

Defining COPD and severe (early onset) COPD using a multi-omics approach.

by Johannes L. L. van Nijnatten

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of:

UTS: Dr Alen Faiz

Dist. Prof Brian Oliver

RUG: Dr Corry-Anke Brandsma

Prof Maarten van den Berge

Prof Wim Timens

University of Technology Sydney

Faculty of Science

January 2025

Certificate of Original Authorship

I, Johannes (Jos) L. L. van Nijnatten, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the School of Life Science, Faculty of Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with the University of Groningen.

This research is supported by the Australian Government Research Training Program.

Format of Thesis Statement

I hereby state that this is a submission of a thesis by compilation.

Publications

Generated from this work

The following publications contribute directly to the understanding and arguments presented in this thesis.

- **Jos van Nijnatten***, Alen Faiz, Wim Timens, Victor Guryev, Dirk-Jan Slebos, Karin Klooster, Jorine Hartman, Tessa Kole², David F. Choy, Michele Grimbaldston, Carrie Rosenberger, Genentech partners, Huib Kerstjens, Corry-Anke Brandsma[†], Maarten van den Berge[†]. 2023. **Unique differentially expressed genes in Bronchial Brushes in severe COPD and a potential biomarker in the nose**. European Respiratory Journal.
- **Jos van Nijnatten***, Corry-Anke Brandsma, Katrina Steiling, Pieter S. Hiemstra, Wim Timens, Maarten van den Berge[†], Alen Faiz[†]. 2022. **High miR203a-3p and miR-375 expression in the airways of smokers with and without COPD**. Scientific Reports

Produced adjunct to the current work

The publications listed below do not directly contribute to the current work but are included as evidence of research productivity and skill development.

- Nicolaas J. Bekker*, Alienke van Pijkeren, Justina C. Wolters, Alejandro Sánchez Brotons, Victor Guryev, Roy R. Woldhuis, Rainer Bischoff, Wynand Alkema, Peter Horvatovich, **Jos van Nijnatten**, Maarten van den Berge, Wim Timens, Corry-Anke Brandsma[†]. (2023) **A proteomics approach to identify COPD-related changes in lung fibroblasts**. American Journal of Physiology - Lung cellular and molecular physiology.
- Senani N.H. Rathnayake*, Benedikt Ditz*, **Jos van Nijnatten**, Tayyaba Sadaf, Philip M. Hansbro, Corry A. Brandsma, Wim Timens, Annemarie van Schadewijk, Peter S. Hiemstra, Nick H.T. ten Hacken, Brian Oliver, Huib A.M. Kerstjens, Maarten van den Berge[†], Alen Faiz[†]. (2022) **Smoking induces shifts in cellular composition and transcriptome within the bronchial mucus barrier**. Respiriology
- Elise M. A. Slob*, Alen Faiz*, **Jos van Nijnatten**, Susanne J. H. Vijverberg, Cristina Longo, Merve Kutlu, Fook Tim Chew, Yang Yie Sio, Esther Herrera-Luis, Antonio Espuela-Ortiz, Javier Perez-Garcia, Maria Pino-Yanes, Esteban G. Burchard, Uroš Potočnik, Mario Gorenjak, Colin Palmer, Cyrielle Maroteau, Steve Turner, Katia Verhamme, Leila Karimi, Somnath Mukhopadhyay, Wim Timens, Pieter S. Hiemstra, Mariëlle W. Pijnenburg, Margaret Neighbors, Michele A. Grimbaldston, Gaik W. Tew, Corry A. Brandsma, Vojko Berce, Hananeh Aliee, Fabian Theis, Don D. Sin, Xuan Li, Maarten van den Berge, Anke H. Maitland-van der Zee, Gerard H. Koppelman[†]. (2022) **Association of bronchial steroid inducible methylation quantitative trait loci with asthma and chronic obstructive pulmonary disease treatment response**. Clinical and Translational Allergy
- Mirjam P. Roffel*, Ilse M. Boudewijn, **Jos van Nijnatten**, Alen Faiz, Corneel J. Vermeulen, Antoon J. van Oosterhout, Karen Affleck, Wim Timens, Ken R. Bracke, Tania Maes, Irene H. Heijink, Corry-Anke Brandsma[†], Maarten van den Berge[†]. (2022) **Identification of asthma-associated microRNAs in bronchial biopsies**. European Respiratory Journal, 59: 2101294
- Hananeh Aliee*, Florian Massip*, Cancan Qi*, Maria Stella de Biase*, **Jos van Nijnatten***, Elin T. G. Kersten*, Nazanin Z. Kermani*, Basil Khuder*, Judith M. Vonk, Roel C. H. Vermeulen, U-BIOPRED study group, Cambridge Lung Cancer Early Detection Programme, INER-Ciencias Mexican Lung Program, Margaret Neighbors, Gaik W. Tew, Michele A. Grimbaldston, Nick H. T. ten Hacken, Sile

Hu, Yike Guo, Xiaoyu Zhang, Kai Sun, Pieter S. Hiemstra, Bruce A. Ponder, Mika J. Mäkelä, Kristiina Malmström, Robert C Rintoul†, Paul A. Reyfman†, Fabian J. Theist†, Corry-Anke Brandsma†, Ian M. Adcock†, Wim Timens†, Cheng-Jian Xu†, Maarten van den Berge†, Roland F. Schwarz†, Gerard H. Koppelman†, M.C. Nawijn†, Alen Faiz†. (2021) **Determinants of expression of SARS-CoV-2 entry-related genes in upper and lower airways.** Allergy

* and † = equal contribution

Presentations and Posters

Oral Presentations

1. **Jos van Nijnatten***, Alen Faiz, Pieter S. Hiemstra, Wim Timens, Cornelis Joseph Vermeulen, Maaïke de Vries, Rosa Faner, Maarten van den Berge[†], Corry-Anke Brandsma[†]. (2023)
Differential DNA Methylation Associated with COPD and COPD Severity. TSANZSRS, Christchurch, NZ.
2. **Jos van Nijnatten***, Alen Faiz, Wim Timens, Victor Guryev, Dirk-Jan Slebos, Karin Klooster, Jorine E. Hartman, Tessa Kole, David F. Choy, Arindam Chakrabarti, Michele Grimaldeston, Carrie M. Rosenberger, Huib Kerstjens, Corry-Anke Brandsma[†] and Maarten van den Berge[†]. (2022)
- 3.

4. **A bronchial gene signature for severe COPD that is retained in** the nose. NAME, Sydney, AU.
5. **Jos van Nijnatten***, Alen Faiz, Wim Timens, Victor Guryev, Dirk-Jan Slebos, Karin Klooster, Jorine E. Hartman, Tessa Kole, David F. Choy, Arindam Chakrabarti, Michele Grimaldeston, Carrie M. Rosenberger, Huib Kerstjens, Corry-Anke Brandsma[†] and Maarten van den Berge[†]. (2022)
- 6.

7. **A bronchial gene signature for severe COPD that is retained in the nose.** TSANZ, Sydney, AU.
8. Jos van Nijnatten*, Kathy Phung, Maarten van den Berge, Corry-Anke Brandsma, Wim Timens[†], Alen Faiz[†]. (2019) **ICS use in COPD changes DNA methylation and targets FKBP5 methylation and expression..** NAME, Sydney, AU.
9. Jos van Nijnatten*, Kathy Phung, Maarten van den Berge, Corry-Anke Brandsma, Wim Timens[†], Alen Faiz[†]. (2019) **ICS use in COPD changes DNA methylation and targets FKBP5 methylation and expression..** TSANZ, Sydney, AU.

Poster Presentations

1. Jos van Nijnatten*, Alen Faiz, Wim Timens, Cornelis Joseph Vermeulen, Maaïke de Vries, Rosa Faner, Maarten van den Berge[†], Corry-Anke Brandsma[†]. 2023. **Differential methylation in COPD bronchial biopsies.** ERS, Milan, IT.
2. Jos van Nijnatten*, Alen Faiz, Wim Timens, Victor Guryev, Dirk-Jan Slebos, Karin Klooster, Jorine E. Hartman, Tessa Kole, David F. Choy, Arindam Chakrabarti, Michele Grimaldeston, Carrie M. Rosenberger, Huib Kerstjens, Corry-Anke Brandsma[†] and Maarten van den Berge[†]. 2022.
- 3.
4. **A bronchial gene signature for severe COPD that is retained in the nose.** ERS, Sydney, AU.
5. Nicolaas J. Bekker, Jos van Nijnatten*, Roy R. Woldhuis, Wierd Kooistra, Wim Timens, Maarten van den Berge, Alen Faiz[†], Corry-Anke Brandsma[†]. 2022. **Differential miRNA expression in SEO-COPD is associated with altered lysosome, vesicular transport and ECM-related pathways.** UMCG/Griac, Groningen, NL.
6. Jos van Nijnatten*, Corry-Anke Brandsma, Pieter Hiemstra, Wim Timens, Maarten Van Den Berge, Alen Faiz[†]. 2020. **Changes in DNA methylation after corticosteroids treatment in COPD patients.** ERS, Online.

Table of Contents

Certificate of Original Authorship	II
Format of Thesis Statement	III
Publications	0.4
Presentations and Posters	0.6
Table of Contents	IX
List of Figures and Tables	X
List of Abbreviations	XIII
Chapter 1 – General Introduction	1.16
Chapter 2 – A bronchial gene signature for severe COPD that is retained in the nose	2.31
Chapter 3 – A lung tissue gene expression profile of severe early onset COPD	3.55
Chapter 4 – Differential miRNA expression in SEO-COPD is associated with altered lysosome, vesicular transport and ECM-related pathways	4.83
Chapter 5 – High miR203a-3p and miR-375 expression in the airways of smokers with and without COPD	5.103
Chapter 6 – Elucidating the association between DNA methylation and COPD in bronchial biopsies	6.124
Chapter 7 – ICS use in COPD changes DNA methylation and targets FKBP5 methylation and expression	7.143
Chapter 8 – Summary	8.160
Chapter 9 – General Discussion	9.165
Chapter 10 – Apendices	174

List of Figures and Tables

Figures

Figure 1.1) Pathophysiological changes in COPD.	1.17
Figure 1.2) GR activation and the interaction with FKBP5.	1.20
Figure 1.3) Function of miRNA in transcriptional regulation.....	1.21
Figure 1.4) The location and type of sample for every type of study that is used in this thesis.	1.24
Figure 2.1) Differential gene expression of bronchial brushes of severe COPD.....	2.37
Figure 2.2) Cellular deconvolution and cell proportions	2.38
Figure 2.3) Reflection of the bronchial severe COPD (sCOPD) signature in the nose	2.40
Figure 2.4) StringDB analysis of the genes unique for severe COPD (sCOPD).....	2.42
Figure 3.1) Identification of a gene signature for SEO-COPD.	3.60
Figure 3.2) Visualization of the gene set enrichment analysis (GSEA) on the SEO-COPD	3.62
Figure 3.3) Visualization of the 18 SEO-COPD leading edge genes	3.62
Figure 3.4) Proportions of cell types calculated using cellular deconvolution	3.63
Figure 3.5) Gene expression of MS4A14 and TNFAIP8L3 associated with genetic variants.....	3.67
Figure 3.6) Immunohistochemical quantification of CD79A in lung tissue	3.69
Figure 4.1) Differential miRNA expression in peripheral lung tissue.....	4.88
Figure 4.2) Visualizations of significant inverse correlations	4.91
Figure 4.3) Pathway enrichment analysis of inversely correlated miR-193b-3p and miR-331-3p	4.92
Figure 4.4) Unbiased pathway enrichment analysis of all significant inversed associations	4.94
Figure 4.5) Associations between predicted lung tissue cell proportions and miRNA expression....	4.95
Figure 5.1) Flow diagram of the study approach.	5.107
Figure 5.2) Differentially expressed microRNAs (miRNAs).....	5.109
Figure 5.3) Significant pathway negatively enrichment by miRNA-203a-3p and miR-375.....	5.112
Figure 6.1) Flow chart of the statistical methods.....	6.127
Figure 6.2) Differential Methylation comparing COPD and non-COPD.	6.129
Figure 6.3) Differential Methylation associated with FEV ₁ % pred (A-C) and RV/TLC % pred (D-F).6.130	
Figure 6.4) [A] Cell proportions calculated by cellular deconvolution stratified by COPD status ...	6.134
Figure 6.5) Protein interaction network using String-DB of all eQTM associated with RV/TLC % pred.	6.135
Figure 7.1) Outline of study methods.....	7.146
Figure 7.2) Differential methylation results.	7.150
Figure 7.3) cg23416081 located in a GR binding site.	7.152
Figure 7.4) The second highest eQTM is cg23416081 and FKBP5.....	7.153
Figure 7.5) A) and B) FKBP5 expression following targeted methylation of cg23416081	7.154
Figure 8.1) Summary of the main results per chapter.	8.161
Figure 9.1) The activated feedback loop after hypomethylation of cg23416081 utilises FKBP5 to block transduction of activated GR.....	9.169

Tables

Table 1.1) GOLD classification of COPD based on FEV ₁ as measure of airflow limitation.	1.16
Table 2.1) Patient demographics.....	2.35
Table 2.2) Top 20 genes more and less expressed in severe COPD (sCOPD)	2.39
Table 3.1) Clinical characteristics of study subjects	3.59
Table 3.2) Top 10 higher/lower expressed genes in the SEO-COPD gene signature.....	3.60
Table 3.3) Known COPD- and lung function-associated single nucleotide polymorphisms	3.66

Table 3.4) SEO-COPD profile genes with expression quantitative trait loci	3.67
Table 4.1) Subject clinical characteristics	4.86
Table 4.2) Top 20 target genes inversely correlated with the differential SEO-COPD miRNAs.....	4.89
Table 4.3) Specific pathways in enriched pathway clusters	4.93
Table 5.1) Patient demographics in relation to smoking status.	5.107
Table 5.2) Differential expression of miRNAs for respiratory healthy controls and subjects with COPD when comparing smoking status.....	5.107
Table 5.3) Selective literature of predicted target genes that were negatively correlated to miR-203a- 3p, miR-375 and miR-31-3p, whilst significant after our meta-analysis (miR-203a-3p and miR-375) or FDR significant (miR-31-3p) after correlation.	5.110
Table 5.4) Significant gene sets negatively affected by the identified miRNAs in COPD patients and asymptomatic participants. Only the top pathways with a family-wise error rate < 0.05 in the COPD patients are shown.....	5.113
Table 6.1) COPD and non-COPD patient demographics.	6.128
Table 6.2) Top differentially methylated sites associated with COPD, lung function and hyperinflation.....	6.131
Table 6.3) Top 25 eQTM's associated with lung function and hyperinflation.....	6.132
Table 6.4) Enriched pathways associated with eQTM's and methylation in hyperinflation.	6.136
Table 7.1) Guide RNA sequences.	7.147
Table 7.2) Patients demographics.	7.148
Table 7.3) Top 20 differentially methylated CpG sites after six months of treatment.....	7.151
Table 7.4) Top 20 differentially methylated CpG sites after six months of treatment.....	7.151
Table 7.5) Top 15 eQTM's identified from a genome-wide eQTM analysis	7.152
Table 7.6) Pathway analysis of eQTM genes.	7.153

Supplementary Figures

Supplementary Figure 2.1) Flowchart of statistical methods.....	2.45
Supplementary Figure 2.2) Heatmap of differential gene expression between non-COPD vs sCOPD	2.46
Supplementary Figure 2.3) Heatmap of differential gene expression between non-COPD vs mCOPD	2.47
Supplementary Figure 2.4) Heatmap of differential gene expression between mCOPD vs sCOPD ..	2.48
Supplementary Figure 2.5) Spearman correlation estimate and p-value of the GSVA score.....	2.49
Supplementary Figure 2.6) Gene expression of FN1 and VEGFA	2.50
Supplementary Figure 2.7) Gene expression of FN1 and VEGFA	2.51
Supplementary Figure 3.1) MS4A14_LD_block_plot.....	3.71
Supplementary Figure 3.2) TNFAIP8L3_LD_block_plot.....	3.72
Supplementary Figure 4.1) Residual data points of differential SEO-COPD miRNAs	4.98
Supplementary Figure 7.1) Design of gRNA sequences on Benchling.....	7.157

Supplementary Tables

Supplementary Table 3.1) Clinical characteristics of eQTL study subjects.....	3.72
Supplementary Table 3.2) SEO-COPD gene profile.	3.72
Supplementary Table 3.3) GSEA pathway analysis.	3.74
Supplementary Table 3.4) Leading edge genes of the 25 pathways.....	3.76
Supplementary Table 3.5) Cell x gene heatmap.....	3.77

Supplementary Table 3.6) List of eQTLs, 15 top.....	3.77
Supplementary Table 3.7) MS4A14 LD Matrix	3.78
Supplementary Table 3.8) TNFAIP8L3 LD Matrix	3.78
Supplementary Table 4.1) eQTM of the three miRNAs	4.98
Supplementary Table 5.1) Correlations of miR-203q-3p and miR-375 to genes in both the COPD and respiratory healthy dataset.	5.117
Supplementary Table 5.2) Significance of enrichment scores for single-cell signatures in named cell types.....	5.117
Supplementary Table 5.3) Enrichment scores of single-cell signatures	5.118
Supplementary Table 5.4) Differential expression analysis of miRNAs in smokers	5.119
Supplementary Table 5.5) Differential expression analysis of miRNAs in smokers	5.119
Supplementary Table 6.1) Validation of the CpG sites associated with COPD and lung function...	6.138
Supplementary Table 6.2) Associations between lung function measurements and cell proportions.	6.138
Supplementary Table 6.3) Associations of the eQTMs (associated with hyperinflation) with cell proportions.	6.139
Supplementary Table 7.1) Top 15 eQTMS identified from a genome-wide eQTM analysis	7.157

List of Abbreviations

AATD	Alpha-1-Anti Trypsin Deficiency
ABCG1	ATP Binding Cassette Subfamily G Member 1
ABHD6	Abhydrolase domain containing 6, acylglycerol lipase
ADAMTS15	ADAM metalloproteinase with thrombospondin type 1 motif 15
AMP	Adenosine Mono-Phosphate
BCL2L11	BCL2 like 11
cAMP	Cyclic AMP
cDNA	Complementary DNA
ChIP	Chromatin Immuno-Precipitation
CLMP	CXADR-like membrane protein
CLR	C-type Lectin Receptor
COPD	Chronic Obstructive Pulmonary Disease
CpG	(Shorthand for) 5'—C—phosphate—G—3'
CT	Computed Tomography
CXCL12	C-X-C motif chemokine ligand 12
DE	Differentially Expressed
DICER	Dicer 1, Ribonuclease III
DMP	Differentially Methylated Site
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
DROSHA	Drosha Ribonuclease III
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor
EMILIN1	Elastin Microfibril Interfacer 1
EQTL	Expression Quantitative Trait Loci
eQTL	Expression Wuantitative Trait Loci
FBLN1	Fibulin-1
FBN1	Fibrillin 1
FDR	False Discovery Rate
FEV ₁ % pred	FEV ₁ percentage of predicted
FEV ₁	Forced Expiratory Volume in one second
FKBP4	FK506-Binding Protein 4
FKBP5	FK506-Binding Protein 5
FPR	Formyl-Peptide Receptor
FRZB	Frizzled Related Protein
FVC	Forced Vital Capacity
GATA6	GATA binding protein 6
GDPR	General Data Protection Regulation
GGA2	Golgi Associated, Gamma Adaptin Ear Containing, ARF Binding Protein 2
GGT5	Gamma-glutamyltransferase 5
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GR	Glucocorticoid Receptor
GRE	Glucocorticoid Response Element
GRIP-1	Glucocorticoid Receptor-Interacting Protein-1
GSAMD	Infinium Global Screening Array + Multi Disease
GSEA	Gene Set Enrichment Analysis
GWAS	Genome-Wide Association Study
HLCA	Human Lung Cell Atlas
HRC	Human Reference Genome

HSP	Heat Shock Proteins
IBM	International Business Machines Corporation
ICS	Inhaled Cortico-Steroids
IGHV1-18	Immunoglobulin Heavy Variable 1-18
IGLV2-11	Immunoglobulin Lambda Variable 2-11
IGLV2-14	Immunoglobulin Lambda Variable 2-14
IHC	Immuno-Histo-Chemical
IKZF3	IKAROS Family Zinc Finger 3
IL	Interleukin
ILD	Interstitial Lung Disease
ITGA7	Integrin subunit alpha 7
KRT4	Keratin 4 3.48,
LABA	Long-Acting beta2-Agonist
LAMA	Long-Acting Muscarinic Antagonist
LARP4B	La ribonucleoprotein 4B
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MAPK	Mitogen-Activated Protein Kinase
miRNA	MicroRNA
MMP	Matrix Metallo-Proteinase
MR	Mineralocorticoid Receptor
mRNA	Messenger RNA
MS4A14	Membrane Spanning 4-Domains A14
MSigDB	Molecular Signatures Database
NF- κ B	Nuclear Factor kappa beta
NLR	NOD-Like Receptor
NLS	nuclear localisation signals
NPC	Nuclear Pore Complex
NR1D1	Nuclear Receptor Subfamily 1 Group D Member 1
PAQR8	Progesterin and adipoQ receptor family member 8
PDGF	Platelet-Derived Growth Factor
PDGFRB	Platelet-derived growth factor receptor beta
PLPP1	Phospholipid Phosphatase 1
PODN	Podocan
PRDM1	PR Domain Zinc Finger Protein 1
PRESTO	PRoteogenomics Early onSeT cOPd lung cohort
PRR	Pattern Recognition Receptor
PTX3	Pentraxin 3
QC	Quality Check
RHOH	Ras homolog family member H
RISC	RNA-Induced Silencing Complex
RLR	RIG-I-Like Receptor
RNA	Ribonucleic Acid
RNA-Seq	RNA sequencing
ROR2	Receptor tyrosine kinase-like orphan receptor 2
RORC	RAR-related orphan receptor C
SABA	Short-Acting beta2-Agonist
SAMA	Short-Acting Muscarinic Antagonist
SD	Standard Deviation
SEO	Severe-Early Onset
SERPINA1	Serpin Family A Member 1

SERPINE2	Serpin Family E Member 2
SGTA	Small Glutamine Rich Tetratricopeptide Repeat Co-Chaperone Alpha
SLC7A7	Solute Carrier Family 7 Member 7
SMOC2	SPARC Related Modular Calcium Binding 2
SNP	Single Nucleotide Polymorphism
SPON1	Spondin-1
SPSS	Statistical Product and Service Solutions
TBC1D2	TBC1 domain family member 2
TERT	Telomerase Reverse Transcriptase
TGF- β	Transforming Growth Factor-beta
TIAM1	T-Lymphoma Invasion And Metastasis-Inducing Protein 1
TLR	Toll-Like Receptor
TNFAIP8L3	TNF Alpha Induced Protein 8 Like 3
TPPP	Tubulin polymerization promoting protein
TTC7A	Tetratricopeptide repeat domain 7A
UAVG	Uitvoeringswet Algemene verordening gegevensbescherming (Dutch law)
UMCG	University Medical Center Groningen
WGBO	Medical Treatment Agreement Act (Dutch law)
ZNF608	Zinc finger protein 608

Chapter 1 – General Introduction

1.1 – Chronic Obstructive Pulmonary Disease (COPD)

COPD is a common pulmonary disease that can result in poor quality of life, and in severe cases, also death. Symptoms of COPD include wheezing, shortness of breath, chest tightness and fatigue, excessive mucus production, cough, and frequent respiratory infections¹. In addition, COPD is associated with weight loss, muscle weakness, and (cardiovascular) comorbidities²⁻⁴.

COPD is characterized by chronic airflow obstruction defined as a post-bronchodilator forced expiratory volume in one second (FEV_1)/ forced vital capacity (FVC) ratio of < 0.7 , also known as Tiffeneau-Index⁵. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) categorised the severity of COPD according to the level of FEV_1 % of predicted (FEV_1 % pred, Table 1.1)^{6, 7}. There is wide heterogeneity between subjects with respect to susceptibility to cigarette smoke^{8, 9}. Some maintain a normal lung function despite a very high number of packyears smoking, whereas others already develop quite severe COPD at an early age with relatively few packyears of smoking. The term Severe-Early Onset (SEO-)COPD is referred to when subjects develop severe airflow obstruction with FEV_1 below 40% of predicted and age below 53-55 years¹⁰⁻¹².

Table 1.1) GOLD classification of COPD based on FEV_1 as measure of airflow limitation. (With always $FEV_1/FVC < 0.7$). Table adapted from “GOLD - Global initiative for chronic obstructive lung disease” (2022).

Gold stage	Common name	Definition based on spirometry
GOLD 1	Mild	$FEV_1 \geq 80$ % predicted
GOLD 2	Moderate	$50 \% \leq FEV_1 < 80$ % predicted
GOLD 3	Severe	$30\% \leq FEV_1 < 50$ % predicted
GOLD 4	Very severe	$FEV_1 < 30\%$ predicted

Smoke exposure is the key risk factor for developing COPD⁷ as it is observed that 80-95% of those developing COPD are current- or ex-smokers¹³. On the other hand, only 10-20% of smokers develop COPD¹⁴⁻¹⁶. Other risk factors include dust and chemicals, exposure to air pollution, and indoor cooking using biomass fuel⁷. Along with environmental exposures, genetic susceptibility plays a role in the risk to develop COPD. Although mendelian mutations in SERPINA1, known as alpha-1-anti trypsin deficiency (AATD)¹⁷ have been identified along with other more common genetic variants, a significant part of the genetic susceptibility to develop COPD remains unknown.

1.2 – Pathological changes in COPD

Pathological changes in COPD include chronic inflammation as a general feature that contributes to the main pathological features chronic bronchitis, airway wall remodelling (including small airways disease and airway wall fibrosis) and emphysema.

Chronic bronchitis is characterized by inflammation of the large airways, which can be driven by repeated injury of the airway epithelium by smoking or other irritants. Upon damage of the epithelial cells, inflammatory cytokines are secreted attracting the influx of immune cells, and basal cells are stimulated to differentiate aberrantly into, on one hand, increased goblet cells and, on the other hand, squamous cell metaplasia. Together these epithelial changes lead to goblet cell hyperplasia, increased mucus production, and reduced mucus clearance obstructing the airways^{18, 19}.

In the more peripheral and small airways, the repeated epithelial injury and the release of pro-inflammatory cytokines and growth factors like transforming growth factor-beta (TGF- β), platelet-derived growth factor (PDGF), and epidermal growth factor (EGF) leads to the activation of

fibroblasts and bronchial smooth muscle cells, which in some cases will respond by producing excessive amounts of collagen and other extracellular matrix proteins²⁰⁻²⁴. The build-up of these molecules will result in the local formation of scar tissue resulting in remodelling and fibrosis of the airways. Previous research has shown there is a loss of small airways in COPD, which may be preceded by obstruction²⁵.

Emphysema is characterized by the destruction of alveolar walls, which one on hand leads to loss of surface area for gas exchange, impairing the ability of the lungs to effectively take in oxygen and remove carbon dioxide, leading to decreased oxygen uptake and heightened respiratory effort (dyspnea)²⁶. On the other hand loss of alveolar walls leads to loss of elastic recoil and reduced tethering of the peripheral airways resulting in airway collapse and obstruction, and hyperinflation or air trapping²⁷ (Figure 1.1).

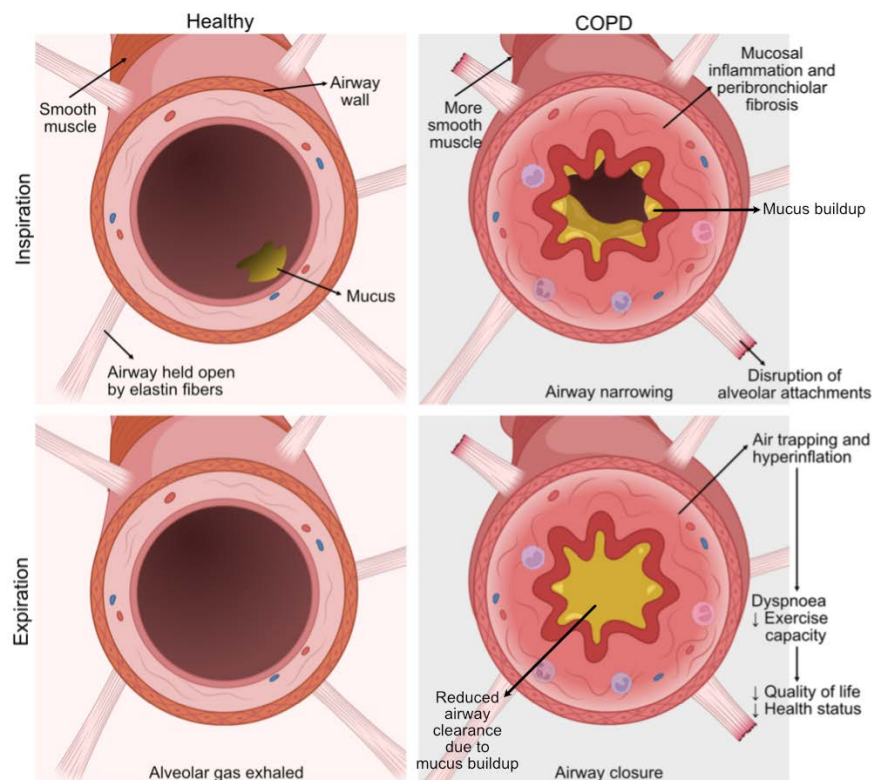


Figure 1.1) Pathophysiological changes in COPD. Healthy individuals can breathe in and out without an obstruction of the airway due to alveolar attachments, such as elastin fibres, that keep the airway open. The airway does not contain a lot of mucus and is not narrowed extensively, causing no obstruction. In individuals suffering from COPD, more smooth muscle mass, mucosal inflammation, fibrosis, and mucus hypersecretion cause narrowing of the airway. During expiration, the airway narrows even more due to the loss of alveolar attachments, and excess mucus can block the airway. Figure adapted from P. Barnes et al. 'Chronic obstructive pulmonary disease' (2015).

1.3 – Chronic inflammation in COPD

Chronic inflammation is an important feature of smoking-induced COPD. Even cigarette-smoking individuals with normal lung function have increased airway inflammation, indicating that inflammation is a usual response of the airway epithelium to smoke²⁸. However, as the disease worsens, the degree of inflammation increases, and more inflammation-related tissue changes occur^{29, 30}.

Smoking exposes the pulmonary system to over 4000 chemicals, toxins, and other pollutants³¹. These irritants activate pattern recognition receptors (PRRs) on airway epithelial cells, such as Toll-

like receptors (TLRs), NOD-like receptors (NLRs), RIG-I-like receptors (RLRs), C-type lectin receptors (CLRs), and formyl-peptide receptors (FPRs)^{32, 33}. For example, activation of TLRs initiates NF-κB and mitogen-activated protein kinase (MAPK) downstream signalling, leading to pro-inflammatory cytokines and type 1 interferon induction³³⁻³⁵. The activation of PRRs will first activate the innate immune response by activating the epithelium and ensuring mucus secretion by secretory cells, such as goblet and club cells³⁶⁻³⁸. It will also lead to the recruitment of neutrophils and macrophages to the lung. Neutrophils secrete proteases, such as elastase, which lead to degradation of peribronchial elastin fibres in the lung parenchyma that normally help to keep the airway open, while other proteases produced in the lungs including matrix metalloproteinases (MMPs), lead to tissue degradation and remodelling³⁹. Later, the adaptive immune system will be triggered, and dendritic cells will present antigens to T and B cells. In more severe COPD, there is an increase of CD4+ T cells in the smaller airways⁴⁰, together with more frequently observed lymphoid aggregates and follicles of B-cells and CD4 T-cells⁴¹⁻⁴⁴. The presence of such follicles has been associated with an autoimmune component in the pathogenesis of COPD, at least in part of the cases^{41, 45}. This all contributes to the main pathological features of COPD and worsening of respiratory symptoms and the progression of COPD.

Smoking cessation reduces, but does not eliminate, airway inflammation⁴⁶. This is in particular observed in COPD as after one year of smoking cessation, airway inflammation persists in bronchial biopsies. However, the number of sputum neutrophils, lymphocytes, interleukin (IL)-8 and eosinophilic-cationic-protein levels even significantly increased at 12 months⁴⁶. In a cross-sectional study, persistent inflammation in COPD, as compared to controls, was observed for even more than 3 years after quitting smoking⁴⁷.

In addition, lower counts of leukocytes in blood have been reported after smoking cessation, and fewer neutrophils and lower or similar levels of macrophages^{48, 49}. However, ex-smokers without COPD have higher counts of natural killer cells compared to never smokers⁵⁰. The fact that the inflammation after smoking cessation is reduced but not eliminated suggests the presence of a prolonged effect of smoking.

Additionally, smoking cessation does not lead to the reduction of smooth muscle mass or fibrosis in both subjects with normal lung function and with COPD, suggesting that these smoking-related changes are irreversible⁵¹. However, previous research has found less goblet cell hyperplasia after smoking cessation and similar levels of squamous cell metaplasia in subjects without COPD⁵¹. Together this indicates that a number of features are reduced during smoking cessation, however inflammation in COPD appears to be retained at some level. Smoking cessation reduces but does not eliminate lung inflammation, improves respiratory symptoms, and it corrects the accelerated decline in lung function to a large extent⁵¹. Despite that, approximately 40% of COPD patients are current smokers⁵².

1.4 – Treatment of COPD

Pharmaceutical treatment of COPD typically includes short- or long-acting beta₂-agonist (SABA or LABA), and short- or long-acting muscarinic antagonist (SAMA or LAMA)⁷. Beta₂-agonists, such as SABA and LABA, stimulate the beta₂ adrenergic receptors, which increases cyclic AMP (cAMP) and produces functional antagonism to bronchoconstriction^{53, 54}. Bronchodilators ease breathing, especially during an exacerbation, by relaxing the smooth muscle and widening the airways, offering symptomatic relief. Muscarinic antagonists, such as SAMA and LAMA, block the bronchoconstrictor effects of acetylcholine on M3 muscarinic receptors expressed on airway smooth muscle cells, and in so also relax the smooth muscle and offer symptomatic relief⁵⁵. Combining bronchodilators with

different mechanisms and durations of action, such as SABA and LAMA, increases symptomatic relief more than a single bronchodilator's increased dosage and decreases side effects. Anti-inflammatory treatment with inhaled corticosteroids (ICS) is frequently prescribed in COPD. ICS reduce exacerbations, improve lung function and quality of life, particularly in those with evidence of eosinophilic inflammation (> 150 eosinophils per microliter). ICS have broad anti-inflammatory effects by reducing the number of mast cells, eosinophils, and T cells⁵⁶ and limit the secretion of pro-inflammatory cytokines by epithelial cells. They also limit the recruitment of inflammatory cells such as eosinophils to the lungs and airways⁵⁶. However not all COPD patients respond effectively to ICS and a subset remain corticosteroid resistant.

Molecular mechanism of ICS and the glucocorticoid receptor (GR)

ICS exert their effects by binding to the glucocorticoid receptor (GR, specifically GR α) and, to a lesser extent, to the mineralocorticoid receptor (MR)^{57, 58}. GR can exist as part of a protein complex, even when not activated. In its inactive state, the GR is typically located in the cytoplasm bound to a multi-protein complex, which includes chaperones (heat shock proteins, E.g. HSP90), co-chaperones (E.g. hsp70-hsp90 organizing protein (Hop)), Small Glutamine Rich Tetratricopeptide Repeat Co-Chaperone Alpha (SGTA), FK506-Binding Protein 4 (FKBP4) and FK506-Binding Protein 5 (FKBP5)⁵⁹. This complex helps to maintain the GR in an inactive conformation and prevents its translocation into the nucleus. Several chaperones and co-chaperones can thereby inhibit, transrepress, or mediate, transactivate, the binding of the activated GR to the Glucocorticoid Response Element (GRE)^{60, 61}. Through transrepression, GR inhibits the activity of key transcription factors, such as NF- κ B and AP-1, which are critical for the expression of pro-inflammatory genes. Transrepression does not require direct binding of GR to DNA⁶². Instead, activated GR interacts with these transcription factors or their co-activators, preventing their translocation to the nucleus or their binding to DNA. GR can also sequester p65⁶³, a subunit of NF- κ B, in the cytoplasm, preventing it from entering the nucleus to initiate the transcription of inflammatory genes. For example, FKBP5 is a negative regulator, a transrepressor, of GR activity by reducing the GR affinity to glucocorticoids (Figure 1.2) by targeting GR for degradation.

Transactivation, however, leads to the transcription of genes by binding to the promotor regions of these target genes. Glucocorticoids transactivate genes by via a cascade of interactions. Upon binding by glucocorticoids, the GR activates, undergoes a structural change, and dissociates from the protein complex, exposing two nuclear localisation signals (NLS), NLS1 and NLS2⁶⁴. NLS1 is responsible for quick nuclear translocation, whereas NLS2 is for slower nuclear translocation⁶⁴. It does this, along with HSP90, by interacting with importin proteins, e.g. importin- α , importin- β and importin-7, which recognise the NLS sequences and form a complex with the GR⁶⁵⁻⁶⁷. This complex then moves through the nuclear pore complex (NPC), which acts as a gateway between the cytoplasm and the nucleus⁶⁸. Once inside the nucleus, the importin proteins release the GR and will be recycled⁶⁹. Activated GR in the nucleus binds to a DNA sequence called the GRE located in the promotor regions of target genes^{70, 71}. This binding leads to the recruitment of co-activators, such as glucocorticoid receptor-interacting protein-1 (GRIP-1), transcription factors, chromatin modulators, and basal transcription factors initiates, transactivating, gene transcription^{70, 71}. Specifically, GR activation has been associated with decreased inflammatory response and pathways. Prolonged activation using ICS can also lead to osteoporosis, diabetes, pneumonia and other adverse effects⁷²⁻⁷⁴.

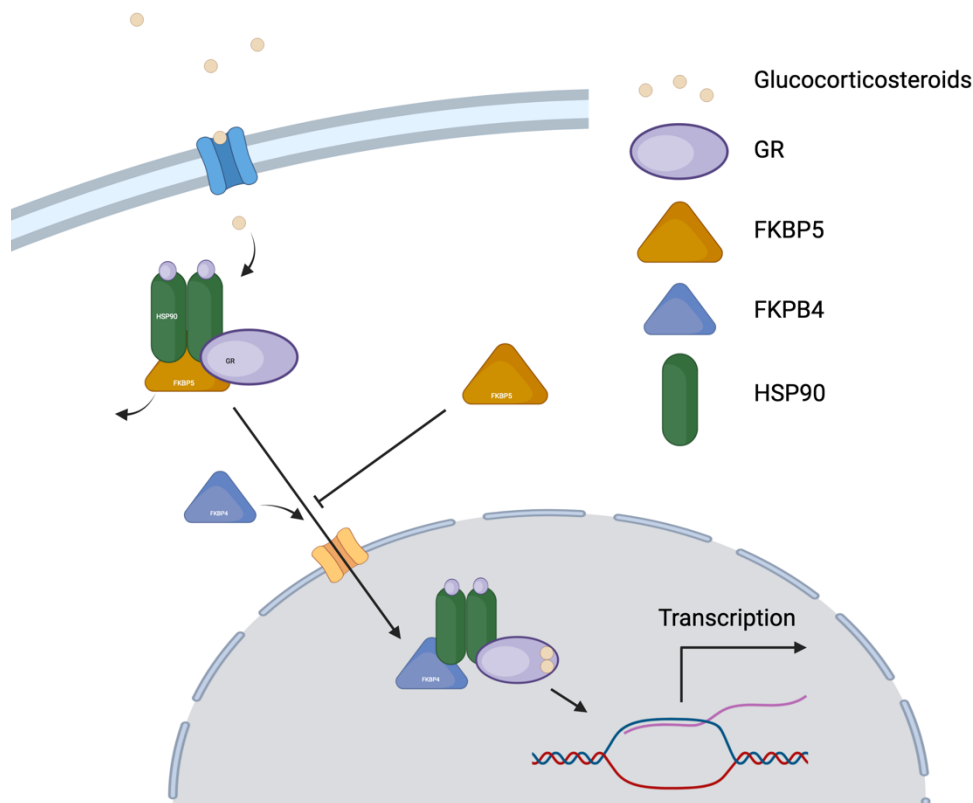


Figure 1.2) GR activation and the interaction with FKBP5. Glucocorticosteroids bind to the GR receptor, detaching FKBP5 from the protein complex and translocating to the nucleus. Here, the activated GR complex, in combination with FKBP4, will attach to a GRE and activate gene transcription. FKBP5, when transcribed, can also act as a negative regulator for gene transcription by blocking the transduction of the activated GR complex.

1.5 – Transcriptomics and the effects of DNA methylation and miRNA expression

The transcriptome is the complete set of RNA molecules and their quantity produced by the cells of an organism at a given state⁷⁵. It includes all the messenger RNA (mRNA) molecules that, after post-transcriptional modification, are translated into proteins⁷⁶. The transcriptome also includes non-coding RNA molecules with various regulatory functions, such as microRNAs (miRNAs) and long-non coding RNAs⁷⁵. The expression of mRNAs is tightly regulated by various factors, including miRNAs. These regulatory mechanisms ensure precise control over protein synthesis, which is crucial for cellular functions, development, and responses to environmental signals. Studying the transcriptome provides insights into gene expression patterns and helps to understand how different genes are activated or suppressed in response to different stimuli in different diseases.

The epigenome is the complete set of marks on the DNA that can influence gene expression without changing the DNA sequence itself⁷⁷. These marks, such as DNA methylation and histone modifications, can affect the accessibility of genes and determine whether they are turned on or off⁷⁷. The epigenome is dynamic and can be influenced by environmental factors, lifestyle choices, and developmental processes, but can also last multiple generations⁷⁸.

DNA Methylation is a process in which a methyl group is added to the DNA-molecule^{79, 80}. This modification can affect gene expression by blocking the binding of transcription factors to the DNA or recruiting proteins that repress gene transcription^{79, 80}. The positions where this can happen is

called a CpG site (shorthand for 5'—C—phosphate—G—3'), where a cytosine is directly followed by a guanine in the sequence of DNA. DNA methylation patterns can be heritable and affect development, aging, and disease⁷⁹.

MicroRNAs (miRNAs) are involved in post-transcriptional regulation of mRNAs and gene translation⁸¹ (Figure 1.3). After being transcribed from DNA, miRNAs undergo processing steps to become mature miRNAs⁸¹. The RNA-induced silencing complex (RISC), consisting of mature miRNA and Argonaute proteins, guides the miRNA to its target mRNA⁸¹. Once bound, the miRNA can trigger mRNA degradation or inhibit translation, depending on the degree of complementarity between the miRNA and mRNA⁸¹. This mechanism allows for gene expression fine-tuning and plays a vital role in various biological processes, including development, differentiation, and disease⁸².

