]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK202165/>
293. Sun YV, Hu YJ. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv Genet* [Internet]. 2016;93:147–90. doi:10.1016/bs.adgen.2015.11.004
294. Padilla-Martínez F, Collin F, Kwasniewski M, Kretowski A. Systematic review of polygenic risk scores for type 1 and type 2 diabetes. *Int J Mol Sci* [Internet]. 2020;21(5):1703. doi:10.3390/ijms21051703
295. Corvol H, Blackman SM, Boëlle PY, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun* [Internet]. 2015;6(1):8382. doi:10.1038/ncomms9382
296. García-Giménez JL, Seco-Cervera M, Tollefsbol TO, et al. Epigenetic biomarkers: current strategies and future challenges for their use in the clinical laboratory. *Crit Rev Clin Lab Sci* [Internet]. 2017;54(7–8):529–50. doi:10.1080/10408363.2017.1410520
297. García-Giménez JL, Romá-Mateo C, Carbonell N, et al. A new mass spectrometry-based method for the quantification of histones in plasma from septic shock patients. *Sci Rep* [Internet]. 2017;7(1):10643. doi:10.1038/s41598-017-10830-z
298. Jansen MP, Knijnenburg T, Reijm EA, et al. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer Res* [Internet]. 2013;73(22):6632–41. doi:10.1158/0008-5472.can-13-0704

299. Gilad S, Lithwick-Yanai G, Barshack I, et al. Classification of the four main types of lung cancer using a microRNA-based diagnostic assay. *J Mol Diag* [Internet]. 2012;14(5):510–7. doi:10.1016/j.jmoldx.2012.03.004
300. Dirks RA, Stunnenberg HG, Marks H. Genome-wide epigenomic profiling for biomarker discovery. *Clin Epigenet* [Internet]. 2016;8(1):122. doi:10.1186/s13148-016-0284-4
301. CF Foundation biorepository 2020 [Internet]. Bethesda (MD): Cystic Fibrosis Foundation; c2020 [cited 2020 Nov 26]. Available from: <https://www.cff.org/Research/Researcher-Resources/Tools-and-Resources/CF-Foundation-Biorepository/>
302. Plasschaert LW, Žilionis R, Choo-Wing R, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nat* [Internet]. 2018;560(7718):377–81. doi:10.1038/s41586-018-0394-6
303. Montoro DT, Haber AL, Biton M, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nat* [Internet]. 2018;560(7718):319–24. doi:10.1038/s41586-018-0393-7
304. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*. 2020 Jun;26(6):842–4.
305. Rilinger J, Kern WV, Duerschmied D, et al. A prospective, randomised, double blind placebo-controlled trial to evaluate the efficacy and safety of tocilizumab in patients with severe COVID-19 pneumonia (TOC-COVID): A structured summary of a study protocol for a randomised controlled trial. *Trials* [Internet]. 2020;21(1):470. doi:10.1186/s13063-020-04447-3
306. Schupp JC, Khanal S, Gomez JL, et al. Single cell transcriptional archetypes of airway inflammation in cystic fibrosis. *Am J Respir Crit Care Med*. 2020 Nov;202(10):1419–29. doi:10.1164/rccm.202004-0991OC
307. Montgomery ST, Dittrich AS, Garratt LW, et al. Interleukin-1 is associated with inflammation and structural lung disease in young children with cystic fibrosis. *J Cyst Fibros* [Internet]. 2018;17(6):715–22. doi:10.1016/j.jcf.2018.05.006
308. Esther CR, Jr., Muhlebach MS, Ehre C, et al. Mucus accumulation in the lungs precedes structural changes and infection in children with cystic fibrosis. *Sci Transl Med* [Internet]. 2019;11(486):eaav3488. doi:10.1126/scitranslmed.aav3488
309. Esther CR, Jr., Hill DB, Button B, et al. Sialic acid-to-urea ratio as a measure of airway surface hydration. *Am J Physiol Lung Cell Mol Physiol* [Internet]. 2017;312(3):L398–404. doi:10.1152/ajplung.00398.2016

310. Wheeler HE, Aquino-Michaels K, Gamazon ER, et al. Poly-omic prediction of complex traits: OmicKriging. *Genet Epidemiol* [Internet]. 2014;38(5):402–15. doi:10.1002/gepi.21808
311. Amiri Roudbar M, Mohammadabadi MR, Ayatollahi Mehrgardi A, et al. Integration of single nucleotide variants and whole-genome DNA methylation profiles for classification of rheumatoid arthritis cases from controls. *Heredity* [Internet]. 2020;124(5):658–74. doi:10.1038/s41437-020-0301-4
312. Vazquez AI, Veturi Y, Behring M, et al. Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles. *Genet* [Internet]. 2016;203(3):1425–38. doi:10.1534/genetics.115.185181
313. Davies JC, Cunningham S, Harris WT, et al. Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-5 years with cystic fibrosis and a CFTR gating mutation (KIWI): an open-label, single-arm study. *Lancet Respir Med* [Internet]. 2016;4(2):107–15. doi:10.1016/s2213-2600(15)00545-7
314. Davies JC, Wainwright CE, Sawicki GS, et al. Ivacaftor in infants aged 4 to <12 months with cystic fibrosis and a gating mutation: results of a 2-part phase 3 clinical trial. *Am J Respir Crit Care Med* [Internet]. 2020. doi:10.1164/rccm.202008-3177OC
315. Rosenfeld M, Wainwright CE, Higgins M, et al. Ivacaftor treatment of cystic fibrosis in children aged 12 to <24 months and with a CFTR gating mutation (ARRIVAL): a phase 3 single-arm study. *Lancet Respir Med* [Internet]. 2018;6(7):545–53. doi:10.1016/s2213-2600(18)30202-9
316. Heijerman HGM, McKone EF, Downey DG, et al. Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial. *Lancet* [Internet]. 2019;394(10212):1940–8. doi:10.1016/s0140-6736(19)32597-8
317. Middleton PG, Mall MA, Dřevínek P, et al. Elexacaftor-tezacaftor-ivacaftor for cystic fibrosis with a single Phe508del allele. *NEJM* [Internet]. 2019;381(19):1809–19. doi:10.1056/NEJMoa1908639
318. Corvol H, Mésinè J, Douksieh IH, et al. SLC26A9 gene is associated with lung function response to ivacaftor in patients with cystic fibrosis. *Front Pharmacol* [Internet]. 2018;9:828. doi:10.3389/fphar.2018.00828
319. Strug LJ, Gonska T, He G, et al. Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. *Hum Mol Genet* [Internet]. 2016;25(20):4590–600. doi:10.1093/hmg/ddw290

320. Our strategic plan: care 2020 [Internet]. Bethesda (MD): Cystic Fibrosis Foundation; c2020 [cited 2020 Nov 26]. Available from: <https://www.cff.org/About-Us/Reports-and-Financials/Our-Strategic-Plan/Care/>

1. Anderson DH. Cystic fibrosis of the pancreas and its relation to celiac disease. *Am J Dis Child* 1938;56(2):344-99.
2. Cystic Fibrosis Australia. Cystic Fibrosis in Australia 2014. 2016
3. Spoonhower KA, Davis PB. Epidemiology of Cystic Fibrosis. *Clinics in chest medicine* 2016;37(1):1-8. doi: 10.1016/j.ccm.2015.10.002 [published Online First: 2016/02/10]
4. Ong T, Marshall SG, Karczeski BA, et al. Cystic Fibrosis and Congenital Absence of the Vas Deferens. In: Pagon RA, Adam MP, Ardinger HH, et al., eds. *GeneReviews(R)*. Seattle (WA)1993.
5. Zielenski J, Rozmahel R, Bozon D, et al. Genomic DNA sequence of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics* 1991;10(1):214-28. [published Online First: 1991/05/01]
6. Vankeerberghen A, Cuppens H, Cassiman JJ. The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2002;1(1):13-29. [published Online First: 2004/10/07]
7. O'Sullivan BP, Freedman SD. Cystic fibrosis. *Lancet (London, England)* 2009;373(9678):1891-904. doi: 10.1016/s0140-6736(09)60327-5 [published Online First: 2009/05/01]
8. Pittman JE, Ferkol TW. The Evolution of Cystic Fibrosis Care. *Chest* 2015;148(2):533-42. doi: 10.1378/chest.14-1997
9. Shanthikumar S, Massie J. A Review of Treatments That Improve Cystic Fibrosis Transmembrane Conductance Regulator Function. *Clinical Medicine Insights: Therapeutics* 2017;9:1179559X17719123.
10. Sawicki GS, Sellers DE, Robinson WM. High treatment burden in adults with cystic fibrosis: challenges to disease self-management. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2009;8(2):91-6. doi: 10.1016/j.jcf.2008.09.007 [published Online First: 2008/10/28]
11. Cytic Fibrosis Foundation Patient Registry. Annual Data Report 2015. 2016
12. United Kingdom Cystic Fibrosis Registry. 2015 Annual Data Report. 2016
13. World Bank Group. Life expectancy at birth, total (years) 2017 [Available from: <http://data.worldbank.org/indicator/SP.DYN.LE00.IN>].
14. Kirkby S, Hayes D, Jr. Pediatric lung transplantation: indications and outcomes. *Journal of thoracic disease* 2014;6(8):1024-31. doi: 10.3978/j.issn.2072-1439.2014.04.27 [published Online First: 2014/08/19]
15. Yusen RD, Christie JD, Edwards LB, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirtieth Adult Lung and Heart-Lung Transplant Report--2013; focus theme: age. *The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation* 2013;32(10):965-78. doi: 10.1016/j.healun.2013.08.007 [published Online First: 2013/09/24]

16. Pillarisetti N, Williamson E, Linnane B, et al. Infection, inflammation, and lung function decline in infants with cystic fibrosis. *American journal of respiratory and critical care medicine* 2011;184(1):75-81. doi: 10.1164/rccm.201011-1892OC [published Online First: 2011/04/16]
17. Regnis JA, Robinson M, Bailey DL, et al. Mucociliary clearance in patients with cystic fibrosis and in normal subjects. *American journal of respiratory and critical care medicine* 1994;150(1):66-71. doi: 10.1164/ajrccm.150.1.8025774 [published Online First: 1994/07/01]
18. Boucher RC. Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. *Annual review of medicine* 2007;58:157-70. doi: 10.1146/annurev.med.58.071905.105316 [published Online First: 2007/01/16]
19. Matsui H, Grubb BR, Tarran R, et al. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* 1998;95(7):1005-15. [published Online First: 1999/01/06]
20. Hoegger MJ, Fischer AJ, McMenimen JD, et al. Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Science (New York, NY)* 2014;345(6198):818-22. doi: 10.1126/science.1255825 [published Online First: 2014/08/16]
21. Quinton PM. Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis. *Lancet (London, England)* 2008;372(9636):415-7. doi: 10.1016/s0140-6736(08)61162-9 [published Online First: 2008/08/05]
22. Zabner J, Smith JJ, Karp PH, et al. Loss of CFTR chloride channels alters salt absorption by cystic fibrosis airway epithelia in vitro. *Molecular cell* 1998;2(3):397-403. [published Online First: 1998/10/17]
23. Smith JJ, Travis SM, Greenberg EP, et al. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* 1996;85(2):229-36. [published Online First: 1996/04/19]
24. Pezzulo AA, Tang XX, Hoegger MJ, et al. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature* 2012;487(7405):109-13. doi: 10.1038/nature11130 [published Online First: 2012/07/06]
25. Schultz A, Puvvadi R, Borisov SM, et al. Airway surface liquid pH is not acidic in children with cystic fibrosis. *Nature communications* 2017;8(1):1409. doi: 10.1038/s41467-017-00532-5 [published Online First: 2017/11/11]
26. Bonfield TL, Konstan MW, Berger M. Altered respiratory epithelial cell cytokine production in cystic fibrosis. *The Journal of allergy and clinical immunology* 1999;104(1):72-8. [published Online First: 1999/07/10]
27. Tirouvanziam R, de Bentzmann S, Hubeau C, et al. Inflammation and infection in naive human cystic fibrosis airway grafts. *American journal of respiratory cell and molecular biology* 2000;23(2):121-7. doi: 10.1165/ajrcmb.23.2.4214 [published Online First: 2000/08/02]
28. Khan TZ, Wagener JS, Bost T, et al. Early pulmonary inflammation in infants with cystic fibrosis. *American journal of respiratory and critical care medicine* 1995;151(4):1075-82. doi: 10.1164/ajrccm.151.4.7697234 [published Online First: 1995/04/01]
29. Rosen BH, Evans TIA, Moll SR, et al. Infection Is Not Required for Mucoinflammatory Lung Disease in CFTR-Knockout Ferrets. *American journal of respiratory and critical care medicine* 2018;197(10):1308-18. doi: 10.1164/rccm.201708-1616OC [published Online First: 2018/01/13]

30. Meyerholz DK, Stoltz DA, Namati E, et al. Loss of cystic fibrosis transmembrane conductance regulator function produces abnormalities in tracheal development in neonatal pigs and young children. *American journal of respiratory and critical care medicine* 2010;182(10):1251-61. doi: 10.1164/rccm.201004-0643OC [published Online First: 2010/07/14]
31. Stoltz DA, Meyerholz DK, Pezzulo AA, et al. Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth. *Science translational medicine* 2010;2(29):29ra31. doi: 10.1126/scitranslmed.3000928 [published Online First: 2010/04/30]
32. Fischer AJ, Singh SB, Adam RJ, et al. Tracheomalacia is associated with lower FEV1 and Pseudomonas acquisition in children with CF. *Pediatric pulmonology* 2014;49(10):960-70. doi: 10.1002/ppul.22922 [published Online First: 2013/10/30]
33. Waters V, Ratjen F. Pulmonary Exacerbations in Children with Cystic Fibrosis. *Annals of the American Thoracic Society* 2015;12 Suppl 2:S200-6. doi: 10.1513/AnnalsATS.201502-098AW [published Online First: 2015/11/26]
34. Flume P, Vandevanter D. Pulmonary Exacerbations. In: Bush A, Bilton D, Hodson M, eds. Hodson and Geddes' Cystic Fibrosis. Fourth ed2016.
35. Quittner AL, Buu A, Messer MA, et al. Development and validation of The Cystic Fibrosis Questionnaire in the United States: a health-related quality-of-life measure for cystic fibrosis. *Chest* 2005;128(4):2347-54. doi: 10.1378/chest.128.4.2347 [published Online First: 2005/10/21]
36. Gee L, Abbott J, Conway SP, et al. Development of a disease specific health related quality of life measure for adults and adolescents with cystic fibrosis. *Thorax* 2000;55(11):946-54. [published Online First: 2000/10/26]
37. Habib AR, Manji J, Wilcox PG, et al. A systematic review of factors associated with health-related quality of life in adolescents and adults with cystic fibrosis. *Annals of the American Thoracic Society* 2015;12(3):420-8. doi: 10.1513/AnnalsATS.201408-393OC [published Online First: 2015/02/03]
38. Bedrossian CW, Greenberg SD, Singer DB, et al. The lung in cystic fibrosis. A quantitative study including prevalence of pathologic findings among different age groups. *Human pathology* 1976;7(2):195-204. [published Online First: 1976/03/01]
39. Kang EY, Miller RR, Muller NL. Bronchiectasis: comparison of preoperative thin-section CT and pathologic findings in resected specimens. *Radiology* 1995;195(3):649-54. doi: 10.1148/radiology.195.3.7753989 [published Online First: 1995/06/01]
40. de Jong PA, Tiddens HA. Cystic fibrosis specific computed tomography scoring. *Proceedings of the American Thoracic Society* 2007;4(4):338-42. doi: 10.1513/pats.200611-175HT [published Online First: 2007/07/27]
41. Loeve M, Hop WC, de Bruijne M, et al. Chest computed tomography scores are predictive of survival in patients with cystic fibrosis awaiting lung transplantation. *American journal of respiratory and critical care medicine* 2012;185(10):1096-103. doi: 10.1164/rccm.201111-2065OC [published Online First: 2012/03/10]
42. Brody AS, Sucharew H, Campbell JD, et al. Computed tomography correlates with pulmonary exacerbations in children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2005;172(9):1128-32. doi: 10.1164/rccm.200407-989OC [published Online First: 2005/08/16]

43. Tepper LA, Utens EM, Caudri D, et al. Impact of bronchiectasis and trapped air on quality of life and exacerbations in cystic fibrosis. *The European respiratory journal* 2013;42(2):371-9. doi: 10.1183/09031936.00137612 [published Online First: 2013/01/15]
44. Loeve M, Gerbrands K, Hop WC, et al. Bronchiectasis and pulmonary exacerbations in children and young adults with cystic fibrosis. *Chest* 2011;140(1):178-85. doi: 10.1378/chest.10-1152 [published Online First: 2010/12/15]
45. Bortoluzzi CF, Volpi S, D'Orazio C, et al. Bronchiectases at early chest computed tomography in children with cystic fibrosis are associated with increased risk of subsequent pulmonary exacerbations and chronic pseudomonas infection. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2014;13(5):564-71. doi: 10.1016/j.jcf.2014.03.006 [published Online First: 2014/04/15]
46. Mott LS, Park J, Murray CP, et al. Progression of early structural lung disease in young children with cystic fibrosis assessed using CT. *Thorax* 2012;67(6):509-16. doi: 10.1136/thoraxjnl-2011-200912 [published Online First: 2011/12/28]
47. de Jong PA, Nakano Y, Hop WC, et al. Changes in airway dimensions on computed tomography scans of children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2005;172(2):218-24. doi: 10.1164/rccm.200410-1311OC [published Online First: 2005/04/16]
48. Nasr SZ, Kuhns LR, Brown RW, et al. Use of computerized tomography and chest x-rays in evaluating efficacy of aerosolized recombinant human DNase in cystic fibrosis patients younger than age 5 years: a preliminary study. *Pediatric pulmonology* 2001;31(5):377-82. [published Online First: 2001/05/08]
49. Hisert KB, Heltshe SL, Pope C, et al. Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections. *American journal of respiratory and critical care medicine* 2017;195(12):1617-28.
50. Ronan NJ, Einarsson GG, Twomey M, et al. CORK study in cystic fibrosis: sustained improvements in ultra-low-dose chest CT scores after CFTR modulation with ivacaftor. *Chest* 2018;153(2):395-403.
51. Davis SD, Fordham LA, Brody AS, et al. Computed tomography reflects lower airway inflammation and tracks changes in early cystic fibrosis. *American journal of respiratory and critical care medicine* 2007;175(9):943-50. doi: 10.1164/rccm.200603-343OC [published Online First: 2007/02/17]
52. Wainwright CE, Vidmar S, Armstrong DS, et al. Effect of bronchoalveolar lavage-directed therapy on *Pseudomonas aeruginosa* infection and structural lung injury in children with cystic fibrosis: a randomized trial. *Jama* 2011;306(2):163-71. doi: 10.1001/jama.2011.954 [published Online First: 2011/07/14]
53. Stick S, Tiddens H, Aurora P, et al. Early intervention studies in infants and preschool children with cystic fibrosis: are we ready? *The European respiratory journal* 2013;42(2):527-38. doi: 10.1183/09031936.00108212 [published Online First: 2013/03/23]
54. Ramsey KA, Ranganathan S. Interpretation of lung function in infants and young children with cystic fibrosis. *Respirology (Carlton, Vic)* 2014;19(6):792-9. doi: 10.1111/resp.12329 [published Online First: 2014/06/21]
55. Kerem E, Reisman J, Corey M, et al. Prediction of mortality in patients with cystic fibrosis. *The New England journal of medicine* 1992;326(18):1187-91. doi: 10.1056/nejm199204303261804 [published Online First: 1992/04/30]