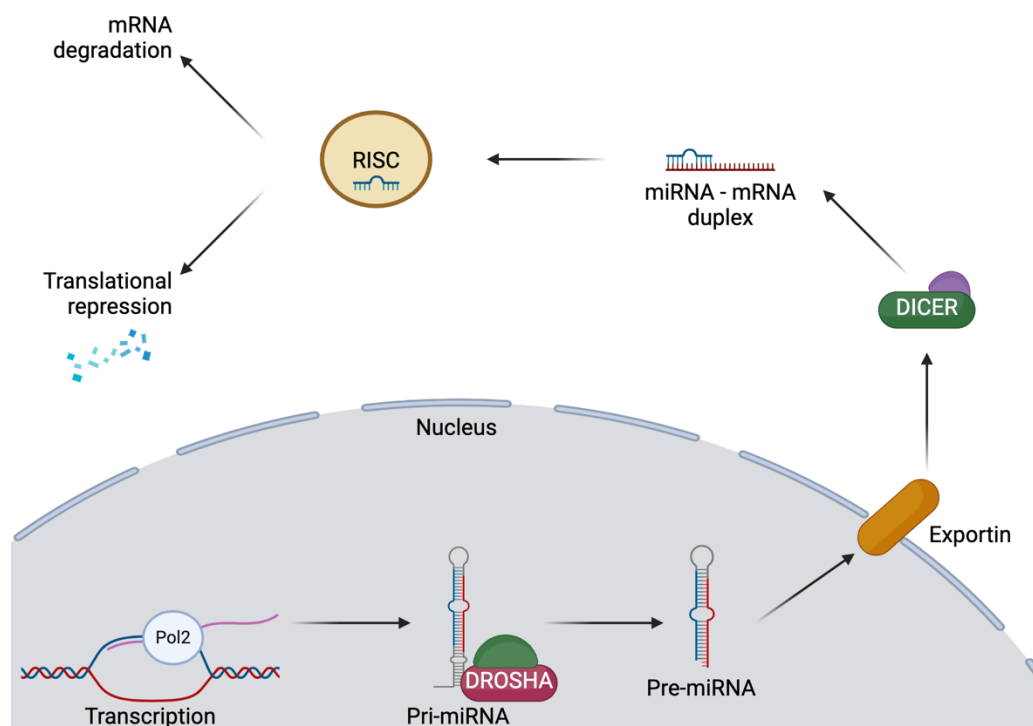


Figure 1.3) Function of miRNA in transcriptional regulation. A miRNA gene gets transcribed and processed via DROSHA and DICER to mature miRNA. This miRNA forms a RISC complex that can degrade mRNA or repress translation.

Understanding the transcriptome can help in complex diseases by identifying key genes and pathways involved in the disease process⁸³. These genes and pathways can serve as potential targets for therapeutic interventions^{84, 85}. Additionally, understanding DNA methylation patterns and miRNA expression can provide insights into gene regulation and help us understand how different environmental factors or genetic variations can influence gene expression⁸³. DNA methylation as well as miRNAs can be used as a biomarker for various diseases and as a target for therapeutic interventions⁸³. We can better understand gene expression and regulation by studying both the transcriptome and the epigenome. In this thesis both approaches are applied to improve our understanding of COPD and its specific subtypes. The specific techniques that were used will now be explained in more detail.

1.6 – Techniques to investigate the transcriptome and epigenome

High-throughput transcriptome sequencing

High-throughput transcriptome sequencing, i.e. RNA sequencing (RNA-Seq) is one of the methods used to gather these vast amounts of data. RNA-Seq is a method to investigate the transcriptome of

tissue or cells by identifying and quantifying the complementary DNA (cDNA) of RNA transcripts⁸⁶. Differences in mRNA levels quantified by RNA-Seq are often used as a surrogate for differences in expression of proteins these mRNA's code for⁸⁷, whereas an increase in miRNA expression can suppress the translation of transcribed genes to proteins and, therefore, alter biology via cell function.

Identifying and quantifying the complementary DNA (cDNA) of RNA transcripts is achieved by extraction of RNA from biological material⁸⁸. Thereafter, RNA species are isolated using a specific protocol, such as a ribo-depletion protocol, to remove ribosomal RNAs⁸⁸. Then, the cDNA synthesis is performed using reverse transcription, after which specific sequencing adaptors are ligated to the ends of the cDNA fragments⁸⁸. Lastly, amplification will be done using PCR, after which the RNA-Seq library will be ready for sequencing using platforms such as the Illumina HiSeq 2500⁸⁸. This generates data that then needs to be quality-checked and mapped to a reference genome while quantifying the reads (per gene)⁸⁸.

Depending on the RNA species, RNA-Seq can be used to detect gene transcripts (bulk RNA-Seq) or micro-RNAs (miRNAs via small RNA-Seq)^{86, 89}. With that, RNA-Seq can be used to find differences in expression, identify alternative splicing, uncover functional relationships between DNA and RNA, detect gene fusions and identify a myriad of noncoding RNAs, including miRNAs⁹⁰⁻⁹³.

The Illumina Methylation Epic bead chip array for DNA methylation

The Illumina MethylationEpic BeadChip array is a method to study DNA methylation in the biological sample of choice by identifying and quantifying the DNA methylation status of pre-defined sites^{94, 95}. Identifying and quantifying the DNA methylation status of pre-defined sites is achieved by extracting and purifying DNA from biological material which then undergoes bisulfite conversion⁹⁵. Thereafter, this DNA is denatured and neutralised, followed by an amplified and fragmented strand of DNA⁹⁵. Thereafter, the amplified and fragmented DNA is added to the bead chip array, after which the fluorescent probes are hybridised into the caught fragments⁹⁵. After a washing step, the array can be read on platforms such as the Illumina iScan. This generates data that then needs to be quality-checked, and failed and cross-reactive probes will be removed⁹⁶. The data will then be processed to values representing the proportion of methylation at a specific site (beta-value) or the log-transformed proportion of methylation at a specific site (M-value)⁹⁷. An increased proportion of methylation for a specific site in a sample means that more cells have the specific methylation site methylated, as a site can only be methylated or unmethylated⁹⁷.

1.7 – Recent work in COPD research using bioinformatics

The use of bioinformatics has been important in recent years in advancing COPD research. Since the onset of large-scale genomic, transcriptomic, epigenetic, and proteomic data in the early 2000s, bioinformatics has played a vital role in analysing and interpreting this wealth of information⁹⁸. By applying computational methods and algorithms, researchers have been able to identify genetic variants associated with COPD susceptibility, explore gene expression patterns in lung tissues, and uncover epigenetic modifications that contribute to disease progression⁹⁹⁻¹⁰¹. Additionally, bioinformatics tools have facilitated the integration of multi-omics data, enabling a comprehensive understanding of the complex molecular mechanisms underlying COPD⁹⁸.

Gene Expression Profiling in COPD

The focus of past transcriptomics research in COPD has aimed to find molecular mechanisms for better understanding the pathogenesis of COPD. One study focussed on finding biomarkers in blood using differential expression analyses¹⁰², and identified 3829 genes associated with emphysema in two cohorts. Additionally, more in-depth analysis using alternative splicing, protein association, and

pathway analysis identified 942 gene isoforms, 260 exons, 714 proteins and 11 biological pathways, revealing key biological pathways of emphysema.

In a similar analysis, using tobacco smoke combined with elastin peptides to induce emphysema in a mouse model, 1159 genes were found to be differentially expressed¹⁰³. The identified genes were involved in immune response pathways and oxidative stress. Though this study was performed in mice, it may shed light on the mechanism behind smoke-induced emphysema.

Another approach that has been used to study emphysema in COPD was to identify features from CT scans and associate these to the genes identified using a differential expression analysis¹⁰⁴.

Integration of medical imaging data with bulk RNA-seq data can provide valuable insights into disease mechanisms and treatment response. However, further research is needed to validate the clinical utility of such associations.

Gene expression profiling can also be performed using microarrays, such as the Affymetrix Human Gene 1.0 ST array. Our group used this method to perform gene expression profiling on nasal epithelial brushes of 31 severe COPD patients and 22 controls to characterize COPD in a non-invasive manner¹⁰⁵. They found 135 genes significantly differentially expressed between severe COPD patients and controls. This signature was then replicated in two independent cohorts. Furthermore, part of the COPD nasal gene expression signature was also present in the bronchus.

miRNA profiling in COPD

Past research has also made progress in uncovering how miRNAs contribute to pulmonary diseases. In particular, small RNA-Seq has been used recently to identify subtypes and biomarkers in both COPD and interstitial lung disease (ILD)¹⁰⁶. They found 255 miRNAs associated with COPD and ILD and identified four miRNA clusters and five sample clusters. Clustering of the samples identified a remarkable overlap between COPD and ILD, showing that some miRNAs have the similar expression pattern between diseases, whereas the integrated network analysis implicated the miR-34/449 family of miRNAs and the NOTCH pathway in both COPD and ILD.

DNA methylation profiling in COPD

Fewer studies have focussed on the interaction of DNA methylation with COPD development or progression. However, a recent study investigated the DNA methylation patterns in cells from the distal airways of COPD patients and control subjects to identify the functional effects of DNA methylation on the parts of the lung that are most affected by the disease¹⁰⁷. 1,155 Bonferroni-significant differentially methylated sites (DMP) were identified, and 7,097 differentially methylated regions (DMR) were identified. Gene ontology analysis of the DNA methylation-associated genes¹⁰⁸ showed enrichment for transcriptional factor activity, anatomical structure development, and regulation of metabolic processes¹⁰⁷. This study showed how differential DNA methylation can be linked to gene expression, and how this leads to an understanding of COPD.

Recent research into COPD using bioinformatics has used a wide variety of methods to find biomarkers and mechanisms underlying disease pathogenesis as well as progression.

Overall, currently all COPD patients receive a one size fits all treatment, indicating all patients receive the same treatment. Not all patients respond to the main line therapy (ICS). Therefore, there is a need to understand the underlying mechanisms of COPD which may help to identify the right therapy for individual patients. By using multi-omics from different locations of the respiratory system (Figure 1.4) and integrating multiple levels of big data, we believe that we will be able to answer these multi-layer complex questions. In addition, determining whether severe COPD is a

simple disease progression or a different entity, will help find markers that may identify patients that are at most risk of developing the most severe form of COPD.

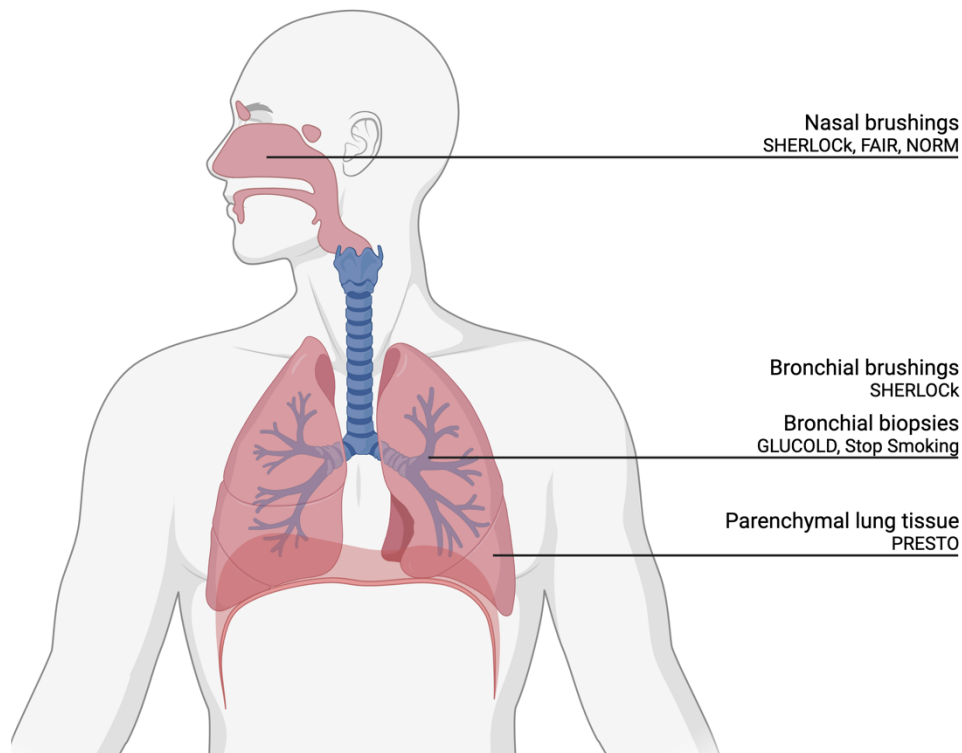


Figure 1.4) The location and type of sample for every type of study that is used in this thesis.

1.8 – Scope of this thesis

We hypothesise that severe COPD and severe early onset are unique phenotypes rather than stages within COPD. Therefore, the aim of this thesis is to find characteristics for COPD in general, and severe COPD and SEO-COPD specifically using a multi-omics approach.

We will first focus on transcriptomics of bronchial brushings to identify genes that are associated with severe COPD. We then move to the lower airways in the parenchyma to answer the same question focussing on transcriptomics and miRNAs. We then look at the effects of the environment on miRNA expression in bronchial biopsies focussing on cigarette smoke. Finally, we focus on epigenetics (DNA methylation). First looking at the difference between COPD and non-COPD, and then focussing on the influence of ICS on DNA methylation.

In **Chapter 2**, we aimed to investigate whether severe COPD represents a clinically distinct disease phenotype that is different from mild or moderate COPD. To this end, we identified unique differentially expressed genes in bronchial brushes of severe COPD patients compared to mild, moderate, and non-COPD controls in the SHERLOCK study (Figure 1.4). Additionally, since bronchial brushings are an invasive way to diagnose and phenotype disease, we investigated, in the same dataset, if and to what extent, the identified COPD-associated bronchial gene signature reflected the nasal epithelium, as this could pave the way for easier obtaining potential biomarkers and verified this in nasal brushings from the FAIR and NORM studies. Thereafter, in **Chapter 3**, we focused on parenchymal lung tissue in severe early onset-COPD in the PRESTO dataset and hypothesized that its pathogenesis differs from common COPD. We determined if this was the case by identifying the unique SEO-COPD gene expression profile compared to control, following the removal of common

COPD genes. We validated the signature using immunohistochemistry. Next, in **Chapter 4**, we investigated, in the same dataset, if and how miRNAs are involved in the regulation of SEO-COPD pathology. We assessed miRNA expression in peripheral lung tissue, comparing SEO-COPD patients with non-COPD control subjects and patients with mild, moderate, and severe COPD using small RNA sequencing. This was followed by identifying the putative target genes and pathways targeted by these differentially expressed miRNAs. Thereafter, in **Chapter 5**, we aimed to elucidate the overall smoking-induced miRNA changes and those specific to COPD. We did investigate this by analysing the influence of current smoking on miRNA expression in bronchial biopsies in two independent datasets (GLUCOLD and NORM) consisting of current and ex-smoking COPD patients and current and non-smoking respiratory controls. In addition, to investigate the downstream effects of these miRNAs affected by smoking on gene expression and pathways, we did correlate these miRNAs to mRNA expression levels in bronchial biopsies from the same subjects. We performed cellular deconvolution and associated the miRNA expression levels to cell proportions and immunohistological stainings to take the effect of cell type composition into account. In **Chapter 6**, we assessed how differences in DNA methylation associated with the presence and severity of COPD. We identified differences in DNA methylation in airway wall biopsies between current smokers with and without COPD and associated these with severity of airflow obstruction and hyperinflation. Additionally, we investigated how the differential DNA methylation affects the expression of genes and gene pathways relevant to COPD pathogenesis. For this we used the GLUCOLD and Stop-Smoking datasets. In **Chapter 7**, we investigated the epigenetic and transcriptomic effects of ICS in bronchial biopsies from the GLUCOLD study. We identified differentially methylated CpG sites directly affected by ICS usage in COPD using ChIP-Seq data. We then assessed how these methylation sites impacted gene expression levels in the airway wall. Finally, we performed epigenetic editing using CRISPR-dCas9-DNMT3A in A549 cells to verify the direct effects of ICS-associated changes in DNA methylation on expression of its gene. To conclude, in **Chapter 8**, we summarised all findings from this thesis, and in **Chapter 9** we discussed the relevance and implications of these findings, and described some of the future perspectives in better understanding COPD when looking at transcriptomics and epigenetics of COPD.

1.9 – References

1. Miravittles, M., et al., *Observational study to characterise 24-hour COPD symptoms and their relationship with patient-reported outcomes: results from the ASSESS study*. Respir Res, 2014. **15**(1): p. 122.
2. Tkac, J., S.F. Man, and D.D. Sin, *Systemic consequences of COPD*. Ther Adv Respir Dis, 2007. **1**(1): p. 47-59.
3. Decramer, M., et al., *Systemic effects of COPD*. Respir Med, 2005. **99 Suppl B**: p. S3-10.
4. Sin, D.D., et al., *Mortality in COPD: Role of comorbidities*. Eur Respir J, 2006. **28**(6): p. 1245-57.
5. Pauwels, R.A., et al., *Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary*. Am J Respir Crit Care Med, 2001. **163**(5): p. 1256-76.
6. Agusti, A., et al., *Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary*. Am J Respir Crit Care Med, 2023. **207**(7): p. 819-837.
7. Global Initiative For Chronic Obstructive Lung Disease. *Global strategy for prevention, diagnosis and management of COPD: 2023 report*. [cited 2023 1 October]; Available from: https://goldcopd.org/wp-content/uploads/2023/03/GOLD-2023-ver-1.3-17Feb2023_WMV.pdf.
8. Franciosi, L., et al., *Susceptibility to COPD: differential proteomic profiling after acute smoking*. PLoS One, 2014. **9**(7): p. e102037.
9. Sorheim, I.C., et al., *Gender differences in COPD: are women more susceptible to smoking effects than men?* Thorax, 2010. **65**(6): p. 480-5.
10. Silverman, E.K., et al., *Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis*. Am J Respir Crit Care Med, 1998. **157**(6 Pt 1): p. 1770-8.
11. Foreman, M.G., et al., *Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPD Gene Study*. Am J Respir Crit Care Med, 2011. **184**(4): p. 414-20.
12. Woldhuis, R.R., et al., *Link between increased cellular senescence and extracellular matrix changes in COPD*. Am J Physiol Lung Cell Mol Physiol, 2020. **319**(1): p. L48-L60.
13. CDC. *The Health Consequences of Smoking - 50 Years of Progress*. Surgeon General's Report 2014; Available from: <https://www.cdc.gov/tobacco/sgr/50th-anniversary/index.htm>.
14. Lundback, B., et al., *Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies*. Respir Med, 2003. **97**(2): p. 115-22.
15. Soriano, J.B. and R. Rodriguez-Roisin, *Chronic obstructive pulmonary disease overview: epidemiology, risk factors, and clinical presentation*. Proc Am Thorac Soc, 2011. **8**(4): p. 363-7.
16. Raherison, C. and P.O. Girodet, *Epidemiology of COPD*. Eur Respir Rev, 2009. **18**(114): p. 213-21.
17. Chappell, S., et al., *Cryptic haplotypes of SERPINA1 confer susceptibility to chronic obstructive pulmonary disease*. Hum Mutat, 2006. **27**(1): p. 103-9.
18. Prescott, E., P. Lange, and J. Vestbo, *Chronic mucus hypersecretion in COPD and death from pulmonary infection*. Eur Respir J, 1995. **8**(8): p. 1333-8.
19. Jing, Y., et al., *NOTCH3 contributes to rhinovirus-induced goblet cell hyperplasia in COPD airway epithelial cells*. Thorax, 2019. **74**(1): p. 18-32.
20. Barnes, P.J., *Small airway fibrosis in COPD*. Int J Biochem Cell Biol, 2019. **116**: p. 105598.
21. Kitaguchi, Y., et al., *Clinical characteristics of combined pulmonary fibrosis and emphysema*. Respirology, 2010. **15**(2): p. 265-71.
22. Chen, G. and N. Khalil, *TGF-beta1 increases proliferation of airway smooth muscle cells by phosphorylation of map kinases*. Respir Res, 2006. **7**(1): p. 2.
23. Karpova, A.Y., et al., *MEK1 is required for PDGF-induced ERK activation and DNA synthesis in tracheal myocytes*. Am J Physiol, 1997. **272**(3 Pt 1): p. L558-65.

24. Krymskaya, V.P., et al., *Phosphatidylinositol 3-kinase mediates mitogen-induced human airway smooth muscle cell proliferation*. Am J Physiol, 1999. **277**(1): p. L65-78.
25. Hogg, J.C., P.D. Pare, and T.L. Hackett, *The Contribution of Small Airway Obstruction to the Pathogenesis of Chronic Obstructive Pulmonary Disease*. Physiol Rev, 2017. **97**(2): p. 529-552.
26. Cooper, C.B., *The connection between chronic obstructive pulmonary disease symptoms and hyperinflation and its impact on exercise and function*. Am J Med, 2006. **119**(10 Suppl 1): p. 21-31.
27. Kim, Y.W., et al., *Resting hyperinflation and emphysema on the clinical course of COPD*. Sci Rep, 2019. **9**(1): p. 3764.
28. Park, G.Y., et al., *Prolonged airway and systemic inflammatory reactions after smoke inhalation*. Chest, 2003. **123**(2): p. 475-80.
29. Hogg, J.C., et al., *The nature of small-airway obstruction in chronic obstructive pulmonary disease*. New England Journal of Medicine, 2004. **350**(26): p. 2645-2653.
30. Dransfield, M.T., et al., *Gender differences in the severity of CT emphysema in COPD*. Chest, 2007. **132**(2): p. 464-70.
31. Hoffmann, D. and I. Hoffmann, *Letters to the Editor - Tobacco smoke components*. Beiträge zur Tabakforschung International/Contributions to Tobacco Research, 1998. **18**(1): p. 49-52.
32. Johnston, S.L., et al., *Airway Epithelial Innate Immunity*. Front Physiol, 2021. **12**: p. 749077.
33. Myszor, I.T. and G.H. Gudmundsson, *Modulation of innate immunity in airway epithelium for host-directed therapy*. Front Immunol, 2023. **14**: p. 1197908.
34. Comer, D.M., et al., *Airway epithelial cell apoptosis and inflammation in COPD, smokers and nonsmokers*. Eur Respir J, 2013. **41**(5): p. 1058-67.
35. Schuliga, M., *NF-kappaB Signaling in Chronic Inflammatory Airway Disease*. Biomolecules, 2015. **5**(3): p. 1266-83.
36. Jia, Z., et al., *EGFR activation-induced decreases in claudin1 promote MUC5AC expression and exacerbate asthma in mice*. Mucosal Immunol, 2021. **14**(1): p. 125-134.
37. Koff, J.L., et al., *Multiple TLRs activate EGFR via a signaling cascade to produce innate immune responses in airway epithelium*. Am J Physiol Lung Cell Mol Physiol, 2008. **294**(6): p. L1068-75.
38. Kanai, K., et al., *Cigarette smoke augments MUC5AC production via the TLR3-EGFR pathway in airway epithelial cells*. Respir Investig, 2015. **53**(4): p. 137-48.
39. Chua, F. and G.J. Laurent, *Neutrophil elastase: mediator of extracellular matrix destruction and accumulation*. Proc Am Thorac Soc, 2006. **3**(5): p. 424-7.
40. Sullivan, A.K., et al., *Oligoclonal CD4+ T cells in the lungs of patients with severe emphysema*. Am J Respir Crit Care Med, 2005. **172**(5): p. 590-6.
41. Rojas-Quintero, J., et al., *Spatial Transcriptomics Resolve an Emphysema-specific Lymphoid Follicle B Cell Signature in COPD*. Am J Respir Crit Care Med, 2023.
42. Hogg, J.C., et al., *The nature of small-airway obstruction in chronic obstructive pulmonary disease*. N Engl J Med, 2004. **350**(26): p. 2645-53.
43. Bonarius, H.P., et al., *Antinuclear autoantibodies are more prevalent in COPD in association with low body mass index but not with smoking history*. Thorax, 2011. **66**(2): p. 101-7.
44. van der Strate, B.W., et al., *Cigarette smoke-induced emphysema: A role for the B cell?* Am J Respir Crit Care Med, 2006. **173**(7): p. 751-8.
45. Cosio, M.G., M. Saetta, and A. Agusti, *Immunologic aspects of chronic obstructive pulmonary disease*. N Engl J Med, 2009. **360**(23): p. 2445-54.
46. Willemse, B.W., et al., *Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers*. Eur Respir J, 2005. **26**(5): p. 835-45.
47. Lapperre, T.S., et al., *Relation between duration of smoking cessation and bronchial inflammation in COPD*. Thorax, 2006. **61**(2): p. 115-21.
48. Abel, G.A., et al., *Effects of biochemically confirmed smoking cessation on white blood cell count*. Mayo Clin Proc, 2005. **80**(8): p. 1022-8.

49. Higuchi, T., et al., *Current cigarette smoking is a reversible cause of elevated white blood cell count: Cross-sectional and longitudinal studies*. Prev Med Rep, 2016. **4**: p. 417-22.
50. Hodge, G., et al., *Enhanced cytotoxic function of natural killer and natural killer T-like cells associated with decreased CD94 (Kp43) in the chronic obstructive pulmonary disease airway*. Respiriology, 2013. **18**(2): p. 369-76.
51. Willemse, B.W., et al., *The impact of smoking cessation on respiratory symptoms, lung function, airway hyperresponsiveness and inflammation*. Eur Respir J, 2004. **23**(3): p. 464-76.
52. Wheaton, A.G., et al., *Employment and Activity Limitations Among Adults with Chronic Obstructive Pulmonary Disease — United States, 2013*. Morbidity and Mortality Weekly Report, 2015. **64**: p. 289 - 295.
53. Robison, G., R. Butcher, and E. Sutherland, *Adenyl cyclase as an adrenergic receptor*. Annals of the New York Academy of Sciences, 1967. **139**(3): p. 703-723.
54. Johnson, M., *Beta2-adrenoceptors: mechanisms of action of beta2-agonists*. Paediatr Respir Rev, 2001. **2**(1): p. 57-62.
55. Alagha, K., et al., *Long-acting muscarinic receptor antagonists for the treatment of chronic airway diseases*. Ther Adv Chronic Dis, 2014. **5**(2): p. 85-98.
56. Barnes, P.J., *Inhaled Corticosteroids*. Pharmaceuticals (Basel), 2010. **3**(3): p. 514-540.
57. Bamberger, C.M., et al., *Glucocorticoid receptor beta, a potential endogenous inhibitor of glucocorticoid action in humans*. J Clin Invest, 1995. **95**(6): p. 2435-41.
58. Meijer, O.C., J. Buurstedde, and M.J. Schaaf, *Corticosteroid receptors in the brain: transcriptional mechanisms for specificity and context-dependent effects*. Cellular and Molecular Neurobiology, 2019. **39**: p. 539-549.
59. Bamberger, C.M., H.M. Schulte, and G.P. Chrousos, *Molecular determinants of glucocorticoid receptor function and tissue sensitivity to glucocorticoids*. Endocr Rev, 1996. **17**(3): p. 245-61.
60. Alvira, S., et al., *Structural characterization of the substrate transfer mechanism in Hsp70/Hsp90 folding machinery mediated by Hop*. Nat Commun, 2014. **5**: p. 5484.
61. Paul, A., et al., *The cochaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing protein alpha) demonstrates regulatory specificity for the androgen, glucocorticoid, and progesterone receptors*. J Biol Chem, 2014. **289**(22): p. 15297-308.
62. Pascual, G. and C.K. Glass, *Nuclear receptors versus inflammation: mechanisms of transrepression*. Trends Endocrinol Metab, 2006. **17**(8): p. 321-7.
63. De Bosscher, K., et al., *Glucocorticoids repress NF-kappaB-driven genes by disturbing the interaction of p65 with the basal transcription machinery, irrespective of coactivator levels in the cell*. Proc Natl Acad Sci U S A, 2000. **97**(8): p. 3919-24.
64. Savory, J.G., et al., *Discrimination between NL1- and NL2-mediated nuclear localization of the glucocorticoid receptor*. Mol Cell Biol, 1999. **19**(2): p. 1025-37.
65. Nigg, E.A., *Nucleocytoplasmic transport: signals, mechanisms and regulation*. Nature, 1997. **386**(6627): p. 779-87.
66. Ohno, M., M. Fornerod, and I.W. Mattaj, *Nucleocytoplasmic transport: the last 200 nanometers*. Cell, 1998. **92**(3): p. 327-36.
67. Gorlich, D. and I.W. Mattaj, *Nucleocytoplasmic transport*. Science, 1996. **271**(5255): p. 1513-8.
68. Davis, L.I., *The nuclear pore complex*. Annu Rev Biochem, 1995. **64**: p. 865-96.
69. Bates, M.K., R.M. Kerr, and ProQuest (Firm), *Nuclear Receptors*. 2011: Nova Science.
70. Voss, T.C., et al., *Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism*. Cell, 2011. **146**(4): p. 544-54.
71. McNally, J.G., et al., *The glucocorticoid receptor: rapid exchange with regulatory sites in living cells*. Science, 2000. **287**(5456): p. 1262-5.
72. Zhang, N., et al., *Risk of Fracture and Osteoporosis in Patients With COPD and Inhaled Corticosteroids Treatment*. Respir Care, 2023. **68**(12): p. 1719-1727.

73. Saeed, M.I., et al., *Use of inhaled corticosteroids and the risk of developing type 2 diabetes in patients with chronic obstructive pulmonary disease*. Diabetes Obes Metab, 2020. **22**(8): p. 1348-1356.
74. Miravittles, M., et al., *Systematic review on long-term adverse effects of inhaled corticosteroids in the treatment of COPD*. Eur Respir Rev, 2021. **30**(160).
75. Pertea, M., *The human transcriptome: an unfinished story*. Genes, 2012. **3**(3): p. 344-360.
76. Melé, M., et al., *The human transcriptome across tissues and individuals*. Science, 2015. **348**(6235): p. 660-665.
77. Bernstein, B.E., A. Meissner, and E.S. Lander, *The mammalian epigenome*. Cell, 2007. **128**(4): p. 669-681.
78. Lange, U.C. and R. Schneider, *What an epigenome remembers*. Bioessays, 2010. **32**(8): p. 659-668.
79. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function*. Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
80. Razin, A. and H. Cedar, *DNA methylation and gene expression*. Microbiol Rev, 1991. **55**(3): p. 451-8.
81. Mohr, A.M. and J.L. Mott, *Overview of microRNA biology*. Semin Liver Dis, 2015. **35**(1): p. 3-11.
82. Pritchard, C.C., H.H. Cheng, and M. Tewari, *MicroRNA profiling: approaches and considerations*. Nat Rev Genet, 2012. **13**(5): p. 358-69.
83. Lee, T.I. and R.A. Young, *Transcriptional regulation and its misregulation in disease*. Cell, 2013. **152**(6): p. 1237-51.
84. Akhurst, R.J. and A. Hata, *Targeting the TGF β signalling pathway in disease*. Nature reviews Drug discovery, 2012. **11**(10): p. 790-811.
85. Kahn, M., *Can we safely target the WNT pathway?* Nature reviews Drug discovery, 2014. **13**(7): p. 513-532.
86. Marguerat, S. and J. Bahler, *RNA-seq: from technology to biology*. Cell Mol Life Sci, 2010. **67**(4): p. 569-79.
87. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses*. Nat Rev Genet, 2012. **13**(4): p. 227-32.
88. Wolf, J.B., *Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial*. Mol Ecol Resour, 2013. **13**(4): p. 559-72.
89. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities*. Nat Rev Genet, 2011. **12**(2): p. 87-98.
90. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
91. Shen, S., et al., *rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data*. Proc Natl Acad Sci U S A, 2014. **111**(51): p. E5593-601.
92. Haas, B.J., et al., *STAR-Fusion: fast and accurate fusion transcript detection from RNA-Seq*. BioRxiv, 2017: p. 120295.
93. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq*. Bioinformatics, 2009. **25**(9): p. 1105-11.
94. Cheung, K., et al., *Correlation of Infinium HumanMethylation450K and MethylationEPIC BeadChip arrays in cartilage*. Epigenetics, 2020. **15**(6-7): p. 594-603.
95. Noble, A.J., et al., *A validation of Illumina EPIC array system with bisulfite-based amplicon sequencing*. PeerJ, 2021. **9**: p. e10762.
96. Nazarenko, T., et al., *Technical and biological sources of unreliability of Infinium type II probes of the Illumina MethylationEPIC BeadChip microarray*. bioRxiv, 2023: p. 2023.03. 14.532595.
97. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. Bioinformatics, 2014. **30**(10): p. 1363-1369.
98. Gauthier, J., et al., *A brief history of bioinformatics*. Brief Bioinform, 2019. **20**(6): p. 1981-1996.

99. Hu, W.P., et al., *Identification of novel candidate genes involved in the progression of emphysema by bioinformatic methods*. Int J Chron Obstruct Pulmon Dis, 2018. **13**: p. 3733-3747.
100. Spira, A., et al., *Gene expression profiling of human lung tissue from smokers with severe emphysema*. Am J Respir Cell Mol Biol, 2004. **31**(6): p. 601-10.
101. Benincasa, G., et al., *Epigenetics and pulmonary diseases in the horizon of precision medicine: a review*. Eur Respir J, 2021. **57**(6).
102. Suryadevara, R., et al., *Blood-based Transcriptomic and Proteomic Biomarkers of Emphysema*. Am J Respir Crit Care Med, 2023.
103. Feng, X., et al., *RNA Sequencing and Related Differential Gene Expression Analysis in a Mouse Model of Emphysema Induced by Tobacco Smoke Combined with Elastin Peptides*. Int J Chron Obstruct Pulmon Dis, 2023. **18**: p. 2147-2161.
104. Chen, J., et al., *Deep Learning Integration of Chest Computed Tomography Imaging and Gene Expression Identifies Novel Aspects of COPD*. Chronic Obstr Pulm Dis, 2023. **10**(4): p. 355-368.
105. Boudewijn, I.M., et al., *Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression*. Respir Res, 2017. **18**(1): p. 213.
106. Pavel, A.B., et al., *Integrative genetic and genomic networks identify microRNA associated with COPD and ILD*. Sci Rep, 2023. **13**(1): p. 13076.
107. Pan, W., et al., *Identification of Potential Differentially-Methylated/Expressed Genes in Chronic Obstructive Pulmonary Disease*. COPD, 2023. **20**(1): p. 44-54.
108. Ringh, M.V., et al., *Tobacco smoking induces changes in true DNA methylation, hydroxymethylation and gene expression in bronchoalveolar lavage cells*. EBioMedicine, 2019. **46**: p. 290-304.

Chapter 2 – A bronchial gene signature for severe COPD that is retained in the nose

2.1 – Publication Declaration

Jos van Nijnatten, Alen Faiz, Wim Timens, Victor Guryev, Dirk-Jan Slebos, Karin Klooster, Jorine E. Hartman, Tessa Kole, David F. Choy, Arindam Chakrabarti, Michele Grimbaldston, Carrie M. Rosenberger, Huib Kerstjens, Corry-Anke Brandsma[†] and Maarten van den Berge[†]. (2023). A bronchial gene signature specific for severe COPD that is retained in the nose. *ERJ Open Research*

Status: Published (doi: 10.1183/23120541.00354-2023)

Author Contributions:

J.v.N. wrote the main manuscript text and performed the analyses; V.G. performed an analysis; M.v.d.B. and A.F. are the primary supervisors, and C.-A.B. and W.T. are the secondary supervisors and as such involved in the analytical procedure and interpretation; M.v.d.B, H.K., D.J.S., W.T., J.E.H., T.K., D.F.C., A.C., M.G., C.M.R., K.K. were involved in obtaining the primary dataset; M.v.d.B. was involved in obtaining the FAIR and NORM datasets; All authors reviewed the manuscript and approved the final version of the manuscript.

Name:	Signature:	Date:
Jos van Nijnatten	Production Note: Signature removed prior to publication.	08-12-2023
Alen Faiz	Production Note: Signature removed prior to publication.	2/12/2024
Wim Timens	Production Note: Signature removed prior to publication.	December 2, 2023
Victor Guryev	Production Note: Signature removed prior to publication.	02 Dec 2023
Dirk-Jan Slebos	Production Note: Signature removed prior to publication.	06-12-2023
Karin Klooster	Production Note: Signature removed prior to publication.	07/dec/2023
Jorine E. Hartman	Production Note: Signature removed prior to publication.	1-12-2023
Tessa Kole	Production Note: Signature removed prior to publication.	5-12-2024
David F. Choy	Production Note: Signature removed prior to publication.	December 2, 2024

Name:	Signature:	Date:
Arindam Chakrabarti	(contact information missing)	
Michele Grimaldeston	Production Note: Signature removed prior to publication.	December 2, 2024
Carrie M. Rosenberger	Production Note: Signature removed prior to publication.	December 8, 2023
Huib Kerstjens	Production Note: Signature removed prior to publication.	3 Dec 2023
Corry-Anke Brandsma	Production Note: Signature removed prior to publication.	07-12-2024
Maarten van den Berge	Production Note: Signature removed prior to publication.	December 5, 2024

2.2 – Abstract

Introduction: A subset of COPD patients develops advanced disease with severe airflow obstruction, hyperinflation and extensive emphysema. We propose that the pathogenesis in these patients differs from mild–moderate COPD and is reflected by bronchial gene expression. The aim of the present study was to identify a unique bronchial epithelial gene signature for severe COPD patients.

Methods: We obtained RNA sequencing data from bronchial brushes from 123 ex-smokers with severe COPD, 23 with mild–moderate COPD and 23 non-COPD controls. We identified genes specific to severe COPD by comparing severe COPD to non-COPD controls, followed by removing genes that were also differentially expressed between mild–moderate COPD and non-COPD controls. Next, we performed a pathway analysis on these genes and evaluated whether this signature is retained in matched nasal brushings.

Results: We identified 219 genes uniquely differentially expressed in severe COPD. Interaction network analysis identified VEGFA and FN1 as the key genes with the most interactions. Genes were involved in extracellular matrix regulation, collagen binding and the immune response. Of interest were 10 genes (VEGFA, DCN, SPARC, COL6A2, MGP, CYR61, ANXA6, LGALS1, C1QA and C1QB) directly connected to fibronectin 1 (FN1). Most of these genes were lower expressed in severe COPD and showed the same effect in nasal brushings.

Conclusions: We found a unique severe COPD bronchial gene signature with key roles for VEGFA and FN1, which was retained in the upper airways. This supports the hypothesis that severe COPD, at least partly, comprises a different pathology and supports the potential for biomarker development based on nasal brushes in COPD.

Keywords: Severe COPD, transcriptomic signature, bronchial, nasal, brush

2.3 – Introduction

COPD is a common, chronic inflammatory lung disease affecting over hundreds of millions of people worldwide^{1,2}. It was the third leading cause of death worldwide in 2019 and is expected to become even more prevalent in the upcoming years³. Characteristics of COPD include irreversible airflow limitation, hypersecretion of mucus and alveolar destruction (emphysema)⁴.

One of the most common risk factors for COPD development is the inhalation of noxious particles. This includes cigarette smoking, second-hand smoke, biomass smoke and air pollution⁵⁻⁷. Cigarette smoke exposes the lung, and specifically the bronchial epithelium, to over 4000 different components⁸, directly causing irritation, mucus hypersecretion and inflammation in the airway. Duration and intensity of smoking have previously been associated with COPD incidence, increasing the risk of developing COPD five-fold so that ~30% of smokers develop COPD⁹⁻¹². Importantly, some people develop COPD faster and to a much more severe extent than others, suggesting an underlying individual susceptibility to the disease¹³.

Most patients develop COPD later in life and only suffer from mild to moderate airflow obstruction (mCOPD), whereas a small subset of patients progresses to advanced disease with severe airflow obstruction, hyperinflation and extensive emphysema or small airways disease (sCOPD)¹⁴. This severe group accounts for a majority of the personal burden as well as societal and economic burden attributed to COPD via healthcare and time lost from work¹⁵. Next, to help with smoking cessation, treatment of the remaining sCOPD patients aims at alleviating the symptoms by providing temporary bronchodilation, while this does not alter the progressive lung function decline characteristic of the disease^{16,17}. More insight into the mechanisms leading to sCOPD and progressive lung function decline is needed to find novel targets for therapeutic intervention.

The aim of the present study was to investigate whether sCOPD represents a clinically distinct disease phenotype. To this end, we aimed to identify unique differentially expressed genes in bronchial brushes of sCOPD patients compared to mCOPD and non-COPD. Additionally, since bronchial brushings are an invasive way to diagnose and phenotype disease, we also investigated whether the identified bronchial gene signature is also represented in the nasal epithelium, providing potential biomarkers for sCOPD in the nose.

2.4 – Methods

Patients and study design

SHERLOCK (An integrative genomic approach to Solve tHe puzzle of sevERe earLy-Onset COPD, ClinicalTrials.gov: NCT04263961 and NCT04023409) is a cross-sectional study without pharmacological intervention performed by the University of Groningen, the Netherlands.

We enrolled 23 non-COPD controls, 23 mCOPD patients (Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 1 and 2) and 123 patients with sCOPD (GOLD stages 3 and 4, with both extensive hyperinflation and emphysema). Participants did not smoke for at least 2 months prior to inclusion in the study and did not have an exacerbation or lung infection 4 weeks before the study.

Subjects underwent bronchoscopy, during which bronchial and nasal brushes were obtained^{18,19}. All patients were fully characterised, i.e., lung function, computed tomography scans, blood and questionnaire data. RNA isolation and the RNA-Seq procedure are outlined in the supplementary methods.

The local medical ethics committees approved the study, and all subjects gave their written informed consent (the SHERLOCK study was approved by the medical ethics committee of the University of Groningen/University Medical Center Groningen, METc 2016/572 and METc 2014/102).

RNA-Sequencing

Total RNA was isolated from the bronchial and nasal brushings using QIAgen AllPrep DNA/RNA/miRNA Universal Mini Kit per manufacturer's instructions. Ribosomal RNA was removed using the RiboZero Magnetic Gold kit. RNA sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA method. Libraries were paired-end sequenced. FastQC (version 0.11.7) was used to determine the raw RNA sequencing data quality. StarAligner (version 2.73a) was used to index and align the raw sequencing data to the human reference genome (version GrCh38). Ensembl (release 100) was used as the gene annotation database. We checked if the expected sexes and recorded sexes matched and if the mapped read counts were similar across samples of the same type. Lastly, we checked for any outliers in a principal component analysis.