56. Miller MR, Crapo R, Hankinson J, et al. General considerations for lung function testing. *The European respiratory journal* 2005;26(1):153-61. doi: 10.1183/09031936.05.00034505 [published Online First: 2005/07/05]
57. Aurora P, Bush A, Gustafsson P, et al. Multiple-breath washout as a marker of lung disease in preschool children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2005;171(3):249-56. doi: 10.1164/rccm.200407-895OC [published Online First: 2004/11/02]
58. Gustafsson PM, De Jong PA, Tiddens HA, et al. Multiple-breath inert gas washout and spirometry versus structural lung disease in cystic fibrosis. *Thorax* 2008;63(2):129-34. doi: 10.1136/thx.2007.077784 [published Online First: 2007/08/07]
59. Horsley AR, Gustafsson PM, Macleod KA, et al. Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax* 2008;63(2):135-40. doi: 10.1136/thx.2007.082628 [published Online First: 2007/08/07]
60. Marostica PJ, Weist AD, Eigen H, et al. Spirometry in 3- to 6-year-old children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2002;166(1):67-71. doi: 10.1164/rccm.200111-056OC [published Online First: 2002/07/02]
61. Ramsey KA, Ranganathan S, Park J, et al. Early respiratory infection is associated with reduced spirometry in children with cystic fibrosis. *American journal of respiratory and critical care medicine* 2014;190(10):1111-6. doi: 10.1164/rccm.201407-1277OC [published Online First: 2014/10/17]
62. Harrison AN, Regelman WE, Zirbes JM, et al. Longitudinal assessment of lung function from infancy to childhood in patients with cystic fibrosis. *Pediatric pulmonology* 2009;44(4):330-9. doi: 10.1002/ppul.20994 [published Online First: 2009/03/11]
63. Belessis Y, Dixon B, Hawkins G, et al. Early cystic fibrosis lung disease detected by bronchoalveolar lavage and lung clearance index. *American journal of respiratory and critical care medicine* 2012;185(8):862-73. doi: 10.1164/rccm.201109-1631OC [published Online First: 2012/02/11]
64. Owens CM, Aurora P, Stanojevic S, et al. Lung Clearance Index and HRCT are complementary markers of lung abnormalities in young children with CF. *Thorax* 2011;66(6):481-8. doi: 10.1136/thx.2010.150375 [published Online First: 2011/03/23]
65. Vermeulen F, Proesmans M, Boon M, et al. Lung clearance index predicts pulmonary exacerbations in young patients with cystic fibrosis. *Thorax* 2014;69(1):39-45. doi: 10.1136/thoraxjnl-2013-203807 [published Online First: 2013/09/12]
66. Milla CE, Ratjen F, Marigowda G, et al. Lumacaftor/ivacaftor in patients aged 6–11 years with cystic fibrosis and homozygous for F508del-CFTR. *American journal of respiratory and critical care medicine* 2017;195(7):912-20.
67. Ratjen F, Hug C, Marigowda G, et al. Efficacy and safety of lumacaftor and ivacaftor in patients aged 6–11 years with cystic fibrosis homozygous for F508del-CFTR: a randomised, placebo-controlled phase 3 trial. *The Lancet Respiratory medicine* 2017;5(7):557-67.
68. Shaw M, Khan U, Clancy JP, et al. Changes in LCI in F508del/F508del patients treated with lumacaftor/ivacaftor: Results from the prospect study. *Journal of Cystic Fibrosis* 2020

69. Amin R, Subbarao P, Jabar A, et al. Hypertonic saline improves the LCI in paediatric patients with CF with normal lung function. *Thorax* 2010;65(5):379-83. doi: 10.1136/thx.2009.125831 [published Online First: 2010/05/04]
70. Robinson PD, Cooper P, Van Asperen P, et al. Using index of ventilation to assess response to treatment for acute pulmonary exacerbation in children with cystic fibrosis. *Pediatric pulmonology* 2009;44(8):733-42. doi: 10.1002/ppul.20956 [published Online First: 2009/07/15]
71. Fuchs SI, Gappa M, Eder J, et al. Tracking Lung Clearance Index and chest CT in mild cystic fibrosis lung disease over a period of three years. *Respiratory medicine* 2014;108(6):865-74. doi: 10.1016/j.rmed.2014.03.011 [published Online First: 2014/04/15]
72. Stanojevic S, Davis SD, Retsch-Bogart G, et al. Progression of Lung Disease in Preschool Patients with Cystic Fibrosis. *American journal of respiratory and critical care medicine* 2017;195(9):1216-25. doi: 10.1164/rccm.201610-2158OC [published Online First: 2016/12/13]
73. Ranganathan SC, Hall GL, Sly PD, et al. Early Lung Disease in Infants and Pre-school Children with Cystic Fibrosis: What Have We Learnt and What Should We Do About It? *American journal of respiratory and critical care medicine* 2016 doi: 10.1164/rccm.201606-1107CI [published Online First: 2016/12/03]
74. Simpson SJ, Mott LS, Esther CR, Jr., et al. Novel end points for clinical trials in young children with cystic fibrosis. *Expert review of respiratory medicine* 2013;7(3):231-43. doi: 10.1586/ers.13.25 [published Online First: 2013/06/06]
75. Long FR, Williams RS, Castile RG. Structural airway abnormalities in infants and young children with cystic fibrosis. *The Journal of pediatrics* 2004;144(2):154-61. doi: 10.1016/j.jpeds.2003.09.026 [published Online First: 2004/02/05]
76. Sly PD, Brennan S, Gangell C, et al. Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *American journal of respiratory and critical care medicine* 2009;180(2):146-52. doi: 10.1164/rccm.200901-0069OC [published Online First: 2009/04/18]
77. Ranganathan SC, Dezateux C, Bush A, et al. Airway function in infants newly diagnosed with cystic fibrosis. *Lancet (London, England)* 2001;358(9297):1964-5. [published Online First: 2001/12/19]
78. Mott LS, Graniel KG, Park J, et al. Assessment of early bronchiectasis in young children with cystic fibrosis is dependent on lung volume. *Chest* 2013;144(4):1193-8. doi: 10.1378/chest.12-2589 [published Online First: 2013/05/18]
79. de Jong PA, Mayo JR, Golmohammadi K, et al. Estimation of cancer mortality associated with repetitive computed tomography scanning. *American journal of respiratory and critical care medicine* 2006;173(2):199-203. doi: 10.1164/rccm.200505-810OC [published Online First: 2005/10/29]
80. Ramsey KA, Rosenow T, Turkovic L, et al. Lung Clearance Index and Structural Lung Disease on Computed Tomography in Early Cystic Fibrosis. *American journal of respiratory and critical care medicine* 2016;193(1):60-7. doi: 10.1164/rccm.201507-1409OC [published Online First: 2015/09/12]
81. Hall GL, Logie KM, Parsons F, et al. Air trapping on chest CT is associated with worse ventilation distribution in infants with cystic fibrosis diagnosed following newborn screening. *PloS one* 2011;6(8):e23932. doi: 10.1371/journal.pone.0023932 [published Online First: 2011/09/03]

82. Ellemunter H, Fuchs SI, Unsinn KM, et al. Sensitivity of Lung Clearance Index and chest computed tomography in early CF lung disease. *Respiratory medicine* 2010;104(12):1834-42. doi: 10.1016/j.rmed.2010.06.010 [published Online First: 2010/07/20]
83. Burrows B, Knudson RJ, Lebowitz MD. The relationship of childhood respiratory illness to adult obstructive airway disease. *The American review of respiratory disease* 1977;115(5):751-60. doi: 10.1164/arrd.1977.115.5.751 [published Online First: 1977/05/01]
84. Stocks J, Hislop A, Sonnappa S. Early lung development: lifelong effect on respiratory health and disease. *The Lancet Respiratory medicine* 2013;1(9):728-42. doi: 10.1016/s2213-2600(13)70118-8 [published Online First: 2014/01/17]
85. Chow CW, Landau LI, Taussig LM. Bronchial mucous glands in the newborn with cystic fibrosis. *European journal of pediatrics* 1982;139(4):240-3. [published Online First: 1982/12/01]
86. Fakhoury K, Durie PR, Levison H, et al. Meconium ileus in the absence of cystic fibrosis. *Archives of disease in childhood* 1992;67(10 Spec No):1204-6. [published Online First: 1992/10/01]
87. Linnane BM, Hall GL, Nolan G, et al. Lung function in infants with cystic fibrosis diagnosed by newborn screening. *American journal of respiratory and critical care medicine* 2008;178(12):1238-44. doi: 10.1164/rccm.200804-551OC [published Online First: 2008/09/13]
88. Hoo AF, Thia LP, Nguyen TT, et al. Lung function is abnormal in 3-month-old infants with cystic fibrosis diagnosed by newborn screening. *Thorax* 2012;67(10):874-81. doi: 10.1136/thoraxjnl-2012-201747 [published Online First: 2012/07/04]
89. Sly PD, Gangell CL, Chen L, et al. Risk factors for bronchiectasis in children with cystic fibrosis. *The New England journal of medicine* 2013;368(21):1963-70. doi: 10.1056/NEJMoa1301725 [published Online First: 2013/05/23]
90. Ranganathan SC, Stocks J, Dezateux C, et al. The evolution of airway function in early childhood following clinical diagnosis of cystic fibrosis. *American journal of respiratory and critical care medicine* 2004;169(8):928-33. doi: 10.1164/rccm.200309-1344OC [published Online First: 2004/02/03]
91. Nguyen TT, Thia LP, Hoo AF, et al. Evolution of lung function during the first year of life in newborn screened cystic fibrosis infants. *Thorax* 2014;69(10):910-7. doi: 10.1136/thoraxjnl-2013-204023 [published Online First: 2013/09/28]
92. Davies G, Thia LP, Stocks J, et al. Minimal change in structural, functional and inflammatory markers of lung disease in newborn screened infants with cystic fibrosis at one year. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2020 doi: 10.1016/j.jcf.2020.01.006 [published Online First: 2020/02/12]
93. Long FR, Williams RS, Adler BH, et al. Comparison of quiet breathing and controlled ventilation in the high-resolution CT assessment of airway disease in infants with cystic fibrosis. *Pediatric radiology* 2005;35(11):1075-80. doi: 10.1007/s00247-005-1541-4 [published Online First: 2005/07/26]
94. Mott LS, Park J, Gangell CL, et al. Distribution of early structural lung changes due to cystic fibrosis detected with chest computed tomography. *The Journal of pediatrics* 2013;163(1):243-8.e1-3. doi: 10.1016/j.jpeds.2012.12.042 [published Online First: 2013/01/30]

95. Stiglbauer R, Schurawitzki H, Eichler I, et al. High resolution CT in children with cystic fibrosis. *Acta radiologica (Stockholm, Sweden : 1987)* 1992;33(6):548-53. [published Online First: 1992/11/01]
96. Demirkazik FB, Ariyurek OM, Ozcelik U, et al. High resolution CT in children with cystic fibrosis: correlation with pulmonary functions and radiographic scores. *European journal of radiology* 2001;37(1):54-9. [published Online First: 2001/03/29]
97. Brody AS, Klein JS, Molina PL, et al. High-resolution computed tomography in young patients with cystic fibrosis: distribution of abnormalities and correlation with pulmonary function tests. *The Journal of pediatrics* 2004;145(1):32-8. doi: 10.1016/j.jpeds.2004.02.038 [published Online First: 2004/07/09]
98. Marchant JM, Masel JP, Dickinson FL, et al. Application of chest high-resolution computer tomography in young children with cystic fibrosis. *Pediatric pulmonology* 2001;31(1):24-9. [published Online First: 2001/02/17]
99. Helbich TH, Heinz-Peer G, Eichler I, et al. Cystic fibrosis: CT assessment of lung involvement in children and adults. *Radiology* 1999;213(2):537-44. doi: 10.1148/radiology.213.2.r99nv04537 [published Online First: 1999/11/07]
100. Maffessanti M, Candusso M, Brizzi F, et al. Cystic fibrosis in children: HRCT findings and distribution of disease. *Journal of thoracic imaging* 1996;11(1):27-38. [published Online First: 1996/01/01]
101. Drumm ML, Konstan MW, Schluchter MD, et al. Genetic modifiers of lung disease in cystic fibrosis. *The New England journal of medicine* 2005;353(14):1443-53. doi: 10.1056/NEJMoa051469 [published Online First: 2005/10/07]
102. Whiting P, Al M, Burgers L, et al. Ivacaftor for the treatment of patients with cystic fibrosis and the G551D mutation: a systematic review and cost-effectiveness analysis. *Health technology assessment (Winchester, England)* 2014;18(18):1-106. doi: 10.3310/hta18180 [published Online First: 2014/03/25]
103. Rowbotham NJ, Smith S, Leighton PA, et al. The top 10 research priorities in cystic fibrosis developed by a partnership between people with CF and healthcare providers. *Thorax* 2017 doi: 10.1136/thoraxjnl-2017-210473 [published Online First: 2017/08/06]
104. Van Goor F, Hadida S, Grootenhuys PD, et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106(44):18825-30. doi: 10.1073/pnas.0904709106 [published Online First: 2009/10/23]
105. Rowe SM, Miller S, Sorscher EJ. Cystic fibrosis. *The New England journal of medicine* 2005;352(19):1992-2001. doi: 10.1056/NEJMra043184
106. Kerem E, Corey M, Kerem BS, et al. The relation between genotype and phenotype in cystic fibrosis--analysis of the most common mutation (delta F508). *The New England journal of medicine* 1990;323(22):1517-22. doi: 10.1056/nejm199011293232203 [published Online First: 1990/11/29]
107. Johansen HK, Nir M, Hoiby N, et al. Severity of cystic fibrosis in patients homozygous and heterozygous for delta F508 mutation. *Lancet (London, England)* 1991;337(8742):631-4. [published Online First: 1991/03/16]
108. Santis G, Osborne L, Knight R, et al. Genetic influences on pulmonary severity in cystic fibrosis. *Lancet (London, England)* 1990;335(8684):294. [published Online First: 1990/02/03]
109. Cystic Fibrosis Genotype-Phenotype Consortium. Correlation between genotype and phenotype in patients with cystic fibrosis. *The New England journal of*

- medicine* 1993;329(18):1308-13. doi: 10.1056/nejm199310283291804
[published Online First: 1993/10/28]
110. Koch C, Cuppens H, Rainisio M, et al. European Epidemiologic Registry of Cystic Fibrosis (ERCF): comparison of major disease manifestations between patients with different classes of mutations. *Pediatric pulmonology* 2001;31(1):1-12. [published Online First: 2001/02/17]
 111. Santis G, Osborne L, Knight RA, et al. Independent genetic determinants of pancreatic and pulmonary status in cystic fibrosis. *Lancet (London, England)* 1990;336(8723):1081-4. [published Online First: 1990/11/03]
 112. Mekus F, Ballmann M, Bronsveld I, et al. Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. *Twin research : the official journal of the International Society for Twin Studies* 2000;3(4):277-93. [published Online First: 2001/07/21]
 113. Vanscoy LL, Blackman SM, Collaco JM, et al. Heritability of lung disease severity in cystic fibrosis. *American journal of respiratory and critical care medicine* 2007;175(10):1036-43. doi: 10.1164/rccm.200608-1164OC [published Online First: 2007/03/03]
 114. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nature reviews Genetics* 2008;9(4):255-66. doi: 10.1038/nrg2322 [published Online First: 2008/03/06]
 115. Collaco JM, Blackman SM, McGready J, et al. Quantification of the relative contribution of environmental and genetic factors to variation in cystic fibrosis lung function. *The Journal of pediatrics* 2010;157(5):802-7.e1-3. doi: 10.1016/j.jpeds.2010.05.018 [published Online First: 2010/06/29]
 116. Waddington C. The Epigenotype. *Endeavour* 1942:18-20.
 117. Deans C, Maggert KA. What do you mean, "epigenetic"? *Genetics* 2015;199(4):887-96.
 118. Haig D. Commentary: The epidemiology of epigenetics. *International journal of epidemiology* 2011;41(1):13-16.
 119. Nanney DL. Epigenetic control systems. *Proceedings of the National Academy of Sciences* 1958;44(7):712-17.
 120. Holliday R. Epigenetics: an overview. *genesis* 1994;15(6):453-57.
 121. Russo VE, Martienssen RA, Riggs AD. Epigenetic mechanisms of gene regulation: Cold Spring Harbor Laboratory Press 1996.
 122. Bird A. Perceptions of epigenetics. *Nature* 2007;447(7143):396.
 123. Feinberg AP. The Key Role of Epigenetics in Human Disease Prevention and Mitigation. *New England Journal of Medicine* 2018;378(14):1323-34.
 124. Luger K, Mäder AW, Richmond RK, et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997;389(6648):251-60. doi: 10.1038/38444
 125. Annunziato A. DNA packaging: nucleosomes and chromatin. *Nature Education* 2008;1(1):26.
 126. Mattick JS, Makunin IV. Non-coding RNA. *Human Molecular Genetics* 2006;15(suppl_1):R17-R29. doi: 10.1093/hmg/ddl046
 127. Friedman RC, Farh KK, Burge CB, et al. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* 2009;19(1):92-105. doi: 10.1101/gr.082701.108 [published Online First: 2008/10/29]
 128. Piatek MJ, Werner A. Endogenous siRNAs: regulators of internal affairs. *Biochem Soc Trans* 2014;42(4):1174-9. doi: 10.1042/BST20140068

129. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 2007;8(3):209-20. doi: 10.1038/nrm2124
130. Siomi MC, Sato K, Pezic D, et al. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 2011;12(4):246-58. doi: 10.1038/nrm3089 [published Online First: 2011/03/24]
131. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* 2018;19(1):45-58. doi: 10.1038/nrm.2017.77
132. Guo M, Schimmel P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol* 2013;9(3):145-53. doi: 10.1038/nchembio.1158
133. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 2014;157(1):77-94. doi: 10.1016/j.cell.2014.03.008
134. Horn PJ, Peterson CL. Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science (New York, NY)* 2002;297(5588):1824-7. doi: 10.1126/science.1074200 [published Online First: 2002/09/14]
135. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science (New York, NY)* 1975;187(4173):226-32. [published Online First: 1975/01/24]
136. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature genetics* 2007;39(4):457-66. doi: 10.1038/ng1990 [published Online First: 2007/03/06]
137. Guo H, Zhu P, Yan L, et al. The DNA methylation landscape of human early embryos. *Nature* 2014;511(7511):606-10. doi: 10.1038/nature13544
138. Sundrani D, Khot V, Joshi S. DNA Methylation for Prediction of Adverse Pregnancy Outcomes. In: Garcia-Gimenez J, ed. *Epigenetic Biomarkers and Diagnostics*. London: Elsevier 2016.
139. Dhe-Paganon S, Syeda F, Park L. DNA methyl transferase 1: regulatory mechanisms and implications in health and disease. *International journal of biochemistry and molecular biology* 2011;2(1):58-66. [published Online First: 2011/10/05]
140. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nature reviews Genetics* 2013;14(3):204-20. doi: 10.1038/nrg3354
141. Ollikainen M, Smith KR, Joo EJ, et al. DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum Mol Genet* 2010;19(21):4176-88. doi: 10.1093/hmg/ddq336
142. Gordon L, Joo JE, Powell JE, et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome research* 2012;22(8):1395-406. doi: 10.1101/gr.136598.111
143. Teh AL, Pan H, Chen L, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome research* 2014;24(7):1064-74. doi: 10.1101/gr.171439.113
144. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* 2013;49(2):359-67. doi: 10.1016/j.molcel.2012.10.016 [published Online First: 2012/11/28]