Statistics

All analyses were performed using R statistical software (version 4.0.2). The normality of the distribution of the data was established using histogram plots. A Kruskal–Wallis test was conducted across groups, and a Mann–Whitney U-test or Wilcoxon signed-rank test was conducted between the groups for non-paired and paired data, respectively. A Benjamini–Hochberg false discovery rate (FDR) was calculated where appropriate. A p-value or FDR ≤ 0.05 was considered significant unless specified otherwise.

Selection of unique genes for sCOPD

To identify genes that are uniquely changed in bronchial brushes of sCOPD patients, we first identified genes that were differentially expressed between sCOPD patients compared to the non-COPD controls. We conducted a linear model using the EdgeR package (version 3.30.3, dependent upon Limma version 3.44.3), correcting for sex, age and smoking pack-years. An FDR of <0.05 and an absolute fold change of >2 was considered statistically significant. To assess if the sCOPD-associated genes also differ in mCOPD compared to non-COPD controls, we conducted a candidate gene approach for the sCOPD genes. Genes were identified as common COPD genes and removed from the sCOPD list when they were also differentially expressed in mCOPD versus non-COPD controls in the same direction. We here used a more lenient FDR significance cut-off of 0.25 to avoid false-negative outcomes. Finally, we checked if the remaining sCOPD genes were also differentially expressed when directly comparing the sCOPD to mCOPD patients using an FDR cut-off of 0.05. Since most sCOPD subjects used high inhaled corticosteroid (ICS) doses, we removed steroid-sensitive genes, as previously identified in the GLUCOLD study, from our analysis²⁰. Additionally, we calculated the cell proportions using cellular deconvolution and adjusted for this in our model. These selection criteria for cell types and ICS genes, as well as replication in the nasal brushings and an independent study, and the pathway analysis, are described in the supplementary materials. Supplementary Figure 2.1 presents an outline of the study.

Removal of cell effect using Cellular Deconvolution

To estimate the proportions of cell types in the bronchial brushes, we performed cellular deconvolution using scRNA-Seq signatures from 600 informative genes from a single cell dataset obtained from bronchial biopsies, including 15 cell types (alveolar macrophages, not alveolar macrophages, arterial endothelial cells, proliferating basal cells, not proliferating basal cells, ciliated cells, dendritic cells, fibroblasts, goblet cells, ionocytes, B cells, T cells, mast cells, monocytes, and submucosal secretory cells) as previously described²¹. We estimated the cell proportions using the

non-negative least squares (NNLS) method on our counts per million (CPM) normalised gene expression data. A new differential expression analysis was performed on the uniquely differentially expressed genes for sCOPD, comparing non-COPD controls versus (very) sCOPD, in which we also corrected for the cell proportions that made up more than two per cent of the cell populations in the samples on average in addition to the previous parameters and were different between COPD groups.

Classification of ICS-sensitive genes

Patients with more sCOPD often used high doses of inhaled corticosteroids and/or systemic prednisolone, providing a potential confounder in our study. To account for this confounder, we identified ICS responsive genes in the airways of COPD patients using the Groningen Leiden Universities and Corticosteroids in Obstructive Lung Disease study (GLUCOLD, ClinicalTrials.gov NCT00158847) as described previously²², and removed these ICS-responsive genes from our sCOPD gene signature. Bronchial biopsy RNA-Seq data of patients (n=19) at baseline and after six months of ICS treatment was compared using EdgeR while correcting for age, sex, current smoking status and packyears. A nominal p-value less than 0.05 was considered statistically significant.

Replication in nasal brushes

To investigate if the identified genes could be replicated in nasal brushes, we used matching nasal brushes of the same patients. We used gene set variation analysis (GSVA, version 1.42) to estimate the changes of the negatively and positively correlated genes separately per disease group²³. Next, we checked if the same genes could be replicated using the same method in an independent sCOPD cohort of nasal brushings (FAIR and NORM cohorts, ClinicalTrials.gov: NCT01351792 and NCT00848406, see below)²⁴.

We then performed an unbiased differential expression analysis using the matched nasal samples and the independent nasal brushings cohort separately, comparing sCOPD and non-COPD controls for the 219 genes unique for sCOPD, and performed a meta-analysis. Pathway analysis and protein interaction on the genes unique for sCOPD and replicated genes Protein interaction analysis and pathway enrichment within our sCOPD gene signature in the bronchus, as well as in the replicated in the nose, was performed using StringDB. We used the default settings of the program to construct the gene network and used the following databases for functional enrichment analysis: Gene Ontology (biological processes, molecular functions, cellular components) and Kegg pathways. The original 219 genes unique for sCOPD were used as input for the StringDB analysis. Additionally, we performed a separate StringDB analysis on the genes that were FDR significant in the meta-analysis of the nasal brushings.

2.5 – Results

Clinical characteristics

In the current study, we investigated the differences in bronchial gene expression between patients with mCOPD (n=23), sCOPD (n=123) and non-COPD controls (n=23). There was no significant difference in age across the groups. However, there was a difference in the male/female ratio (non-COPD: 52% male, mCOPD: 78%, sCOPD: 29%) and pack-years (non-COPD: 31.1±20.6 pack-years (mean±sd), mCOPD: 66.6±62.9, sCOPD 39.1±18.3). We corrected for these two confounding factors in our models. Additionally, non-COPD participants did not use ICS, while mCOPD participants used 291±527µg beclomethasone equivalent, and sCOPD used 620±877µg. Clinical characteristics of included subjects are presented in Table 2.1.

Table 2.1) Patient demographics

	Non-COPD controls	COPD Patients	
		mCOPD	sCOPD
Patients n	23	23	123
Sex, male, n (%)	12 (52)	18 (78)	36 (29)
Age, years, median (IQR)	60 (52,63)	62 (57,65)	60 (56,64)
Smoking status = Ex-smoker (%)	23 (100)	23 (100)	122 (99)
Packyears (median (IQR))	28 (14,48)	56 (23,74)	36 (30,46)
Years of cessation (median (IQR))	9 (2,21)	9 (5,18)	5 (2,9)
FEV₁ %pred (median (IQR))	102 (90,110)	77 (68,89)	24 (20,29)
FEV₁/FVC (%) (median (IQR))	0.73 (0.70,0.77)	0.55 (0.5,0.59)	0.28 (0.24,0.32)
RV %pred (median (IQR))	106 (95,111)	110 (102,115)	239 (215,260)
ICS before inclusion = true (%)	0 (0.0)	11 (48)	108 (99)
SGRQ Part 1 Q2 (median (IQR))	5 (5,5)	5 (3,5)	4 (2,5)
FEV ₁ : forced expiratory volume in 1 second; FVC: forced vital capacity; ICS: Inhaled Corticoid Steroids			

Identification of genes common for COPD and unique for sCOPD

We identified 435 genes differentially expressed between sCOPD patients and non-COPD controls (FDR <0.05, fold change (FC) $\geq \pm 2$). Of these, 213 genes showed a higher and 222 genes showed a lower expression in sCOPD patients. A volcano plot and heatmap are depicted in Figure 2.1a and Supplementary Figure 2.2, respectively.

Next, we performed a differential gene expression between non-COPD controls and mCOPD participants. Here we took a more lenient FDR cut-off of FDR <0.25 and identified 123 genes differentially expressed in both mCOPD and sCOPD, which should thus not be considered unique for sCOPD (see Supplementary Table 2.1). A volcano plot and a heatmap are depicted in Figure 2.1b and Supplementary Figure 2.3, respectively. After removing these 123 genes from the 435 sCOPD gene list, we were left with 312 potentially unique genes for sCOPD.

Next, we directly compared sCOPD versus mCOPD and found that 285 of the 312 genes were differentially expressed between mCOPD and sCOPD (FDR <0.05). Of these, 118 genes were higher expressed in sCOPD compared to mCOPD, while 167 genes were lower expressed. A volcano plot and a heatmap are depicted in Figure 2.1c and Supplementary Figure 2.4, respectively. The top three remaining higher and lower expressed genes are shown in Figure 2.1d–i.

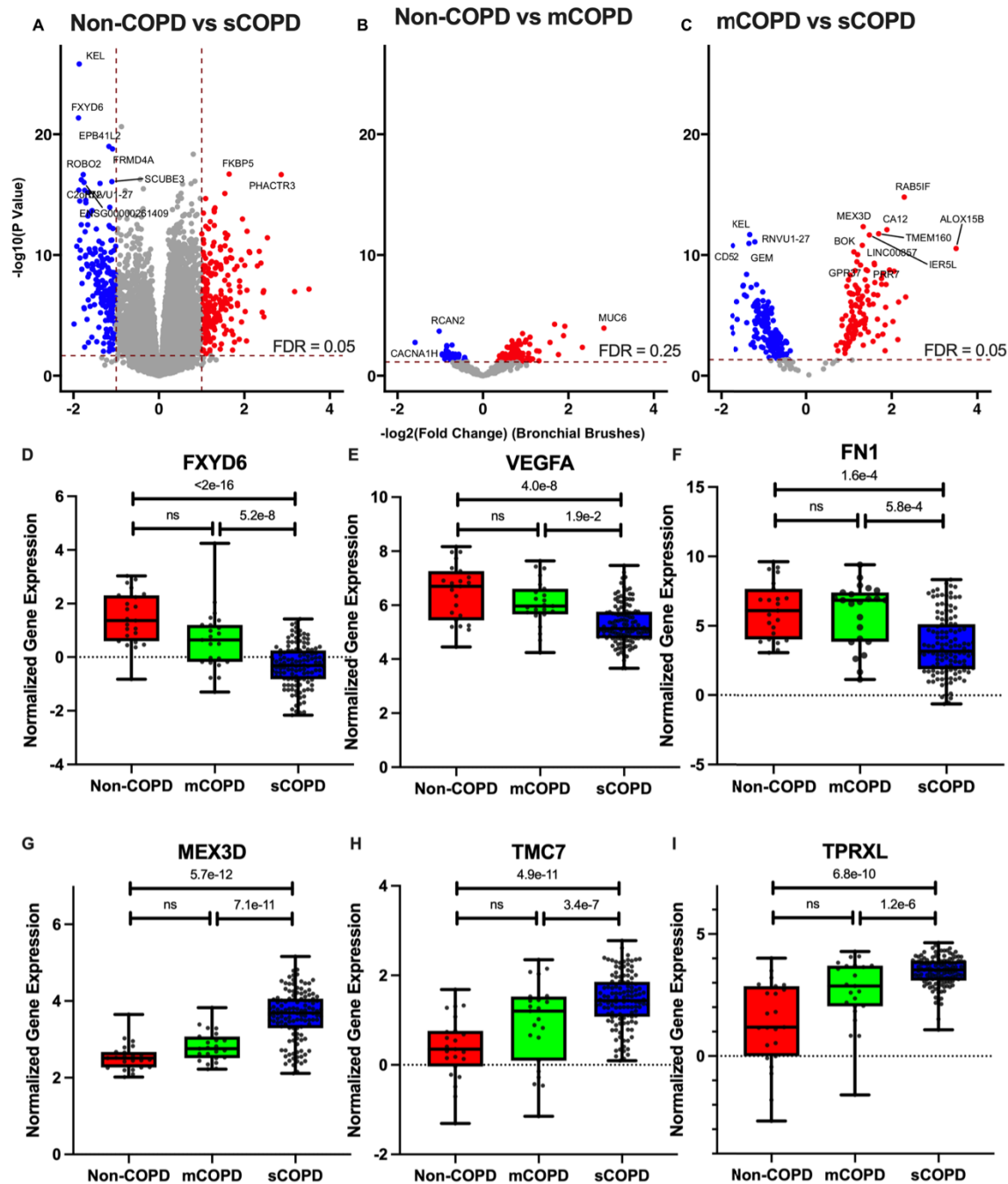


Figure 2.1) Differential gene expression of bronchial brushes of severe COPD (sCOPD) and non-COPD controls. a) Differential expression analysis between sCOPD compared to non-COPD controls. Biased differential analysis between b) mild to moderate COPD (mCOPD) compared with non-COPD controls and c) sCOPD and mCOPD. d–f) The three higher and g–i) lower expressed genes between sCOPD compared to non-COPD controls.

Effects of cell-type proportions

We then investigated whether there was a difference in cellular composition between the three patient groups. Within our bronchial brushings, we found goblet cells to be the most common cell type (median: 86.5% (IQR: 74.2–93.7)), followed by ciliated cells (11.7% (5.9–19.9)) and the basal cells (0.0% (0.0–2.6)) (Figure 2.2a). When comparing sCOPD to non-COPD controls, we found that there was a significantly lower proportion of ciliated and basal and a higher proportion of goblet cells (Figure 2.2b–d). Adjusting for basal, ciliated and goblet cell proportions in our first differential

expression analysis showed that 262 out of 285 sCOPD-associated genes (92%) were not affected by cell proportions (see Supplementary Table 2.2).

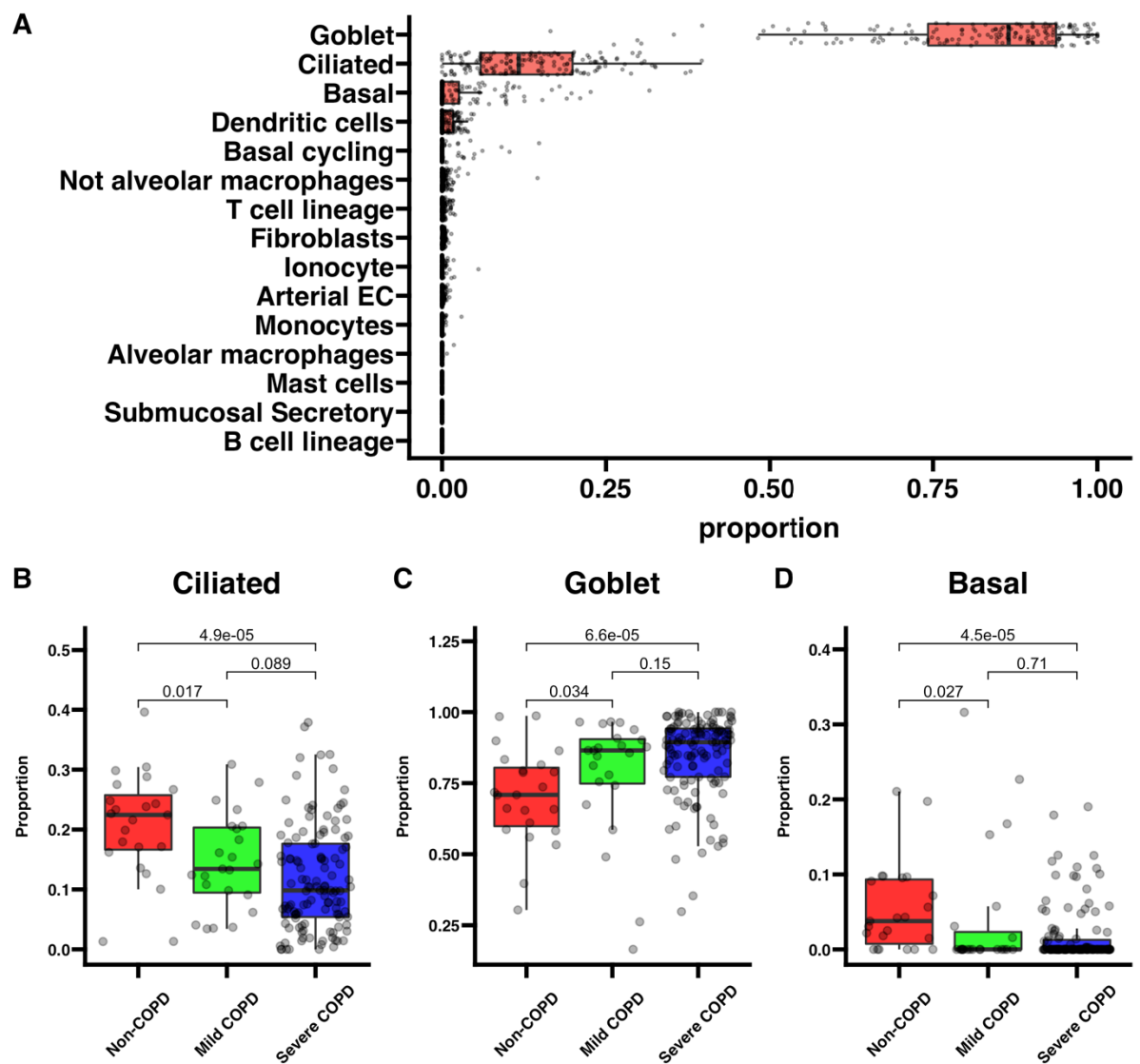


Figure 2.2) Cellular deconvolution and cell proportions. Cellular deconvolution of bronchial brushings using non-negative least squares regression (NNLS). a) Overall composition of bronchial brushings. Separation of cell-type composition by disease status for b) ciliated, c) goblet and d) basal. A Kruskal–Wallis test was conducted over all the groups and a Mann–Whitney U-test was conducted between the groups. EC: endothelial cells.

The effect of inhaled corticosteroids

We used an existing dataset of bronchial biopsies obtained from COPD patients before and after 6 months of treatment with either ICS ± long-acting β_2 -agonists (LABA) or placebo to define ICS-responsive genes. This resulted in a list of 2691 ICS-response genes, of which 43 were present in our sCOPD gene signature. The complete list of ICS-responsive genes can be found in Supplementary Table 2.3. These 43 genes were therefore removed from our sCOPD gene signature (see Supplementary Table 2.4).

Specific bronchial epithelium gene signature for sCOPD

Our final specific bronchial epithelium gene signature for sCOPD consisted of 219 genes that are uniquely differentially expressed in sCOPD compared to non-COPD controls. Of these 219 genes, 104 genes were higher expressed in sCOPD, with the top 10 most significant genes being: MEX3D, LINC00857, CEACAM5, TMC7, FNDC10, TPRXL, NETO2, SERPINB5, CALML3 and MUC12. A total of 115 genes were lower in sCOPD, with the 10 most significant genes being: FXYD6, GGTA1P, GEM, CPED1, KCNJ5, VEGFA, JAKMIP2, DOK2, KMO and GPR174. A list of the top 10 genes more and less expressed in sCOPD can be found in Table 2.2; the complete list of 219 genes can be found in Supplementary Table 2.5.

Table 2.2) Top 20 genes more and less expressed in severe COPD (sCOPD) that were specific for sCOPD, not influenced by basal, ciliated and secretory cell types and 6 months of inhaled corticosteroid usage

Ensembl Gene ID	HGNC symbol	Log(FC)	Log(CPM)	F	P-value	FDR
ENSG00000237523	LINC00857	1.41	9.11	80.38	1.35E-15	1.24E-13
ENSG00000181588	MEX3D	1.44	11.90	81.52	9.28E-16	1.24E-13
ENSG00000105388	CEACAM5	1.87	16.44	78.04	2.90E-15	2.00E-13
ENSG00000228594	FNDC10	1.23	9.11	74.92	8.19E-15	3.77E-13
ENSG00000170537	TMC7	1.35	9.72	74.97	8.06E-15	3.77E-13
ENSG00000180438	TPRXL	1.92	11.73	72.28	1.99E-14	7.86E-13
ENSG00000171208	NETO2	1.67	10.67	69.99	4.35E-14	1.33E-12
ENSG00000206075	SERPINB5	1.34	14.10	68.89	6.35E-14	1.75E-12
ENSG00000178363	CALML3	2.48	11.34	68.40	7.52E-14	1.89E-12
ENSG00000205277	MUC12	1.50	11.07	66.74	1.34E-13	3.08E-12
ENSG00000137726	FXYD6	-1.58	8.63	66.43	1.49E-13	3.16E-12
ENSG00000204136	GGTA1P	-1.34	8.11	57.48	3.62E-12	4.17E-11
ENSG00000164949	GEM	-1.57	10.22	55.19	8.41E-12	8.60E-11
ENSG00000106034	CPED1	-1.24	8.67	47.87	1.32E-10	9.87E-10
ENSG00000120457	KCNJ5	-1.33	8.34	42.97	8.93E-10	5.48E-09
ENSG00000112715	VEGFA	-1.23	13.97	39.65	3.35E-09	1.52E-08
ENSG00000176049	JAKMIP2	-1.18	8.64	39.43	3.67E-09	1.63E-08
ENSG00000147443	DOK2	-1.28	9.16	39.02	4.32E-09	1.86E-08
ENSG00000117009	KMO	-1.33	8.93	37.56	7.82E-09	3.13E-08
ENSG00000147138	GPR174	-1.45	9.32	37.38	8.41E-09	3.32E-08

HGNC: HUGO Gene Nomenclature Committee gene symbol; FC: fold change; CPM: counts per million; FDR: false discovery rate.

Representation of the sCOPD signature in nasal brushings

We assessed whether our bronchial epithelium sCOPD gene signature was also present in nasal brushings. Here we first compared the bronchial signature in matched nasal brushings that were collected at the same visit. Using gene set variation (GSVA) analysis, we demonstrated a significant representation of the bronchial gene signature in the nasal brushes that were lower in sCOPD compared to non-COPD controls and mCOPD, while no significant representation was observed for the gene set that was higher in sCOPD (Figure 2.3a and b).

Next, we checked whether our results showed similar representation in an independent nasal gene expression dataset comparing sCOPD to control. (Clinical characteristics of included subjects are presented in Supplementary Table 2.6.) The bronchial gene signature that was lower in sCOPD was similarly represented in that dataset, whereas the bronchial gene set that was higher was not (Figure 2.3c and d).

We then performed a meta-analysis on the 219 genes unique for sCOPD in the matched nasal brushings and the independent nasal brushings dataset to assess similar representation in the nose on the gene level. In this meta-analysis, 83 genes in both datasets were significantly associated with sCOPD (meta-FDR <0.05) in the same direction in both nasal cohorts. 42 genes were lower expressed in the sCOPD patients compared to controls (top five genes: EPB41L2, FRMD4A, GGTA1P, GEM and CPED1), and 41 genes were higher expressed (top five genes: MEX3D, TMEM132B, PCSK1N, PRR7 and MESP1) (see Supplementary Table 2.7).

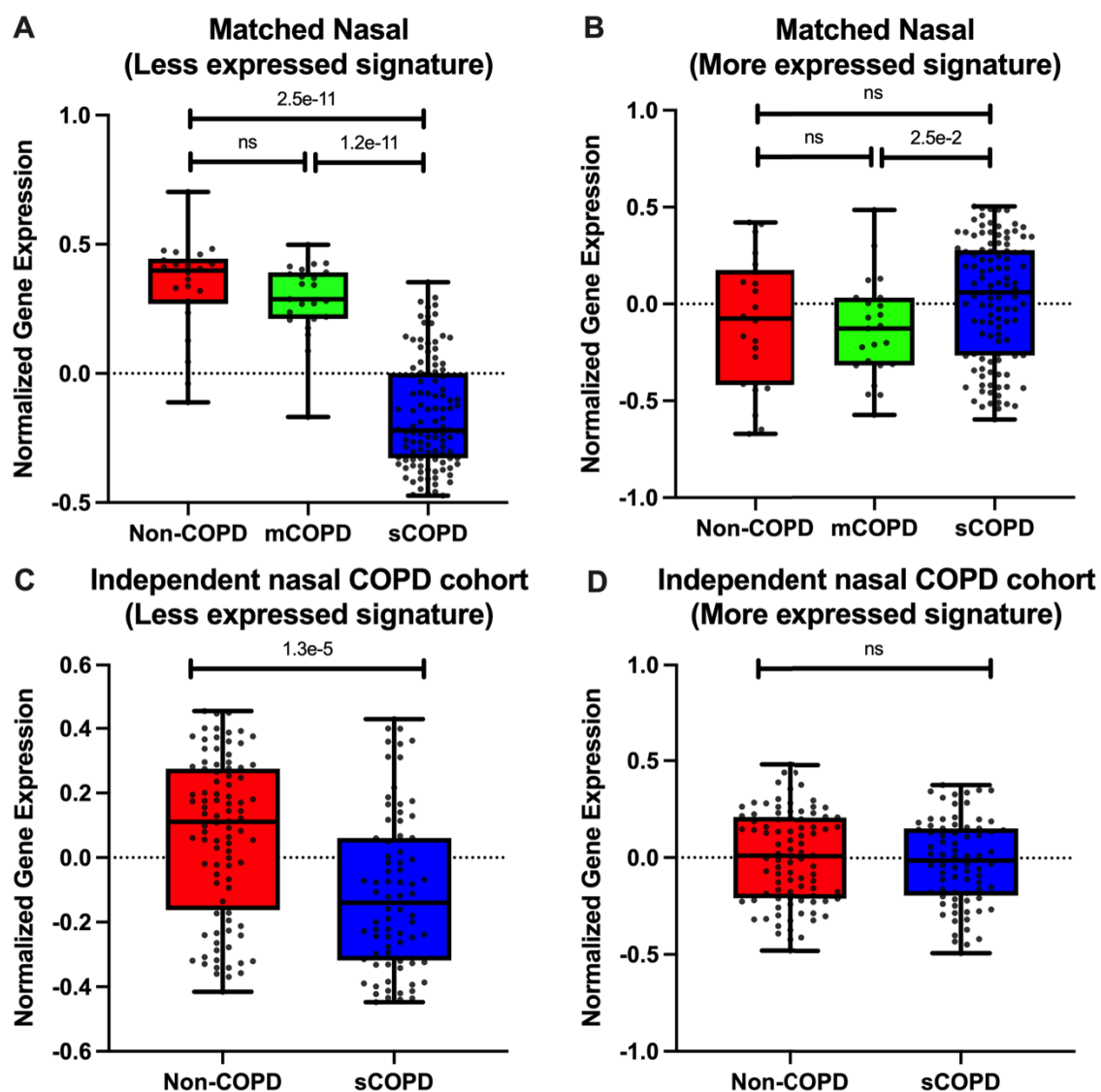


Figure 2.3) Reflection of the bronchial severe COPD (sCOPD) signature in the nose. Comparison of the sCOPD signature a) less and b) more expressed in matched nasal brushings. Reflection of the sCOPD signature c) less and d) more expressed in nasal brushing from sCOPD (n=76) and non-COPD controls (n=92). Reflection of the sCOPD signature in normal COPD

(n=38) and non-COPD controls (n=30). A Mann–Whitney U-test was conducted ($p < 0.05$ was considered significant). mCOPD: mild to moderate COPD.

Pathway analysis

StringDB analysis using default settings of the 219 unique sCOPD genes demonstrated significant enrichment of pathways related to the extracellular matrix (ECM), ECM binding and collagen (see Supplementary Table 2.8). Spearman correlation of the GSVA enrichment scores of these pathways with single cell proportions showed a positive correlation between basal and ciliated cells with ECM and collagen-related pathways, and a negative correlation with goblet cells (see Supplementary Figure 2.5).

Within the StringDB network, FN1 and VEGFA were key regulatory genes, both with 25 connections (see Supplementary Table 2.9). 15 connecting genes overlapped between the two key regulatory genes (SPARC, TWIST1, LIF, SEMA3E, FOS, PTHLH, PECAM1, ABCB1, BDNF, CEACAM5, CX3CR1, CYR61, DCN, DKK1 and EGR1). Both key regulatory networks were related to tissue development, while the network surrounding FN1 was also involved with collagen and ECM binding. The overlapping networks showed enrichment in the regulation of developmental processes, tissue development and cell division.

Functionally related to FN1 was a network surrounding DCN. DCN, lower expressed in sCOPD, was at the centre of a cluster of eight genes (COL6A2, SPARCL1, C1QA, B3GAT1, FAGFA, FN1, SPARC and MGP), together involved in collagen binding and the ECM (Figure 2.4). We used the human single cell lung atlas to determine which cell types expressed FN1 and VEGFA most²⁵. FN1 is expressed by many cells in the lung, including (alveolar) fibroblasts, (alveolar) macrophages, and endothelial and smooth muscle cells, and is lowly expressed in epithelial cells, whereas VEGFA is mostly expressed in airway basal cells, goblet cells and to a lesser extent in other cells (Supplementary Figure 2.6).

Other regulatory genes in the network with >10 connections were FOS (15 connections), EGR1 (14 connections), FCGR3A (13 connections), and BDNF and NR4A1 (both 12 connections). The StringDB analysis of the 83 genes also represented in the nose showed a very similar network with again FN1 and VEGFA as the key genes with the most connections (both 16 connected genes) and enrichment of pathways involved in the ECM, and collagen binding, cell adhesion and cell signalling, all in the same direction as the same enriched pathway in the bronchus (Supplementary Figure 2.7).

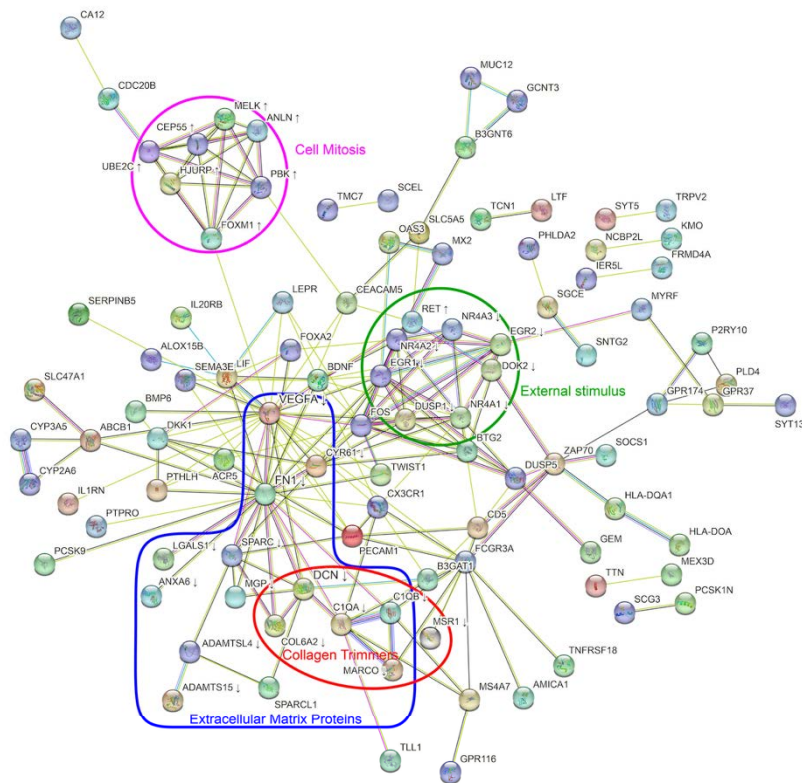


Figure 2.4) StringDB analysis of the genes unique for severe COPD (sCOPD). Arrows indicate if the gene was more (↑) or less (↓) expressed in sCOPD. The purple grouping contains genes involved in cell mitosis; the green grouping contains genes involved in external stimuli; the blue grouping contains genes involved in the extracellular matrix; the red grouping contains genes involved with collagen trimmers.

Bootstrapping in the sCOPD subgroup

Since the number of subjects in the sCOPD group was much larger compared to mCOPD and non-COPD controls, we performed bootstrapping in the sCOPD subgroup. To this end, we randomly sampled 23 sCOPD cases with sCOPD and compared their expression of the 435 sCOPD-associated genes to the expression in non-COPD controls. We performed 1000 iterations and found that on average, 63.2% of the 435 genes were replicated.

2.6 – Discussion

We identified a specific bronchial epithelial gene expression signature for sCOPD, consisting of 219 genes. This sCOPD signature is different from mCOPD, supporting our hypothesis that sCOPD represents a distinct disease phenotype. Pathway analyses demonstrated that sCOPD-associated genes are mainly involved in immune response, developmental processes and ECM binding. Protein-interaction networks indicate VEGFA and FN1 as potential key drivers in sCOPD. Additionally, the gene signature that was lower in sCOPD in bronchial brushes was also represented in matched nasal brushings as well as nasal samples from an independent sCOPD cohort. Of interest, the signature-related gene set that was present in both nasal cohorts was again centred around VEGFA and FN1.

The two key genes driving the sCOPD gene signature were VEGFA and FN1, both lower expressed in sCOPD. VEGFA, expressed in basal and goblet cells²⁵, is a key growth factor for building lung architecture²⁶ that needs VEGFR2, expressed in endothelial cells, to form pulmonary capillaries²⁷, VEGFA is less expressed in the lower respiratory tract of smokers and even lower in smokers with COPD²⁸. It was previously shown that loss of VEGFA leads to endothelial cell apoptosis and is associated with emphysema²⁸⁻³², a hallmark of sCOPD. This might be due to reduced blood supply

from small capillaries associated with loss of alveolar septa. Our findings suggest a role for VEGFA in the development of sCOPD, not only in the alveoli, as shown by previous studies²⁸, but also in the bronchus and the nose. Therefore, the lower VEGFA expression in bronchial and nasal brushes could reflect similar changes in the parenchyma leading to emphysematous destruction and lack of alveolar and endothelial repair due to the lack of VEGFA.

The other key regulatory gene identified in our sCOPD gene signature was FN1. FN1 is a glycoprotein expressed by many cells in the lung, including fibroblasts, monocytes, and endothelial and smooth muscle cells, and is lowly expressed in epithelial cells²⁵. FN1 is important during the development of the lung, barely detectable during adulthood in healthy lungs³³, but highly upregulated during tissue repair³³. Therefore, lower FN1 expression in the bronchial and nasal brushes in sCOPD patients may reflect a disturbed epithelial repair response and may reflect similar events in the parenchyma leading to a lack of repair leading to emphysematous tissue destruction present in these patients. Of interest is the connection between FN1 and VEGFA in our network since it has been shown that FN1, when bound to VEGFA, is necessary to promote VEGFA-induced endothelial cell proliferation and migration³⁴. Next to FN1 and VEGFA, the protein-interaction network included 13 genes involved in ECM organisation and 11 specifically involved in collagen binding (HMCN2, LGALS1, C1QB, C1QA, ANXA6, SPARCL1, DCN, MGP, SPARC, COCH and FN1). Five of these genes (C1QB, ANXA6, DCN, MGP and SPARC) were previously found to interact with FN1³⁵⁻³⁷, and two genes, SPARC and DCN, also interacted with VEGFA. SPARC, lower expressed in sCOPD, encodes for the secreted protein osteonectin and is lowly expressed in basal cells and mainly in endothelial cells, fibroblasts and macrophages²⁵. It binds to VEGFA and interferes with its binding to VEGFR1³⁸ and, in doing so, inhibits the proliferation of endothelial cells. SPARC also has a role in the regulation of secretion rates of fibronectin^{39, 40}. One of the genes that interact with SPARC, FN1 and VEGFA is DCN. This gene encodes for decorin, is lower expressed in sCOPD and is mainly expressed in basal cells and fibroblasts²⁵. It is a protein with an important role in collagen cross-linking and fibrillogenesis. Decreased DCN expression in the lung might affect collagen tensile strength and binding of ECM proteins, resulting in changed cell fate and function. Lastly, ANXA6, lower expressed in sCOPD, was previously found to inhibit the secretion of FN1⁴¹, adding another mechanism to control fibronectin function and angiogenesis.

Clearly, part of the sCOPD-related gene expression changes is also present in the nose, providing support for the use of nasal brushes as a proxy for the lung. Here we identified a gene signature in the nose that reflects bronchial changes specific to sCOPD. This study confirms and extends what previous findings show that severe COPD can be differentiated from non-COPD controls using nasal gene expression. Furthermore, our finding of an overlapping signature in the bronchus and the nose supports the theory of a single transcriptional profile throughout the airways^{42, 43}. However, it should be noted that not all genes are concordantly expressed. We only replicated this finding for genes with a lower expression in sCOPD. Our findings are important as it suggests that the nose may serve as an easily accessible biomarker in COPD, at least for a subset of genes and biological pathways.

There were some limitations to the current study. One of the limitations was the size of the non-COPD and mCOPD groups, which could have influenced the results. However, bootstrapping of the sampling of the sCOPD group showed similar results as, on average, 63.2% of the genes could be replicated. Furthermore, most patients with sCOPD used high ICS doses, which may affect gene expression levels. Because of the uneven distribution, with non-COPD controls not using ICS and mild-moderate COPD patients using much lower doses, it is difficult to adjust for this variable in the statistical analyses. To account for this possible confounding factor, we removed all genes previously shown to be sensitive to treatment with corticosteroids, as previously identified in a placebo-

controlled longitudinal study⁴⁴. This way, we made sure that the higher ICS dose in sCOPD did not lead to the identification of sCOPD-associated differentially expressed genes. However, we could not rule out the possibility that we removed genes relevant to the development of sCOPD. Additionally, in line with previous observations, a higher percentage of the sCOPD patients was female⁴⁵. Although we adjusted for this in our linear models, we cannot entirely rule out the possibility that some of the observed sCOPD-associated gene expression differences are due to an imbalance in sex. Lastly, although transcriptomic data are a good indicator for changes on the protein level in most cases, it is not a replacement. Future studies are needed to investigate the protein levels of FN1 and VEGFA, which could be done in nasal epithelial lining fluids collected from severe COPD patients and controls. This exploration would further clarify how FN1 and VEGFA are involved in the pathogenesis of severe COPD. In addition, it would be of interest to assess their utility as biomarkers for disease severity and progression.

In conclusion, we found a unique sCOPD gene signature that was indicative of an abnormal epithelial repair response, impaired fibroblast function and decreased angiogenesis, which was retained throughout the airways in the nose. This supports the hypothesis that sCOPD comprises a partly different pathology compared to the majority of patients with mCOPD driving the specific disease phenotype. Moreover, as part of the sCOPD-related gene expression, changes are also present in the nose, supporting the potential for biomarker development based on nasal brushes in COPD.

2.7 – Funding

Funded by the PPP-allowance by Health Holland (LSH) and the UMCG. This collaboration project is co-financed by the Ministry of Economic Affairs and Climate Policy by means of the PPP-allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships.

2.8 – Supplementary Material

Methods

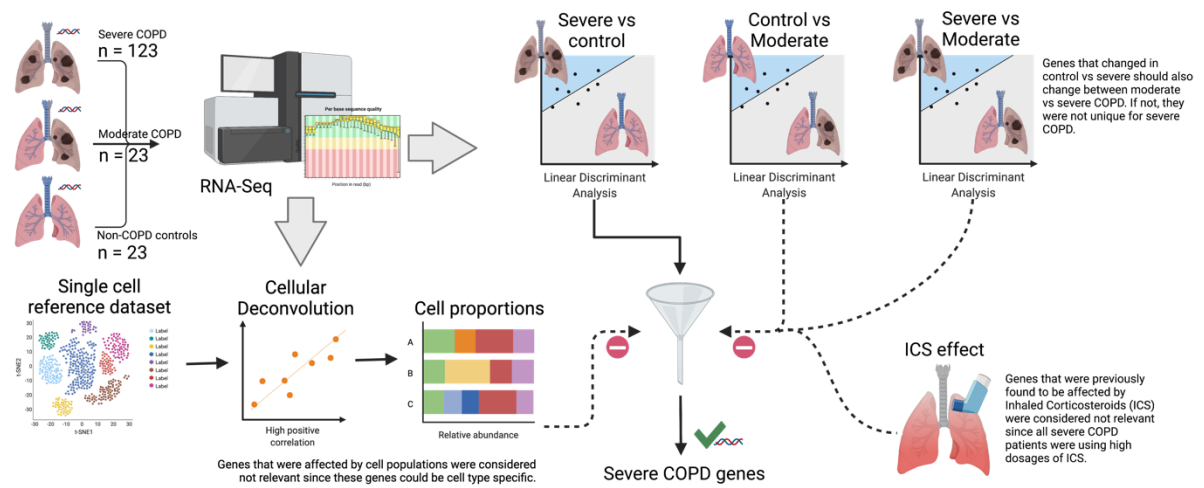
Description of the COPD replication cohort for nasal brushings

The FAIR cohort (ClinicalTrials.gov: NCT01351792) consists of COPD patients who participated in a multicenter, randomised, longitudinal study. Participants were excluded if they were diagnosed with asthma, were pregnant, were treated with long-term oxygen therapy, had a clinically unstable concurrent disease (as judged by the investigator), had a COPD exacerbation within two months prior to the first study visit and reversibility of the FEV1>15% and >200mL of initial FEV1.

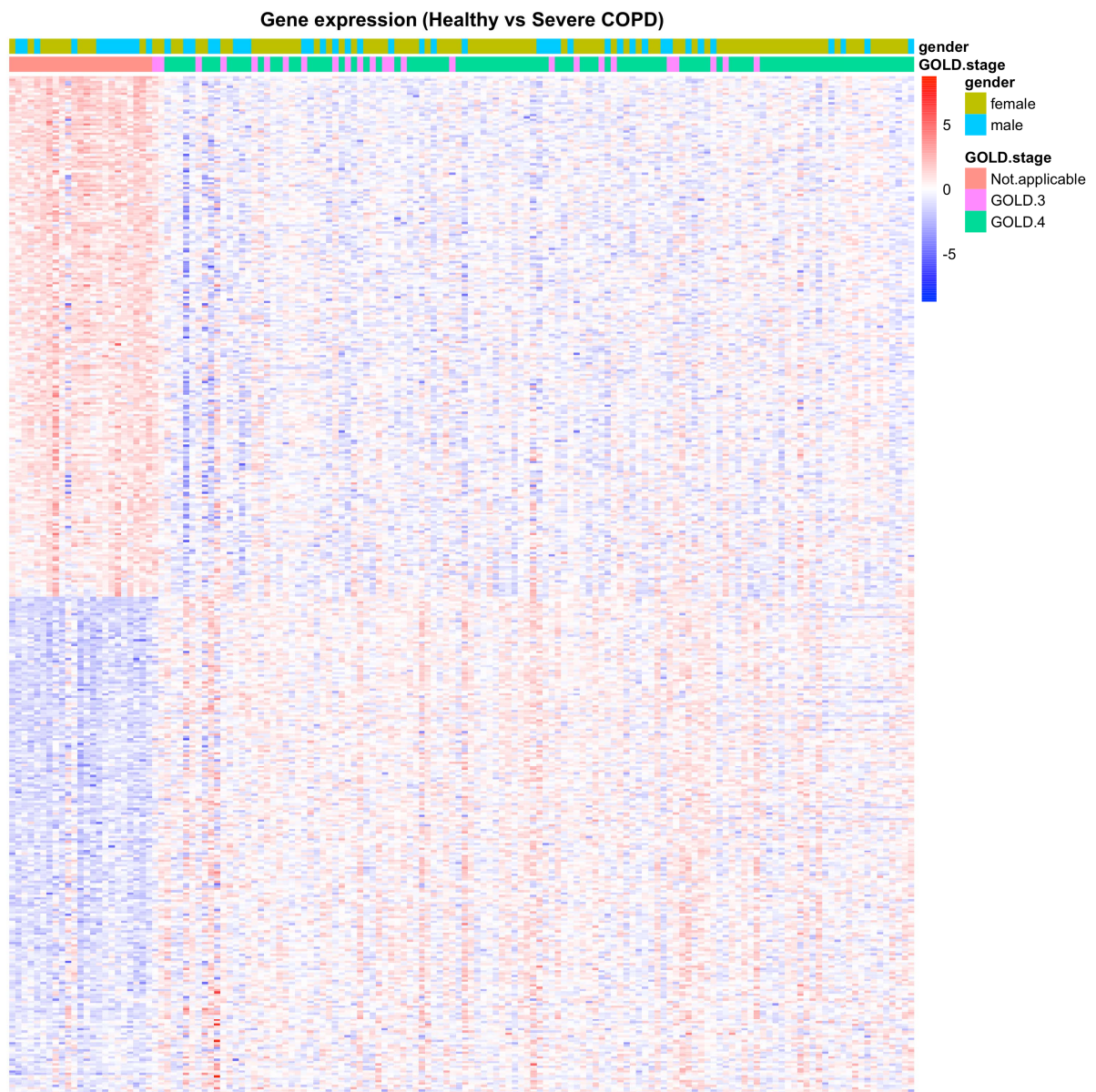
Description of the non-COPD replication cohort for the nasal brushings

The NORM cohort (ClinicalTrials.gov: NCT00848406) consists of non-COPD participants who participated in a cross-sectional study and had a pulmonary function, defined by an FEV1/FVC > lower limit of normal, absence of bronchial hyperresponsiveness and reversibility of the FEV1% predicted to salbutamol <10%. Exclusion criteria were the use of inhaled or oral corticosteroids, upper respiratory tract infection, clinically unstable concurrent disease (as judged by the investigator) or pregnancy.