145. Gervin K, Hammero M, Akselsen HE, et al. Extensive variation and low heritability of DNA methylation identified in a twin study. *Genome research* 2011;21(11):1813-21. doi: 10.1101/gr.119685.110 [published Online First: 2011/09/29]
146. Bierne H, Hamon M, Cossart P. Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med* 2012;2(12):a010272. doi: 10.1101/cshperspect.a010272
147. Cole J, Morris P, Dickman MJ, et al. The therapeutic potential of epigenetic manipulation during infectious diseases. *Pharmacology & therapeutics* 2016;167:85-99. doi: 10.1016/j.pharmthera.2016.07.013 [published Online First: 2016/08/16]
148. Kopp BT, Ortega-García JA, Sadreameli SC, et al. The Impact of Secondhand Smoke Exposure on Children with Cystic Fibrosis: A Review. *International journal of environmental research and public health* 2016;13(10) doi: 10.3390/ijerph13101003 [published Online First: 2016/10/19]
149. Guerrero-Preston R, Goldman LR, Brebi-Mieville P, et al. Global DNA hypomethylation is associated with in utero exposure to cotinine and perfluorinated alkyl compounds. *Epigenetics* 2014;5(6):539-46. doi: 10.4161/epi.5.6.12378
150. Ivorra C, Fraga MF, Bayon GF, et al. DNA methylation patterns in newborns exposed to tobacco in utero. *J Transl Med* 2015;13:25. doi: 10.1186/s12967-015-0384-5
151. Joubert BR, Haberg SE, Nilsen RM, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environmental health perspectives* 2012;120(10):1425-31. doi: 10.1289/ehp.1205412 [published Online First: 2012/08/02]
152. Rauschert S, Melton PE, Burdge G, et al. Maternal Smoking During Pregnancy Induces Persistent Epigenetic Changes Into Adolescence, Independent of Postnatal Smoke Exposure and Is Associated With Cardiometabolic Risk. *Frontiers in genetics* 2019;10:770. doi: 10.3389/fgene.2019.00770 [published Online First: 2019/10/17]
153. Joubert BR, Håberg SE, Bell DA, et al. Maternal Smoking and DNA Methylation in Newborns: In Utero Effect or Epigenetic Inheritance? *Cancer Epidemiology Biomarkers & Prevention* 2014;23(6):1007-17. doi: 10.1158/1055-9965.epi-13-1256
154. Breton CV, Byun HM, Wenten M, et al. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *American journal of respiratory and critical care medicine* 2009;180(5):462-7. doi: 10.1164/rccm.200901-0135OC
155. Wiklund P, Karhunen V, Richmond RC, et al. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. *Clinical Epigenetics* 2019;11(1):97. doi: 10.1186/s13148-019-0683-4
156. Corey M, McLaughlin FJ, Williams M, et al. A comparison of survival, growth, and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *Journal of clinical epidemiology* 1988;41(6):583-91. [published Online First: 1988/01/01]
157. Voisin S, Almen MS, Moschonis G, et al. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. *Eur J Hum Genet* 2015;23(5):654-62. doi: 10.1038/ejhg.2014.139

158. Figueroa V, Milla C, Parks EJ, et al. Abnormal lipid concentrations in cystic fibrosis. *The American journal of clinical nutrition* 2002;75(6):1005-11. doi: 10.1093/ajcn/75.6.1005 [published Online First: 2002/05/31]
159. Woestenenk JW, Schulkes DA, Schipper HS, et al. Dietary intake and lipid profile in children and adolescents with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2017;16(3):410-17. doi: 10.1016/j.jcf.2017.02.010
160. Vaughan WJ, Lindgren FT, Whalen JB, et al. Serum lipoprotein concentrations in cystic fibrosis. *Science (New York, NY)* 1978;199(4330):783-6. [published Online First: 1978/02/17]
161. Wijnands KP, Obermann-Borst SA, Steegers-Theunissen RP. Early life lipid profile and metabolic programming in very young children. *Nutrition, metabolism, and cardiovascular diseases : NMCD* 2015;25(6):608-14. doi: 10.1016/j.numecd.2015.02.010 [published Online First: 2015/04/05]
162. de la Rocha C, Perez-Mojica JE, Leon SZ, et al. Associations between whole peripheral blood fatty acids and DNA methylation in humans. *Sci Rep* 2016;6:25867. doi: 10.1038/srep25867
163. Cassidy L, Morton A, Etherington C, et al. 205 Folate deficiency in adult patients at a regional cystic fibrosis (CF) centre. *Journal of Cystic Fibrosis* 2014;13:S99.
164. Chakrabarty B, Kabra SK, Gulati S, et al. Peripheral neuropathy in cystic fibrosis: a prevalence study. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2013;12(6):754-60. doi: 10.1016/j.jcf.2013.01.005
165. Maqbool A, Schall JI, Mascarenhas MR, et al. Vitamin B(12) status in children with cystic fibrosis and pancreatic insufficiency. *J Pediatr Gastroenterol Nutr* 2014;58(6):733-8. doi: 10.1097/MPG.0000000000000313
166. Jacob RA, Gretz DM, Taylor PC, et al. Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *The Journal of nutrition* 1998;128(7):1204-12. doi: 10.1093/jn/128.7.1204 [published Online First: 1998/07/03]
167. Rampersaud GC, Kauwell GP, Hutson AD, et al. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *The American journal of clinical nutrition* 2000;72(4):998-1003. doi: 10.1093/ajcn/72.4.998 [published Online First: 2000/09/30]
168. Haggarty P, Hoad G, Campbell DM, et al. Folate in pregnancy and imprinted gene and repeat element methylation in the offspring. *The American journal of clinical nutrition* 2013;97(1):94-9. doi: 10.3945/ajcn.112.042572 [published Online First: 2012/11/16]
169. Fryer AA, Nafee TM, Ismail KM, et al. LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. *Epigenetics* 2009;4(6):394-8. [published Online First: 2009/09/17]
170. Perng W, Rozek LS, Mora-Plazas M, et al. Micronutrient status and global DNA methylation in school-age children. *Epigenetics* 2012;7(10):1133-41. doi: 10.4161/epi.21915 [published Online First: 2012/08/25]
171. Quittner AL, Goldbeck L, Abbott J, et al. Prevalence of depression and anxiety in patients with cystic fibrosis and parent caregivers: results of The International Depression Epidemiological Study across nine countries. *Thorax* 2014;69(12):1090-7. doi: 10.1136/thoraxjnl-2014-205983

172. Smith BA, Modi AC, Quittner AL, et al. Depressive symptoms in children with cystic fibrosis and parents and its effects on adherence to airway clearance. *Pediatric pulmonology* 2010;45(8):756-63. doi: 10.1002/ppul.21238
173. Hilliard ME, Eakin MN, Borrelli B, et al. Medication beliefs mediate between depressive symptoms and medication adherence in cystic fibrosis. *Health Psychology* 2015;34(5):496.
174. Havermans T, Colpaert K, Dupont LJ. Quality of life in patients with Cystic Fibrosis: association with anxiety and depression. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2008;7(6):581-4. doi: 10.1016/j.jcf.2008.05.010 [published Online First: 2008/08/12]
175. Riekert KA, Bartlett SJ, Boyle MP, et al. The association between depression, lung function, and health-related quality of life among adults with cystic fibrosis. *Chest* 2007;132(1):231-7. doi: 10.1378/chest.06-2474 [published Online First: 2007/07/13]
176. McGowan PO, Sasaki A, D'Alessio AC, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* 2009;12(3):342-8. doi: 10.1038/nn.2270
177. Esposito EA, Jones MJ, Doom JR, et al. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Development and psychopathology* 2016;28(4pt2):1385-99. doi: 10.1017/s0954579416000055 [published Online First: 2016/10/26]
178. Serpeloni F, Radtke K, de Assis SG, et al. Grandmaternal stress during pregnancy and DNA methylation of the third generation: an epigenome-wide association study. *Transl Psychiatry* 2017;7(8):e1202. doi: 10.1038/tp.2017.153
179. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016;17:61. doi: 10.1186/s13059-016-0926-z
180. van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nature communications* 2016;7:11115. doi: 10.1038/ncomms11115
181. Yousefi M, Karmaus W, Zhang H, et al. The methylation of the LEPR/LEPROT genotype at the promoter and body regions influence concentrations of leptin in girls and BMI at age 18 years if their mother smoked during pregnancy. *International journal of molecular epidemiology and genetics* 2013;4(2):86-100. [published Online First: 2013/07/23]
182. Klengel T, Mehta D, Anacker C, et al. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci* 2013;16(1):33-41. doi: 10.1038/nn.3275
183. Robinson WH, Lindstrom TM, Cheung RK, et al. Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nature reviews Rheumatology* 2013;9(5):267-76. doi: 10.1038/nrrheum.2013.14 [published Online First: 2013/02/20]
184. Ideozu JE, Zhang X, Pan A, et al. Increased Expression of Plasma-Induced ABCC1 mRNA in Cystic Fibrosis. *International journal of molecular sciences* 2017;18(8):1752.
185. Magalhães M, Rivals I, Claustres M, et al. DNA methylation at modifier genes of lung disease severity is altered in cystic fibrosis. *Clinical epigenetics* 2017;9(1):19.