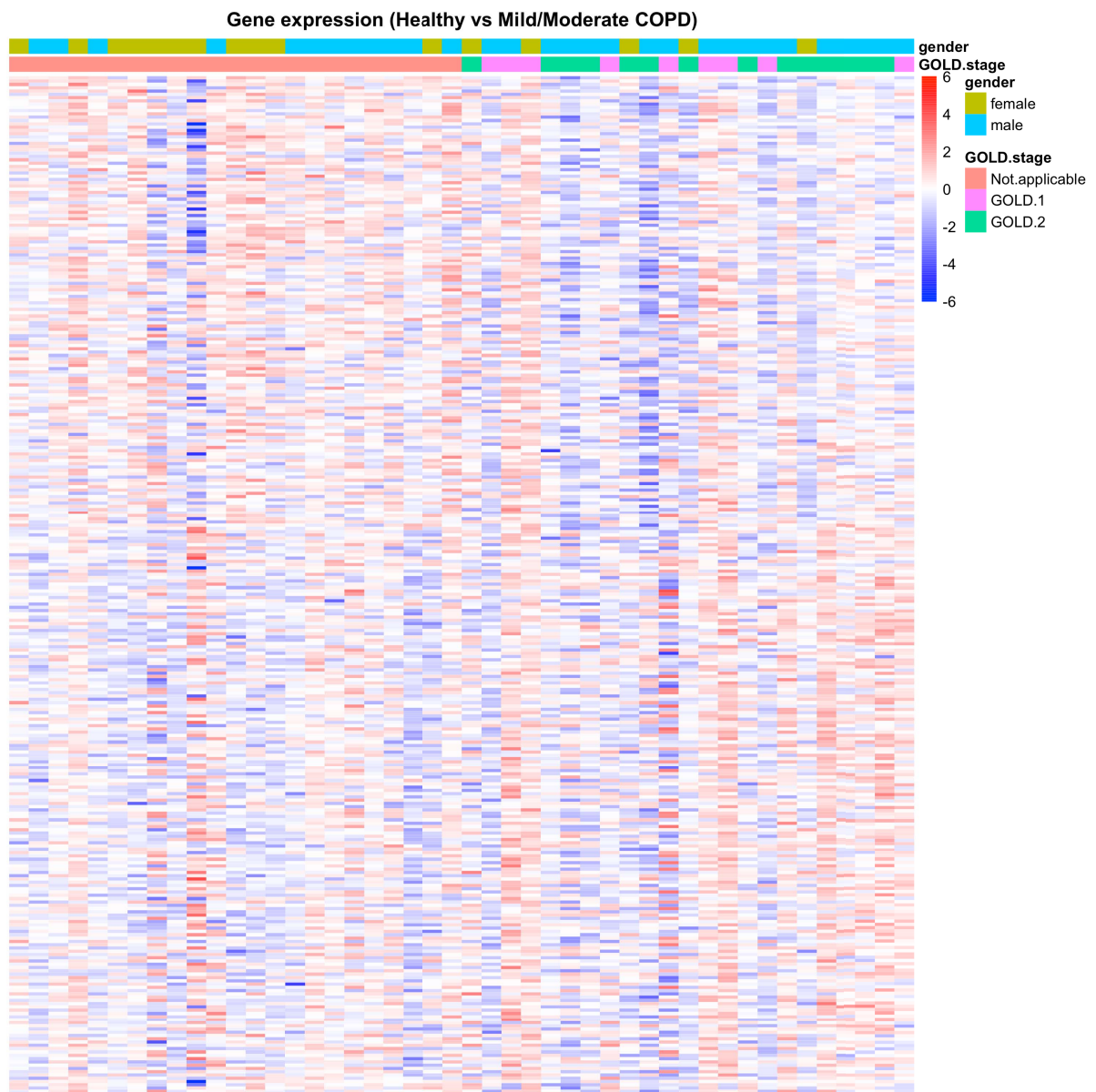
Figures



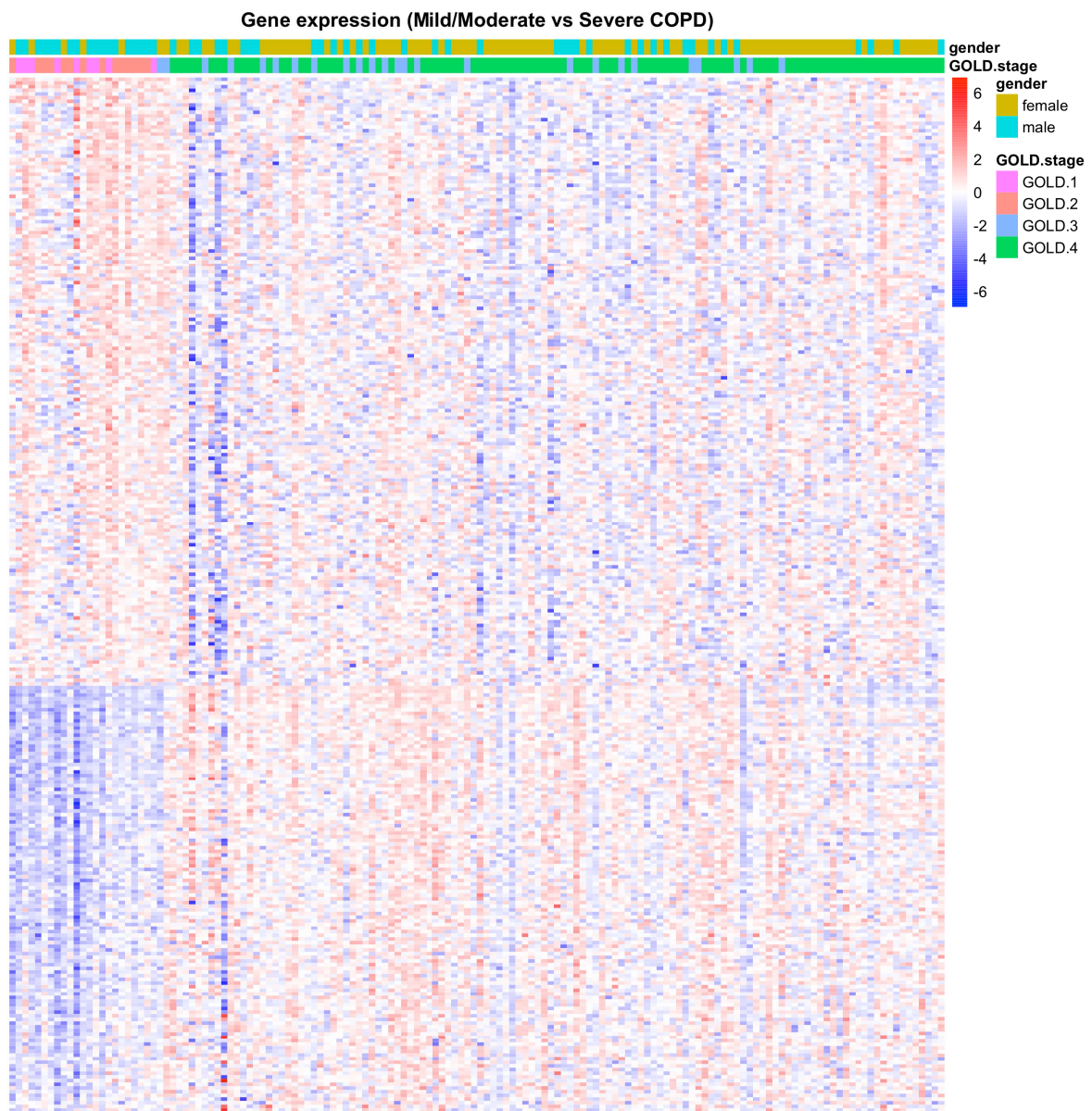
Supplementary Figure 2.1) Flowchart of statistical methods



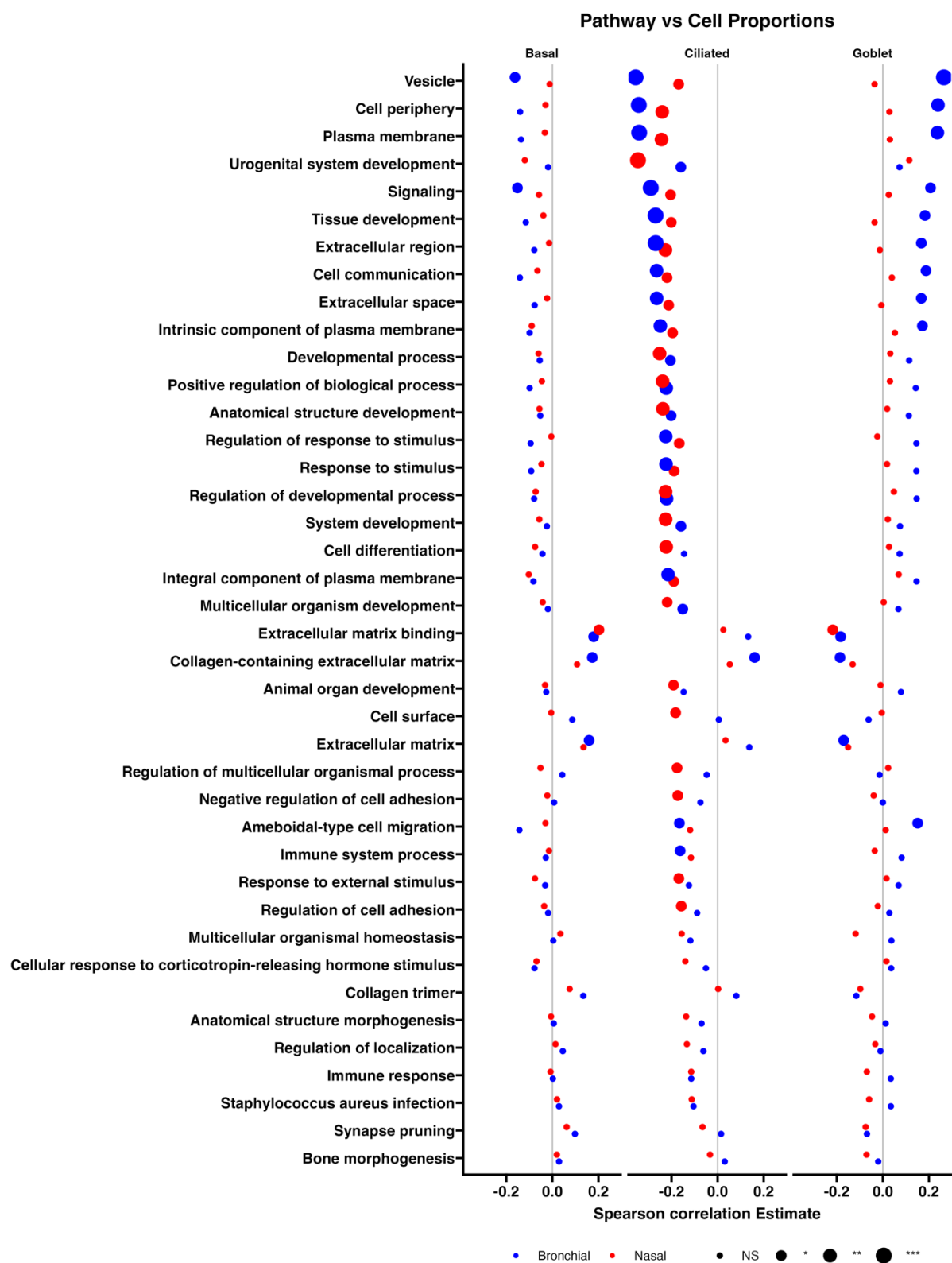
Supplementary Figure 2.2) Heatmap of differential gene expression between non-COPD vs sCOPD



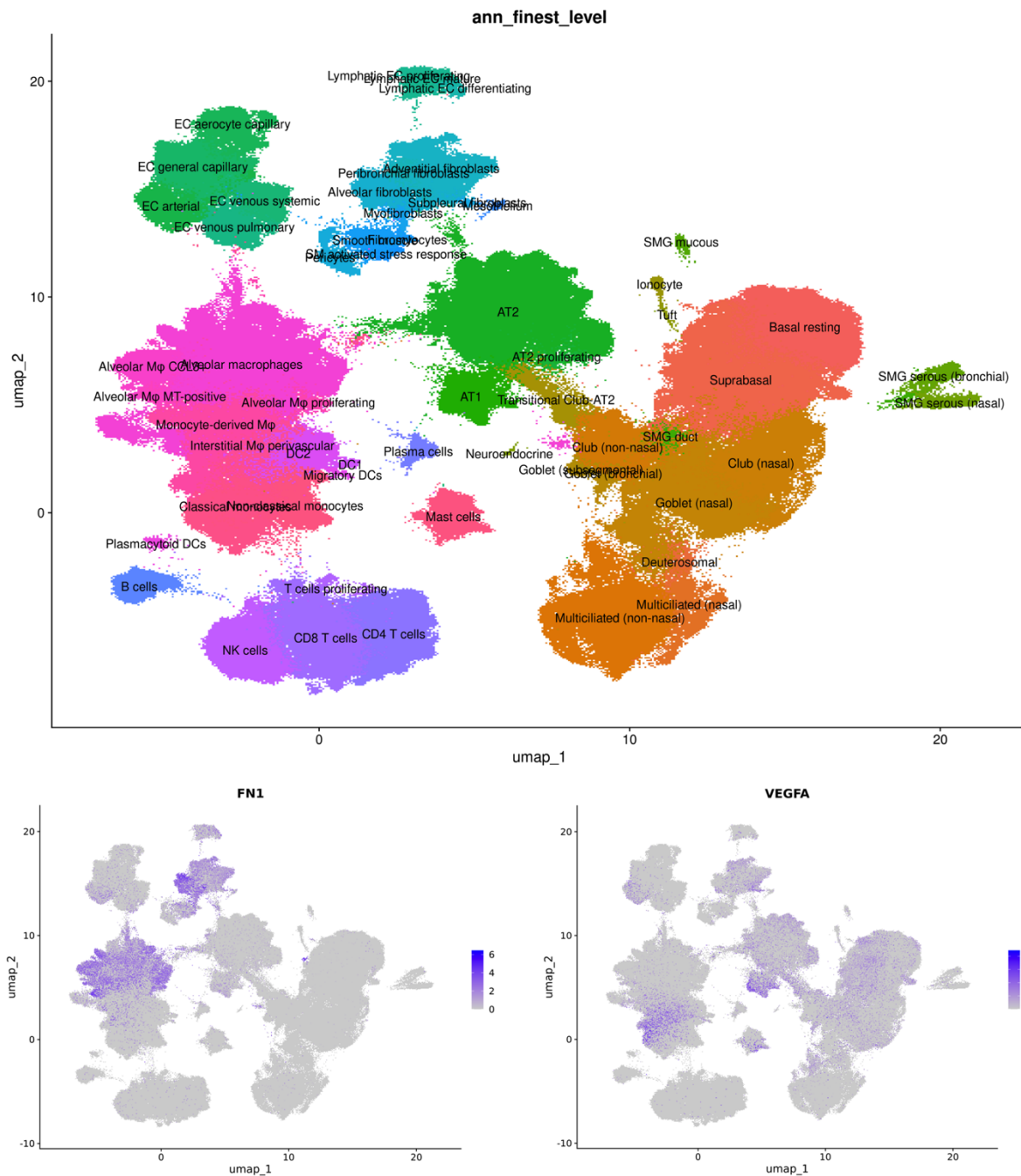
Supplementary Figure 2.3) Heatmap of differential gene expression between non-COPD vs mCOPD



Supplementary Figure 2.4) Heatmap of differential gene expression between mCOPD vs sCOPD



Supplementary Figure 2.5) Spearman correlation estimate and p-value of the GSVA score of the significant pathways versus the cellular proportions in bronchial brushes and nasal brushes.



Supplementary Figure 2.7) Gene expression of FN1 and VEGFA in an independent single-cell dataset (human lung cell atlas)

Tables

See <http://www.GitHub.com/vanNijnatten/PhD-Thesis> .

2.9 – References

1. Vogelmeier, C.F., et al., *Goals of COPD treatment: Focus on symptoms and exacerbations*. Respir Med, 2020. **166**: p. 105938.
2. Eisner, M.D., et al., *An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease*. Am J Respir Crit Care Med, 2010. **182**(5): p. 693-718.
3. WHO, *Global status report on noncommunicable diseases 2014*. 2010: WHO.
4. Hogg, J.C., *Pathophysiology of airflow limitation in chronic obstructive pulmonary disease*. Lancet, 2004. **364**(9435): p. 709-21.
5. Fletcher, C. and R. Peto, *The natural history of chronic airflow obstruction*. Br Med J, 1977. **1**(6077): p. 1645-8.
6. Goldklang, M.P., S.M. Marks, and J.M. D'Armiento, *Second hand smoke and COPD: lessons from animal studies*. Front Physiol, 2013. **4**: p. 30.
7. Kurmi, O.P., et al., *COPD and chronic bronchitis risk of indoor air pollution from solid fuel: a systematic review and meta-analysis*. Thorax, 2010. **65**(3): p. 221-8.
8. Hoffmann, D. and I. Hoffmann, *Letters to the Editor - Tobacco smoke components*. Beiträge zur Tabakforschung International/Contributions to Tobacco Research, 1998. **18**(1): p. 49-52.
9. Chang, J.T., et al., *Prediction of COPD risk accounting for time-varying smoking exposures*. PLoS One, 2021. **16**(3): p. e0248535.
10. Raherison, C. and P.O. Girodet, *Epidemiology of COPD*. Eur Respir Rev, 2009. **18**(114): p. 213-21.
11. Lange, P., et al., *Effects of smoking and changes in smoking habits on the decline of FEV1*. European Respiratory Journal, 1989. **2**(9): p. 811-816.
12. Silverman, E.K., et al., *Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis*. Am J Respir Crit Care Med, 1998. **157**(6 Pt 1): p. 1770-8.
13. Silverman, E.K., *Genetic Epidemiology of COPD*. Chest, 2002. **121**(3 Suppl): p. 1S-6S.
14. Barnes, P.J., S.D. Shapiro, and R.A. Pauwels, *Chronic obstructive pulmonary disease: molecular and cellular mechanisms*. Eur Respir J, 2003. **22**(4): p. 672-88.
15. Wouters, E.F., *The burden of COPD in The Netherlands: results from the Confronting COPD survey*. Respir Med, 2003. **97 Suppl C**: p. S51-9.
16. Fromer, L. and C.B. Cooper, *A review of the GOLD guidelines for the diagnosis and treatment of patients with COPD*. Int J Clin Pract, 2008. **62**(8): p. 1219-36.
17. Miravitlles, M., et al., *A review of national guidelines for management of COPD in Europe*. Eur Respir J, 2016. **47**(2): p. 625-37.
18. Imkamp, K., et al., *Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci*. J Allergy Clin Immunol, 2018. **142**(1): p. 314-317 e15.
19. Reddy, K.D., et al., *Current Smoking Alters Gene Expression and DNA Methylation in the Nasal Epithelium of Patients with Asthma*. Am J Respir Cell Mol Biol, 2021. **65**(4): p. 366-377.
20. van den Berge, M., et al., *Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity*. Thorax, 2014. **69**(1): p. 14-23.
21. Aliee, H., et al., *Determinants of expression of SARS-CoV-2 entry-related genes in upper and lower airways*. Allergy, 2022. **77**(2): p. 690-694.
22. Tasena, H., et al., *microRNA-mRNA regulatory networks underlying chronic mucus hypersecretion in COPD*. Eur Respir J, 2018. **52**(3).
23. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
24. Faiz, A., et al., *Identifying a nasal gene expression signature associated with hyperinflation and treatment response in severe COPD*. Sci Rep, 2020. **10**(1): p. 17415.

25. Deprez, M., et al., *A Single-Cell Atlas of the Human Healthy Airways*. Am J Respir Crit Care Med, 2020. **202**(12): p. 1636-1645.
26. Laddha, A.P. and Y.A. Kulkarni, *VEGF and FGF-2: Promising targets for the treatment of respiratory disorders*. Respir Med, 2019. **156**: p. 33-46.
27. Yamamoto, H., et al., *Epithelial-vascular cross talk mediated by VEGF-A and HGF signaling directs primary septae formation during distal lung morphogenesis*. Dev Biol, 2007. **308**(1): p. 44-53.
28. Kawamoto, T., et al., *Evaluation of the severity of small airways obstruction and alveolar destruction in chronic obstructive pulmonary disease*. Respir Med, 2018. **141**: p. 159-164.
29. Soltani, A., et al., *Inhaled corticosteroid normalizes some but not all airway vascular remodeling in COPD*. Int J Chron Obstruct Pulmon Dis, 2016. **11**: p. 2359-2367.
30. Kasahara, Y., et al., *Endothelial cell death and decreased expression of vascular endothelial growth factor and vascular endothelial growth factor receptor 2 in emphysema*. Am J Respir Crit Care Med, 2001. **163**(3 Pt 1): p. 737-44.
31. Kasahara, Y., et al., *Inhibition of VEGF receptors causes lung cell apoptosis and emphysema*. J Clin Invest, 2000. **106**(11): p. 1311-9.
32. Breen, E.C., et al., *Impaired pulmonary defense against Pseudomonas aeruginosa in VEGF gene inactivated mouse lung*. J Cell Physiol, 2013. **228**(2): p. 371-9.
33. Dean, D.C., *Expression of the fibronectin gene*. Am J Respir Cell Mol Biol, 1989. **1**(1): p. 5-10.
34. Wijelath, E.S., et al., *Heparin-II domain of fibronectin is a vascular endothelial growth factor-binding domain: enhancement of VEGF biological activity by a singular growth factor/matrix protein synergism*. Circ Res, 2006. **99**(8): p. 853-60.
35. Bing, D.H., et al., *Fibronectin binds to the C1q component of complement*. Proc Natl Acad Sci U S A, 1982. **79**(13): p. 4198-201.
36. Iruela-Arispe, M.L., et al., *Type I collagen-deficient Mov-13 mice do not retain SPARC in the extracellular matrix: Implications for fibroblast function*. Developmental Dynamics, 1996. **207**(2): p. 171-183.
37. Schmidt, G., H. Hausser, and H. Kresse, *Interaction of the small proteoglycan decorin with fibronectin. Involvement of the sequence NKISK of the core protein*. Biochem J, 1991. **280** (Pt 2): p. 411-4.
38. Kupprion, C., K. Motamed, and E.H. Sage, *SPARC (BM-40, osteonectin) inhibits the mitogenic effect of vascular endothelial growth factor on microvascular endothelial cells*. J Biol Chem, 1998. **273**(45): p. 29635-40.
39. Kamihagi, K., et al., *Osteonectin/SPARC regulates cellular secretion rates of fibronectin and laminin extracellular matrix proteins*. Biochem Biophys Res Commun, 1994. **200**(1): p. 423-8.
40. Yusuf, N., et al., *SPARC was overexpressed in human endometrial cancer stem-like cells and promoted migration activity*. Gynecol Oncol, 2014. **134**(2): p. 356-63.
41. Garcia-Melero, A., et al., *Annexin A6 and Late Endosomal Cholesterol Modulate Integrin Recycling and Cell Migration*. J Biol Chem, 2016. **291**(3): p. 1320-35.
42. Boudewijn, I.M., et al., *Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression*. Respir Res, 2017. **18**(1): p. 213.
43. Steiling, K., et al., *A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment*. Am J Respir Crit Care Med, 2013. **187**(9): p. 933-42.
44. Lapperre, T.S., et al., *Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: a randomized trial*. Ann Intern Med, 2009. **151**(8): p. 517-27.
45. Hardin, M., et al., *Sex-Based Genetic Association Study Identifies CELSR1 as a Possible Chronic Obstructive Pulmonary Disease Risk Locus among Women*. Am J Respir Cell Mol Biol, 2017. **56**(3): p. 332-341.

Chapter 3 – A lung tissue gene expression profile of severe early onset COPD

Nicolaas J. Bekker, Jos van Nijnatten, Peter L. Horvatovich, Justina C. Wolters, Victor Guryev, Tatiana Karp, Roy R. Woldhuis, Wim Timens, Maarten van den Berge[†], Corry-Anke Brandsma[†]

Status: Not published

3.1 – Abstract

Severe early-onset (SEO-)COPD represents the most severe form of COPD that manifests at an earlier age than common COPD. We hypothesize that this is driven by a different disease pathogenesis and aim to determine whether SEO-COPD has a distinct gene expression profile.

RNA sequencing was performed on lung tissue from 18 SEO-COPD, 20 common COPD and 30 non-COPD control subjects. A gene signature for SEO-COPD was created by comparing SEO-COPD to the control, followed by removal of genes that also differed between common COPD and control. This was followed by GSEA pathway enrichment, cellular deconvolution, GWAS catalog interrogation, eQTL analysis and immunohistochemical validation.

One hundred twenty-one genes were differentially expressed in SEO-COPD versus controls, of which 105 remained after the removal of common COPD genes. Pathway analysis showed enrichment of B-cell-mediated adaptive immunity, chemotaxis and ECM pathways among genes upregulated in SEO-COPD, including B-cell marker CD79A and ECM component FBLN1 as leading-edge genes. CD79A-positive immune infiltrate numbers were increased in SEO-COPD lung sections. In the GWAS catalog, we identified 15 SEO-COPD genes with COPD- or lung function-associated single nucleotide polymorphisms (SNPs), including FBLN1 and FBN1, while eQTLs were identified for 6 SEO-COPD-associated genes, including MS4A14 and TNFAIP8L3.

We identified a lung gene expression profile for SEO-COPD, with increased B-cell immunity and ECM organization signature, including genetic risk variants and eQTLs. These findings support a role for autoimmunity and aberrant tissue repair in the SEO-COPD phenotype, as well as genetic predisposition.

Keywords: Severe early-onset COPD, COPD, transcriptomics, B-cells, ECM, emphysema, genetic risk factors

3.2 – Introduction

Chronic obstructive pulmonary disease (COPD) is a common chronic disease characterized by the presence of airflow limitation that is not fully reversible after treatment with bronchodilators¹. COPD results from long-term gene-environment reactions that inflict damage to the lungs, most commonly caused by noxious aerial particles such as smoke exposure and air pollution¹.

Of particular interest is the subgroup of patients with severe early-onset (SEO-)COPD. This is a subset of patients that present with the most severe form of COPD at a relatively early age²⁻⁴. SEO-COPD presents a high pathological and societal burden to patients, heavily affecting their quality of life, with lung transplantation often being the main treatment option. Yet much of the pathology underlying SEO-COPD development remains unclear. We propose that SEO-COPD represents a distinct sub-group of COPD with a (partially) different underlying pathogenesis compared to later-onset COPD (referred to as 'common' COPD from here on).

Genetic risk factors and early-life events have been proposed to be associated with higher susceptibility to develop COPD¹, possibly explaining why, in some individuals, the disease manifests at an earlier age and/or progresses more rapidly in response to smoking. Indeed, known risk factors for SEO-COPD are race/ethnicity, sex, maternal smoking and maternal COPD, and genetic risk factors like TERT mutation, SERPINE2 polymorphism and alpha-1 anti-trypsin deficiency²⁻⁶. Yet it is suggested that SEO-COPD susceptibility cannot be assigned to a single gene or polymorphism but is more likely attributed to a combination of multiple genes⁶. This emphasizes the need for a better understanding of the molecular drivers of SEO-COPD.

Our aim was to identify whether SEO-COPD subjects have a distinct lung tissue gene expression profile, to provide new insight into SEO-COPD pathogenesis and to reveal new potential molecular drivers of SEO-COPD. To achieve this, the transcriptomes of peripheral lung tissue from non-COPD (n=30) and SEO-COPD (n=18) subjects were sequenced and compared to identify a list of differentially expressed genes in SEO-COPD. This was followed by removing genes from the list that were also differentially regulated between non-COPD and common COPD subjects (n=20) so that a profile of gene expression changes specific to SEO-COPD would remain. Next, the profile was characterized through pathway enrichment analysis, expression quantitative trait loci (eQTL) analysis and interrogation with known single nucleotide polymorphism (SNP) risk variants to establish whether the SEO-COPD gene expression changes may be genetically driven.

3.3 – Methods

Tissue acquisition

Peripheral lung tissue was isolated from leftover lung surgery material of non-COPD control and COPD subjects, forming the PRoteogenomics Early onSet cOPd lung cohort (PRESTO, n=117 subjects total). The cohort consisted of never-smokers, current smokers and ex-smoker controls, stage II, III, IV COPD subjects, SEO-COPD subjects, and alpha-1 antitrypsin-deficiency (AATD) COPD subjects.

For the current RNA-Seq analysis, subjects were selected from the cohort to form three groups based on forced expiratory volume (FEV₁), forced vital capacity (FVC), smoking history and age using previously defined definitions^{2,3,5,7}. Never smokers and subjects with alpha-1 anti-trypsin deficiency were excluded.

Controls: FEV₁/FVC ratio >70%, age ≥ 40 years, no other lung disease history except lung tumours, current /ex-smoking (n=30).

SEO-COPD: FEV₁/FVC ratio <70%, FEV₁ percentage of predicted value (FEV₁ % pred) < 40%, age ≤ 55 years, ex-smoking (n=18).

Common COPD: FEV₁/FVC ratio <70%, FEV₁ % pred 40–80%, age ≥ 40 years, current/ex-smoking (n=20).

Lung tissue material was available from the date that patients underwent lung surgery. To exclude the possibility of including very severe COPD patients in the common COPD group that actually had SEO-COPD but were older than 55 years at the time of lung surgery, we only included patients with FEV₁ % pred >40% in the common COPD group.

This project's methods are in agreement with the Research Code of the University Medical Center Groningen⁸ and Dutch national ethical and professional guidelines⁹. Lung tissue material used in this study was not subject to the Medical Research Human Subjects Act in the Netherlands by virtue of all biological samples being from archival materials that are exempt from consent in compliance with applicable laws and regulations (Dutch laws: Medical Treatment Agreement Act (WGBO) art 458 / GDPR art 9/ UAVG art 24). In addition, review and approval by the UMCG institutional ethical review board had been acquired. Clinical characteristics and samples of included subjects were deidentified through pseudonymized identification codes prior to experimentation to guard patient privacy.

RNA sequencing

DNA and RNA were isolated from lung sample sections with AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Cat#80224) according to manufacturer protocol. RNA samples underwent total transcriptomic sequencing on an Illumina NovaSeq 6000 (PE 150) at GenomeScan B.V.

Transcriptomics

Quality control of raw RNA sequencing data was done with FastQC (version 0.11.8). Transcriptomic reads were aligned to Ensembl and HGNC IDs (GRCh38 genome reference, Ensembl gene build 100, annotation file: Homo_sapiens.GRCh38.100.gtf) and quantified with STAR aligner (version 2.7.3a) as described previously in supplemental methods¹⁰. Duplicate genes and lowly expressed genes were filtered out. Data was log₂-transformed and library sizes were normalized with the Trimmed Mean of M-values (TMM) method in edgeR (version 3.30.3).

Statistical methods

Transcriptome data was analysed with edgeR (version 3.30.3, dependent on Limma version 3.44.3) in R (version 4.0.2).

Gene expression was compared between SEO-COPD and non-COPD controls in a linear regression model adjusting for age and sex. A Benjamini-Hochberg false discovery rate (FDR) < 0.05 cutoff was used to determine significance. Significant SEO-COPD-associated genes were then used as input for a biased analysis using the same linear model. First we compared common COPD and non-COPD controls. Genes that were nominally significantly differentially expressed in the same direction in both common COPD and SEO-COPD were excluded (P < 0.05, same fold change direction). Second, since smoking status (current- or ex) was unequally distributed across the groups, we additionally assessed the effect of smoking status on SEO-COPD-associated genes by comparing current-smoker controls with ex-smokers controls. Volcano plots were made with the ggplot2 package in R.

Pathway enrichment

GSEA^{11, 12} (version 4.3.2), and Cytoscape Enrichment Map^{13, 14} (version 3.9.0), pathway network visualization were made as described previously¹⁵, using the C5 human gene set collection from MSigDB as database: c5.all.v2022.1.Hs.symbols.gmt^{11, 16}.

Cellular deconvolution

Cellular deconvolution was performed with CIBERSORTx¹⁷, using the transcriptomic results as gene expression input and the Human Lung Cell Atlas¹⁸ (HLCA) as the database for cell type gene expression signatures. Cell types that had null values in >60% of subjects were excluded, as were interstitial macrophages and capillary endothelial cells per the protocol's benchmark tests. Differences in cell proportions between disease groups were assessed with Mann-Whitney U. Associations between cell proportions and expression levels of SEO-COPD-associated genes were determined with Spearman rho correlation, using $\rho > 0.6$ and $p < 0.05$ as the cut-off for persuasive association.

Genome-wide association studies (GWAS) catalog

Lists of SNPs associated with the traits "FEV₁", "FEV₁/FVC", or "chronic obstructive pulmonary disease" were retrieved from the GWAS Catalog¹⁹, and overlap was assessed between these lists and the list of SEO-COPD-specific genes. Lists were downloaded on 16/02/2024.

eQTL

DNA samples were genotyped on Infinium Global Screening Array + Multi Disease (GSAMD-24v3.0). Standard QC was performed on samples and markers. SNPs with a low genotype rate (<90%), minor allele frequency (MAF) < 0.05 and deviation from Hardy-Weinberg equilibrium $p < 10^{-5}$ were excluded. Samples with a low genotype rate (<90%), more than 3SD away from the mean heterozygosity rate, that did not show consistent information between reported sex and genotypes on the X chromosome were excluded. For samples with observed identity-by-descent rate > 0.25, one sample from the pair was excluded. Imputation was performed using the Human Reference Genome (HRC) r1.1 2016 (GRCh37/hg19) reference panel on the Michigan Imputation Server²⁰. After imputation, SNPs with MAF < 0.05 and Estimated Imputation Accuracy (r^2) < 0.3 were excluded. For eQTL analysis, all PRESTO cohort subjects for which both RNA and genotype data were available and passed quality control were used (n=99 total). Lowly expressed genes were filtered out. Gene counts were log₂-transformed and library sizes were normalized with the TMM method in edgeR. SNPs located 1Mb up-/downstream the transcription start site were analysed for each SEO-COPD gene. An additive linear model in R package MatrixEQTL²¹ was used to identify eQTLs by correlating gene expression with genotype across all samples. The covariates were sex, age, principal components 1 and 2 (from genotyped data), and smoking status. The EigenMT method was utilized to calculate the number of independent genetic variants tested for association per gene²², and the sum of all independent variants was used for Bonferroni multiple testing correction (Bonferroni adjusted $p < 0.05$).

Linkage disequilibrium (LD) between SNPs was assessed with LDmatrix²³ using all European populations. SNPs with r^2 value > 0.8 were considered in LD.

Immunohistochemistry

Lung tissue sections of SEO-COPD subjects (ages 42-55) and ex-smoker controls (ages 42-65) from a separate cohort were used for immunohistochemical (IHC) staining through standard IHC protocol. CD79A staining was done with anti-CD79a SP18 rabbit antibody (Ventana CONFIRM, Cat#790-4432) and FBLN1 staining with anti-FBLN1 CL0337 mouse antibody (Abcam, Cat#ab21156). Stained sections were scanned at 0.23 $\mu\text{m}/\text{pixel}$ resolution (40x) using a Hamamatsu Nanozoomer 2.0 HT. The area and intensity of CD79A and FBLN1 staining in total lung tissue or airways were measured with the SlideJ plugin²⁴ in ImageJ²⁵ (version 1.52e), as described previously²⁶. Area% and mean intensity were calculated in R with the package reshape, using the following formulae as described previously²⁶: Area%: (Number of target staining-positive pixels / number of pixels in total tissue) * 100

Mean intensity: 255 – (Sum of intensities of target staining-positive pixels / total number of target staining-positive pixels)

The number of CD79A-positive inflammatory infiltrates was counted manually, where a minimum of 50 CD79A-positive cells clustered closely together was required to count as an infiltrate. Mann-Whitney U test and mixed linear models were done in IBM SPSS Statistics 28.

3.4 – Results

Clinical characteristics of subjects

Clinical characteristics of the subjects included in the RNA sequencing analysis are presented in Table 3.1. SEO-COPD and common COPD subjects significantly differed from the non-COPD controls with respect to age and smoking status ($P < 0.05$). Therefore, age and sex correction were applied to linear regression analyses, and the effect of current smoking on gene expression was tested separately.

Clinical characteristics of the subjects in the eQTL analysis are depicted in Supplementary Table 3.1.

Table 3.1) Clinical characteristics of study subjects

Characteristic	Control	Severe Early Onset (SEO-)COPD	Common COPD
Subjects, N	30	18	20
Sex, N	15 f – 15 m	13 f – 5 m ¹	5 f – 15 m
Age	61.5 (45 – 74)	51.5 (39 – 55) ²	68 (47 – 79) ²
Smoking status, N	17 CS – 13 ES	0 CS – 18 ES ²	4 CS – 16 ES ²
Pack years	35.5 (10.0 – 75.0)	28 (6.0 – 54.0)	39 (6.5 – 130)
FEV ₁ % pred	84.25 (59.8 – 116)	17.25 (12.0 – 37.0) ²	56.7 (42.0 – 75.4) ²
FEV ₁ /FVC % pred	76.5 (71.0 – 86.6)	29.15 (20.5 – 50.0) ²	54.1 (40.0 – 67.9) ²

N = number, FEV₁ % of pred = forced expiratory volume percentage of predicted value, FEV₁/FVC % pred = FEV₁ / forced vital capacity ratio using percent predicted values. ¹Characteristic significantly differed between SEO-COPD or common COPD compared to non-COPD control, based on Mann Whitney U or Chi² test. ²Characteristics significantly differed between SEO-COPD and the control group in non-parametric one-way ANOVA ($P < 0.05$). All measurements are in “median (range)”, unless otherwise specified.

Identification of a gene signature for SEO-COPD

After quality control and filtering for low abundant genes, 16,930 genes remained as input for the differential expression analysis. When comparing SEO-COPD to non-COPD controls, we identified 121 significantly differentially expressed (DE) genes, of which 80 were higher, and 41 were lower expressed in SEO-COPD (Figure 3.1A, FDR < 0.05). To limit our gene signature to those that are specific for SEO-COPD, we excluded genes that were also differentially expressed between common COPD and control, using a lenient cut-off, i.e. nominal p-value < 0.05 . Of the 121 genes, 16 were nominally significantly differently expressed in common COPD versus control, with the same direction of effect as SEO-COPD versus control (Figure 3.1B) and therefore removed from our SEO-COPD associated gene list. Thus, 105 DE genes remained that were specific for SEO-COPD: 68 genes with a higher and 37 genes with a lower expression in SEO-COPD compared to non-COPD controls. The top 10 higher and lower expressed genes are depicted in Table 3.2; the full list can be found in Supplementary Table 3.2. PLPP1 was the most significantly higher expressed gene in SEO-COPD, and TIAM1 was the most significantly lower expressed gene. Meanwhile, PTX3 had the greatest positive fold change, and KRT4 had the greatest negative fold change.

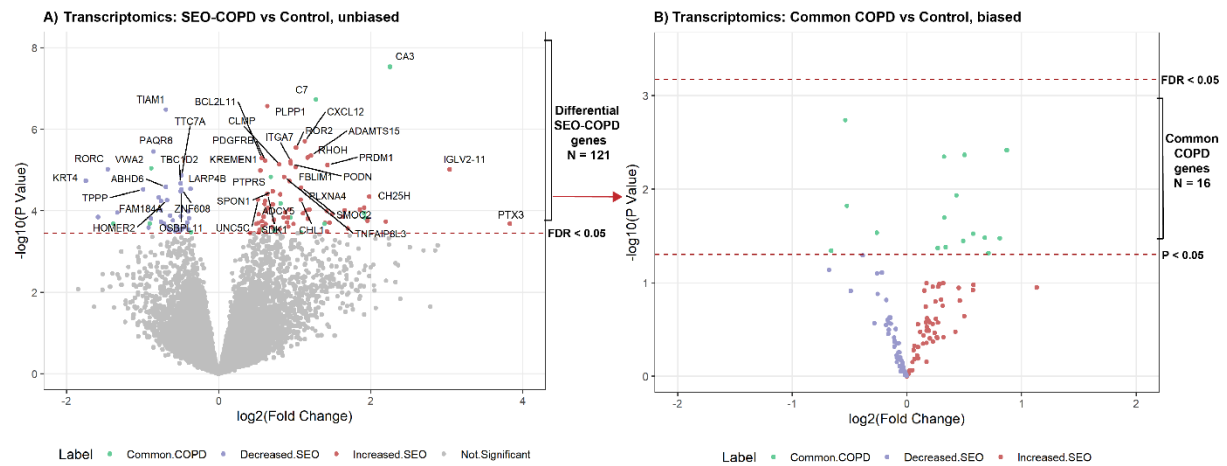


Figure 3.1) Identification of a gene signature for SEO-COPD. A) Volcano plot comparing SEO-COPD (n=18) to non-COPD control (n=30). Of 16,930 genes, 121 genes were significantly lower (blue) or higher (red) differentially expressed in SEO-COPD compared to controls (FDR < 0.05). B) Volcano plot comparing common COPD (n=20) to non-COPD control (n=30) using the 121 differential SEO-COPD genes from plot A as input. Sixteen genes were nominally significant (p < 0.05) in common COPD versus control (green) and were thus not specific to SEO-COPD. Thus, 105 SEO-COPD-associated genes remained for further analysis (blue/red).

Table 3.2) Top 10 higher/lower expressed genes in the SEO-COPD gene signature, compared to non-COPD control.

Higher expressed				
Gene symbol	Gene name	P value	False discovery rate	Fold change
PLPP1	Phospholipid phosphatase 1	$2.72 \cdot 10^{-7}$	$1.41 \cdot 10^{-3}$	1.56
CXCL12	C-X-C motif chemokine ligand 12	$1.96 \cdot 10^{-6}$	0.01	2.20
ROR2	Receptor tyrosine kinase-like orphan receptor 2	$2.81 \cdot 10^{-6}$	0.01	2.03
ADAMTS15	ADAM metalloproteinase with thrombospondin type 1 motif 15	$4.46 \cdot 10^{-6}$	0.01	2.32
RHOH	Ras homolog family member H	$4.97 \cdot 10^{-6}$	0.01	2.25
BCL2L11	BCL2 like 11	$5.08 \cdot 10^{-6}$	0.01	1.48
PDGFRB	Platelet-derived growth factor receptor beta	$5.99 \cdot 10^{-6}$	0.01	1.53
ITGA7	Integrin subunit alpha 7	$6.08 \cdot 10^{-6}$	0.01	1.93
PODN	Podocan	$7.00 \cdot 10^{-6}$	0.01	1.93
CLMP	CXADR-like membrane protein	$7.29 \cdot 10^{-6}$	0.01	1.74
Lower expressed				
Gene symbol	Gene name	P	False discovery rate	Fold change
TIAM1	TIAM Rac1 associated GEF 1	$3.33 \cdot 10^{-7}$	$1.41 \cdot 10^{-3}$	0.62
PAQR8	Progestin and adipoQ receptor family member 8	$3.53 \cdot 10^{-6}$	0.01	0.55
RORC	RAR-related orphan receptor C	$9.48 \cdot 10^{-6}$	0.01	0.37
TTC7A	Tetratricopeptide repeat domain 7A	$1.35 \cdot 10^{-5}$	0.01	0.71
KRT4	Keratin 4	$1.84 \cdot 10^{-5}$	0.01	0.30
TBC1D2	TBC1 domain family member 2	$2.11 \cdot 10^{-5}$	0.01	0.71
ABHD6	Abhydrolase domain containing 6, acylglycerol lipase	$2.60 \cdot 10^{-5}$	0.02	0.62
LARP4B	La ribonucleoprotein 4B	$2.91 \cdot 10^{-5}$	0.02	0.78

TPPP	Tubulin polymerization promoting protein	$2.97 \cdot 10^{-5}$	0.02	0.50
ZNF608	Zinc finger protein 608	$2.99 \cdot 10^{-5}$	0.02	0.71

As smoking status was a potential confounder in our study, we next assessed whether current smoking significantly influenced the expression of SEO-COPD-associated genes by directly comparing their expression between current and ex-smoking controls. None of the SEO-COPD genes were FDR significant between current and ex-smokers, while four genes were nominally significantly higher ($p < 0.05$; GGT5, SPON1, NR1D1 and FRZB) in current- versus ex-smoking controls. Since the direction is the same in SEO-COPD versus control, with current smokers only in the control group, the difference in SEO-COPD is unlikely to be driven by current smoking.

Enrichment of the SEO-COPD genes shows a strong signal for immune, chemotaxis and ECM-related pathways

To assess which biological pathways are overrepresented in the SEO-COPD gene signature, GSEA was performed on all genes ($n = 16,930$) ranked based on significance and fold change when comparing SEO-COPD versus control. This resulted in 402 pathways significantly enriched among genes higher expressed in SEO-COPD and 4 pathways significantly enriched among genes lower expressed in SEO-COPD (Supplementary Table 3.3). These pathways were mapped and annotated by Cytoscape into a network of 73 clusters of closely related pathways based on their number of shared genes (Figure 3.2). The six largest clusters in the network included pathways pertaining to the adaptive immune system (N pathways = 84), organ / tissue development ($N = 77$), Heart, lung & artery abnormalities ($N = 34$), chemotaxis & cytokine signalling ($N = 33$), muscle and kidney development ($N = 30$) and extracellular matrix (ECM) organization ($N = 15$).

We subsequently looked at which of the 105 SEO-COPD genes were among the leading-edge genes in the 25 pathways with the highest normalized enrichment score (NES) to gain insight into the possible functions of the 105 SEO-COPD genes in the most prominently enriched pathways in the GSEA. This identified 18 SEO-COPD leading edge genes (Supplementary Table 3.4). These included IGLV2-11, IGLV2-14, IGHV1-18, CD79A, PTX3, GATA6, IKZF3, PRDM1 and RHOH which were all enriched in pathways of the adaptive immune system cluster (19 of top 25 pathways), CXCL12, ITGA7, IGHV1-18 and CD79A which were enriched in pathways of the chemotaxis and cytokine signalling cluster (2 of top 25), ADAMTS15, SMOC2 and FBN1 being enriched in pathways of glycosaminoglycan binding cluster (2 of top 25) and PODN, SPON1, EMILIN1, FBLN1 and FBN1 being enriched in pathways of the ECM organization cluster (1 of top 25) (Figure 3.3).

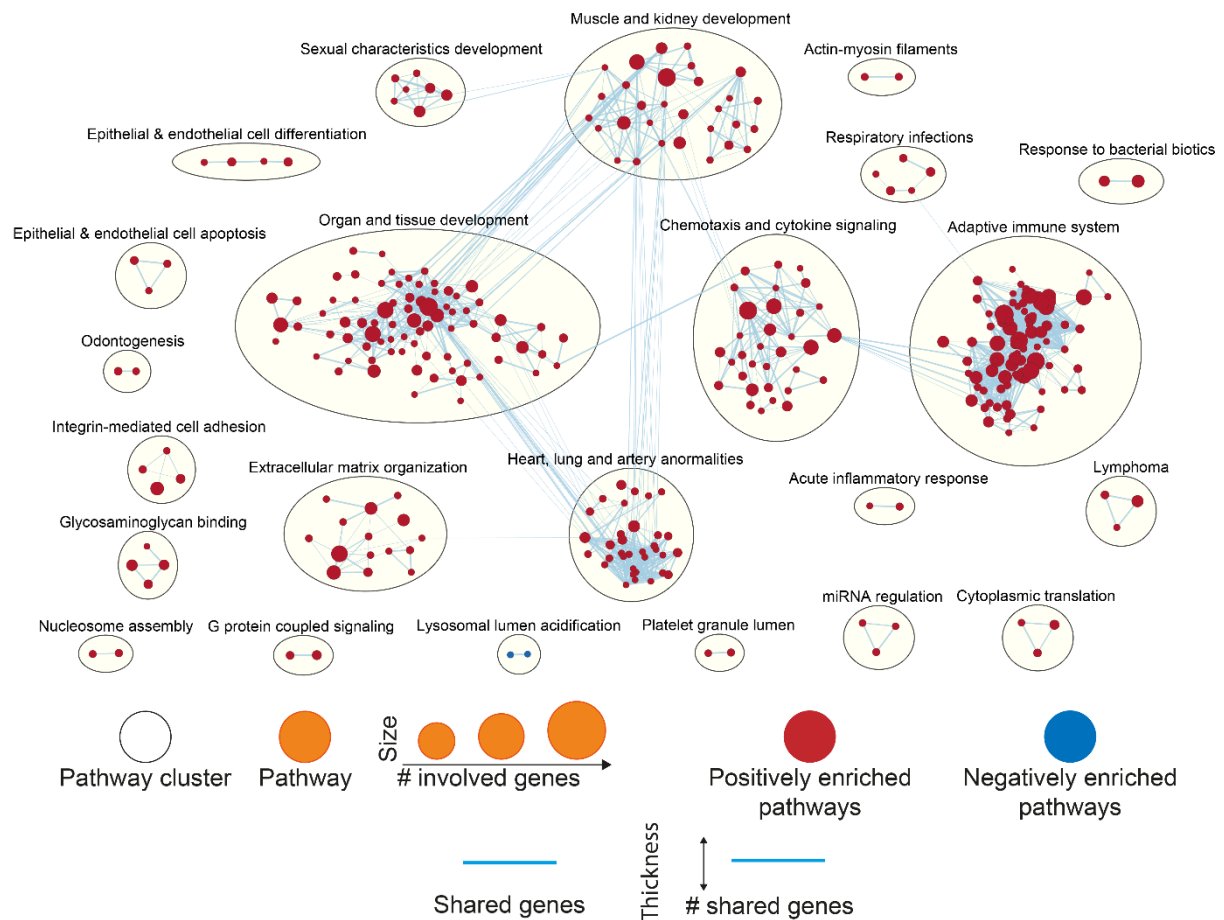


Figure 3.2) Visualization of the gene set enrichment analysis (GSEA) on the SEO-COPD versus control comparison ($n=16,930$). Node: pathway; size denotes the number of involved genes, connections mean two pathways share genes, line thickness indicates the number of overlapping genes, red indicates positively enriched pathways in SEO-COPD and blue negatively enriched pathways. Circles: pathway clusters; names summarize pathways in the cluster. Only FDR-significant pathways were included. Single unconnected pathways were omitted.

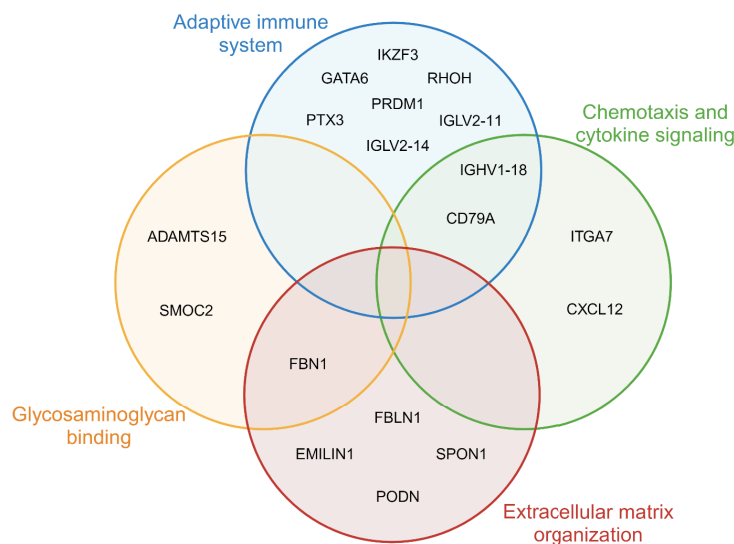


Figure 3.3) Visualization of the 18 SEO-COPD leading edge genes enriched among the top pathways identified in GSEA, divided among four pathway clusters (see Figure 3.2). The four clusters of closely related pathways pertain to the adaptive immune system (blue), chemotaxis and cytokine signalling (green), extracellular matrix organization (red) or

glycosaminoglycan binding (yellow). Adapted from “Venn Diagram with Fill 2 (Layout)” by BioRender.com (2024). Retrieved from <https://app.biorender.com/biorender-templates>.

Cellular deconvolution predicts differential cell proportions of alveolar macrophages and fibroblasts in SEO-COPD

Since the differential gene expression observed in SEO-COPD lung tissue may be affected by differences in cellular composition, cellular deconvolution was performed on the gene expression data of all subjects against single-cell data from the HLCA²⁷. Ten cell types passed the minimum expression threshold in our dataset (Figure 3.4A), and their predicted proportions were compared between SEO-COPD and control. Of those, alveolar macrophages (Figure 3.4B) showed a significantly lower proportion in SEO-COPD tissue, whereas fibroblast lineage cells (Figure 3.4C) showed a significantly higher proportion in SEO-COPD compared to control lung tissue.

For the proportion of macrophages, we identified significant associations ($FDR < 0.05$, $\rho > 0.6$) with 10 SEO-COPD genes, the most significant being GGA2, SLC7A7 and ABCG1 (Figure 3.4D). Seven of these genes had lower gene expression in SEO-COPD versus control and had a positive association with alveolar macrophages (Supplementary Table 3.5), indicating that the lower gene expression in SEO-COPD (Figure 3.1) might be related to the lower alveolar macrophage cell proportion (Figure 3.4B).

For the proportion of cells of the fibroblast lineage, we identified association with 8 SEO-COPD genes (Figure 3.4E), of which seven (SPON1, FBN1, BICC1, UNC5C, LSAMP, SLC16A2 and FBLN1) had higher gene expression and a positive association with fibroblast lineage proportion (Supplementary Table 3.5), thus their higher gene expression in SEO-COPD versus control could be related to the higher fibroblast lineage cell proportion (Figure 3.1, Figure 3.4C). The expression of the remaining 87 SEO-COPD genes was not associated with cell type proportions. There was no overlap in genes associated with fibroblast or alveolar macrophages (Supplementary Table 3.5).

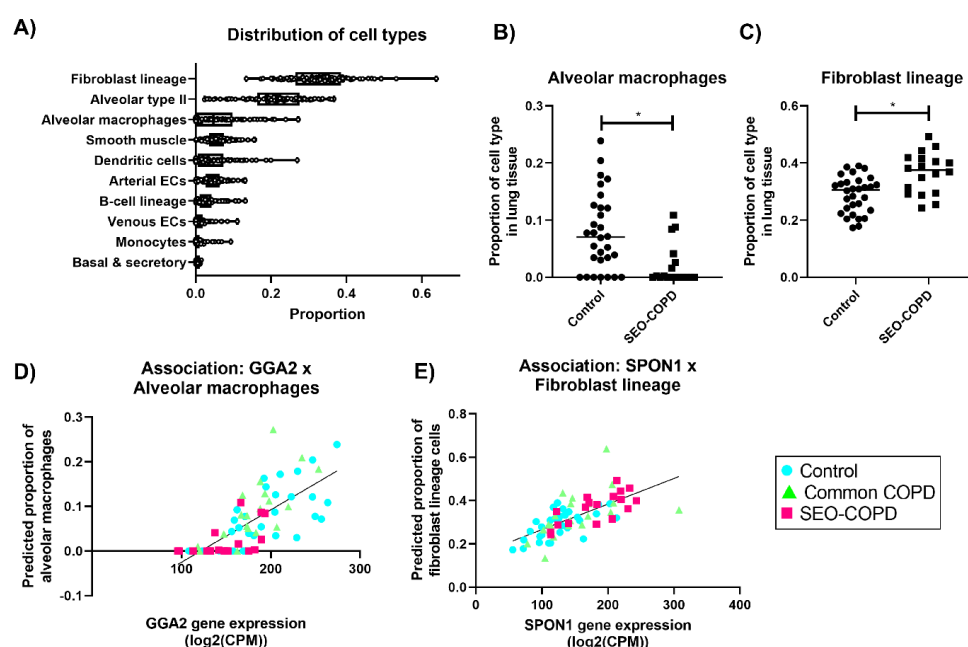


Figure 3.4) Proportions of cell types calculated using cellular deconvolution and associations with genes. A) Boxplots depicting the relative proportions of predicted cell types detected in the lung tissue gene expression data of all subjects (n=68). B/C) Visualization of the difference in alveolar macrophage (B) and fibroblast lineage (C) proportions between disease groups (MWU, $p < 0.05$). D/E) Association plots of the most significant gene \times cell associations: D) GGA2 \times alveolar

macrophages, and E) SPON1 × fibroblast lineage cells (Spearman rho, $p < 0.05$, $\rho > 0.6$). Predicted cell proportion on the y-axis and gene expression on the x-axis, using all subjects. * $p < 0.05$. ECs = endothelial cells.

SEO-COPD signature genes harbour known genetic risk variants and strong, novel eQTLs

Next, we determined whether our SEO-COPD gene signature harbours genetic risk variants known to be associated with COPD susceptibility and lung function by checking our gene list in the GWAS Catalog (cross-reference date: 16/02/2024). Two of the 105 SEO-COPD genes harbour COPD-associated SNPs, while twelve harbour lung function-associated SNPs and one has both COPD and lung function SNPs (Table 3.3). These included the immune and chemotaxis-associated genes GATA6 and ITGA7 and the ECM-associated SEO-COPD leading edge genes FBLN1 and FBN1.

In addition, an eQTL analysis with all PRESTO cohort subjects (Supplementary Table 3.1) was performed for each of the 105 SEO-COPD genes to determine the association between gene expression and genetic variants. A total of 138 significant eQTLs, linked to six (MS4A14, TNFAIP8L3, OSCAR, TMEM108, TLN2, and RAB3IL1) of the 105 SEO-COPD genes were identified (Bonferroni adjusted $p < 0.05$, Table 3.4, Supplementary Table 3.6). I.e., these six genes had many genetic variants that could affect their gene expression levels.

MS4A14 had the most significant and highest number of 117 eQTLs. This was followed by TNFAIP8L3 with 12 eQTLs. For MS4A14, three different SNPs were previously identified to be associated with lung function (Table 3.3). We used LDmatrix to check whether the 117 MS4A14 eQTLs and 3 GWAS SNPs were in high LD, i.e. whether these genetic variants were frequently co-inherited and thus represented the same genetic signal ($r^2 > 0.8$, Supplementary Table 3.7). We identified one large LD block containing the MS4A14 eQTL SNPs, which separated from the GWAS SNPs that were not in LD (Supplementary Figure 3.1). The eQTLs of TNFAIP8L3, OSCAR, TMEM108 and TLN2 were all in high LD among each other, though TNFAIP8L3 presented two separate blocks of high LD variants (See <http://www.GitHub.com/vanNijnatten/PhD-Thesis/>).

Supplementary Table 3.8, Supplementary Figure 3.2).

The most significant eQTLs for MS4A14 and TNFAIP8L3 were rs4939347 and rs12906304, respectively (Supplementary Table 3.6). Gene expression of MS4A14 and TNFAIP8L3 was plotted against the genetic variants of their top eQTLs to assess how the variants associated with gene expression (Figure 3.5A/C) and how these variants were distributed among SEO-COPD and control subjects (Figure 3.5B/D).

Rs4939347 had three allelic variants in the MS4A14 gene: CC, CG and GG. MS4A14 expression was lower in CG variants compared to GG variants, and CC variants were lower than both CG and GG (Figure 3.5A), indicating the C allele is a risk allele for lower MS4A14 expression. In line with the lower MS4A14 gene expression in SEO-COPD (Figure 3.5E), these patients tended to have a higher proportion of subjects with CG variants than the controls (Figure 3.5B) (Pearson χ^2 , $p = 0.053$).

Rs12906304 had the allelic variants CC, CT and TT in the TNFAIP8L3 gene. The CC variant showed higher TNFAIP8L3 expression compared to CT, which in turn was higher than TT (Figure 3.5C), indicating the C allele as a risk allele for higher TNFAIP8L3 expression. The proportion of the CC variant was not significantly different ($p = 0.359$) between SEO-COPD subjects and controls (Figure 3.5D) despite SEO-COPD having higher TNFAIP8L3 expression than controls (Figure 3.5F).

None of the SEO-COPD eQTLs overlapped with the SEO-COPD leading edge genes, nor were they among the top 10 most significantly expressed higher and lower genes.

Immunohistochemical quantification of CD79A and FBLN1 in lung tissue

To confirm the presence and localization of FBLN1 and CD79A in lung tissue, lung tissue sections obtained from age- and smoking-matched SEO-COPD (n=12) and control subjects (n=9) were stained with IHC. FBLN1 and CD79A were selected because FBLN1 was one of the SEO-COPD leading edge genes in the ECM organization pathway cluster with an SNP associated with FEV₁/FVC^{28, 29} and CD79A, a B-cell marker, was an SEO-COPD leading-edge gene involved in the adaptive immune response pathway cluster that together with 3 immunoglobulin genes (IGHV1-18, IGLV2-11, IGLV2-14) points towards a role for B-cells in SEO-COPD³⁰⁻³².

For FBLN1, no differences were observed in positively stained areas or staining intensity between SEO-COPD and controls when assessing the staining in the whole tissue sections nor when we specifically measured around the airways (data not shown). Meanwhile, CD79A was clearly expressed in B-cells, both in single B-cells in parenchyma and also in inflammatory infiltrates in SEO-COPD whole tissue sections (Figure 3.6A/B). There was no difference in the CD79A positively stained area between SEO-COPD and control subjects (Figure 3.6D), but SEO-COPD subjects had significantly more CD79A-positive inflammatory infiltrates in the whole tissue than controls (Figure 3.6C) and significantly higher CD79A staining intensity (Figure 3.6E) compared to controls, in line with the increased gene expression of CD79A in SEO-COPD (Figure 3.6F).

Table 3.3) Known COPD- and lung function-associated single nucleotide polymorphisms in the SEO-COPD gene profile

SNP	Gene	Full gene name	SNP traits(s)	Accession ID	Author
rs143053231	ITGA7	Integrin Subunit Alpha 7	FEV ₁ /FVC	GCST003264	³³
rs10808265	PLXNA4	Plexin A4	FEV ₁	GCST001444	³⁴
rs4093840 rs2062432	ADCY5	Adenylate Cyclase 5	COPD, Lung disease	GCST007692 GCST90016594 GCST90016588 GCST90016591 GCST90016585 GCST90016590	^{35, 36}
rs9387626 rs12664251	FAM184 A	Family With Sequence Similarity 184 Member A	FEV ₁ /FVC COPD	GCST007080 GCST90244098	^{37, 38}
rs2667376	HOMER 2	Homer Scaffold Protein 2	FEV ₁ /FVC	GCST007080	³⁷
rs35312258 rs1404568	SLC9A9	Solute Carrier Family 9 Member A9	FEV ₁	GCST004107	³⁹
rs551851 rs79864933	CDON	Cell Adhesion Associated, Oncogene Regulated	FEV ₁ /FVC FEV ₁	GCST007080 GCST90244092	^{37, 40}
rs12709669 rs1985511	GATA6	GATA Binding Protein 6	COPD, Lung disease FEV ₁ /FVC	GCST007656 GCST007080 GCST007431	^{37, 40, 41}
rs138753278 rs75368561 rs77253690	MS4A14	Membrane Spanning 4-Domains A14	FEV ₁	GCST003262	³³
rs17462641	FBN1	Fibrillin 1	FEV ₁ /FVC	GCST007080	³⁷
rs10853191 rs28478891	PIEZO2	Piezo Type Mechanosensitive Ion Channel Component 2	FEV ₁	GCST003263	³³
rs11702924	FBLN1	Fibulin 1	FEV ₁ /FVC	GCST007080	³⁷
rs13082924	PDZRN3	PDZ Domain Containing Ring Finger 3	FEV ₁ /FVC	GCST007080	³⁷
rs10007052	RNF150	Ring Finger Protein 150	COPD, Lung disease	GCST001737	³⁶
rs2229712	RPS6KA 1	Ribosomal Protein S6 Kinase A1	FEV ₁	GCST90270081	⁴²

List of genes from the SEO-COPD profile with COPD- or lung function-associated single nucleotide polymorphism (SNPs), gathered from the GWAS catalog. Genes are ordered from top to bottom based on significance in the SEO-COPD gene profile.

Table 3.4) SEO-COPD profile genes with expression quantitative trait loci

Gene symbol	Full gene name	# significant eQTLs	FDR (top eQTL)
MS4A14	Membrane Spanning 4-Domains A14	117	$1.50 \cdot 10^{-6}$
TNFAIP8L3	TNF Alpha Induced Protein 8 Like 3	12	0.0159
OSCAR	Osteoclast Associated Ig-like Receptor	2	0.0274
TMEM108	Transmembrane Protein 108	3	0.0415
TLN2	Talin 2	3	0.0415
RAB3IL1	RAB3A Interacting Protein Like 1	1	0.0464

List of genes from the SEO-COPD profile with significant expression quantitative gene loci (eQTLs). Genes are ordered from top to bottom based on their most significant eQTL in the eQTL list (Supplemental Table 3.6).

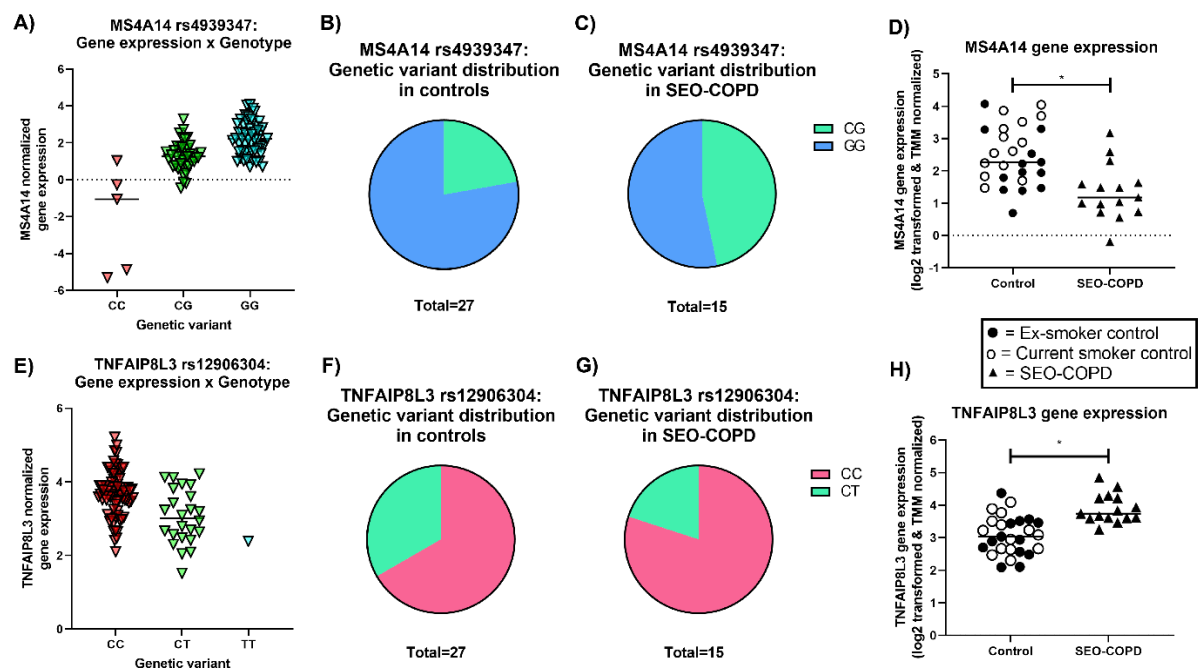


Figure 3.5) Gene expression of MS4A14 and TNFAIP8L3 associated with genetic variants of their top eQTLs. A/E) Gene expression of MS4A14 (A) and TNFAIP8L3 (E) per genotype of their most significant expression quantitative trait locus (eQTL) in the eQTL list: rs4939347 and rs12906304 respectively (Supplementary Table 3.6) List of eQTLs, 15 top), including all PRESTO subjects (n = 99). B/C/F/G) Pie charts showing the distribution of MS4A14 rs4939347 and TNFAIP8L3 rs12906304 genotypes in control (current- and ex-smoker) (B/F) and SEO-COPD subjects (C/G). D/H) Visualization of differential gene expression of MS4A14 (D) and TNFAIP8L3 (H) comparing SEO-COPD versus control (current- and ex-smoker). Gene expression values are transcriptomic read counts, TMM normalized, and log₂-transformed. Due to a lack of genotype data, 3 SEO-COPD and 3 controls were excluded from the eQTL analysis.



Figure 3.6) Immunohistochemical quantification of CD79A in lung tissue. Quantification of immunohistochemical (IHC) staining results of CD79A in whole lung tissue sections. A-B) Examples of CD79A staining in lung parenchyma of an SEO-COPD patient (A) and a non-COPD control (B), single B-cells indicated with cyan arrows and CD79A-positive infiltrates with magenta arrows. C) Difference in number of CD79A-positive inflammatory infiltrates between control (n = 9) and SEO-COPD (n = 12) subjects. D-E) Difference in CD79A-positive staining area (D) and staining intensity (E) between control (n = 7) and SEO-COPD (n = 12) subjects across whole lung tissue. F) Visualization of CD79A gene expression difference between SEO-COPD and controls in PRESTO cohort. Ex-smoker SEO-COPD are used as filled-in squares, ex-smoker controls are used as filled-in circles, and current smoker controls are used as empty circles. In the area and intensity quantification, two control subjects were excluded due to technical problems. * = $P < 0.05$ based on Mann Whitney U. Images taken at 0.23 $\mu\text{m}/\text{pixel}$ resolution.

3.5 – Discussion

In the current manuscript, we identified 105 genes that were specifically differentially expressed in SEO-COPD but not common COPD, compared to non-COPD controls. SEO-COPD-associated genes were enriched for several biological pathways relevant to the disease, including activation of the adaptive immune system, chemotaxis and ECM organization, indicating a role for B-cell-driven autoimmunity and aberrant ECM repair in SEO-COPD and supporting the hypothesis of a distinct underlying disease pathogenesis. Furthermore, eQTL analysis identified strong associations between genetic variants and the expression of six SEO-COPD genes, and GWAS interrogation showed the presence of previously identified COPD and lung function-associated genetic variants among the SEO-COPD gene signature, supporting genetic susceptibility in SEO-COPD and providing potential molecular drivers.

Of interest, the higher expression of CD79A, the B-cell developmental genes CXCL12 and IKZF3^{43, 44}, the immunoglobulin genes IGLV2-11, IGLV2-14 and IGHV1-18, and other genes in the cytokine signalling and B-cell-mediated adaptive immunity pathways all provide a clear signal for upregulation of B-cell lineage activity in SEO-COPD. This is further supported by our finding of a higher number of CD79A-positive inflammatory infiltrates as well as the higher intensity of CD79A staining in SEO-COPD lung tissue compared to controls. The lack of difference in the total area of CD79A staining indicates that B-cells in SEO-COPD are more clustered in infiltrates, while in controls, more single B-cells are present in the parenchyma.

We and others have previously demonstrated excessive B-cell quantities and the formation of tertiary lymphoid follicles in the lung tissue of patients with COPD^{32, 45-48}. Combined with the detection of auto-antibodies in COPD^{49, 50}, the involvement of a B-cell-driven auto-immune component in COPD pathology was proposed^{32, 43, 45, 48, 49}. It is hypothesized that B-cells contribute to COPD pathology by infiltrating the lung to form B-cell follicles against invasive elements, eventually generating autoantibodies against self-antigens from damaged lung tissue and degraded ECM fragments^{48, 49, 51}, which then perpetuate chronic inflammation, lung tissue damage and emphysema^{32, 43, 50, 52}. Notably, in literature, there are observations of a subpopulation phenotype of emphysematous COPD that is characterized by B-cells and autoimmunity symptoms, which include high presence of antinuclear antibodies, overexpression of BAFF in B-cells, alveolar macrophages and lymphoid follicles, as well as an abundance of B-cell follicles with an oligoclonal antigen-specific response against ECM components such as elastin^{32, 45, 48-52}. Since the SEO-COPD gene signature and subsequent CD79A protein quantification align with the traits of this auto-immune phenotype of severe COPD, our findings support a role for autoimmunity in SEO-COPD pathology or may even suggest that the identity of this auto-immune severe COPD phenotype is in fact SEO-COPD, but this requires further investigation.

Next, the elevated gene expression of ECM organization and elastic fibre formation genes FBLN1⁵³, FBN1⁵⁴ and EMILIN1⁵⁵ corresponded to our previous findings where we demonstrated increased gene expression of elastic fibre ECM genes FBLN5, ELN, LTBP2 and MFAP4 in COPD lung tissue genome-wide gene expression profiling⁵⁶ and increased protein expression of FBLN5 in SEO-COPD lung tissue proteomics¹⁰. It was suggested in both studies that the paradoxical upregulation of elastogenesis genes in COPD against the elastic fibre and ECM destruction seen in emphysema is a sign⁵⁷ tissue repair that is ineffective or insufficient to keep up with the rate of tissue destruction^{10, 56, 58}. *Nota bene*, abnormal elastic fibre assembly is a significant contributor to emphysema⁵⁹. Therefore, it is possible that the upregulation of FBLN1, FBN1, EMILIN1 and other ECM-related genes has an effect on dysregulated elastic fibre formation and ECM repair in SEO-COPD that contributes to emphysema. The additional upregulation of ECM protease ADAMTS15⁶⁰ in SEO-COPD would be in line with the idea that the rate of ECM repair cannot keep up with the rate of degradation. Moreover, the IHC results showed no significant difference in FBLN1-positive area between SEO-COPD subjects and controls, which may support the hypothesis that the increased gene expression of FBLN1 is a signal of attempted lung tissue repair that is not sufficient to show an increase in FBLN1 on protein level. Yet, at the same time, the higher expression of FBLN1, FBN1 and SPON1 was correlated with higher fibroblast lineage proportions in SEO-COPD versus control, which could indicate that higher ECM gene expression is resulting from higher relative proportions or activation of (myo)fibroblasts in SEO-COPD⁵⁶. Multi-omics and single-cell approaches are needed to confirm this.

In the GWAS data, we additionally identified that FBLN1 and FBN1 harbour genetic risk variants for airway obstruction (FEV₁/FVC), suggesting a potential genetic factor underlying our ECM-related gene expression changes in SEO-COPD.

Our eQTL analysis showed that the expression of 6 SEO-COPD genes was regulated by eSNPs, the strongest effects being observed for MS4A14 and TNFAIP8L3. MS4A14 belongs to the MS4A protein family, which is often differentially expressed in leukocytes in various diseases, playing a role in their immune sensing and activation⁶¹. MS4A14 itself is a gene that is described to be mainly expressed in testes and myeloid cells, particularly monocyte-macrophage lineage cells⁶². MS4A14 expression is high in monocytes and declines with macrophage differentiation⁶², though in our results, MS4A14 was not associated with alveolar macrophage proportions (Supplementary Table 3.5) Cell x gene heatmap). Notably, MS4A14 was the only gene in the eQTL analysis to also harbour three lung function-associated SNPs identified in previous GWAS. This suggests that MS4A14 SNPs may be involved in SEO-COPD pathogenesis because they have an impact on lung function, which warrants further investigation in larger SEO-COPD cohorts.

TNFAIP8L3 is part of the tumour necrosis factor- α -induced protein 8 (TNFAIP8/TIPE) family of genes, a family with high homology that is implicated in immune homeostasis, inflammation regulation and chronic inflammatory diseases, particularly TNFAIP8L2⁶³⁻⁶⁵. TNFAIP8L3 itself is detected in the lung and has functions in phospholipid signalling, anti-apoptosis and tumorigenesis, but is otherwise not thoroughly characterized⁶⁴⁻⁶⁸. TNFAIP8L3 could perhaps share inflammation regulation functions with TNFAIP8L2 based on homology. In fact, in mouse models of acute lung injury, higher expression of TNFAIP8L3 attenuated inflammation and lung damage via LXR α signalling^{65, 69}. Thus, higher TNFAIP8L3 expression in SEO-COPD could denote an attempt to decrease inflammation in the diseased lung, suggesting TNFAIP8L3 has a role as an inflammatory regulator in SEO-COPD.

A unique feature of this study is the specific focus on SEO-COPD as a subset of COPD that is likely driven by a different underlying pathogenesis than common COPD. Fundamentally, the SEO-COPD group was younger and had more severe disease than the common COPD group, making it difficult

to untwine whether the pathological changes in SEO-COPD were caused by ageing, early disease onset or disease severity. Historical and longitudinal data will be necessary to further deconvolute these complex effects. The arbitrary age and lung function thresholds were, however, necessary to distinguish common and SEO-COPD from each other.

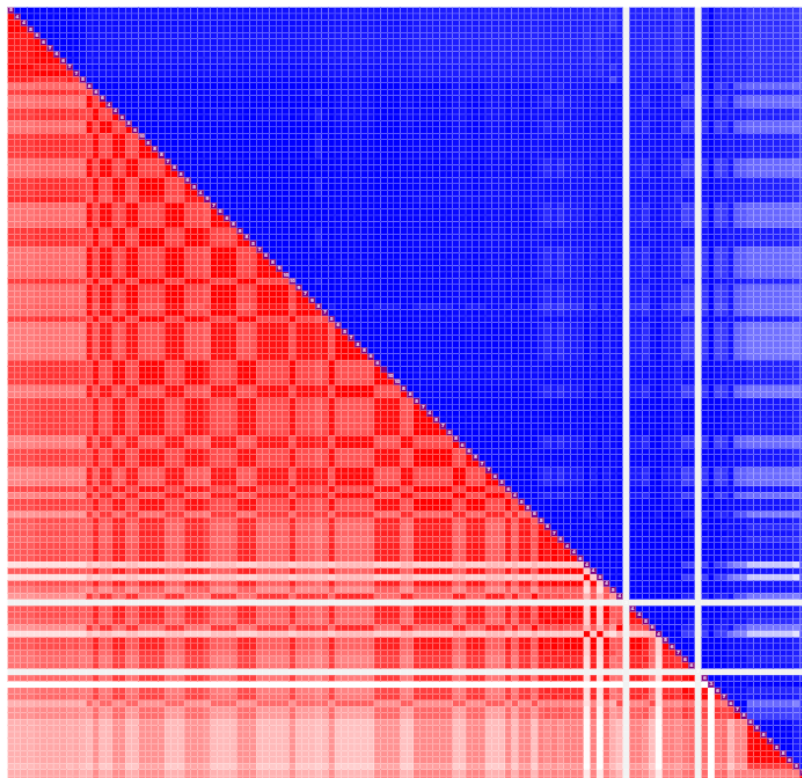
To conclude, we identified a gene expression profile of 105 genes in SEO-COPD pathology with indications for functions in the adaptive immune response pointing towards a B-cell-driven auto-immune phenotype and increased ECM repair and remodelling genes likely reflecting a failed attempt to repair chronic tissue damage. At least 21 of the 105 genes are linked to genetic risk factors via GWAS or eQTLs supporting genetic susceptibility of SEO-COPD and potential molecular drivers of SEO-COPD that need further investigation. Single-cell approaches connecting the key transcript and protein changes to genetic risk variants, multi-omics studies to study different layers of regulation, as well as longitudinal studies to document changes over time are warranted to further define the pathological drivers of SEO-COPD.

3.6 – Funding

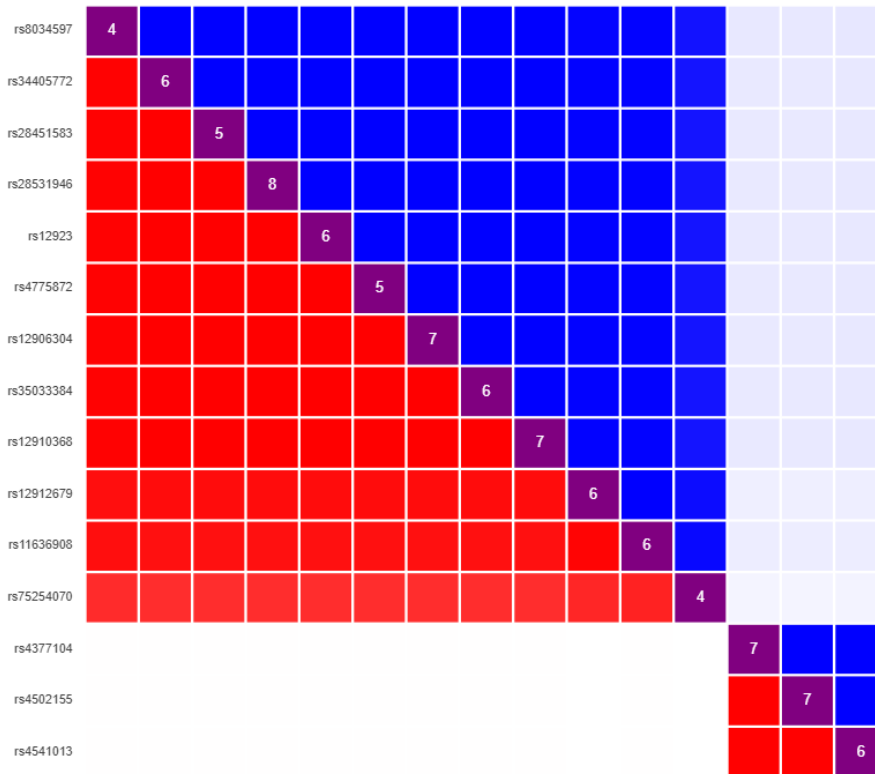
Funded by the PPP-allowance by Health Holland (LSH) and the UMCG. This collaboration project is co-financed by the Ministry of Economic Affairs and Climate Policy by means of the PPP-allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships.

3.7 – Supplementary Material

Figures



Supplementary Figure 3.1) MS4A14_LD_block_plot



Supplementary Figure 3.2) TNFAIP8L3_LD_block_plot

Tables

Supplementary Table 3.1) Clinical characteristics of eQTL study subjects.

Characteristic	PRESTO
Subjects, N	99
Sex, N	50 f – 49 m
Age, median (range)	60 (39 – 82)
Smoking status, N	18 CS – 68 ES – 13 NS
Disease status, N	39 non-COPD – 60 COPD

N = number, CS = Current smoker, ES = Ex-smoker, NS = Never smoker. Only characteristics used as covariates for eQTL analysis are shown.

Supplementary Table 3.2) SEO-COPD gene profile.