186. Magalhães M, Tost J, Pineau F, et al. Dynamic changes of DNA methylation and lung disease in cystic fibrosis: lessons from a monogenic disease. *Epigenomics* 2018;10(8):1131-45. doi: 10.2217/epi-2018-0005 [published Online First: 2018/07/28]
187. Michels KB, Binder AM, Dedeurwaerder S, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nature methods* 2013;10(10):949-55. doi: 10.1038/nmeth.2632 [published Online First: 2013/10/01]
188. Brennan S, Sly PD, Gangell CL, et al. Alveolar macrophages and CC chemokines are increased in children with cystic fibrosis. *The European respiratory journal* 2009;34(3):655-61. doi: 10.1183/09031936.00178508 [published Online First: 2009/04/24]
189. Peterson-Carmichael SL, Harris WT, Goel R, et al. Association of lower airway inflammation with physiologic findings in young children with cystic fibrosis. *Pediatric pulmonology* 2009;44(5):503-11. doi: 10.1002/ppul.21044
190. Reinhardt N, Chen CI, Loppow D, et al. Cellular profiles of induced sputum in children with stable cystic fibrosis: comparison with BAL. *The European respiratory journal* 2003;22(3):497-502. doi: 10.1183/09031936.03.00043603
191. Armstrong DS, Grimwood K, Carlin JB, et al. Lower airway inflammation in infants and young children with cystic fibrosis. *American journal of respiratory and critical care medicine* 1997;156(4 Pt 1):1197-204. doi: 10.1164/ajrccm.156.4.96-11058 [published Online First: 1997/11/14]
192. Marguet C, Jouen-Boedes F, Dean TP, et al. Bronchoalveolar cell profiles in children with asthma, infantile wheeze, chronic cough, or cystic fibrosis. *American journal of respiratory and critical care medicine* 1999;159(5 Pt 1):1533-40. doi: 10.1164/ajrccm.159.5.9805028 [published Online First: 1999/05/06]
193. Ratjen F, Rietschel E, Griesse M, et al. Fractional analysis of bronchoalveolar lavage fluid cytology in cystic fibrosis patients with normal lung function. Bronchoalveolar lavage for the evaluation of anti-inflammatory treatment (BEAT) study group. *The European respiratory journal* 2000;15(1):141-5. doi: 10.1183/09031936.00.15114100 [published Online First: 2000/03/04]
194. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* 2012;13:86. doi: 10.1186/1471-2105-13-86 [published Online First: 2012/05/10]
195. Garratt LW, Kicic A, Robertson C, et al. The AREST CF experience in biobanking - More than just tissues, tubes and time. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2017;16(5):622-27. doi: 10.1016/j.jcf.2017.08.003 [published Online First: 2017/08/15]
196. Teschendorff AE, Zheng SC. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* 2017;9(5):757-68. doi: 10.2217/epi-2016-0153 [published Online First: 2017/05/19]
197. Lin X, Tan JYL, Teh AL, et al. Cell type-specific DNA methylation in neonatal cord tissue and cord blood: a 850K-reference panel and comparison of cell types. *Epigenetics* 2018;13(9):941-58. doi: 10.1080/15592294.2018.1522929 [published Online First: 2018/09/21]

198. Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 2013;8(3):290-302. doi: 10.4161/epi.23924 [published Online First: 2013/02/22]
199. Teschendorff AE, Gao Y, Jones A, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nature communications* 2016;7:10478. doi: 10.1038/ncomms10478 [published Online First: 2016/01/30]
200. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(30):10604-09. doi: 10.1073/pnas.0500398102
201. Bakulski KM, Feinberg JL, Andrews SV, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* 2016;11(5):354-62. doi: 10.1080/15592294.2016.1161875 [published Online First: 2016/03/29]
202. Zou J, Lippert C, Heckerman D, et al. Epigenome-wide association studies without the need for cell-type composition. *Nature methods* 2014;11(3):309-11. doi: 10.1038/nmeth.2815 [published Online First: 2014/01/28]
203. Rahmani E, Zaitlen N, Baran Y, et al. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nature methods* 2016;13(5):443-5. doi: 10.1038/nmeth.3809 [published Online First: 2016/03/29]
204. Shanthikumar S, Burton M, Saffery R, et al. Single-Cell Flow Cytometry Profiling of BAL in Children. *American journal of respiratory cell and molecular biology* 2020;63(2):152-59. doi: 10.1165/rcmb.2019-0453MA [published Online First: 2020/03/26]
205. Chandler JD, Margaroli C, Horati H, et al. Myeloperoxidase oxidation of methionine associates with early cystic fibrosis lung disease. *The European respiratory journal* 2018;52(4) doi: 10.1183/13993003.01118-2018 [published Online First: 2018/09/08]
206. Levy L, Juvet SC, Boonstra K, et al. Sequential broncho-alveolar lavages reflect distinct pulmonary compartments: clinical and research implications in lung transplantation. *Respiratory Research* 2018;19(1):102. doi: 10.1186/s12931-018-0786-z
207. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, . 2014 [Available from: <http://www.r-project.org/>].
208. Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data. *F1000Research* 2016;5:1281. doi: 10.12688/f1000research.8839.3 [published Online First: 2017/04/26]
209. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics (Oxford, England)* 2017;33(4):558-60. doi: 10.1093/bioinformatics/btw691 [published Online First: 2016/12/31]
210. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80 [published Online First: 2004/10/06]

211. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 2012;4(3):325-41. doi: 10.2217/epi.12.21 [published Online First: 2012/06/14]
212. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 2016;17(1):208. doi: 10.1186/s13059-016-1066-1 [published Online First: 2016/10/09]
213. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* 2018;19(1):64. doi: 10.1186/s13059-018-1448-7 [published Online First: 2018/05/31]
214. Blériot C, Chakarov S, Ginhoux F. Determinants of Resident Tissue Macrophage Identity and Function. *Immunity* 2020;52(6):957-70. doi: 10.1016/j.immuni.2020.05.014 [published Online First: 2020/06/20]
215. Tan SY, Krasnow MA. Developmental origin of lung macrophage diversity. *Development (Cambridge, England)* 2016;143(8):1318-27. doi: 10.1242/dev.129122 [published Online First: 2016/03/10]
216. Bassler K, Fujii W, Kapellos TS, et al. Alterations of multiple alveolar macrophage states in chronic obstructive pulmonary disease. *bioRxiv* 2020
217. Chen Y, Armstrong DA, Salas LA, et al. Genome-wide DNA methylation profiling shows a distinct epigenetic signature associated with lung macrophages in cystic fibrosis. *Clin Epigenetics* 2018;10(1):152. doi: 10.1186/s13148-018-0580-2 [published Online First: 2018/12/12]
218. Wikenius E, Moe V, Smith L, et al. DNA methylation changes in infants between 6 and 52 weeks. *Sci Rep* 2019;9(1):17587. doi: 10.1038/s41598-019-54355-z [published Online First: 2019/11/28]
219. Tan Q, Heijmans BT, Hjelmborg JV, et al. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. *Int J Epidemiol* 2016;45(4):1146-58. doi: 10.1093/ije/dyw132 [published Online First: 2016/08/09]
220. Arneson D, Yang X, Wang K. MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents. *Communications biology* 2020;3(1):422. doi: 10.1038/s42003-020-01146-2 [published Online First: 2020/08/05]
221. Li M, Zou D, Li Z, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Research* 2018;47(D1):D983-D88. doi: 10.1093/nar/gky1027
222. Braga FAV, Kar G, Berg M, et al. A cellular census of healthy lung and asthmatic airway wall identifies novel cell states in health and disease. *bioRxiv* 2019:527408.
223. Yang IV, Konigsberg I, MacPhail K, et al. DNA Methylation Changes in Lung Immune Cells Are Associated with Granulomatous Lung Disease. *American journal of respiratory cell and molecular biology* 2019;60(1):96-105. doi: 10.1165/rcmb.2018-0177OC [published Online First: 2018/08/25]
224. Tricas L, Echeverria A, Blanco MA, et al. Flow cytometry counting of bronchoalveolar lavage leukocytes with a new profile of monoclonal antibodies combination. *Cytometry B Clin Cytom* 2012;82(2):61-6. doi: 10.1002/cyto.b.20630

225. Hodge SJ, Hodge GL, Holmes M, et al. Flow cytometric characterization of cell populations in bronchoalveolar lavage and bronchial brushings from patients with chronic obstructive pulmonary disease. *Cytometry B Clin Cytom* 2004;61(1):27-34. doi: 10.1002/cyto.b.20020
226. Ringh MV, Hagemann-Jensen M, Needhamsen M, et al. Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells. *EBioMedicine* 2019;46:290-304. doi: 10.1016/j.ebiom.2019.07.006 [published Online First: 2019/07/16]
227. Breeze CE, Paul DS, van Dongen J, et al. eFORGE: A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data. *Cell reports* 2016;17(8):2137-50. doi: 10.1016/j.celrep.2016.10.059 [published Online First: 2016/11/17]
228. Meyer KC, Raghu G, Baughman RP, et al. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *American journal of respiratory and critical care medicine* 2012;185(9):1004-14. doi: 10.1164/rccm.201202-0320ST [published Online First: 2012/05/03]
229. Calabrese F, Lunardi F, Pezzuto F, et al. Are There New Biomarkers in Tissue and Liquid Biopsies for the Early Detection of Non-Small Cell Lung Cancer? *Journal of clinical medicine* 2019;8(3) doi: 10.3390/jcm8030414 [published Online First: 2019/03/29]
230. Saglani S, Nicholson AG, Scallan M, et al. Investigation of young children with severe recurrent wheeze: any clinical benefit? *The European respiratory journal* 2006;27(1):29-35. doi: 10.1183/09031936.06.00030605 [published Online First: 2006/01/03]
231. Shanthikumar S, Neeland MN, Saffery R, et al. Gene modifiers of cystic fibrosis lung disease: A systematic review. *Pediatric pulmonology* 2019;54(9):1356-66. doi: 10.1002/ppul.24366 [published Online First: 2019/05/30]
232. Sly PD, Wainwright CE. Diagnosis and early life risk factors for bronchiectasis in cystic fibrosis: a review. *Expert review of respiratory medicine* 2016;10(9):1003-10. doi: 10.1080/17476348.2016.1204915 [published Online First: 2016/06/23]
233. Emond MJ, Louie T, Emerson J, et al. Exome Sequencing of Phenotypic Extremes Identifies CAV2 and TMC6 as Interacting Modifiers of Chronic Pseudomonas aeruginosa Infection in Cystic Fibrosis. *PLoS genetics* 2015;11(6):e1005273. doi: 10.1371/journal.pgen.1005273 [published Online First: 2015/06/06]
234. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nature genetics* 2012;44(8):886-9. doi: 10.1038/ng.2344 [published Online First: 2012/07/10]
235. Randolph JJ. Online Kappa Calculator [Computer software] 2916 [Available from: <http://justusrandolph.net/kappa/>].
236. Wu Z, Aryee MJ. Subset quantile normalization using negative control features. *Journal of computational biology : a journal of computational molecular cell biology* 2010;17(10):1385-95. doi: 10.1089/cmb.2010.0049 [published Online First: 2010/10/27]
237. Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics (Oxford, England)* 2005;21(9):2067-75. doi: 10.1093/bioinformatics/bti270 [published Online First: 2005/01/20]