Rank	Ensembl ID	Gene symbol	P value	FDR	logFC	logCPM	F	FC
1	ENSG00000067113	PLPP1	2.72E-07	0.001	0.641	5.2944	35.6539	1.560
2	ENSG00000156299	TIAM1	3.33E-07	0.001	-0.693	5.6130	34.9708	0.618
3	ENSG00000107562	CXCL12	1.96E-06	0.007	1.136	6.4680	29.2302	2.198
4	ENSG00000169071	ROR2	2.81E-06	0.008	1.018	4.0462	28.1214	2.025
5	ENSG00000170915	PAQR8	3.53E-06	0.008	-0.856	4.3851	27.4220	0.552
6	ENSG00000166106	ADAMTS15	4.46E-06	0.008	1.213	5.2659	26.7188	2.319
7	ENSG00000168421	RHOH	4.97E-06	0.008	1.170	3.9680	26.3924	2.250
8	ENSG00000153094	BCL2L11	5.08E-06	0.008	0.569	5.1613	26.3295	1.483
9	ENSG00000113721	PDGFRB	5.99E-06	0.008	0.613	8.4410	25.8371	1.529
10	ENSG00000135424	ITGA7	6.08E-06	0.008	0.947	5.2008	25.7940	1.928
11	ENSG00000174348	PODN	7.00E-06	0.008	0.950	6.5388	25.3772	1.931
12	ENSG00000166250	CLMP	7.29E-06	0.008	0.799	5.1315	25.2580	1.739
13	ENSG00000057657	PRDM1	7.52E-06	0.008	1.431	5.7351	25.1675	2.696
14	ENSG00000162458	FBLIM1	8.56E-06	0.009	1.014	5.8685	24.7872	2.019
15	ENSG00000143365	RORC	9.48E-06	0.009	-1.455	4.5384	24.4881	0.365

Rank	Ensembl ID	Gene symbol	P value	FDR	logFC	logCPM	F	FC
16	ENSG00000211668	IGLV2-11	9.74E-06	0.009	3.037	4.2179	24.4108	8.210
17	ENSG00000183762	KREMEN1	1.04E-05	0.009	0.549	6.1852	24.2240	1.463
18	ENSG00000068724	TTC7A	1.35E-05	0.011	-0.489	6.8186	23.4725	0.713
19	ENSG00000221866	PLXNA4	1.48E-05	0.011	0.862	4.4368	23.2058	1.817
20	ENSG00000170477	KRT4	1.84E-05	0.013	-1.747	3.9573	22.5862	0.298
21	ENSG00000183578	TNFAIP8L3	1.87E-05	0.013	0.931	3.5023	22.5464	1.907
22	ENSG00000095383	TBC1D2	2.11E-05	0.014	-0.500	6.6084	22.2037	0.707
23	ENSG00000163686	ABHD6	2.60E-05	0.016	-0.691	4.0371	21.6238	0.619
24	ENSG00000112562	SMOC2	2.66E-05	0.016	1.085	4.8525	21.5610	2.121
25	ENSG00000107929	LARP4B	2.91E-05	0.016	-0.368	7.3108	21.3117	0.775
26	ENSG00000171368	TPPP	2.97E-05	0.016	-0.988	6.4046	21.2502	0.504
27	ENSG00000168916	ZNF608	2.99E-05	0.016	-0.489	5.8993	21.2340	0.712
28	ENSG00000144909	OSBPL11	3.31E-05	0.017	-0.497	5.8013	20.9511	0.709
29	ENSG00000262655	SPON1	3.32E-05	0.017	0.714	7.2133	20.9474	1.641
30	ENSG00000146555	SDK1	3.90E-05	0.019	0.641	5.7788	20.5020	1.559
31	ENSG00000173175	ADCY5	3.94E-05	0.019	0.811	4.7997	20.4740	1.754
32	ENSG00000138135	CH25H	4.50E-05	0.021	1.980	3.4538	20.1144	3.944
33	ENSG00000111879	FAM184A	4.72E-05	0.022	-0.785	4.8515	19.9852	0.580
34	ENSG00000182168	UNC5C	5.33E-05	0.023	0.518	4.9611	19.6550	1.432
35	ENSG00000134121	CHL1	5.37E-05	0.023	1.080	5.0287	19.6340	2.113
36	ENSG00000103942	HOMER2	5.48E-05	0.023	-0.678	4.2768	19.5817	0.625
37	ENSG00000181804	SLC9A9	5.69E-05	0.023	0.610	4.3480	19.4787	1.527
38	ENSG00000064309	CDON	5.79E-05	0.023	-0.757	4.9868	19.4327	0.592
39	ENSG00000144868	TMEM108	6.82E-05	0.026	0.605	4.1844	18.9981	1.521
40	ENSG00000074047	GLI2	6.91E-05	0.026	0.712	3.8729	18.9631	1.638
41	ENSG00000178860	MSC	8.42E-05	0.031	1.915	3.3081	18.4377	3.772
42	ENSG00000147100	SLC16A2	8.66E-05	0.031	0.650	4.4844	18.3645	1.569
43	ENSG00000099998	GGT5	9.26E-05	0.031	0.932	4.5991	18.1884	1.909
44	ENSG00000143297	FCRL5	9.42E-05	0.031	1.857	2.5093	18.1427	3.623
45	ENSG00000162998	FRZB	9.43E-05	0.031	1.193	3.9301	18.1401	2.286
46	ENSG00000183742	MACC1	9.60E-05	0.031	-0.761	5.9034	18.0923	0.590
47	ENSG00000126368	NR1D1	9.67E-05	0.031	1.657	4.7299	18.0736	3.154
48	ENSG00000175906	ARL4D	9.70E-05	0.031	1.150	3.2151	18.0651	2.220
49	ENSG00000064300	NGFR	1.01E-04	0.032	1.421	2.4738	17.9590	2.678
50	ENSG00000005486	RHBDD2	1.03E-04	0.032	-0.398	6.3379	17.9094	0.759
51	ENSG00000141448	GATA6	1.04E-04	0.032	0.621	6.2875	17.8764	1.538
52	ENSG00000166928	MS4A14	1.12E-04	0.033	-1.332	2.4302	17.6919	0.397
53	ENSG00000138080	EMILIN1	1.12E-04	0.033	0.878	7.4151	17.6827	1.837
54	ENSG00000019549	SNAI2	1.16E-04	0.033	1.110	3.5142	17.5963	2.158
55	ENSG00000198832	SELENOM	1.22E-04	0.034	0.527	5.1019	17.4732	1.440
56	ENSG00000118849	RARRES1	1.23E-04	0.034	1.493	4.9657	17.4414	2.814
57	ENSG00000136868	SLC31A1	1.33E-04	0.035	-0.637	6.4139	17.2493	0.643
58	ENSG00000008869	HEATR5B	1.36E-04	0.035	-0.498	6.3389	17.1879	0.708
59	ENSG00000116717	GADD45A	1.38E-04	0.035	1.632	4.9712	17.1447	3.100
60	ENSG00000122870	BICC1	1.40E-04	0.035	0.595	4.8825	17.1145	1.510
61	ENSG00000273540	AGBL1	1.43E-04	0.035	-1.585	4.5839	17.0506	0.333
62	ENSG00000172346	CSDC2	1.44E-04	0.035	0.912	3.6120	17.0397	1.882
63	ENSG00000161405	IKZF3	1.48E-04	0.036	0.856	5.6215	16.9627	1.811
64	ENSG00000072121	ZFYVE26	1.52E-04	0.036	-0.388	6.7602	16.9002	0.764
65	ENSG00000173918	C1QTNF1	1.56E-04	0.036	1.175	5.5945	16.8349	2.258
66	ENSG00000182771	GRID1	1.57E-04	0.036	-0.888	2.9280	16.8128	0.540
67	ENSG00000160801	PTH1R	1.68E-04	0.038	0.729	3.7639	16.6428	1.657
68	ENSG00000106789	CORO2A	1.71E-04	0.038	-0.602	5.2989	16.5938	0.659
69	ENSG00000211666	IGLV2-14	1.75E-04	0.039	1.958	4.7234	16.5374	3.885
70	ENSG00000144712	CAND2	1.80E-04	0.039	0.572	3.8817	16.4687	1.487
71	ENSG00000211945	IGHV1-18	1.84E-04	0.039	2.199	3.8264	16.4131	4.591
72	ENSG00000144959	NCEH1	1.85E-04	0.039	-0.756	5.9592	16.4026	0.592
73	ENSG00000213694	S1PR3	1.91E-04	0.039	0.901	4.9846	16.3239	1.867

Rank	Ensembl ID	Gene symbol	P value	FDR	logFC	logCPM	F	FC
74	ENSG00000166741	NNMT	1.94E-04	0.039	1.461	7.0976	16.2833	2.753
75	ENSG00000103365	GGA2	1.95E-04	0.039	-0.412	7.4651	16.2670	0.751
76	ENSG00000053702	NRIP2	1.99E-04	0.039	1.399	1.2570	16.2127	2.636
77	ENSG00000253159	PCDHGA12	2.02E-04	0.039	0.520	3.7010	16.1761	1.434
78	ENSG00000137266	SLC22A23	2.06E-04	0.039	-0.716	6.2528	16.1230	0.609
79	ENSG00000163661	PTX3	2.09E-04	0.039	3.829	4.7464	16.0840	14.207
80	ENSG00000166147	FBN1	2.10E-04	0.039	0.619	8.0198	16.0731	1.535
81	ENSG00000124772	CPNE5	2.12E-04	0.039	0.987	2.5338	16.0554	1.983
82	ENSG00000185565	LSAMP	2.12E-04	0.039	0.494	5.9024	16.0483	1.408
83	ENSG00000182985	CADM1	2.37E-04	0.043	-0.558	7.5846	15.7705	0.679
84	ENSG00000154864	PIEZO2	2.43E-04	0.043	0.575	6.3662	15.7105	1.490
85	ENSG00000218336	TENM3	2.45E-04	0.043	0.916	3.8772	15.6883	1.886
86	ENSG00000184730	APOBR	2.47E-04	0.043	-0.683	4.5406	15.6691	0.623
87	ENSG00000156026	MCU	2.51E-04	0.043	-0.444	4.9234	15.6260	0.735
88	ENSG00000167994	RAB3IL1	2.52E-04	0.043	0.706	3.5485	15.6192	1.632
89	ENSG00000170909	OSCAR	2.57E-04	0.043	-0.922	4.6308	15.5735	0.528
90	ENSG00000105369	CD79A	2.71E-04	0.045	1.699	3.0873	15.4412	3.247
91	ENSG00000137824	RMDN3	2.78E-04	0.046	-0.484	6.1203	15.3712	0.715
92	ENSG00000071246	VASH1	2.82E-04	0.046	-0.490	5.9859	15.3384	0.712
93	ENSG00000151468	CCDC3	2.83E-04	0.046	0.803	5.0955	15.3291	1.744
94	ENSG00000160179	ABCG1	2.87E-04	0.046	-0.520	6.6498	15.2946	0.697
95	ENSG00000077942	FBLN1	2.92E-04	0.046	0.730	8.9302	15.2476	1.659
96	ENSG00000117676	RPS6KA1	2.93E-04	0.046	-0.395	6.6967	15.2407	0.760
97	ENSG00000121440	PDZRN3	2.98E-04	0.046	0.537	5.5255	15.1989	1.451
98	ENSG00000155465	SLC7A7	2.99E-04	0.046	-0.646	5.7613	15.1960	0.639
99	ENSG00000067365	METTL22	3.08E-04	0.047	-0.503	4.0886	15.1206	0.705
100	ENSG00000156675	RAB11FIP1	3.15E-04	0.047	-0.570	9.2640	15.0609	0.674
101	ENSG00000176692	FOXC2	3.23E-04	0.048	1.426	1.4278	14.9968	2.686
102	ENSG00000120820	GLT8D2	3.42E-04	0.049	0.532	3.5061	14.8584	1.446
103	ENSG00000007237	GAS7	3.47E-04	0.049	0.421	7.1355	14.8245	1.339
104	ENSG00000170153	RNF150	3.51E-04	0.050	0.611	3.8775	14.7918	1.527
105	ENSG00000171914	TLN2	3.57E-04	0.050	0.466	5.1687	14.7536	1.381

Supplementary Table 3.3) GSEA pathway analysis. (Top 50 pathways only.)

Name	Size	NES	FDR q-val	Rank at max	Leading Edge		
					Tags	List	Signal
GOCC_IMMUNOGLOBULIN_COMPLEX	73	2.95	0.00E+00	1515	86%	10%	95%
GOMF_ANTIGEN_BINDING	112	2.75	0.00E+00	1611	60%	10%	66%
GOCC_IMMUNOGLOBULIN_COMPLEX_CIRCULATING	41	2.68	0.00E+00	1332	85%	9%	93%
GOBP_PHAGOCYTOSIS_RECOGNITION	61	2.64	0.00E+00	1332	66%	9%	71%
GOMF_IMMUNOGLOBULIN_RECEPTOR_BINDING	43	2.62	0.00E+00	1332	81%	9%	89%
GOBP_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_I MMUNOGLOBULIN	72	2.61	0.00E+00	1332	60%	9%	65%
GOBP_HUMORAL_IMMUNE_RESPONSE	178	2.60	0.00E+00	2053	48%	13%	55%
GOBP_COMPLEMENT_ACTIVATION	78	2.60	0.00E+00	1332	60%	9%	66%
GOBP_B_CELL_RECEPTOR_SIGNALING_PATHWAY	88	2.57	0.00E+00	1332	51%	9%	56%
GOBP_REGULATION_OF_B_CELL_ACTIVATION	156	2.56	0.00E+00	1375	40%	9%	43%
GOBP_POSITIVE_REGULATION_OF_B_CELL_ACTIVATION	112	2.52	0.00E+00	1375	45%	9%	49%
GOCC_BLOOD_MICROPARTICLE	96	2.47	0.00E+00	1804	48%	12%	54%
GOBP_B_CELL_ACTIVATION	263	2.34	0.00E+00	1971	36%	13%	41%
GOMF_GLYCOSAMINOGLYCAN_BINDING	175	2.32	0.00E+00	1936	39%	12%	44%
GOMF_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT	131	2.31	0.00E+00	2367	48%	15%	56%
GOBP_B_CELL_MEDIATED_IMMUNITY	154	2.30	0.00E+00	1354	35%	9%	38%
GOBP_ADAPTIVE_IMMUNE_RESPONSE	436	2.27	0.00E+00	1779	36%	11%	39%
GOBP_MEMBRANE_INVAGINATION	101	2.27	0.00E+00	1332	41%	9%	44%
GOBP_ANTIMICROBIAL_HUMORAL_RESPONSE	62	2.26	0.00E+00	2053	44%	13%	50%
GOCC_EXTERNAL_SIDE_OF_PLASMA_MEMBRANE	340	2.25	0.00E+00	2102	41%	13%	46%
GOBP_CELL_RECOGNITION	154	2.23	0.00E+00	1765	38%	11%	43%
GOBP_IMMUNOGLOBULIN_PRODUCTION	136	2.23	0.00E+00	1520	31%	10%	34%
GOBP_ANTIGEN_RECEPTOR_MEDIATED_SIGNALING_PATHWAY	199	2.23	0.00E+00	1792	38%	11%	42%

Name	Size	NES	FDR q-val	Rank at max	Leading Edge		
					Tags	List	Signal
GOMF_CYTOKINE_BINDING	118	2.22	0.00E+00	2366	43%	15%	51%
GOMF_HEPARIN_BINDING	132	2.22	0.00E+00	1936	39%	12%	45%
GOBP_SMOOTH_MUSCLE_CELL_PROLIFERATION	129	2.21	0.00E+00	2609	47%	17%	56%
HP_DILATATION_OF_THE_CEREBRAL_ARTERY	43	2.19	0.00E+00	1254	40%	8%	43%
GOBP_REGULATION_OF_LYMPHOCYTE_ACTIVATION	469	2.17	4.15E-05	1876	30%	12%	33%
GOCC_EXTERNAL_ENCAPSULATING_STRUCTURE	423	2.16	7.97E-05	1982	34%	13%	38%
GOBP_GLOMERULUS_DEVELOPMENT	61	2.15	1.55E-04	1418	39%	9%	43%
GOMF_INTEGRIN_BINDING	140	2.15	1.87E-04	2167	40%	14%	46%
GOBP_MUSCLE_CELL_PROLIFERATION	180	2.14	1.82E-04	2609	42%	17%	49%
GOBP_RENAL_SYSTEM_VASCULATURE_DEVELOPMENT	29	2.11	3.52E-04	1040	41%	7%	44%
GOCC_COLLAGEN_CONTAINING_EXTRACELLULAR_MATRIX	331	2.11	3.41E-04	1982	36%	13%	40%
HP_PYLONEPHRITIS	18	2.11	3.32E-04	1004	44%	6%	47%
GOBP_POSITIVE_REGULATION_OF_CELL_ACTIVATION	379	2.11	3.54E-04	2029	31%	13%	35%
HP_ABNORMAL_CEREBRAL_ARTERY_MORPHOLOGY	62	2.09	5.02E-04	1254	31%	8%	33%
HP_ABNORMAL_ASCENDING_AORTA_MORPHOLOGY	22	2.08	6.11E-04	1719	55%	11%	61%
HP_ABNORMAL_SPINAL_MENINGEAL_MORPHOLOGY	22	2.08	6.25E-04	1213	59%	8%	64%
GOBP_ACTIVATION_OF_IMMUNE_RESPONSE	309	2.08	6.09E-04	1792	32%	11%	35%
GOBP_EMBRYONIC_SKELETAL_SYSTEM_MORPHOGENESIS	63	2.07	8.21E-04	1882	33%	12%	38%
HP_TRANSIENT_ISCHEMIC_ATTACK	46	2.07	8.29E-04	2272	41%	15%	48%
GOBP_RESPONSE_TO_CHEMOKINE	72	2.07	8.09E-04	1798	38%	11%	42%
HP_PERIPHERAL_ARTERIAL_STENOSIS	27	2.06	8.97E-04	1213	41%	8%	44%
GOBP_IMMUNE_RESPONSE_REGULATING_CELL_SURFACE_RECEPTOR_SIGNALING_PATHWAY	268	2.06	9.28E-04	1894	32%	12%	36%
GOBP_CHONDROCYTE_DIFFERENTIATION	96	2.05	1.11E-03	2042	36%	13%	42%
GOBP_PRODUCTION_OF_MOLECULAR_MEDIATOR_OF_IMMUNE_RESPONSE	232	2.05	1.11E-03	1666	29%	11%	32%
GOBP_DEFENSE_RESPONSE_TO_BACTERIUM	202	2.05	1.16E-03	2194	38%	14%	44%
GOMF_COLLAGEN_BINDING	63	2.05	1.14E-03	2525	46%	16%	55%

Supplementary Table 3.4) Leading edge genes of the 25 pathways that are most enriched in the SEO-COPD vs control comparison

#1) GOCC: Immunoglobulin complex*	#2) GOMF: Antigen binding*	#3) GOCC: Immunoglobulin complex circulating*	#4) GOBP: Phagocytosis recognition*	#5) GOMF: Immunoglobulin receptor binding*
IGLV2-11 IGLC2 IGLV2-14 IGHV1-18 CD79A IGLV6-57 IGKV1-27 IGLV1-40 <i>+55 more</i>	IGLV2-11 IGLC2 IGLV2-14 IGHV1-18 IGLV6-57 IGLV1-40 IGHA1 IGKV1-17 <i>+59 more</i>	IGLC2 IGHV1-18 IGHA1 IGHG3 IGHV2-70 IGHV3-73 JCHAIN IGHV3-21 <i>+27 more</i>	IGLC2 IGHV1-18 PTX3 IGHA1 MFGE8 IGHG3 IGHV2-70 IGHV3-73 <i>+32 more</i>	IGLC2 IGHV1-18 IGHA1 IGHG3 IGHV2-70 IGHV3-73 JCHAIN IGHV3-21 <i>+27 more</i>
#6) GOBP: Humoral immune response mediated by circulating immunoglobulin*	#7) GOBP: Humoral immune response*	#8) GOBP: Complement activation*	#9) GOBP: B-cell receptor signaling pathway*	#10) GOBP: Regulation of B-cell activation*
C7 C1S IGLC2 IGHV1-18 SERPING1 C1R IGHA1 IGHG3 <i>+35 more</i>	C7 C1S GATA6 IGLC2 IGHV1-18 SERPING1 BCL2 C1R <i>+78 more</i>	C7 C1S IGLC2 IGHV1-18 SERPING1 C1R IGHA1 A2M <i>+39 more</i>	IGLC2 IGHV1-18 CD79A BMX BCL2 IGHA1 MS4A1 CTLA4 <i>+37 more</i>	IGLC2 IKZF3 IGHV1-18 MZB1 BCL2 IGHA1 TNFAIP3 CTLA4 <i>+54 more</i>
#11) GOBP: Positive regulation of B-cell activation*	#12) GOCC: Blood microparticle	#13) GOBP: B-cell activation*	#14) GOMF: Glycosaminoglycan binding§	#15) GOMF: Extracellular matrix structural constituent‡
IGLC2 IGHV1-18 BCL2 IGHA1 TNFRSF13C IGHG3 IGHV2-70 IGHV3-73 <i>+42 more</i>	C1S IGLC2 SERPING1 C1R IGHA1 IGKV1-17 A2M HSPA2 <i>+38 more</i>	IGLC2 IKZF3 IGHV1-18 CD79A LAX1 MZB1 BCL2 IGHA1 <i>+87 more</i>	ADAMTS15 PTPRS SMOC2 FGF7 FBN1 SULF1 CCN5 FGFR4 <i>+61 more</i>	PODN SPON1 EMILIN1 FBN1 FBLN1 COL6A1 SRPX2 MFGE8 <i>+55 more</i>
#16) GOBP: B-cell mediated immunity*	#17) GOBP: Adaptive immune response*	#18) GOBP: Membrane invagination*	#19) GOBP: Antimicrobial humoral response*	#20) GOCC: External side of plasma membrane†
C7 C1S IGLC2 IGHV1-18 SERPING1 C1R IGHA1 IGHG3 <i>+46 more</i>	C7 PRDM1 IGLV2-11 C1S IGLC2 IGLV2-14 IGHV1-18 CD79A <i>+147 more</i>	RHOH IGLC2 IGHV1-18 IGHA1 MFGE8 THBS1 IGHG3 IGHV2-70 <i>+33 more</i>	GATA6 IGHA1 CXCL10 RARRES2 JCHAIN TF IGHM CXCL11 <i>+7 more</i>	CXCL12 ITGA7 IGLC2 IGHV1-18 CD79A CXCR4 CD69 GFRA1 <i>+131 more</i>
#21) GOBP: Cell recognition*	#22) GOBP: Immunoglobulin production*	#23) GOBP: Antigen receptor mediated signaling pathway*	#24) GOMF: Cytokine binding†	#25) GOMF: Heparin binding§
IGLC2 IGHV1-18 PTX3 CXCR4 IGHA1 MFGE8 CCL19 IGHG3 <i>+52 more</i>	IGLV2-11 IGLV2-14 IGLV6-57 IGKV1-27 MZB1 IGLV1-40 IGKV1-17 IGLV3-10 <i>+34 more</i>	IGLC2 IGHV1-18 CD79A LAX1 BMX BCL2 IGHA1 CD3E <i>+67 more</i>	CXCR4 CHRD1 A2M THBS1 TGFB3 LTBP1 CCR7 IL1R1 <i>+43 more</i>	ADAMTS15 PTPRS SMOC2 FGF7 FBN1 CCN5 FGFR4 SAA1 <i>+44 more</i>

Leading edge genes of the 25 pathways with the highest normalized enrichment score (NES) from the SEO-COPD vs control gene set enrichment analysis. Rank in GSEA result list given for each pathway. Leading edge genes in pathways ordered on rank, only first eight shown per pathway. Genes that overlap with the 105 SEO-COPD gene signature are marked in bold.

*Blue: pathway is in the Adaptive immune system cluster. †Green: pathway is in the Chemotaxis and cytokine signaling cluster.

‡Red: pathway is in the Extracellular matrix organization cluster. §Yellow: pathway is in the Glycosaminoglycan binding cluster.

See Figure 3.2 for clusters and Figure 3.3 for corresponding cluster colors and leading edge overlap.

Supplementary Table 3.5) Cell x gene heatmap

Gene symbol	Ensembl ID	FDR SEO-COPD vs Control	Log(FC) SEO-COPD vs Control	Alveolar macrophages	Fibroblast lineage
CH25H	ENSG00000138135	0.021	1.980	-0.65	
GADD45A	ENSG00000116717	0.035	1.632	-0.62	
NNMT	ENSG00000166741	0.039	1.461	-0.61	
FBLN1	ENSG00000077942	0.046	0.730		0.61
SPON1	ENSG00000262655	0.017	0.714		0.76
SLC16A2	ENSG00000147100	0.031	0.650		0.63
FBN1	ENSG00000166147	0.039	0.619		0.71
BICC1	ENSG00000122870	0.035	0.595		0.66
UNC5C	ENSG00000182168	0.023	0.518		0.67
LSAMP	ENSG00000185565	0.039	0.494		0.66
RPS6KA1	ENSG00000117676	0.046	-0.395	0.62	
GGA2	ENSG00000103365	0.039	-0.412	0.77	
RMDN3	ENSG00000137824	0.046	-0.484	0.69	
VASH1	ENSG00000071246	0.046	-0.490	0.61	
TBC1D2	ENSG00000095383	0.014	-0.500		-0.62
ABCG1	ENSG00000160179	0.046	-0.520	0.74	
SLC7A7	ENSG00000155465	0.046	-0.646	0.77	
OSCAR	ENSG00000170909	0.043	-0.922	0.70	
Total significant correlations (FDR < 0.05, rho > 0.6)				10	8

Supplementary Table 3.6) List of eQTLs, 15 top

SNP	Rs-id	Gene name	Beta	T stat	P-value	P adj bonf	N	Consequence type
11:60127398:C:G	rs4939347	MS4A14	1.41	7.55	3.08E-11	7.07E-08	16	intron_variant
11:60128955:A:G	rs11603207	MS4A14	1.41	7.55	3.08E-11	7.07E-08	16	intron_variant
11:60118636:C:T	rs10897054	MS4A14	-1.39	-7.49	4.16E-11	9.56E-08	16	intron_variant
11:60191241:C:T	rs2120181	MS4A14	-1.41	-7.48	4.26E-11	9.79E-08	16	intergenic_variant
11:60181876:A:G	rs1443242	MS4A14	1.40	7.46	4.62E-11	1.06E-07	16	intron_variant
11:60174709:C:G	rs4939354	MS4A14	1.40	7.46	4.63E-11	1.06E-07	16	intron_variant
11:60174896:C:T	rs4938943	MS4A14	1.40	7.46	4.63E-11	1.06E-07	16	intron_variant
11:60179032:C:T	rs10792274	MS4A14	1.40	7.46	4.63E-11	1.06E-07	16	intron_variant
11:60180254:A:G	rs7940689	MS4A14	1.40	7.46	4.63E-11	1.06E-07	16	intron_variant
11:60181627:T:G	rs1443241	MS4A14	1.40	7.46	4.64E-11	1.07E-07	16	intron_variant
11:60768953:G:A	rs10750951	RAB31L1	-0.38	-4.59	1.43E-05	3.28E-02	15	intron_variant
15:50572398:C:T	rs12906304	TNFAIP8L3	-0.67	-4.92	3.82E-06	8.78E-03	20	intron_variant
15:50566473:T:C	rs28451583	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.79E-03	20	intergenic_variant
15:50564358:G:C	rs8034597	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.80E-03	20	intergenic_variant
15:50565246:G:A	rs34405772	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.80E-03	20	intergenic_variant
15:50572601:A:G	rs12910368	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.81E-03	20	intron_variant
15:50570036:A:G	rs12923	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.81E-03	20	3_prime_UTR_variant
15:50571444:T:C	rs4775872	TNFAIP8L3	-0.67	-4.92	3.83E-06	8.81E-03	20	intron_variant
15:50568637:T:G	rs28531946	TNFAIP8L3	-0.67	-4.92	3.84E-06	8.81E-03	20	regulatory_region_variant
15:50605585:G:A	rs11636908	TNFAIP8L3	-0.67	-4.91	3.85E-06	8.86E-03	20	intron_variant
15:50589400:G:A	rs12912679	TNFAIP8L3	-0.67	-4.91	3.93E-06	9.03E-03	20	intron_variant
15:61799148:T:G	rs4377104	TLN2	-0.21	-4.62	1.22E-05	2.81E-02	26	intron_variant
15:61799175:G:C	rs4502155	TLN2	-0.21	-4.62	1.25E-05	2.88E-02	26	intron_variant
15:61799437:T:C	rs4541013	TLN2	-0.21	-4.62	1.26E-05	2.89E-02	26	splice_region_variant
19:54863258:G:A	rs2287826	OSCAR	-0.96	-4.74	7.71E-06	1.77E-02	30	non_coding_transcript_exon_variant
19:54863805:C:T	rs79026632	OSCAR	-0.96	-4.74	7.76E-06	1.78E-02	30	non_coding_transcript_exon_variant
3:132644562:C:T	rs59241569	TMEM108	-0.28	-4.63	1.22E-05	2.80E-02	27	intergenic_variant
3:132644775:C:A	rs73205602	TMEM108	-0.28	-4.63	1.22E-05	2.80E-02	27	intergenic_variant
3:132644815:G:A	rs62290371	TMEM108	-0.28	-4.62	1.22E-05	2.81E-02	27	intergenic_variant

Supplementary Table 3.7) MS4A14 LD Matrix

See <http://www.GitHub.com/vanNijnatten/PhD-Thesis/>.

Supplementary Table 3.8) TNFAIP8L3 LD Matrix

See <http://www.GitHub.com/vanNijnatten/PhD-Thesis/>.

3.8 – References

1. Global Initiative for Chronic Obstructive Lung Disease, I., *GOLD - Global initiative for chronic obstructive lung disease*. 2021.
2. Silverman, E.K., et al., *Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis*. Am J Respir Crit Care Med, 1998. **157**(6 Pt 1): p. 1770-8.
3. Foreman, M.G., et al., *Early-Onset Chronic Obstructive Pulmonary Disease Is Associated with Female Sex, Maternal Factors, and African American Race in the COPD Gene Study*. American Journal of Respiratory and Critical Care Medicine, 2011. **184**(4): p. 414-420.
4. Stanley, S.E., et al., *Telomerase mutations in smokers with severe emphysema*. J Clin Invest, 2015. **125**(2): p. 563-70.
5. Qiao, D., et al., *Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease*. Am J Respir Crit Care Med, 2016. **193**(12): p. 1353-63.
6. Demeo, D.L., et al., *The SERPINE2 gene is associated with chronic obstructive pulmonary disease*. Am J Hum Genet, 2006. **78**(2): p. 253-64.
7. Woldhuis, R.R., et al., *Link between increased cellular senescence and extracellular matrix changes in COPD*. American Journal of Physiology-Lung Cellular and Molecular Physiology, 2020. **319**(1): p. L48-L60.
8. Research code UMCG. 2024 [cited 2024 2024-02-02]; Available from: <https://umcgresearch.org/w/research-code-umcg>.
9. Coreon. *The COREON Code of Conduct for Health Research is available in translation*. 2023 [cited 2023 2023-06-21T10:36:51+00:00]; Available from: <https://www.coreon.org/the-coreon-code-of-conduct-for-health-research-is-available-in-translation/>, <https://www.coreon.org/gedragcode-gezondheidsonderzoek/>.
10. Woldhuis, R.R., et al., *Cellular Senescence Affects ECM Regulation in COPD Lung Tissue*. BioRxiv, 2023.
11. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
12. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes*. Nat Genet, 2003. **34**(3): p. 267-73.
13. Merico, D., et al., *Enrichment map: a network-based method for gene-set enrichment visualization and interpretation*. PLoS One, 2010. **5**(11): p. e13984.
14. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
15. Reimand, J., et al., *Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap*. Nat Protoc, 2019. **14**(2): p. 482-517.
16. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0*. Bioinformatics, 2011. **27**(12): p. 1739-40.
17. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry*. Nat Biotechnol, 2019. **37**(7): p. 773-782.
18. Sikkema, L., et al., *An integrated cell atlas of the lung in health and disease*. Nat Med, 2023. **29**(6): p. 1563-1577.
19. Sollis, E., et al., *The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource*. Nucleic Acids Res, 2023. **51**(D1): p. D977-D985.
20. Das, S., et al., *Next-generation genotype imputation service and methods*. Nat Genet, 2016. **48**(10): p. 1284-1287.
21. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations*. Bioinformatics, 2012. **28**(10): p. 1353-8.
22. Davis, J.R., et al., *An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants*. Am J Hum Genet, 2016. **98**(1): p. 216-24.

23. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants*. Bioinformatics, 2015. **31**(21): p. 3555-3557.
24. Della Mea, V., et al., *SlideJ: An ImageJ plugin for automated processing of whole slide images*. PLoS One, 2017. **12**(7): p. e0180540.
25. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis*. Nature methods, 2012. **9**(7): p. 676-682.
26. Ngassie, M.L.K., et al., *Age-associated differences in the human lung extracellular matrix*. American Journal of Physiology-Lung Cellular and Molecular Physiology, 2023.
27. Schiller, H.B., et al., *The human lung cell atlas: a high-resolution reference map of the human lung in health and disease*. American journal of respiratory cell and molecular biology, 2019. **61**(1): p. 31-41.
28. Liu, G., et al., *Fibulin-1 regulates the pathogenesis of tissue remodeling in respiratory diseases*. JCI insight, 2016. **1**(9).
29. Liu, G., et al., *Fibulin-1c regulates transforming growth factor- β activation in pulmonary tissue fibrosis*. JCI insight, 2019. **4**(16).
30. Mason, D.Y., et al., *CD79a: a novel marker for B-cell neoplasms in routinely processed tissue samples*. 1995.
31. Polverino, F., A. Sam, and S. Guerra, *COPD: to be or not to be, that is the question*. The American journal of medicine, 2019. **132**(11): p. 1271-1278.
32. van der Strate, B.W., et al., *Cigarette smoke-induced emphysema: a role for the B cell?* American journal of respiratory and critical care medicine, 2006. **173**(7): p. 751-758.
33. Lutz, S.M., et al., *A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry*. BMC genetics, 2015. **16**: p. 1-11.
34. Imboden, M., et al., *Genome-wide association study of lung function decline in adults with and without asthma*. Journal of allergy and clinical immunology, 2012. **129**(5): p. 1218-1228.
35. Sakornsakolpat, P., et al., *Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations*. Nature genetics, 2019. **51**(3): p. 494-505.
36. Kim, W., et al., *Genome-wide gene-by-smoking interaction study of chronic obstructive pulmonary disease*. American journal of epidemiology, 2021. **190**(5): p. 875-885.
37. Kichaev, G., et al., *Leveraging polygenic functional enrichment to improve GWAS power*. The American Journal of Human Genetics, 2019. **104**(1): p. 65-75.
38. Cosentino, J., et al., *Inference of chronic obstructive pulmonary disease with deep learning on raw spirometry identifies new genetic loci and improves risk models*. Nature Genetics, 2023. **55**(5): p. 787-795.
39. Forno, E., et al., *Overweight, obesity, and lung function in children and adults—a meta-analysis*. The Journal of Allergy and Clinical Immunology: In Practice, 2018. **6**(2): p. 570-581. e10.
40. Shrine, N., et al., *Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk*. Nature genetics, 2023. **55**(3): p. 410-422.
41. Zhu, Z., et al., *Genetic overlap of chronic obstructive pulmonary disease and cardiovascular disease-related traits: a large-scale genome-wide cross-trait analysis*. Respiratory research, 2019. **20**: p. 1-14.
42. Sinkala, M., et al., *A genome-wide association study identifies distinct variants associated with pulmonary function among European and African ancestries from the UK Biobank*. Communications Biology, 2023. **6**(1): p. 49.
43. Polverino, F., et al., *B cells in chronic obstructive pulmonary disease: moving to center stage*. Am J Physiol Lung Cell Mol Physiol, 2016. **311**(4): p. L687-L695.

44. Georgopoulos, K., S. Winandy, and N. Avitahl, *The role of the Ikaros gene in lymphocyte development and homeostasis*. Annual review of immunology, 1997. **15**(1): p. 155-176.
45. Polverino, F., et al., *A novel insight into adaptive immunity in chronic obstructive pulmonary disease: B cell activating factor belonging to the tumor necrosis factor family*. American journal of respiratory and critical care medicine, 2010. **182**(8): p. 1011-1019.
46. Hogg, J.C., et al., *The nature of small-airway obstruction in chronic obstructive pulmonary disease*. New England Journal of Medicine, 2004. **350**(26): p. 2645-2653.
47. Gosman, M.M., et al., *Increased number of B-cells in bronchial biopsies in COPD*. Eur Respir J, 2006. **27**(1): p. 60-4.
48. Polverino, F., et al., *B cell-activating factor. an orchestrator of lymphoid follicles in severe chronic obstructive pulmonary disease*. American journal of respiratory and critical care medicine, 2015. **192**(6): p. 695-705.
49. Bonarius, H., et al., *Antinuclear autoantibodies are more prevalent in COPD in association with low body mass index but not with smoking history*. Thorax, 2011. **66**(2): p. 101-107.
50. Brandsma, C.A., et al., *Differential switching to IgG and IgA in active smoking COPD patients and healthy controls*. Eur Respir J, 2012. **40**(2): p. 313-21.
51. Brusselle, G.G., et al., *Lymphoid follicles in (very) severe COPD: beneficial or harmful?* Eur Respir J, 2009. **34**(1): p. 219-30.
52. Brandsma, C., et al., *The search for autoantibodies against elastin, collagen and decorin in COPD*. European Respiratory Journal, 2011. **37**(5): p. 1289-1292.
53. Roark, E.F., et al., *The association of human fibulin-1 with elastic fibers: an immunohistological, ultrastructural, and RNA study*. Journal of Histochemistry & Cytochemistry, 1995. **43**(4): p. 401-411.
54. Baldwin, A.K., et al., *Elastic fibres in health and disease*. Expert reviews in molecular medicine, 2013. **15**: p. e8.
55. Adamo, C.S., et al., *EMILIN1 deficiency causes arterial tortuosity with osteopenia and connects impaired elastogenesis with defective collagen fibrillogenesis*. The American Journal of Human Genetics, 2022. **109**(12): p. 2230-2252.
56. Brandsma, C.A., et al., *A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD*. Thorax, 2015. **70**(1): p. 21-32.
57. Khalenkow, D., et al., *Alternative Splicing Is a Major Factor Shaping Transcriptome Diversity in Mild and Severe COPD*. Am J Respir Cell Mol Biol, 2024.
58. Deslee, G., et al., *Elastin expression in very severe human COPD*. Eur Respir J, 2009. **34**(2): p. 324-331.
59. Mecham, R.P., *Elastin in lung development and disease pathogenesis*. Matrix biology, 2018. **73**: p. 6-20.
60. Dancevic, C.M., et al., *Biosynthesis and expression of a disintegrin-like and metalloproteinase domain with thrombospondin-1 repeats-15: a novel versican-cleaving proteoglycanase*. Journal of Biological Chemistry, 2013. **288**(52): p. 37267-37276.
61. Mattioli, I., A. Mantovani, and M. Locati, *The tetraspan MS4A family in homeostasis, immunity, and disease*. Trends in Immunology, 2021. **42**(9): p. 764-781.
62. Silva-Gomes, R., et al., *Differential expression and regulation of MS4A family members in myeloid cells in physiological and pathological conditions*. Journal of Leukocyte Biology, 2022. **111**(4): p. 817-836.
63. Bordoloi, D., et al., *TIPE family of proteins and its implications in different chronic diseases*. International journal of molecular sciences, 2018. **19**(10): p. 2974.
64. Gu, Z., et al., *Regulatory roles of tumor necrosis factor- α -induced protein 8 like-protein 2 in inflammation, immunity and cancers: A review*. Cancer management and research, 2020: p. 12735-12746.

65. Padmavathi, G., et al., *Novel tumor necrosis factor- α induced protein eight (TNFAIP8/TIPE) family: Functions and downstream targets involved in cancer progression*. Cancer letters, 2018. **432**: p. 260-271.
66. Cui, J., et al., *Identical expression profiling of human and murine TIPE3 protein reveals links to its functions*. Journal of Histochemistry & Cytochemistry, 2015. **63**(3): p. 206-216.
67. Niture, S., J. Moore, and D. Kumar, *TNFAIP8: inflammation, immunity and human diseases*. Journal of cellular immunology, 2019. **1**(2): p. 29.
68. Fayngerts, S.A., et al., *TIPE3 is the transfer protein of lipid second messengers that promote cancer*. Cancer Cell, 2014. **26**(4): p. 465-78.
69. Song, J., et al., *TIPE3 protects mice from lipopolysaccharide-induced acute lung injury*. Transpl Immunol, 2023. **77**: p. 101799.

Chapter 4 – Differential miRNA expression in SEO-COPD is associated with altered lysosome, vesicular transport and ECM-related pathways

Nicolaas J. Bekker, Jos van Nijnatten, Roy R. Woldhuis, Wierd Kooistra, Wim Timens, Maarten van den Berge, Alen Faiz[†], Corry-Anke Brandsma[†]

Status: Not published

4.1 – Abstract

Rationale: Severe early-onset (SEO-)COPD manifests at early age with high disease burden; why certain individuals are more susceptible to manifesting severe COPD early is still largely unknown. Previously we identified a differential expression pattern on RNA level in SEO-COPD. Supplementing this pattern with miRNA level data gives further insight in underlying molecular mechanisms of SEO-COPD pathology.

Methods: Small RNA sequencing was used to determine differential miRNA expression in peripheral lung tissue samples between non-COPD (control) (n=31) and SEO-COPD subjects (n=18), followed by excluding genes that were differential between control and common COPD subjects (n=21). Both analyses were corrected for age and sex. Differential miRNAs were characterized by determining inverse correlations between the miRNAs and their targets genes, and exploring the biological pathways of the target genes.

Results: Three significantly differentially expressed miRNAs were found in SEO-COPD subjects: miR-193b-3p and miR-202-5p had higher expression and miR-331-3p had lower expression in SEO-COPD compared to control, and all three were not different between common COPD and control. MiR-193b-3p had 35 inversely correlated target genes, that were enriched for pathways related to vesicular and endosomal membranes, secretory/transport vesicles and lysosomal activity, including transmembrane proteolysis genes NCSTN and APM1 involved in Notch and Wnt signaling. MiR-331-3p had 41 inversely correlated gene targets, that were enriched for ECM organization and morphogenesis pathways, including ECM genes COL6A1, COL6A2 and ELN. No inversely correlated gene targets were identified for miR-202-5p.

Conclusions: We identified three differentially expressed miRNAs specific to SEO-COPD in lung tissue. Based on their target genes and functions, miR-193b-3p may be involved in regulation of vesicular and lysosomal activity, while miR-331-3p may be a regulator of aberrant ECM repair in SEO-COPD. By directly connecting differential miRNA expression to target gene expression we generated important new insights into the potential involvement of these miRNAs in SEO-COPD pathogenesis.

Keywords: miRNA, COPD, Severe Early Onset COPD, Small RNA sequencing, Lung

4.2 – Introduction

Chronic obstructive pulmonary disease (COPD) is a common chronic disease characterized by airflow obstruction and symptoms of cough, dyspnea, and mucus hypersecretion. It has an estimated global prevalence of 10.3% and is ranked as the 3rd leading cause of death¹. Risk factors for COPD include environmental exposures such as smoking, air pollution, and occupational hazards. Genetics also play an important role and may explain the large interindividual differences in susceptibility to develop COPD¹. A subset of patients is particularly susceptible and develops very severe COPD already at a relatively young age (≤ 55 years). This group is referred to as severe early-onset (SEO-)COPD^{2,3}. We propose that SEO-COPD represents a sub-group with a specific disease pathogenesis, that is different from COPD with late(r)-onset of disease that is more common, which is referred to as common COPD here. Previously we identified a lung tissue gene expression signature specific to SEO-COPD and not seen in common COPD⁴. This gene signature included genes involved in adaptive immune responses as well as active ECM repair and remodeling, likely reflecting ongoing, chronic tissue damage.