238. Phipson B, Lee S, Majewski IJ, et al. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. *The annals of applied statistics* 2016;10(2):946-63. doi: 10.1214/16-aos920 [published Online First: 2017/04/04]
239. 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.
240. Gagnon-Bartsch JA, Jacob L, Speed TP. Removing unwanted variation from high dimensional data with negative controls. *Berkeley: Tech Reports from Dep Stat Univ California* 2013:1-112.
241. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014;15(2):R31. doi: 10.1186/gb-2014-15-2-r31 [published Online First: 2014/02/06]
242. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics (Oxford, England)* 2012;13(3):539-52. doi: 10.1093/biostatistics/kxr034 [published Online First: 2011/11/22]
243. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47. doi: 10.1093/nar/gkv007 [published Online First: 2015/01/22]
244. Rosenow T, Oudraad MC, Murray CP, et al. PRAGMA-CF. A Quantitative Structural Lung Disease Computed Tomography Outcome in Young Children with Cystic Fibrosis. *American journal of respiratory and critical care medicine* 2015;191(10):1158-65. doi: 10.1164/rccm.201501-0061OC [published Online First: 2015/03/11]
245. Breuer O, Caudri D, Stick S, et al. Predicting disease progression in cystic fibrosis. *Expert review of respiratory medicine* 2018;12(11):905-17. doi: 10.1080/17476348.2018.1519400 [published Online First: 2018/09/04]
246. Tsai PC, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* 2015;44(4):1429-41. doi: 10.1093/ije/dyv041 [published Online First: 2015/05/15]
247. Graw S, Henn R, Thompson JA, et al. pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). *BMC bioinformatics* 2019;20(1):218. doi: 10.1186/s12859-019-2804-7 [published Online First: 2019/05/01]
248. Rosenow T, Mok LC, Turkovic L, et al. The cumulative effect of inflammation and infection on structural lung disease in early cystic fibrosis. *The European respiratory journal* 2019;54(1) doi: 10.1183/13993003.01771-2018 [published Online First: 2019/04/27]
249. Turkovic L, Caudri D, Rosenow T, et al. Structural determinants of long-term functional outcomes in young children with cystic fibrosis. *The European respiratory journal* 2020;55(5) doi: 10.1183/13993003.00748-2019 [published Online First: 2020/03/07]
250. Cystic Fibrosis Foundation. 2019 Annual Data Report. Bethesda, Maryland, 2019.
251. Stick SM, Brennan S, Murray C, et al. Bronchiectasis in infants and preschool children diagnosed with cystic fibrosis after newborn screening. *The Journal of pediatrics* 2009;155(5):623-8.e1. doi: 10.1016/j.jpeds.2009.05.005 [published Online First: 2009/07/21]

252. Maksimovic J, Gagnon-Bartsch JA, Speed TP, et al. Removing unwanted variation in a differential methylation analysis of Illumina HumanMethylation450 array data. *Nucleic Acids Res* 2015;43(16):e106. doi: 10.1093/nar/gkv526 [published Online First: 2015/05/21]
253. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 2017;45(4):e22. doi: 10.1093/nar/gkw967 [published Online First: 2016/12/08]
254. Brotman Baty Institute. Descartes Atlas of Human Gene Expression During Development 2020 [Available from: <https://descartes.brotmanbaty.org/>].
255. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nature genetics* 2013;45(6):580-85.
256. Carbon S, Ireland A, Mungall CJ, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics (Oxford, England)* 2009;25(2):288-9. doi: 10.1093/bioinformatics/btn615 [published Online First: 2008/11/27]
257. Hsu F, Kent WJ, Clawson H, et al. The UCSC Known Genes. *Bioinformatics (Oxford, England)* 2006;22(9):1036-46. doi: 10.1093/bioinformatics/btl048 [published Online First: 2006/02/28]
258. Yates AD, Achuthan P, Akanni W, et al. Ensembl 2020. *Nucleic Acids Res* 2020;48(D1):D682-d88. doi: 10.1093/nar/gkz966 [published Online First: 2019/11/07]
259. Zerbino DR, Wilder SP, Johnson N, et al. The ensembl regulatory build. *Genome biology* 2015; 16. <http://europepmc.org/abstract/MED/25887522>
<https://doi.org/10.1186/s13059-015-0621-5>
<https://europepmc.org/articles/PMC4407537>
<https://europepmc.org/articles/PMC4407537?pdf=render> (accessed 2015/03//).
260. Brotman Baty Institute. Descartes Atlas of Human Chromatin Accessibility During Development 2020 [Available from: <https://descartes.brotmanbaty.org/>].
261. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 2018;46(D1):D794-d801. doi: 10.1093/nar/gkx1081 [published Online First: 2017/11/11]
262. Martens JH, Stunnenberg HG. BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* 2013;98(10):1487-9. doi: 10.3324/haematol.2013.094243 [published Online First: 2013/10/05]
263. Pliner HA, Packer JS, McFaline-Figueroa JL, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Molecular cell* 2018;71(5):858-71.e8. doi: 10.1016/j.molcel.2018.06.044 [published Online First: 2018/08/07]
264. Hazra A, Gogtay N. Biostatistics Series Module 4: Comparing Groups - Categorical Variables. *Indian journal of dermatology* 2016;61(4):385-92. doi: 10.4103/0019-5154.185700 [published Online First: 2016/08/12]
265. Breheny P, Stromberg A, Lambert J. p-Value Histograms: Inference and Diagnostics. *High-throughput* 2018;7(3) doi: 10.3390/ht7030023 [published Online First: 2018/09/12]
266. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518(7539):317-30. doi: 10.1038/nature14248 [published Online First: 2015/02/20]
267. Kuznetsova T, Prange KHM, Glass CK, et al. Transcriptional and epigenetic regulation of macrophages in atherosclerosis. *Nature reviews Cardiology*

- 2020;17(4):216-28. doi: 10.1038/s41569-019-0265-3 [published Online First: 2019/10/04]
268. Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;583(7818):699-710. doi: 10.1038/s41586-020-2493-4 [published Online First: 2020/07/31]
 269. Ensembl. Regulatory features: EMBL-EBI; 2020 [Available from: https://asia.ensembl.org/info/genome/funcgen/regulatory_features.html.
 270. Rhie SK, Schreiner S, Witt H, et al. Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. *Science advances* 2018;4(12):eaav8550. doi: 10.1126/sciadv.aav8550 [published Online First: 2018/12/18]
 271. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. *Experimental & molecular medicine* 2015;47(6):e166. doi: 10.1038/emm.2015.33 [published Online First: 2015/06/06]
 272. Zerbino DR, Johnson N, Juetteman T, et al. Ensembl regulation resources. *Database* 2016;2016 doi: 10.1093/database/bav119
 273. Meng Y, Lin T, Liang S, et al. Value of DNA methylation in predicting curve progression in patients with adolescent idiopathic scoliosis. *EBioMedicine* 2018;36:489-96. doi: 10.1016/j.ebiom.2018.09.014 [published Online First: 2018/09/23]
 274. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 2010;5(9):1315-6. doi: 10.1097/JTO.0b013e3181ec173d [published Online First: 2010/08/26]
 275. Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta paediatrica (Oslo, Norway : 1992)* 2007;96(5):644-7. doi: 10.1111/j.1651-2227.2006.00178.x [published Online First: 2007/03/23]
 276. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics* 2018;50(9):1219-24. doi: 10.1038/s41588-018-0183-z [published Online First: 2018/08/15]
 277. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols* 2020;15(9):2759-72. doi: 10.1038/s41596-020-0353-1 [published Online First: 2020/07/28]
 278. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine* 2020;12(1):44. doi: 10.1186/s13073-020-00742-5 [published Online First: 2020/05/20]
 279. Onengut-Gumuscu S, Chen WM, Robertson CC, et al. Type 1 Diabetes Risk in African-Ancestry Participants and Utility of an Ancestry-Specific Genetic Risk Score. *Diabetes care* 2019;42(3):406-15. doi: 10.2337/dc18-1727 [published Online First: 2019/01/20]
 280. Escott-Price V, Shoai M, Pither R, et al. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiology of aging* 2017;49:214.e7-14.e11. doi: 10.1016/j.neurobiolaging.2016.07.018 [published Online First: 2016/09/07]
 281. European Society of Human Genetics. Individual access to genomic disease risk factors has a beneficial impact on lifestyles ScienceDaily: ScienceDaily; 2018 [Available from: www.sciencedaily.com/releases/2018/06/180615185408.htm accessed 19 November 2020 2020.