In the current manuscript, we aimed to expand these findings to gain insight in the role of microRNAs (miRNAs) in SEO-COPD. MicroRNAs are highly conserved, small, non-coding RNAs with an average length of 22 nucleotides, transcribed by RNA polymerase II from miRNA genes or from the introns or exons of other host genes⁵. Their primary function is gene silencing through post-transcriptional regulation, thereby influencing a multitude of biological processes such as proliferation, differentiation, apoptosis, development, survival, stress-resistance, tumorigenesis and the immune system^{6,7}.

Differential expression and dysregulation of miRNAs has been detected and implicated in multiple diseases, including respiratory conditions such as COPD^{6,7}.

In the current manuscript, we assessed differential miRNA expression in peripheral lung tissue comparing SEO-COPD patients with non-COPD control subjects and patients with common COPD using small RNA sequencing. This was followed by identifying the putative target genes and pathways affected by these differentially expressed miRNAs to elucidate their potential role in SEO-COPD pathogenesis.

4.3 – Methods

Subject inclusion

Peripheral lung tissue was used from COPD patients and non-COPD controls present in our PRESTO (PRoteogenomics Early onSeT cOpd) cohort⁴. Lung tissue was included from the following groups; non-COPD control (n=31, FEV₁/FVC ratio >70%), common COPD (n=21, FEV₁/FVC ratio <70% and FEV₁ % pred 40% - 80%) and severe early-onset (SEO-)COPD (n=18, FEV₁/FVC ratio < 70%, FEV₁ % pred < 40% and age ≤ 55 years). SEO-COPD definition was based on previous publications^{2,3,8}. To exclude the possibility of including very severe COPD patients in the common COPD group that actually had early onset COPD but received lung transplantation at an age above 55, we only included patients with FEV₁ % pred >40% in the common COPD group.

This project was conducted in accordance with the Research Code of the University Medical Center Groningen⁹ and the Dutch national code of conduct¹⁰. The use of left-over lung tissue in this study was not subject to Medical Research Human Subjects Act in the Netherlands, as confirmed by a statement of the Medical Ethical Committee of the University Medical Center Groningen and therefore exempt from consent according to national laws (Dutch laws: Medical Treatment Agreement Act (WGBO) art 458 / GDPR art 9/ UAVG art 24). All donor material and clinical

information were deidentified prior to experimental procedures, blinding any identifiable information to the investigators.

Small-RNA-sequencing & RNA-sequencing

Total RNA was isolated from peripheral lung tissue using QIAgen AllPrep DNA/RNA/miRNA Universal Kit per manufacturer's instructions⁴. The miRNA sequencing libraries were prepared from the RNA samples using Bioo Scientific NEXTflex Small RNA-Seq Kit V3 according to manufacturer's instructions. GenomeScan B.V. depleted ribosomal RNA and executed paired-end sequencing of total RNA and miRNA libraries via an Illumina NovaSeq6000.

FastQC (version 0.11.7) was used to determine the raw RNA sequencing data quality. For the small-RNA sequencing, adapters were removed from the sequences using TrimGalore (version 0.6.7), 4N ends were cut, and the identical sequences were collapsed. Bowtie2¹¹ (version 2.3.4.2) was used to align the raw sequences to miRNA reference sequences¹² (miRBase version 22.1). We checked if the mapped read counts were similar across samples of the same type, and we checked for any outliers in a principal component analysis.

RNA-Seq data was retrieved from the same samples and processed as previously described⁴.

Differential miRNA expression analysis

The miRNA data was normalized through log₂ transformation of miRNA expression counts per million. Differential expression analysis was performed using linear models in EdgeR¹³ (version 3.40.2, dependent on limma version 3.54.2, in R version 4.2.2) while correcting for sex and age. A Benjamini-Hochberg FDR was calculated where appropriate. An FDR less than 0.05 was considered significant unless specified differently. Only miRNAs with a median expression of ≥ 10 read counts were included. To identify miRNAs that are differentially expressed in SEO-COPD, miRNA expression levels of all miRNAs were first compared between SEO-COPD and non-COPD control. Next, the differentially expressed miRNAs identified in SEO-COPD were compared between non-COPD and common COPD to determine whether they were also differentially expressed in common COPD and thereby not distinct for SEO-COPD. The miRNAs that were nominally significantly ($P < 0.05$) differentially expressed in common COPD with the same direction of change as control vs SEO-COPD were excluded from follow-up analyses. Hereafter, the expression of the remaining miRNAs was compared between the current and ex-smoker non-COPD controls using a nominal significance cutoff ($P < 0.05$) to exclude miRNAs that were affected by current smoking.

Correlation between differentially expressed miRNAs and putative target gene expression

Predicted and validated target genes for the differentially expressed miRNAs in SEO-COPD were retrieved from online databases with multiMiR¹⁴ (version 1.20.0) using the default settings. Correlation analysis between differentially expressed miRNAs and gene expression was performed using MatrixEQTL¹⁵ (version 2.3) for predicted and validated target genes as well as unbiased for all genes, and on all subjects with matched miRNA and mRNA expression data ($n=66$). Gene expression was normalized by log₂ transforming the gene expression counts per million. Lowly expressed genes were removed from the analysis. We corrected for disease status, sex and age.

Pathway analysis

To predict the impact of the genes influenced directly or indirectly by the miRNA on biological pathways, StringDB¹⁶ (version 12) was used to create a protein interaction map of the miRNA target genes that were inversely correlated with the expression of the miRNA, i.e. target gene expression was lower when miRNA expression was higher or vice versa. All interaction sources were enabled

and a minimum required interaction score of 0.400 was used. Further, g:Profiler¹⁷ was used to determine enrichment of biological pathways among the miRNA target genes (ordered query, data sources: GO molecular function, GO cellular component, GO biological process; g:Profiler version e110_eg57_p18_4b54a898, database updated on 14/09/2023), which was then visualized as a pathway network in Cytoscape^{18, 19} (ver3.9.0) as described previously²⁰. For both analyses there were two inputs: one consisted of only the validated/predicted target genes, the other consisted of all genes from the transcriptomic output. In both cases only genes were used that were inversely correlated to the differential miRNAs.

Relation between miRNAs and cell types

To determine whether differential miRNA expression was associated with cell-type composition, we used our previously generated cell type deconvolution results that were obtained from the lung tissue RNA-Seq data to predict cell proportions of the different cell types⁴. These data were used to assess the association between differential miRNA expression and cell type proportions using Spearman correlation.

Using the R (version 4.1.2) miRSCAPE²¹ package (version 1.0) and the Human Lung Cell Atlas²² (HLCA) as the single-cell reference data the relative expression level of the differential miRNAs for specific lung cell types was inferred. The matched bulk miRNA and mRNA data were processed via the default bulkTransform function, and the bulk miRNA was log₂-normalized.

4.4 – Results

Clinical Characteristics

In the current study, we investigated the differences in miRNA expression in peripheral lung tissue between patients with SEO-COPD (n=18) and non-COPD controls (n=31) and between common COPD (n=21) and controls. The clinical characteristics of the included subjects are presented in Table 4.1. As expected and described previously^{2, 3}, the SEO-COPD group was younger and consisted of more females than the non-COPD controls. We corrected for these two confounding factors in our models. Additionally, there were relatively more current smokers in the non-COPD control group than in the COPD groups. To address this, we determined if the differentially expressed miRNAs identified in SEO-COPD were affected by current smoking. This was done in the control group only, since there were no current smokers in the SEO-COPD group.

Table 4.1) Subject clinical characteristics

Characteristic	N	Control N = 31	Severe Early Onset (SEO)-COPD N = 18	Common COPD N = 21	p-value control vs SEO-COPD (Mann Whitney U)
Sex, N	70	17 f – 14 m	13 f – 5 m	5 f – 16 m	0.023
Age, Median (range)	70	61 (45 – 76)	52 (39 – 55)	69 (47 – 79)	< 0.001
Smoking status, N	70	18 CS – 13 ES	0 CS – 18 ES	5 CS – 16 ES	< 0.001
Pack years, Median (range)	70	35.0 (10.0 – 75.0)	28.0 (6.0 – 54.0)	40 (6.5 – 130.0)	0.050
FEV ₁ % pred, Median (range)	70	84.3 (59.8 – 125)	17.25 (12.0 – 37.0)	58.6 (45.0 – 77.0)	< 0.001
FEV ₁ /FVC %, Median (range)	70	76.0 (71.0 – 86.6)	29.2 (20.5 – 50.0)	53.0 (40.0 – 67.9)	< 0.001

f = female, m = male, CS = current smoker, ES = ex-smoker. All clinical characteristics of controls and SEO-COPD compared with Mann Whitney U.

Identification of three SEO-COPD miRNAs

After QC and filtering, 495 miRNAs remained for differential miRNA expression analysis. We identified four differentially expressed (DE) miRNAs between SEO-COPD and non-COPD control (FDR < 0.05). MiR-331-3p was significantly lower expressed, and miR-193b-3p, miR-202-5p and miR-1246 were significantly higher expressed in SEO-COPD compared to non-COPD control lung tissue. A volcano plot is depicted in Figure 4.1A, and boxplots showing the differences in expression are shown in Figure 4.1B-E. It was noted that the differential expression of hsa-miR-1246 was likely driven by several outliers in both the miRNA expression data (Figure 4.1E) and the residual data (Supplementary Figure 4.1). Therefore, it was removed from our miRNA signature.

Next, the expression of the three remaining miRNAs was assessed in patients with common COPD versus non-COPD controls. None of the three SEO-COPD miRNAs were significant in this comparison ($P < 0.05$) indicating their differential expression to be specific for SEO-COPD. To account for the imbalance with respect to smoking status between the groups as a possible confounder, we determined whether the expression of the three SEO-COPD miRNAs was affected by current smoking, which was not the case.

Identification of miRNA correlated target genes

MultimiR was then used to identify the validated and predicted target genes in the matched RNA sequencing data for the three SEO-COPD-associated miRNAs. The total number of genes in the RNA sequencing data was 17050. From there, the total number of validated and predicted gene targets were respectively 3747 for miR-193b-3p, 1999 genes for miR-331-3p, and 1877 for miR-202-5p. For each of the SEO-COPD miRNAs, correlation analysis was performed between the miRNA and the target gene expression. Since miRNAs generally silence their target genes, we only focused on the correlations where gene expression was lower when miRNA expression was higher or vice versa (i.e. inverse correlations).

For miR-193b-3p we identified 35 significant predicted/validated target gene correlations where lower gene expression was correlated with higher miR-193b-3p expression. These included NCSTN, ALDH2 and GLB1 as the most significant correlations (Table 4.2A, Figure 4.2A-C). For miR-331-3p, we identified 41 significant predicted/validated target gene correlations where higher gene expression was correlated with lower miR-331-3p expression. Among these KREMEN1, ADRA2C, and EEF1D were most significant (Table 4.2B, Figure 4.2D-F). For miR-202-5p we did not find any significant correlation with target genes. Full lists of inversely correlated predicted/validated miRNA target genes can be found in Supplementary Table 4.1.

A) Differential miRNA expression; SEO-COPD vs non-COPD control

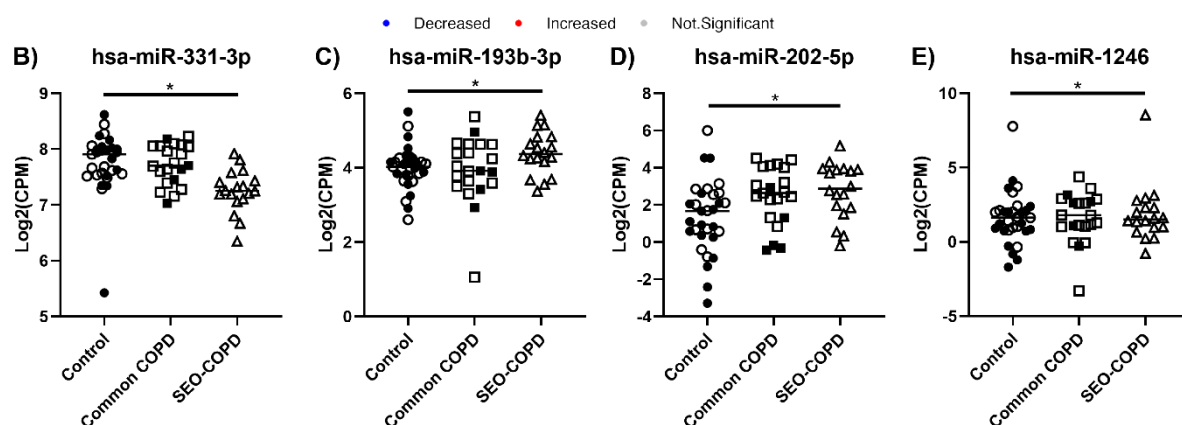
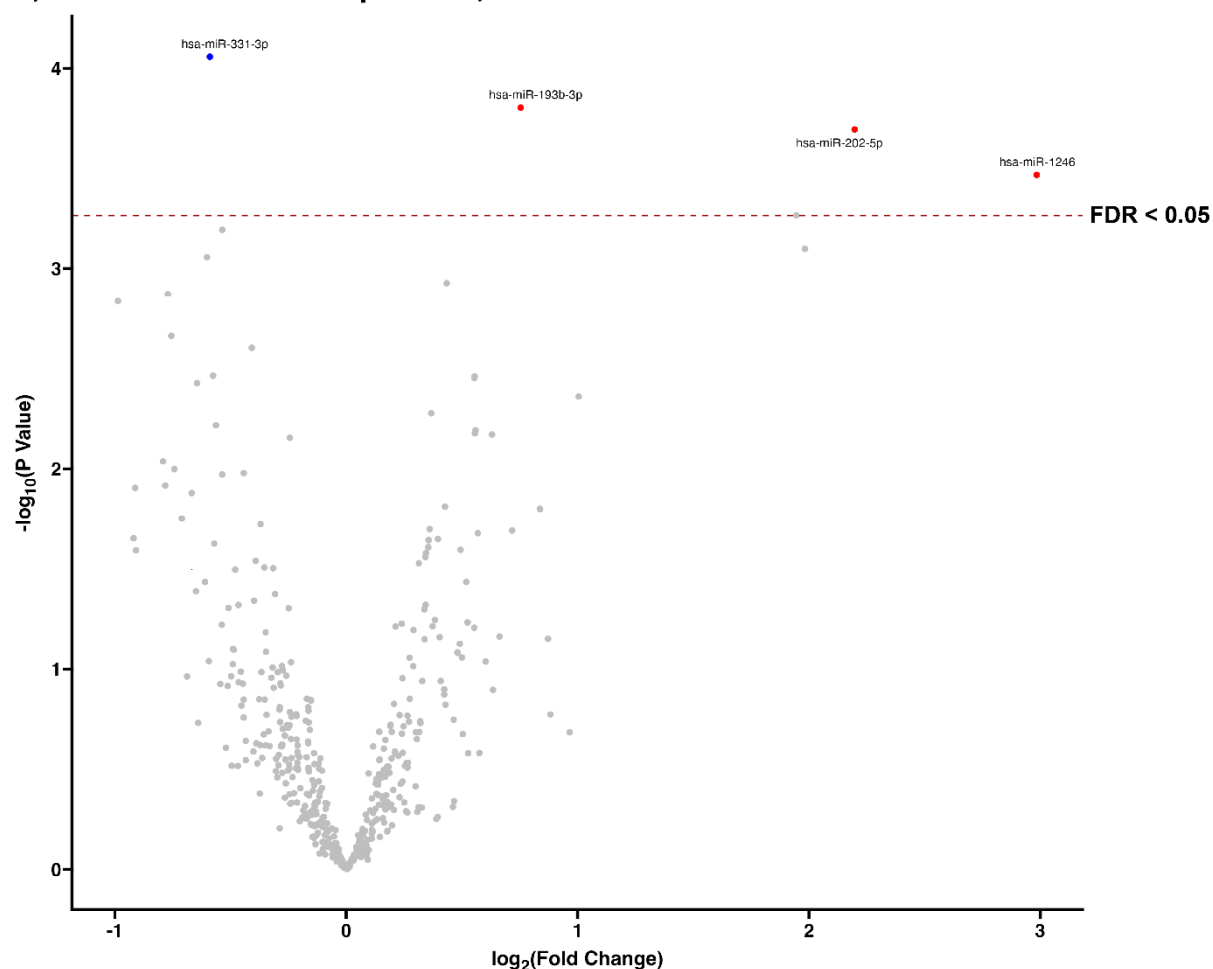


Figure 4.1) Differential miRNA expression in peripheral lung tissue of SEO-COPD subjects vs non-COPD controls. A) Differential miRNA expression profiling of control vs SEO-COPD subjects based on linear regression analysis (correction: age and sex). 495 miRNAs were detected and passed filtering steps, of which four were significantly differentially expressed in SEO-COPD subjects compared to controls (Linear regression, FDR < 0.05). B-E) miRNA expression visualization of the four differential miRNAs: hsa-miR-331 (B), hsa-miR-193b-3p (C), hsa-miR-202-5p (D) and hsa-miR-1246 (E). * = Statistical significance (FDR < 0.05), based on linear regression analysis. Open symbols are ex-smoker subjects, closed symbols are current smoker subjects.

Identification of pathways targeted by SEO-COPD miRNAs

Protein-protein interactions and enrichment of biological pathways among the significant predicted and validated target genes of miR-193b-3p and miR-331-3p were analyzed and visualized to assess the pathways that are targeted by SEO-COPD miRNAs.

For miR-193b-3p, the 35 inversely correlated genes were enriched in three interconnected pathway clusters pertaining to vesicular and endosomal membranes, transport vesicles and secretory vesicles respectively (Figure 4.3A, Table 4.3A). These pathways included the genes NCSTN, GLB1, MON1B, RAB5B, GRN and APH1A, which were all also connected in the protein-protein interaction network (Figure 4.3B).

For miR-331-3p, pathway enrichment analysis demonstrated the 41 inversely correlated miR-331-3p genes to be enriched in a network of extracellular matrix (ECM) pathways (Figure 4.3C, Table 4.3B), as well as pathways related to morphogenesis and (regulation of) anatomical development and size. The enrichment of ECM pathways is also apparent in the protein-protein interaction network which shows interaction between the ECM related genes COL4A1, COL4A2, ELN and CSPG4 (Figure 4.3D). In Table 4.2, the enriched pathways are indicated for the top 20 inversely correlated genes for miR-193b-3p and miR-331-3p.

In addition, to confirm the pathway analysis results in wider context and to gain insight on the associated pathways of the miRNAs outside of their known target genes, we also performed an unbiased pathway analysis on all inversely correlated genes of miR-193b-3p and miR-331-3p, not just those that were identified as predicted/validated target genes in MultiMir. Of the 17050 total genes, miR-193b-3p had 109 genes that were significantly lower expressed with higher miR-193b-3p expression. MiR-331-3p had 351 genes that were significantly higher expressed with lower miR-331-3p expression. Similar to the results based on just the predicted/validated miRNA target genes, unbiased pathway enrichment analysis of the 109 miR-193b-3p inversely correlated genes showed a network of pathways relating to transport, secretory and synaptic vesicles, in addition to lysosomes microbody membranes (Figure 4.4A). For miR-331-3p this analysis resulted in a large network of pathways again including several pathways related to ECM and fiber organization in addition to pathways related to development of muscle, heart and vasculature, regulation of growth and apoptosis, and the adaptive immune system (Figure 4.4B).

Table 4.2) Top 20 target genes inversely correlated with the differential SEO-COPD miRNAs

Gene ID	FDR	Beta	Predicted / validated target	Enriched pathway cluster
hsa-miR-193b-3p (Expression in SEO-COPD: ↑)				
NCSTN	0.001	-0.21	Validated	Vesicular and endosomal membranes, secretory vesicles, transport vesicles
ALDH2	0.012	-0.34	Predicted	-
GLB1	0.019	-0.28	Validated	Secretory vesicles
MON1B	0.019	-0.16	Validated	Vesicular and endosomal membranes
HNRNPUL1	0.019	-0.15	Validated	-
PTPN9	0.019	-0.15	Validated	-
RAB5B	0.020	-0.15	Validated	Vesicular and endosomal membranes, secretory vesicles, transport vesicles
GRN	0.021	-0.40	Validated	Vesicular and endosomal membranes, secretory vesicles
SNRPD3	0.024	-0.19	Validated	-
APH1A	0.025	-0.15	Validated	Vesicular and endosomal membranes, secretory vesicles, transport vesicles
DCAF7	0.026	-0.16	Validated	-
TGOLN2	0.026	-0.15	Validated	Vesicular and endosomal membranes, transport vesicles

Gene ID	FDR	Beta	Predicted / validated target	Enriched pathway cluster
TPRG1L	0.028	-0.19	Validated	Vesicular and endosomal membranes, secretory vesicles, transport vesicles
ARSD	0.030	-0.20	Validated	-
MRPS16	0.036	-0.19	Validated	-
WNK1	0.036	-0.15	Validated	-
CSTF1	0.036	-0.20	Validated	-
MRPS18B	0.039	-0.20	Validated	-
TRIM14	0.042	-0.28	Validated	-
SPN	0.042	-0.38	Predicted	-
hsa-miR-331-3p (Expression in SEO-COPD: ↓)				
KREMEN1	0.002	-0.44	Validated	-
ADRA2C	0.002	-1.05	Validated	Regulation of anatomical structure size
EEF1D	0.006	-0.24	Validated	-
FOXC1	0.006	-1.21	Predicted	Apoptotic process involved in morphogenesis, regulation of anatomical structure size
BCL2L11	0.007	-0.42	Validated	Apoptotic process involved in morphogenesis, mammary gland development
COL6A2	0.007	-0.51	Validated	Extracellular matrix
ZNF703	0.011	-0.75	Validated	Mammary gland development
UNC45B	0.012	-1.00	Validated	-
RNF122	0.013	-0.90	Predicted	-
ISYNA1	0.013	-0.45	Validated	-
CSPG4	0.013	-0.61	Predicted	Extracellular matrix
GAS7	0.013	-0.36	Predicted	-
IGFBP5	0.013	-0.72	Validated	Mammary gland development
NCS1	0.014	-0.86	Validated	-
COL6A1	0.014	-0.50	Validated	Extracellular matrix
BTG2	0.014	-0.96	Validated	-
JSRP1	0.015	-1.59	Predicted	-
ATP8B2	0.021	-0.33	Validated	-
NTN1	0.021	-0.75	Validated	Extracellular matrix, regulation of anatomical structure size, mammary gland development
SPTB	0.021	-0.40	Validated	Regulation of anatomical structure size

Table of the 20 most significant inverse correlations between miRNA expression and gene expression for miR-193b-3p (A) and miR-331-3p (B). Pathway clusters from Figure 4.3 that the genes are enriched in are listed.

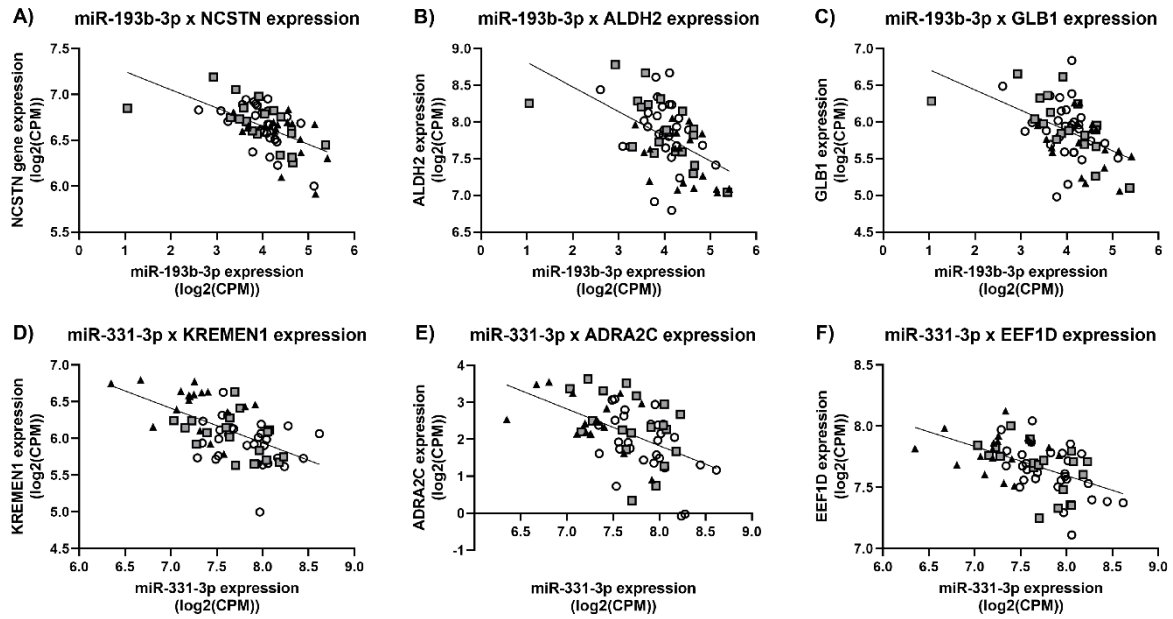
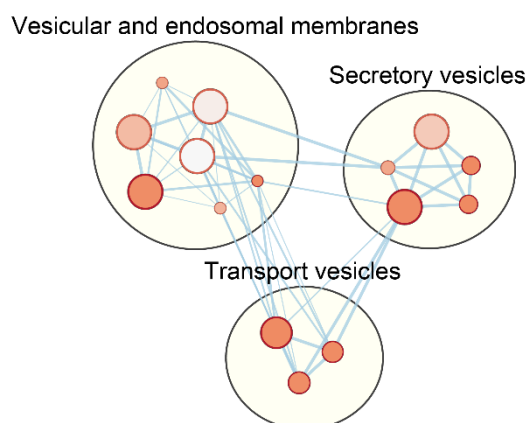
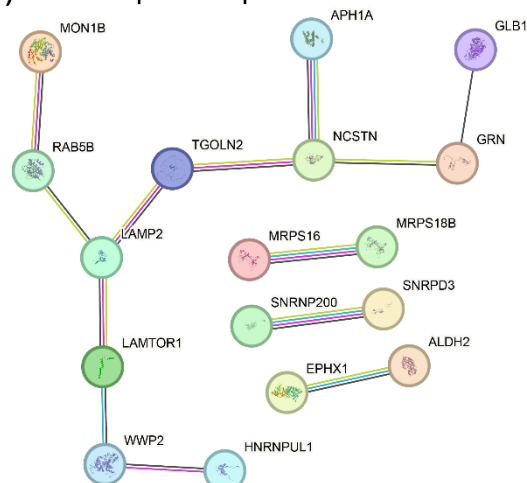


Figure 4.2) Visualizations of significant inverse correlations between the differential SEO-COPD miRNAs and their predicted/validated target genes. A-C) Most significant correlations for miR-193b-3p. D-F) Most significant correlations for miR-331-3p. See Table 4.2A/B for details. Circles are controls, squares are common COPD and triangles are SEO-COPD subjects.

A) miR-193b-3p: Visualization of enriched pathways

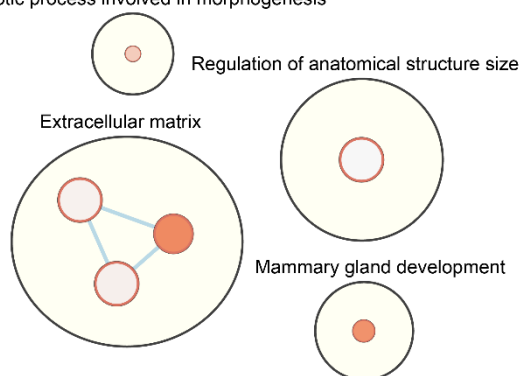


B) miR-193b-3p: Protein-protein interactions



C) miR-331-3p: Visualisation of enriched pathways

Apoptotic process involved in morphogenesis



D) miR-331-3p: Protein-protein interactions

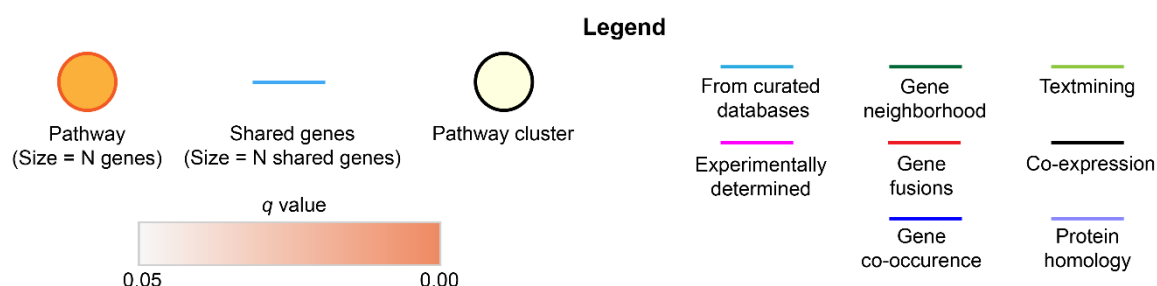
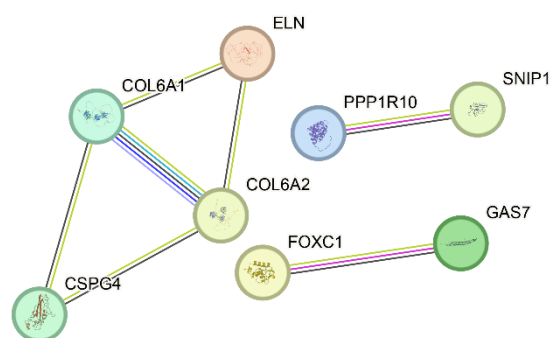


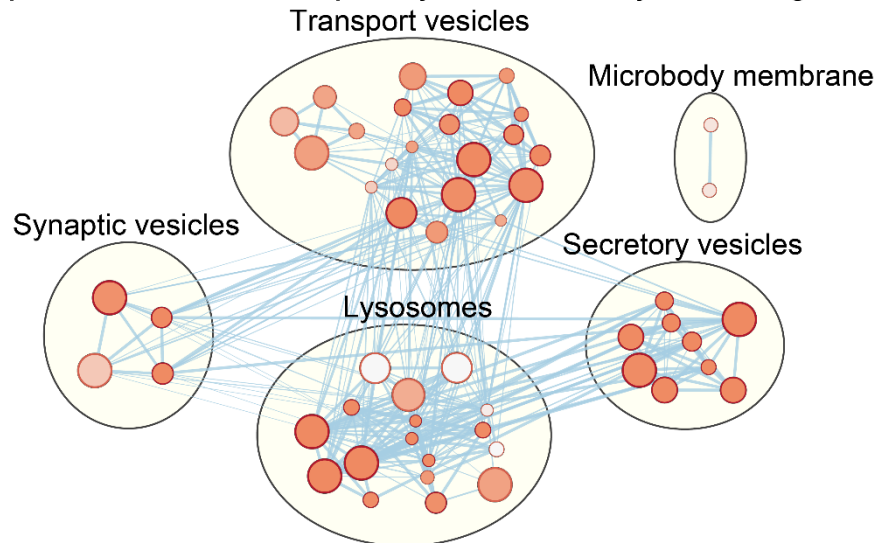
Figure 4.3) Pathway enrichment analysis of inversely correlated miR-193b-3p and miR-331-3p predicted & validated target genes. A/C) G:Profiler/Cytoscape pathway enrichment analyses visualize significant pathways (FDR $q < 0.05$) that the negatively correlated miRNA target genes (validated and predicted) are enriched in and whether these pathways share genes. Together the pathways form pathway clusters of related biological functions that give a suggestion to the role of the miRNA and its target genes in SEO-COPD pathology. Genes enriched in each pathway cluster are given in Table 4.2A/B. Exact pathways in the clusters listed in Table 4.3A/B. B/D) STRING protein-protein interaction networks depict whether the inversely correlated miRNA target genes have interactions (minimum confidence score = 0.040) based on online resources (see interaction types in legend), which could denote their function.

Table 4.3) Specific pathways in enriched pathway clusters

Pathway	Pathway cluster	FDR	Genes
A) hsa-miR-193b-3p (Expression in SEO-COPD: ↑)			
Primary lysosome	Secretory vesicles	4.23x10 ⁻⁴	NCSTN, GLB1, GRN, LAMTOR1, LAMP2
Azurophil granule	Secretory vesicles	4.23x10 ⁻⁴	NCSTN, GLB1, GRN, LAMTOR1, LAMP2
Transport vesicle	Transport vesicles	0.001	NCSTN, RAB5B, APH1A, TGOLN2, TPRG1L
Gamma-secretase complex	Vesicular and endosomal membranes	0.001	NCSTN, APH1A
Synaptic vesicle	Transport vesicles	0.001	NCSTN, RAB5B, APH1A, TPRG1L
Secretory vesicle	Secretory vesicles	0.001	NCSTN, GLB1, RAB5B, GRN, APH1A, TPRG1L, LAMTOR1, EBAG9, LAMP2
Exocytic vesicle	Transport vesicles	0.002	NCSTN, RAB5B, APH1A, TPRG1L
Endosome	Vesicular and endosomal membranes	0.002	NCSTN, MON1B, RAB5B, GRN, APH1A, TGOLN2
Notch receptor processing	Vesicular and endosomal membranes	0.013	NCSTN, APH1A
Azurophil granule membrane	Secretory vesicles	0.014	NCSTN, LAMTOR1, LAMP2
Membrane protein intracellular domain proteolysis	Vesicular and endosomal membranes	0.019	NCSTN, APH1A
Endosome membrane	Vesicular and endosomal membranes	0.022	NCSTN, RAB5B, APH1A, LAMTOR1, LAMP2, SNX17
Secretory granule	Secretory vesicles	0.029	NCSTN, GLB1, RAB5B, GRN, LAMTOR1, EBAG9, LAMP2
Cytoplasmic vesicle membrane	Vesicular and endosomal membranes	0.045	NCSTN, RAB5B, APH1A, TGOLN2, TPRG1L, LAMTOR1, LAMP2, SNX17
Vesicle membrane	Vesicular and endosomal membranes	0.050	NCSTN, RAB5B, APH1A, TGOLN2, TPRG1L, LAMTOR1, LAMP2, SNX17
B) hsa-miR-331-3p (Expression in SEO-COPD: ↓)			
Collagen-containing extracellular matrix	Extracellular matrix	0.008	COL6A2, CSPG4, COL6A1, NTN1, ELN, HNRNPM
Mammary gland development	Mammary gland development	0.010	BCL2L11, ZNF703, IGFBP5, NTN1
Apoptotic process involved in morphogenesis	Apoptotic process involved in morphogenesis	0.026	FOXC1, BCL2L11
Extracellular matrix	Extracellular matrix	0.035	COL6A2, CSPG4, COL6A1, NTN1, ELN, HNRNPM
External encapsulating structure	Extracellular matrix	0.036	COL6A2, CSPG4, COL6A1, NTN1, ELN, HNRNPM
Regulation of anatomical structure size	Regulation of anatomical structure size	0.037	ADRA2C, FOXC1, NTN1, SPTB, ELN, CDC42EP4

Table of pathways that make up the pathway clusters shown in Figure 4.3A/C, with their corresponding enriched genes.

A) miR-193b-3p: Visualization of enriched pathways with all inversely associated genes



B) miR-331-3p: Visualization of enriched pathways with all inversely associated genes

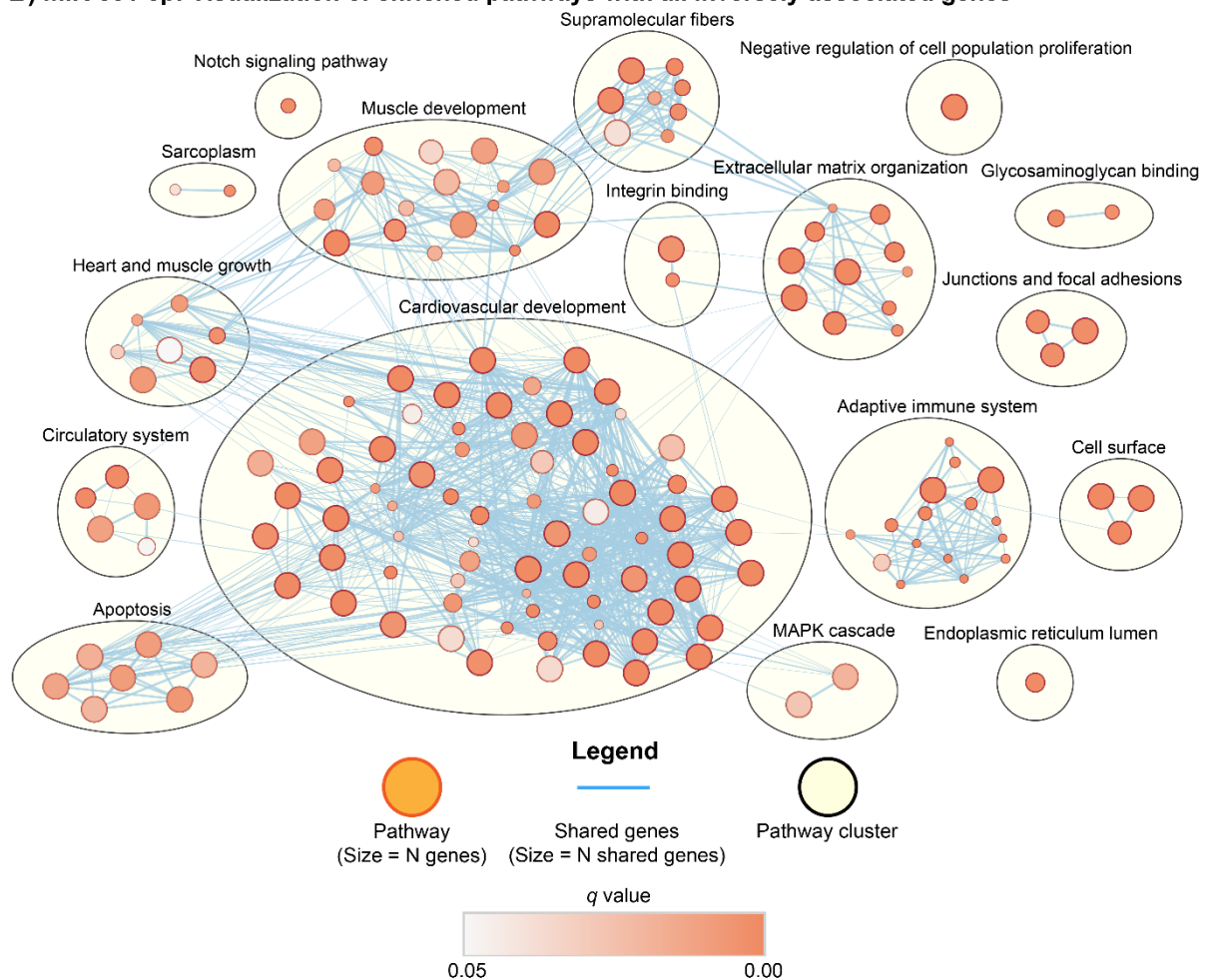


Figure 4.4) Unbiased pathway enrichment analysis of all significant inversed associations with miR-193b-3p and miR-331-3p genes. A/B) G:Profiler/Cytoscape pathway enrichment analyses visualize significant pathways (FDR $q < 0.05$) that all inversely correlated miRNA genes from the RNA-seq (not just predicted and validated genes) are enriched in and whether these pathways share genes. Together the pathways form pathway clusters of related biological functions that give a suggestion to the role of the miRNA and its target genes in SEO-COPD pathology.

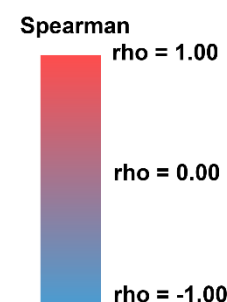
MiRNA expression and association with cell-type proportions

To assess the potential effect of differences in cell type composition in lung tissue on the DE SEO-COPD miRNAs, previously generated cellular deconvolution data was used. Previously we showed that the most abundant cell types present in our lung tissue samples were alveolar type II cells, fibroblast lineage cells, and alveolar macrophages, with the latter two having significantly different relative proportions between SEO-COPD and control groups (higher in SEO-COPD for fibroblasts, lower in SEO-COPD for alveolar macrophages)⁴.

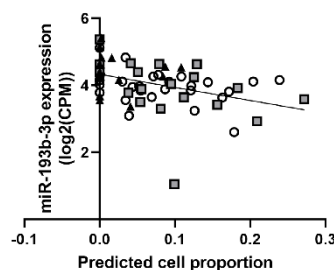
When correlating the expression of miR-193-3p, miR-331-3p and miR-202-5p with the predicted cell type proportions, we found multiple significant (FDR < 0.05) correlations. Higher miR-193b-3p expression was correlated with lower alveolar macrophage proportions (Figure 4.5A-B). Lower miR-331-3p expression was correlated with lower proportions of alveolar macrophages and higher proportions of fibroblast lineage cells, B-cell lineage cells and smooth muscle cells (Figure 4.5A, C-D).

A) Significant miRNA expression x predicted cell proportion associations

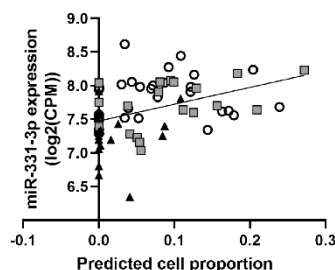
Cell type	hsa-miR-193b-3p	hsa-miR-331-3p	hsa-miR-202-5p
Alveolar macrophages	-0.42	0.45	-0.23
Alveolar type II cells	-0.24	0.28	-0.15
B cell lineage	0.08	-0.37	0.14
Basal & Secretory	0.11	0.10	-0.04
Dendritic cells	-0.01	0.09	0.04
Arterial endothelial cells	0.21	-0.22	0.15
Venous endothelial cells	0.19	-0.27	0.05
Fibroblast lineage	0.18	-0.38	0.19
Monocytes	0.12	-0.29	0.10
Smooth muscle	0.23	-0.43	0.18



B) miR-193b-3p x Alveolar macrophages



C) miR-331-3p x Alveolar macrophages



D) miR-331-3p x Fibroblast lineage

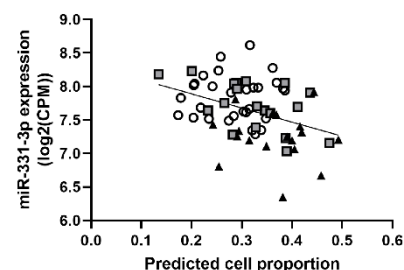


Figure 4.5) Associations between predicted lung tissue cell proportions and miRNA expression for miR-193b-3p, miR-331-3p and miR-202-5p. A) Significant associations (cut-off: FDR < 0.05) between the three DE SEO-COPD miRNAs and the detected cell types in the miRNA transcriptomic data. Association values calculated with Spearman rho. Significant associations are bolded and depicted in a gradient of blue to red to show whether miRNA expression was correlated with respectively lower or higher proportions of a cell type (see legend). Non-significant associations in grey. *Cell type has significantly differential predicted proportions between control and SEO-COPD (FDR < 0.05). B-D) Significant associations (FDR < 0.05) between miR-331-3p and miR-193b-3p with differential proportioned cell types alveolar macrophages and fibroblast lineage cells. Only subjects with both miRNA and RNA data used (n=66). Circles are controls, squares are common COPD and triangles are SEO-COPD subjects.