282. Lecamwasam A, Novakovic B, Meyer B, et al. DNA methylation profiling identifies epigenetic differences between early versus late stages of diabetic chronic kidney disease. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2020 doi: 10.1093/ndt/gfaa226 [published Online First: 2020/11/05]
283. Margaroli C, Garratt LW, Horati H, et al. Elastase Exocytosis by Airway Neutrophils Is Associated with Early Lung Damage in Children with Cystic Fibrosis. *American journal of respiratory and critical care medicine* 2019;199(7):873-81. doi: 10.1164/rccm.201803-0442OC [published Online First: 2018/10/04]
284. Ernst WL, Shome K, Wu CC, et al. VAMP-associated Proteins (VAP) as Receptors That Couple Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Proteostasis with Lipid Homeostasis. *The Journal of biological chemistry* 2016;291(10):5206-20. doi: 10.1074/jbc.M115.692749 [published Online First: 2016/01/08]
285. Ogilvie V, Passmore M, Hyndman L, et al. Differential global gene expression in cystic fibrosis nasal and bronchial epithelium. *Genomics* 2011;98(5):327-36. doi: 10.1016/j.ygeno.2011.06.008 [published Online First: 2011/07/16]
286. Pattison SH, Gibson DS, Johnston E, et al. Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa*. *The European respiratory journal* 2017;50(1) doi: 10.1183/13993003.01569-2016 [published Online First: 2017/07/07]
287. Wright FA, Strug LJ, Doshi VK, et al. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. *Nature genetics* 2011;43(6):539-46. doi: 10.1038/ng.838 [published Online First: 2011/05/24]
288. National Children's Research Centre. SHIELD CF (The Study of Host Immunity and Early Lung Disease in Cystic Fibrosis) Dublin, Ireland [Available from: <https://www.nationalchildrensresearchcentre.ie/research-innovation/research-areas/immunity-infection/cystic-fibrosis/shield-cf/>].
289. Cystic Fibrosis Foundation. Drug Development Pipeline 2020 [Available from: <https://www.cff.org/trials/pipeline> accessed 26/11/2020 2020.
290. Liu X, Ren J, Luo N, et al. Comprehensive DNA methylation analysis of tissue of origin of plasma cell-free DNA by methylated CpG tandem amplification and sequencing (MCTA-Seq). *Clin Epigenetics* 2019;11(1):93. doi: 10.1186/s13148-019-0689-y [published Online First: 2019/06/27]
291. Liu M, Oxnard G, Klein E, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology* 2020
292. Chambers DC, Carew AM, Lukowski SW, et al. Transcriptomics and single-cell RNA-sequencing. *Respirology (Carlton, Vic)* 2019;24(1):29-36. doi: 10.1111/resp.13412 [published Online First: 2018/09/29]
293. The Human Protein Atlas. HMOX2 2020 [Available from: <https://www.proteinatlas.org/ENSG00000103415-HMOX2/antibody#PA> accessed 26/11/2020 2020.
294. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials. Evolution of Translational Omics: Lessons Learned and the Path Forward Washington (DC), USA: National Academies Press (US); 2012 [Available from: <https://www.ncbi.nlm.nih.gov/books/NBK202165/> accessed 26/11/2020 2020.

295. Sun YV, Hu YJ. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Advances in genetics* 2016;93:147-90. doi: 10.1016/bs.adgen.2015.11.004 [published Online First: 2016/02/27]
296. Graustein AD, Berrington WR, Buckingham KJ, et al. Inflammasome Genetic Variants, Macrophage Function, and Clinical Outcomes in Cystic Fibrosis. *American journal of respiratory cell and molecular biology* 2021 doi: 10.1165/rcmb.2020-0257OC [published Online First: 2021/04/14]
297. Padilla-Martínez F, Collin F, Kwasniewski M, et al. Systematic Review of Polygenic Risk Scores for Type 1 and Type 2 Diabetes. *Int J Mol Sci* 2020;21(5) doi: 10.3390/ijms21051703 [published Online First: 2020/03/07]
298. Corvol H, Blackman SM, Boëlle PY, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nature communications* 2015;6:8382. doi: 10.1038/ncomms9382 [published Online First: 2015/09/30]
299. García-Giménez JL, Seco-Cervera M, Tollefsbol TO, et al. Epigenetic biomarkers: Current strategies and future challenges for their use in the clinical laboratory. *Critical reviews in clinical laboratory sciences* 2017;54(7-8):529-50. doi: 10.1080/10408363.2017.1410520 [published Online First: 2017/12/12]
300. Jansen MP, Knijnenburg T, Reijm EA, et al. Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. *Cancer research* 2013;73(22):6632-41. doi: 10.1158/0008-5472.can-13-0704 [published Online First: 2013/11/19]
301. Gilad S, Lithwick-Yanai G, Barshack I, et al. Classification of the four main types of lung cancer using a microRNA-based diagnostic assay. *The Journal of molecular diagnostics : JMD* 2012;14(5):510-7. doi: 10.1016/j.jmoldx.2012.03.004 [published Online First: 2012/07/04]
302. Dirks RA, Stunnenberg HG, Marks H. Genome-wide epigenomic profiling for biomarker discovery. *Clin Epigenetics* 2016;8:122. doi: 10.1186/s13148-016-0284-4 [published Online First: 2016/11/30]
303. Cystic Fibrosis Foundation. CF Foundation Biorepository 2020 [Available from: <https://www.cff.org/Research/Researcher-Resources/Tools-and-Resources/CF-Foundation-Biorepository/> accessed 26/11/2020 2020.
304. Plasschaert LW, Žilionis R, Choo-Wing R, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018;560(7718):377-81. doi: 10.1038/s41586-018-0394-6 [published Online First: 2018/08/03]
305. Montoro DT, Haber AL, Biton M, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 2018;560(7718):319-24. doi: 10.1038/s41586-018-0393-7 [published Online First: 2018/08/03]
306. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nature medicine* 2020:1-3.
307. Rilinger J, Kern WV, Duerschmied D, et al. A prospective, randomised, double blind placebo-controlled trial to evaluate the efficacy and safety of tocilizumab in patients with severe COVID-19 pneumonia (TOC-COVID): A structured summary of a study protocol for a randomised controlled trial. *Trials* 2020;21(1):470. doi: 10.1186/s13063-020-04447-3 [published Online First: 2020/06/05]

308. Schupp JC, Khanal S, Gomez JL, et al. Single Cell Transcriptional Archetypes of Airway Inflammation in Cystic Fibrosis. *American journal of respiratory and critical care medicine*;0(ja):null. doi: 10.1164/rccm.202004-0991OC
309. Montgomery ST, Dittrich AS, Garratt LW, et al. Interleukin-1 is associated with inflammation and structural lung disease in young children with cystic fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* 2018;17(6):715-22. doi: 10.1016/j.jcf.2018.05.006 [published Online First: 2018/06/10]
310. Esther CR, Jr., Muhlebach MS, Ehre C, et al. Mucus accumulation in the lungs precedes structural changes and infection in children with cystic fibrosis. *Science translational medicine* 2019;11(486) doi: 10.1126/scitranslmed.aav3488 [published Online First: 2019/04/05]
311. Esther CR, Jr., Hill DB, Button B, et al. Sialic acid-to-urea ratio as a measure of airway surface hydration. *American journal of physiology Lung cellular and molecular physiology* 2017;312(3):L398-L404. doi: 10.1152/ajplung.00398.2016 [published Online First: 2017/01/08]
312. Wheeler HE, Aquino-Michaels K, Gamazon ER, et al. Poly-omic prediction of complex traits: OmicKriging. *Genetic epidemiology* 2014;38(5):402-15. doi: 10.1002/gepi.21808 [published Online First: 2014/05/07]
313. Amiri Roudbar M, Mohammadabadi MR, Ayatollahi Mehrgardi A, et al. Integration of single nucleotide variants and whole-genome DNA methylation profiles for classification of rheumatoid arthritis cases from controls. *Heredity* 2020;124(5):658-74. doi: 10.1038/s41437-020-0301-4 [published Online First: 2020/03/05]
314. Vazquez AI, Veturi Y, Behring M, et al. Increased Proportion of Variance Explained and Prediction Accuracy of Survival of Breast Cancer Patients with Use of Whole-Genome Multiomic Profiles. *Genetics* 2016;203(3):1425-38. doi: 10.1534/genetics.115.185181 [published Online First: 2016/05/01]
315. Davies JC, Cunningham S, Harris WT, et al. Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-5 years with cystic fibrosis and a CFTR gating mutation (KIWI): an open-label, single-arm study. *The Lancet Respiratory medicine* 2016;4(2):107-15. doi: 10.1016/s2213-2600(15)00545-7 [published Online First: 2016/01/25]
316. Davies JC, Wainwright CE, Sawicki GS, et al. Ivacaftor in Infants Aged 4 to <12 Months With Cystic Fibrosis and a Gating Mutation: Results of a 2-Part Phase 3 Clinical Trial. *American journal of respiratory and critical care medicine* 2020 doi: 10.1164/rccm.202008-3177OC [published Online First: 2020/10/08]
317. Rosenfeld M, Wainwright CE, Higgins M, et al. Ivacaftor treatment of cystic fibrosis in children aged 12 to <24 months and with a CFTR gating mutation (ARRIVAL): a phase 3 single-arm study. *The Lancet Respiratory medicine* 2018;6(7):545-53. doi: 10.1016/s2213-2600(18)30202-9 [published Online First: 2018/06/11]
318. Heijerman HGM, McKone EF, Downey DG, et al. Efficacy and safety of the elexacaftor plus tezacaftor plus ivacaftor combination regimen in people with cystic fibrosis homozygous for the F508del mutation: a double-blind, randomised, phase 3 trial. *Lancet (London, England)* 2019;394(10212):1940-48. doi: 10.1016/s0140-6736(19)32597-8 [published Online First: 2019/11/05]
319. Middleton PG, Mall MA, Dřevínek P, et al. Elexacaftor-Tezacaftor-Ivacaftor for Cystic Fibrosis with a Single Phe508del Allele. *The New England journal of*

- medicine* 2019;381(19):1809-19. doi: 10.1056/NEJMoa1908639 [published Online First: 2019/11/08]
320. Corvol H, Mésinè J, Douksieh IH, et al. SLC26A9 Gene Is Associated With Lung Function Response to Ivacaftor in Patients With Cystic Fibrosis. *Frontiers in pharmacology* 2018;9:828. doi: 10.3389/fphar.2018.00828 [published Online First: 2018/08/25]
321. Strug LJ, Gonska T, He G, et al. Cystic fibrosis gene modifier SLC26A9 modulates airway response to CFTR-directed therapeutics. *Hum Mol Genet* 2016;25(20):4590-600. doi: 10.1093/hmg/ddw290 [published Online First: 2017/02/09]
322. Cystic Fibrosis Foundation. Our Strategic Plan: Care 2020 [cited 2020 26/11/2020]. Available from: <https://www.cff.org/About-Us/Reports-and-Financials/Our-Strategic-Plan/Care/>.