4.5 – Discussion

In the current study we identified three SEO-COPD miRNAs that are specifically differentially expressed in SEO-COPD lung tissue compared to non-COPD controls. The availability of genome-wide gene expression data generated from the same peripheral lung tissue material enabled us to directly assess the downstream functional effects of these three miRNAs. This analysis revealed several target genes and pathways biologically relevant to SEO-COPD, including formation and structure of lysosomes, secretion and transport of vesicles, extracellular matrix, fiber organization and tissue

development. Our findings generate important new insights into the potential involvement of these miRNAs in SEO-COPD pathogenesis.

One of the SEO-COPD-associated miRNAs we identified was miR-193b-3p. Mir-193-3p generally targets genes involved in the regulation of proliferation, cell-cycle and apoptosis, and as such has been previously associated with various cancer types^{23, 24}. In the current study, the enrichment of vesicle-related pathways among the miR-193b-3p target genes suggests that miRNA-193b-3p plays a role in regulating vesicular transport and secretion in the lung. This may be important to SEO-COPD as these can affect numerous biological processes to maintain homeostasis, such as growth, differentiation, apoptosis, inflammation and ECM remodeling^{7, 25, 26}. The higher expression of miR-193b-3p in SEO-COPD was associated with lower expression of vesicular transport/secretion genes, including NCSTN, APH1A, RAB5B, GLB1, GRN, TGOLN2, MON1B, LAMTOR1 and LAMP2. These genes share functions in membrane composition, vesicle transport or lysosomal function. NCSTN and APH1A are vital components of the gamma-secretase complex, an transmembrane cleaving protease with over 140 target substrates including ones involved in activating the Notch and Wnt signaling pathways^{27, 28}. Notch and Wnt signaling are vital for lung development, homeostasis and repair^{29, 30}, thus silencing of these two genes by miR-193b-3p may affect lung tissue repair. Next, GLB1, GRN, LAMTOR1 and LAMP2 all have roles in lysosomal function³¹⁻³³, with GLB1 previously being associated with COPD³¹. The inhibition of lysosomes and autophagy can lead to insufficient clearance of damaged organelles, proteins and aggregates, contributing to ER stress, oxidative stress and cellular senescence^{34, 35}. Some studies suggest that insufficient autophagy can drive COPD pathogenesis³⁴. Additionally, lysosomes are key components of (pathogenic) inflammatory systems like the NLRP3 inflammasome, a process GRN and LAMTOR1 play roles in^{32, 33}.

Further study will be necessary to determine how these vesicular and lysosomal genes are involved in SEO-COPD pathology.

In the current study, miR-331-3p showed lower expression in SEO-COPD compared to control lung tissue. This somewhat aligns with findings in asthma where the levels of circulating miR-331-3p were associated with better lung function³⁶. We show a lower expression of miR-331-3p in SEO-COPD to be associated with higher expression of genes involved in ECM formation: COL6A1, COL6A2, ELN, CSPG4 and NTN1, which were also enriched in the ECM pathways. This suggests miR-331-3p may play a role in disturbed ECM homeostasis and lung tissue remodeling in SEO-COPD^{4, 37}. These findings are in line with previous research from our group and others, where expression of ECM genes were found to be altered in (SEO-)COPD, including higher expression of ELN, COL6A1 and COL6A2^{4, 8, 37-40}. The composition and turnover of the ECM is heavily altered in COPD, particularly in subjects with more severe disease as reflected by on one hand more small airway wall thickening and on the other hand seemingly opposite emphysema^{37, 38, 41, 42}. The upregulation of ECM genes in our results could indicate small airway thickening, but since the peripheral lung samples are mainly composed of parenchyma³⁷ the SEO-COPD expression signature is likely representative of the emphysematous alveolar lung tissue. The increased ECM gene expression may thus be an attempt at lung tissue repair in SEO-COPD lungs that is either aberrant or unable to compensate for the rate of ECM breakdown^{37, 38}. Therefore, our current data suggests that miR-331-3p may be involved in regulating this repair response. Moreover, this may imply there is a regulatory system in place that responds to lung tissue damage by decreasing miR-331-3p expression to increase lung repair. Determining what factor exactly triggers this downregulation of miR-331-3p (e.g. destruction, cytokines, proteases, ECM fragments) and why it falls short of adequate repair in emphysema requires further research.

In addition to the ECM genes, three of the significantly correlated genes for miR-331-3p were also part of the differential gene expression profile in SEO-COPD found in the RNA-Seq analysis⁴, namely KREMEN1, BCL2L11 and GAS7. The enriched pathways in the miR-331-3p pathway enrichment analysis with all genes (Figure 4.3B) also closely matched the gene set enrichment analysis results of the differential SEO-COPD in the RNA-seq⁴. Both analyses demonstrated similar pathway clusters for the adaptive immune system, ECM, glycosaminoglycan binding, and morphogenesis. This could underline miR-331-3p and its target genes being lynchpins for the differential gene expression found in the SEO-COPD phenotype.

Due to miR-202-5p having no significant associations with any of the detected genes in the RNA transcriptome it was challenging to ascertain its role in SEO-COPD pathology based on RNA-Seq data. Aside from miR-202-5p's functions in tumor suppression⁴³, it is also involved in inflammatory processes, including TGFβ signaling regulation^{44, 45}, regulatory T-cell differentiation⁴⁶, and induction of either apoptosis, pro-inflammatory factor release or M2 polarization in macrophages^{47, 48}. Furthermore, miR-202-5p had increased expression in leukocytes of subjects with allergic rhinitis^{46, 47}. Taken together, expression of miR-202-5p could either be involved in promoting or attenuating inflammatory reactions in the diseased lung. As we found no correlations between miR-202-5p and gene expression, the regulatory effects of miR-202-5p likely take place at the post-translational level only affecting protein expression, which was not taken into account in this study.

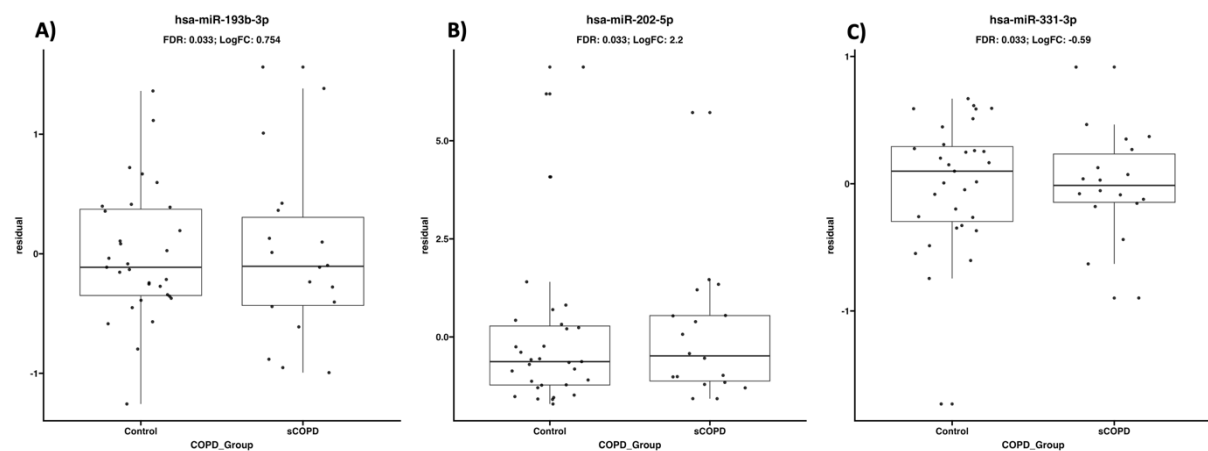
This study was specifically focused on SEO-COPD representing an intriguing subgroup of very susceptible COPD patients, likely with a different underlying disease pathology than common COPD. These patients were, by definition, younger and had more severe disease than the common COPD group, providing a challenge to disentangle whether the differences in SEO-COPD were related to severity of disease, early onset of disease, or age. More historical and longitudinal data is needed to further study this. Setting an actual age cut-off to define the SEO-COPD group remains arbitrary but needed as a best possible attempt to discriminate between groups in the type of analysis performed in this study.

To conclude, we hereby identified a differential miRNA signature in SEO-COPD of miR-202-5p, miR-193b-3p and miR-331-3p, of which, miR-193b-3p may be involved in regulation of vesicular and lysosomal activity, while miR-331-3p may be a regulator of aberrant ECM repair in SEO-COPD pathology.

4.6 – Supplementary Material

Figures

Supplementary Figure 4.1) Residual data points of differential SEO-COPD miRNAs



Tables

Supplementary Table 4.1) eQTM of the three miRNAs

miRNA	HGNC Symbol	Statistic	FDR	N in group	Group Rank
hsa-miR-193b-3p	NCSTN	-5.52	1.39E-03	1936	1
hsa-miR-331-3p	KREMEN1	-5.10	2.28E-03	1302	1
hsa-miR-331-3p	ADRA2C	-4.97	2.39E-03	1302	1
hsa-miR-331-3p	FOXC1	-4.59	5.76E-03	1302	2
hsa-miR-331-3p	EEF1D	-4.62	5.76E-03	1302	3
hsa-miR-331-3p	BCL2L11	-4.45	7.17E-03	1302	3
hsa-miR-331-3p	COL6A2	-4.38	7.45E-03	1302	5
hsa-miR-331-3p	ZNF703	-4.25	1.08E-02	1302	6
hsa-miR-193b-3p	ALDH2	-4.65	1.15E-02	1936	7
hsa-miR-331-3p	UNC45B	-4.16	1.19E-02	1302	2
hsa-miR-331-3p	RNF122	-4.10	1.32E-02	1302	8
hsa-miR-331-3p	IGFBP5	-4.00	1.32E-02	1302	9
hsa-miR-331-3p	CSPG4	-4.03	1.32E-02	1302	9
hsa-miR-331-3p	ISYNA1	-4.04	1.32E-02	1302	9
hsa-miR-331-3p	GAS7	-4.00	1.32E-02	1302	9
hsa-miR-331-3p	BTG2	-3.93	1.43E-02	1302	9
hsa-miR-331-3p	NCS1	-3.95	1.43E-02	1302	14
hsa-miR-331-3p	COL6A1	-3.93	1.43E-02	1302	14
hsa-miR-331-3p	JSRP1	-3.87	1.50E-02	1302	14
hsa-miR-193b-3p	MON1B	-4.23	1.92E-02	1936	17
hsa-miR-193b-3p	HNRNPUL1	-4.18	1.92E-02	1936	3
hsa-miR-193b-3p	PTPN9	-4.18	1.92E-02	1936	3
hsa-miR-193b-3p	GLB1	-4.28	1.92E-02	1936	3
hsa-miR-193b-3p	RAB5B	-4.09	2.02E-02	1936	3
hsa-miR-193b-3p	GRN	-4.06	2.07E-02	1936	7
hsa-miR-331-3p	ATP8B2	-3.73	2.07E-02	1302	8
hsa-miR-331-3p	NTN1	-3.72	2.07E-02	1302	18
hsa-miR-331-3p	SPTB	-3.71	2.07E-02	1302	19
hsa-miR-331-3p	TSKU	-3.66	2.24E-02	1302	19
hsa-miR-193b-3p	SNRPD3	-3.99	2.42E-02	1936	21
hsa-miR-193b-3p	APH1A	-3.96	2.49E-02	1936	9
hsa-miR-331-3p	USP2	-3.62	2.50E-02	1302	10

miRNA	HGNC Symbol	Statistic	FDR	N in group	Group Rank
hsa-miR-193b-3p	TGOLN2	-3.92	2.56E-02	1936	22
hsa-miR-193b-3p	DCAF7	-3.93	2.56E-02	1936	11
hsa-miR-331-3p	ELN	-3.58	2.60E-02	1302	11
hsa-miR-331-3p	DENND5B	-3.58	2.60E-02	1302	23
hsa-miR-331-3p	CDC42EP4	-3.54	2.70E-02	1302	23
hsa-miR-331-3p	ZCCHC24	-3.54	2.70E-02	1302	25
hsa-miR-193b-3p	TPRG1L	-3.86	2.80E-02	1936	25
hsa-miR-331-3p	ERF	-3.51	2.86E-02	1302	13
hsa-miR-193b-3p	ARSD	-3.79	3.02E-02	1936	27
hsa-miR-331-3p	CDKN2D	-3.46	3.18E-02	1302	14
hsa-miR-331-3p	MIDN	-3.44	3.30E-02	1302	28
hsa-miR-331-3p	PHLDB3	-3.41	3.39E-02	1302	29
hsa-miR-331-3p	KRT15	-3.40	3.40E-02	1302	30
hsa-miR-331-3p	IGLV1-44	-3.39	3.41E-02	1302	31
hsa-miR-193b-3p	MRPS16	-3.66	3.64E-02	1936	32
hsa-miR-193b-3p	WNK1	-3.63	3.64E-02	1936	15
hsa-miR-193b-3p	CSTF1	-3.63	3.64E-02	1936	15
hsa-miR-193b-3p	MRPS18B	-3.58	3.87E-02	1936	15
hsa-miR-331-3p	HNRNPM	-3.33	3.93E-02	1302	18
hsa-miR-193b-3p	TRIM14	-3.52	4.16E-02	1936	33
hsa-miR-193b-3p	SPN	-3.52	4.16E-02	1936	19
hsa-miR-193b-3p	ESYT1	-3.50	4.19E-02	1936	19
hsa-miR-193b-3p	SNRNP200	-3.50	4.19E-02	1936	21
hsa-miR-193b-3p	LAMTOR1	-3.49	4.28E-02	1936	21
hsa-miR-193b-3p	YIPF3	-3.47	4.30E-02	1936	23
hsa-miR-193b-3p	SEPHS1	-3.47	4.30E-02	1936	24
hsa-miR-331-3p	LRRC10B	-3.25	4.34E-02	1302	24
hsa-miR-331-3p	PPP1R10	-3.27	4.34E-02	1302	34
hsa-miR-331-3p	MEPCE	-3.25	4.34E-02	1302	34
hsa-miR-331-3p	KCNE4	-3.24	4.45E-02	1302	34
hsa-miR-193b-3p	EBAG9	-3.45	4.46E-02	1936	37
hsa-miR-193b-3p	HOMEZ	-3.44	4.46E-02	1936	26
hsa-miR-193b-3p	KIAA0100	-3.43	4.57E-02	1936	26
hsa-miR-193b-3p	ELOA	-3.41	4.64E-02	1936	28
hsa-miR-331-3p	GRASP	-3.20	4.67E-02	1302	29
hsa-miR-331-3p	TEAD3	-3.19	4.67E-02	1302	38
hsa-miR-331-3p	SNIP1	-3.20	4.67E-02	1302	38
hsa-miR-193b-3p	LAMP2	-3.36	4.68E-02	1936	38
hsa-miR-193b-3p	SNX17	-3.36	4.68E-02	1936	30
hsa-miR-193b-3p	RALBP1	-3.36	4.68E-02	1936	30
hsa-miR-193b-3p	WWP2	-3.38	4.68E-02	1936	30
hsa-miR-193b-3p	ILRUN	-3.36	4.68E-02	1936	30
hsa-miR-193b-3p	EPHX1	-3.38	4.68E-02	1936	30
hsa-miR-331-3p	DMPK	-3.17	4.95E-02	1302	30
hsa-miR-202-5p	NA	NA	NA	NA	41

4.7 – References

1. Global Initiative for Chronic Obstructive Lung Disease, I., *GOLD - Global initiative for chronic obstructive lung disease*. 2021.
2. Silverman, E.K., et al., *Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis*. Am J Respir Crit Care Med, 1998. **157**(6 Pt 1): p. 1770-8.
3. Foreman, M.G., et al., *Early-Onset Chronic Obstructive Pulmonary Disease Is Associated with Female Sex, Maternal Factors, and African American Race in the COPD Gene Study*. American Journal of Respiratory and Critical Care Medicine, 2011. **184**(4): p. 414-420.
4. Bekker, N., et al., *A lung tissue gene expression profile of severe early onset COPD*. ERJ, 2022.
5. Schell, S.L. and Z.S.M. Rahman, *miRNA-Mediated Control of B Cell Responses in Immunity and SLE*. Front Immunol, 2021. **12**: p. 683710.
6. Menon, A., et al., *miRNA: A Promising Therapeutic Target in Cancer*. Int J Mol Sci, 2022. **23**(19).
7. Wu, J., Y. Ma, and Y. Chen, *Extracellular vesicles and COPD: foe or friend?* J Nanobiotechnology, 2023. **21**(1): p. 147.
8. Woldhuis, R.R., et al., *Link between increased cellular senescence and extracellular matrix changes in COPD*. American Journal of Physiology-Lung Cellular and Molecular Physiology, 2020. **319**(1): p. L48-L60.
9. Research code UMCG. 2024 [cited 2024 2024-02-02]; Available from: <https://umcgresearch.org/w/research-code-umcg>.
10. Coreon. *The COREON Code of Conduct for Health Research is available in translation*. 2023 [cited 2023 2023-06-21T10:36:51+00:00]; Available from: <https://www.coreon.org/the-coreon-code-of-conduct-for-health-research-is-available-in-translation/>, <https://www.coreon.org/gedragcode-gezondheidsonderzoek/>.
11. Langmead, B., et al., *Scaling read aligners to hundreds of threads on general-purpose processors*. Bioinformatics, 2019. **35**(3): p. 421-432.
12. Griffiths-Jones, S., et al., *miRBase: microRNA sequences, targets and gene nomenclature*. Nucleic Acids Res, 2006. **34**(Database issue): p. D140-4.
13. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139-40.
14. Ru, Y., et al., *The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations*. Nucleic Acids Res, 2014. **42**(17): p. e133.
15. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations*. Bioinformatics, 2012. **28**(10): p. 1353-8.
16. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets*. Nucleic Acids Res, 2021. **49**(D1): p. D605-D612.
17. Kolberg, L., et al., *g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update)*. Nucleic Acids Res, 2023. **51**(W1): p. W207-W212.
18. Merico, D., et al., *Enrichment map: a network-based method for gene-set enrichment visualization and interpretation*. PLoS One, 2010. **5**(11): p. e13984.
19. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. Genome Res, 2003. **13**(11): p. 2498-504.
20. Reimand, J., et al., *Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap*. Nat Protoc, 2019. **14**(2): p. 482-517.
21. Olgun, G., V. Gopalan, and S. Hannenhalli, *miRSCAPE - inferring miRNA expression from scRNA-seq data*. iScience, 2022. **25**(9): p. 104962.
22. Sikkema, L., et al., *An integrated cell atlas of the lung in health and disease*. Nat Med, 2023. **29**(6): p. 1563-1577.

23. Kober, P., et al., *Epigenetic Downregulation of Hsa-miR-193b-3p Increases Cyclin D1 Expression Level and Cell Proliferation in Human Meningiomas*. Int J Mol Sci, 2023. **24**(17).
24. Roth, S.A., et al., *MicroRNA-193b-3p represses neuroblastoma cell growth via downregulation of Cyclin D1, MCL-1 and MYCN*. Oncotarget, 2018. **9**(26): p. 18160-18179.
25. Berumen Sanchez, G., et al., *Extracellular vesicles: mediators of intercellular communication in tissue injury and disease*. Cell Commun Signal, 2021. **19**(1): p. 104.
26. Gomez, N., et al., *Extracellular vesicles and chronic obstructive pulmonary disease (COPD): a systematic review*. Respir Res, 2022. **23**(1): p. 82.
27. Hur, J.Y., *gamma-Secretase in Alzheimer's disease*. Exp Mol Med, 2022. **54**(4): p. 433-446.
28. Malla, R.R. and P. Kiran, *Tumor microenvironment pathways: Cross regulation in breast cancer metastasis*. Genes Dis, 2022. **9**(2): p. 310-324.
29. Aros, C.J., C.J. Pantoja, and B.N. Gomperts, *Wnt signaling in lung development, regeneration, and disease progression*. Commun Biol, 2021. **4**(1): p. 601.
30. Tsao, P.N., et al., *Epithelial Notch signaling regulates lung alveolar morphogenesis and airway epithelial integrity*. Proc Natl Acad Sci U S A, 2016. **113**(29): p. 8242-7.
31. John, C., et al., *Genetic Associations and Architecture of Asthma-COPD Overlap*. Chest, 2022. **161**(5): p. 1155-1166.
32. Saeedi-Boroujeni, A., et al., *Progranulin (PGRN) as a regulator of inflammation and a critical factor in the immunopathogenesis of cardiovascular diseases*. J Inflamm (Lond), 2023. **20**(1): p. 1.
33. Tsujimoto, K., et al., *The lysosomal Ragulator complex activates NLRP3 inflammasome in vivo via HDAC6*. EMBO J, 2023. **42**(1): p. e111389.
34. Racanelli, A.C., A.M.K. Choi, and M.E. Choi, *Autophagy in chronic lung disease*. Prog Mol Biol Transl Sci, 2020. **172**: p. 135-156.
35. Wang, Q., et al., *Advances in the investigation of the role of autophagy in the etiology of chronic obstructive pulmonary disease: A review*. Medicine (Baltimore), 2023. **102**(47): p. e36390.
36. Kho, A.T., et al., *Circulating MicroRNAs: Association with Lung Function in Asthma*. PLoS One, 2016. **11**(6): p. e0157998.
37. Brandsma, C.A., et al., *Recent advances in chronic obstructive pulmonary disease pathogenesis: from disease mechanisms to precision medicine*. J Pathol, 2020. **250**(5): p. 624-635.
38. Brandsma, C.A., et al., *A large lung gene expression study identifying fibulin-5 as a novel player in tissue repair in COPD*. Thorax, 2015. **70**(1): p. 21-32.
39. Abdillahi, S.M., et al., *Collagen VI Is Upregulated in COPD and Serves Both as an Adhesive Target and a Bactericidal Barrier for Moraxella catarrhalis*. J Innate Immun, 2015. **7**(5): p. 506-17.
40. Deslee, G., et al., *Elastin expression in very severe human COPD*. Eur Respir J, 2009. **34**(2): p. 324-331.
41. Bihlet, A.R., et al., *Biomarkers of extracellular matrix turnover are associated with emphysema and eosinophilic-bronchitis in COPD*. Respir Res, 2017. **18**(1): p. 22.
42. Stolz, D., et al., *Systemic Biomarkers of Collagen and Elastin Turnover Are Associated With Clinically Relevant Outcomes in COPD*. Chest, 2017. **151**(1): p. 47-59.
43. Ghafouri-Fard, S., et al., *The Impact of lncRNAs and miRNAs in Regulation of Function of Cancer Stem Cells and Progression of Cancer*. Front Cell Dev Biol, 2021. **9**: p. 696820.
44. Mody, H.R., et al., *miR-202 Diminishes TGFbeta Receptors and Attenuates TGFbeta1-Induced EMT in Pancreatic Cancer*. Mol Cancer Res, 2017. **15**(8): p. 1029-1039.
45. Gu, S., et al., *Retinal pigment epithelial cells secrete miR-202-5p-containing exosomes to protect against proliferative diabetic retinopathy*. Exp Eye Res, 2020. **201**: p. 108271.
46. Wang, L., et al., *MiR-202-5p/MATN2 are associated with regulatory T-cells differentiation and function in allergic rhinitis*. Hum Cell, 2019. **32**(4): p. 411-417.

47. Wang, L., et al., *MiR-202-5p Promotes M2 Polarization in Allergic Rhinitis by Targeting MATN2*. *Int Arch Allergy Immunol*, 2019. **178**(2): p. 119-127.
48. Xu, F., F. Yao, and Y. Ning, *MicroRNA-202-5p-dependent inhibition of Bcl-2 contributes to macrophage apoptosis and atherosclerotic plaque formation*. *Gene*, 2023. **867**: p. 147366.

Chapter 5 – High miR203a-3p and miR-375 expression in the airways of smokers with and without COPD

5.1 – Publication Declaration

Jos van Nijnatten, Corry-Anke Brandsma, Katrina Steiling, Pieter S Hiemstra, Wim Timens, Maarten van den Berge[†] and Alen Faiz[†] (2022). High miR203a-3p and miR-375 expression in the airways of smokers with and without COPD. *Scientific Reports*

Status: Published (doi: 10.1038/s41598-022-09093-0)

Author Contributions:

J.v.N. wrote the main manuscript text and performed the analyses; K.S. performed the analysis in the independent dataset; M.v.d.B. and A.F. are the primary supervisors, and C.-A.B. and W.T. are the secondary supervisors and as such involved in the analytical procedure and interpretation; M.v.d.B. was involved in obtaining the NORM dataset; W.T. and P.H. were involved in obtaining the GLUCOLD dataset. All authors reviewed the manuscript and approved the final version of the manuscript.

Name:	Signature:	Date:
Jos van Nijnatten	Production Note: Signature removed prior to publication.	08-12-2023
Corry-Anke Brandsma	Production Note: Signature removed prior to publication.	07-12-2023
Katrina Steiling	Production Note: Signature removed prior to publication.	12/1/2013
Pieter S Hiemstra	Production Note: Signature removed prior to publication.	1 december 2023
Wim Timens	Production Note: Signature removed prior to publication.	December 2, 2023
Maarten van den Berge	Production Note: Signature removed prior to publication.	December 5, 2024
Alen Faiz	Production Note: Signature removed prior to publication.	2/12/2024

5.2 – Abstract

Smoking is a leading cause of chronic obstructive pulmonary disease (COPD). It is known to have a significant impact on gene expression and (inflammatory) cell populations in the airways involved in COPD pathogenesis. In this study, we investigated the impact of smoking on the expression of miRNAs in healthy and COPD individuals. We aimed to elucidate the overall smoking-induced miRNA changes and those specific to COPD. In addition, we investigated the downstream effects on regulatory gene expression and the correlation to cellular composition. We performed a genome-wide miRNA expression analysis on a dataset of 40 current- and 22 ex-smoking COPD patients and a

dataset of 35 current- and 38 non-smoking respiratory healthy controls and validated the results in an independent dataset. miRNA expression was then correlated with mRNA expression in the same patients to assess potential regulatory effects of the miRNAs. Finally, cellular deconvolution analysis was used to relate miRNAs changes to specific cell populations. Current smoking was associated with increased expression of three miRNAs in the COPD patients and 18 miRNAs in the asymptomatic smokers compared to respiratory healthy controls. In comparison, four miRNAs were lower expressed with current smoking in asymptomatic controls. Two of the three smoking-related miRNAs in COPD, miR-203a-3p and miR-375, were also higher expressed with current smoking in COPD patients and the asymptomatic controls. The other smoking-related miRNA in COPD patients, i.e. miR-31-3p, was not present in the respiratory healthy control dataset. miRNA-mRNA correlations demonstrated that miR-203a-3p, miR-375 and also miR-31-3p expression were negatively associated with genes involved in pro-inflammatory pathways and positively associated with genes involved in the xenobiotic pathway. Cellular deconvolution showed that higher levels of miR-203a-3p were associated with higher proportions of proliferating-basal cells and secretory (club and goblet) cells and lower levels of fibroblasts, luminal macrophages, endothelial cells, B-cells, amongst other cell types. MiR-375 expression was associated with lower levels of secretory cells, ionocytes and submucosal cells, but higher levels of endothelial cells, smooth muscle cells, and mast cells, amongst other cell types. In conclusion, we identified two smoking-induced miRNAs (miR-375 and miR-203a-3p) that play a role in regulating inflammation and detoxification pathways, regardless of the presence or absence of COPD. Additionally, in patients with COPD, we identified miR-31-3p as a miRNA induced by smoking. Our identified miRNAs should be studied further to unravel which smoking-induced inflammatory mechanisms are reactive and which are involved in COPD pathogenesis.

Keywords: Epigenetics, Risk factors

5.3 – Introduction

Smoking is one of the leading causes of preventable death worldwide, and despite the well-known health effects, over 15% of the world's population still smoked in 2019, according to the WHO. The smoking of cigarettes exposes the lung to more than 4000 components¹. As a result, it is strongly associated with the development of lung cancer^{2,3} and other respiratory diseases^{4,5}, including chronic obstructive pulmonary disease (COPD). Characteristics of COPD are an inflammatory response in the airways and lung parenchyma, associated with progressive, irreversible airflow limitation, dyspnea, hypersecretion of mucus, and alveolar destruction, i.e., emphysema.

We have previously shown that current smoking affects levels of gene expression and DNA methylation in bronchial biopsies⁶. However, current smoking on microRNA expression in the human airway wall is not well studied so far^{7,8}. MicroRNAs (miRNA) are small non-coding RNA transcripts (approximately 17–27 nucleotides long), which interact with the Argonaute protein (AGO protein) in the RNA-induced silencing complex (RISC), leading to cleavage or silencing of a target mRNA⁹. Most human protein-coding genes contain miRNA target sites¹⁰. A single miRNA can influence the expression of hundreds of genes¹¹ and regulate a broad range of biological processes, such as cell proliferation, apoptosis, and the immune system^{12,13}. Because of this, deregulation of miRNA function is associated with numerous diseases, including COPD¹⁴⁻¹⁶.

We also have previously shown that miRNAs play an essential role in mediating inflammatory responses and corticoid effects in asthma and COPD¹⁷⁻²⁰. However, although several studies have

assessed the impact of smoking on miRNA expression, very few have focused on the airway wall in an unbiased method²¹. Here, we investigated the influence of current smoking on miRNA expression in bronchial biopsies in two independent datasets consisting of current and ex-smoking COPD patients and current and non-smoking respiratory healthy controls to elucidate the overall smoking-induced miRNA changes and those specific to COPD. In addition, we correlated miRNA to matched mRNA expression to identify genes and pathways influenced by the miRNAs affected by smoking to investigate the downstream effects on gene expression. Finally, to investigate the relation with cellular composition, differential miRNAs expression was associated with gene signatures of specific cell populations and immunohistological staining of specific cell types in matched biopsies.

5.4 – Methods

Subjects

Affymetrix GeneChip miRNA 1.0 Array and RNA-Seq data from baseline bronchial biopsies from COPD patients and respiratory healthy controls were used for this study as described previously^{17, 22}.

COPD patients participated in the Groningen Leiden Universities and Corticosteroids in Obstructive Lung Disease study (GLUCOLD, ClinicalTrials.gov NCT00158847) (n = 62). Respiratory healthy controls participated in the Study to Obtain Normal Values of Inflammatory Variables from Healthy Subjects (NORM study, ClinicalTrials.gov NCT00848406, n = 73). In- and exclusion criteria and more extensive characteristics of subjects participating in the GLUCOLD and NORM studies were previously described^{23, 24}.

Briefly, for the GLUCOLD study, all patients were current or ex-smokers with > 10 pack years. Participants had irreversible airflow with postbronchodilator forced expiration volume in one second (FEV₁) < 80% predicted and postbronchodilator FEV₁/forced vital capacity (FVC) below 70%. Patients took no oral corticosteroids for at least three months before starting the study and did not use inhaled corticosteroids for at least six months prior to the study. Patients classified as ex-smokers did not smoke at least one year prior to the study.

Participants of the NORM study had normal lung function (FEV₁ > 80% and FEV₁/FVC > 70%) and no bronchial hyperresponsiveness to methacholine (PC₂₀ > 16 mg/ml). Participants defined as non-smokers did not smoke during the year leading up to the study and did not smoke more than one year in total, with a total of < 0.5 packyears.

The local medical ethics committees approved all studies, and all subjects gave their written informed consent (the NORM study was approved by the ethics committee of the UMCG, and the GLUCOLD study was approved by the same committee and the ethics committee of the LUMC)²³⁻²⁵. All methods were performed in accordance with the relevant guidelines and regulations.

Statistics

To identify differentially expressed miRNAs in bronchial biopsies of current and ex-smoking COPD patients and current and non-smoking healthy participants, we used a linear model using Limma (R-package version 3.40.6, R statistical software version 3.6). We used current smoking status as a categorical variable, adjusting for age and sex. A Benjamini–Hochberg corrected p-value (False Discovery Rate (FDR)) of < 0.05 was considered statistically significant.

We performed the identification of potential targets of differentially expressed miRNAs by directly correlating miRNA expression to normalized²⁶ gene expression data available from the same bronchial biopsies. We correlated significant differentially expressed miRNAs to all expressed mRNAs using the Pearson Correlation Coefficient. Significant correlations between miRNA

expression levels and gene expression levels were determined using Psych (R-package version 1.8.12), and we performed a meta-analysis on the results of both studies. Significant correlations with an absolute R-value above 0.25 were considered biologically relevant.

We verified the smoking-related expression changes of the identified miRNAs using an independent miRNA dataset (Illumina HiSeq) obtained from bronchial brushings from 30 COPD patients and 30 non-COPD controls²⁷. COPD was defined in this dataset by $FEV_1/FVC < 70\%$ and $FEV_1 < 80\%$ predicted and non-COPD controls were defined as $FEV_1/FVC \geq 70\%$ or $FEV_1/FVC < 70\%$ and $FEV_1\%$ predicted $> 80\%$. We used a linear model and smoking status as a categorical variable while correcting for age, sex and pack years. We analyzed the COPD and non-COPD participants separately.

miRNA predicted target approach

We calculated the most likely target for each miRNA using miRNetap (R-package version 1.18.0). The default settings were used to calculate the geometric mean of ranks from the databases/algorithms PicTar, Diana, TargetScan, miRanda, and miRDB. The minimum number of sources required for a potential target was 2.

Pathway analysis

To predict which pathways are influenced by smoking-induced miRNAs, we performed Gene Set Enrichment Analysis (GSEA version 4.0.3) on the mRNAs positively and negatively correlated with our differentially expressed smoking-associated miRNAs. A separate analysis was done for the negatively correlated predicted target genes only. We compared the ranked lists to BioCarta and KEGG gene set databases. We then performed a meta-analysis on the pathways found in both studies and adjusted the p-value using Benjamini-Hochberg's method.

Association of miRNAs with composite scores of cell-type-specific genes based on GSVA

We used cell-type-specific gene signatures based on single-cell data for 15 cell types (activated endothelium, B-cells, basal cells, proliferating basal cells, ciliated cells, club cells and goblet cells, fibroblasts, inflammatory dendritic cells, ionocytes, luminal macrophages, mast cells, neutrophils, smooth muscle cells, and submucosal cells)²⁸. We then performed Gene Set Variation Analysis (GSVA²⁹, R-package version 1.32.0) to calculate the composite scores for each cell type. We then correlated each composite score to relevant miRNAs using the Pearson correlation coefficient.

Evaluation of miRNAs with the presence of cell populations in matched bronchial biopsies

For the differentially expressed miRNAs, we assessed the correlation between miRNA expression and numbers of eosinophils, neutrophils, macrophages, mast cells, CD3 or CD4 or CD8 positive T-cells, and percentage of Periodic acid–Schiff stain (PAS)-positive goblet cells in paraffin-embedded bronchial biopsies from the same patients taken at the same location in the lung. These data were obtained from previously published data sets for the GLUCOLD³⁰ and NORM²⁵ studies.

A flow diagram showing the analysis approach is provided in Figure 5.1.

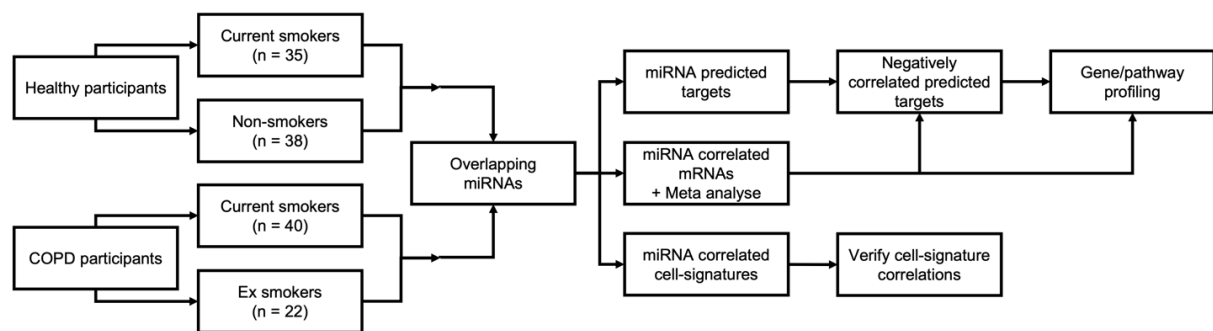


Figure 5.1) Flow diagram of the study approach. Two studies were used: the GLUCOLD study with current or ex-smoker COPD patients and the NORM study with current or non-smoking respiratory healthy controls. A linear regression model was used to identify smoking-associated miRNAs. The overlapping two miRNAs were used for further analysis of COPD patients. A meta-analysis of Pearson correlations was performed to identify miRNA associated mRNA. The predicted targets were obtained, and a list of negatively correlated predicted targets was compiled. Pathway analysis was performed using this list as well as on all the negative and positively correlated mRNAs. Negatively correlated predicted targets were repeated in the NORM. Additionally, miRNA expression was correlated to cell type signatures and cell populations.

5.5 – Results

In the current study, we investigated smoking-related miRNA changes in 62 COPD patients (ex-smokers, n = 22, and current smokers, n = 40) and 73 respiratory healthy individuals (non-smokers, n = 38 and current smokers, n = 35). The subject characteristics of the two studies are presented in Table 5.1) Patient demographics in relation to smoking status.. There was a significant difference in sex between the COPD patients and respiratory healthy individuals (Chi-Squared p-value = 3E-4), as well as a significant difference in age (Mann–Whitney U p-value = 2E-11).

Table 5.1) Patient demographics in relation to smoking status.

	COPD patients; GLUCOLD		Respiratory Healthy controls; NORM	
	Ex-smokers	Current smokers	Non-smokers	Current smokers
N	22	40	38	35
Age mean +/- SD	63 +/- 8	59 +/- 7	38 +/- 19	41 +/- 15
Sex male n (%)	20 (90.9)	33 (82.5)	20 (52.6)	20 (57.1)
FEV ₁ % predicted	55.9 +/- 10.7	54.7 +/- 9.5	101.4 +/- 12.2	99.1 +/- 9.4
FEV ₁ /FVC %	50.1 +/- 9.3	46.8 +/- 8.5	80.5 +/- 6.6	78.0 +/- 6.2
Age quit smoking mean +/- SD	56 +/- 10			
Cigarettes per day mean +/- SD	0	18.5 +/- 10.2	0	15.0 +/- 7

FEV₁: forced expiratory volume in 1 second; FVC: forced vital capacity

Changes in miRNA expression associated with current smoking

We identified three miRNAs (miR-31-3p, miR-203a-3p, and miR-375) differentially expressed between current- and ex-smoking COPD patients (FDR < 0.05). All three miRNAs showed a higher level of expression in current- compared to ex-smokers (Table 5.2). A volcano plot and heatmap are depicted in Figure 5.2A,B.

Table 5.2) Differential expression of miRNAs for respiratory healthy controls and subjects with COPD when comparing smoking status.

miRNA id	COPD			Asymptomatic		
	Log ₂ (FC)	P-value	FDR	Log ₂ (FC)	P-value	FDR
hsa-miR-31-3p	1.18	1.16E-04	1.21E-02	NA	NA	NA
hsa-miR-203a-3p	1.50	1.17E-04	1.21E-02	0.83	2.63E-04	1.15E-02
hsa-miR-375	1.02	1.64E-04	1.21E-02	1.21	9.22E-08	9.10E-06
hsa-miR-200b-3p	0.70	2.65E-03	9.80E-02	0.49	2.28E-03	4.50E-02
hsa-miR-183-5p	0.75	8.10E-03	1.07E-01	0.84	1.37E-08	1.80E-06
hsa-miR-31-5p	0.46	1.96E-02	1.68E-01	0.65	4.10E-04	1.62E-02
hsa-miR-200c-3p	0.35	2.40E-02	1.96E-01	0.76	1.21E-05	7.94E-04
hsa-miR-200a-5p	0.55	3.06E-02	2.27E-01	0.68	1.22E-03	3.38E-02
hsa-miR-182-5p	0.35	4.16E-02	2.58E-01	0.76	3.50E-09	1.38E-06
hsa-miR-200a-3p	0.32	6.45E-02	2.87E-01	0.49	1.26E-03	3.38E-02
hsa-miR-708-5p	0.49	6.51E-02	2.87E-01	0.69	1.12E-03	3.38E-02
hsa-miR-149-5p	-0.35	1.79E-01	4.32E-01	0.54	1.43E-03	3.38E-02
hsa-miR-181b-5p	0.15	3.65E-01	6.31E-01	0.41	1.37E-03	3.38E-02
hsa-miR-574-5p	0.24	4.21E-01	6.54E-01	0.65	1.91E-03	4.18E-02
hsa-miR-331-3p	0.14	4.79E-01	6.94E-01	0.55	4.96E-04	1.78E-02
hsa-miR-130b-3p	0.11	5.04E-01	7.22E-01	1.09	1.64E-04	9.06E-03
hsa-miR-181a-5p	0.05	7.70E-01	8.81E-01	0.47	2.17E-03	4.50E-02
hsa-miR-126-3p	-0.03	8.99E-01	9.45E-01	-0.38	2.65E-03	4.76E-02
hsa-miR-10b-5p	-0.01	9.81E-01	9.91E-01	-0.47	1.83E-04	9.06E-03
hsa-miR-3065-5p	NA	NA	NA	2.18	8.79E-09	1.74E-06
hsa-miR-3065-3p	NA	NA	NA	1.98	3.36E-06	2.66E-04
hsa-miR-126-5p	NA	NA	NA	-0.53	1.45E-03	3.38E-02
hsa-miR-1468-5p	NA	NA	NA	-1.10	2.62E-03	4.76E-02

FC: Fold Change; FDR: False Discovery Rate (Benjamini Hochberg corrected p-value)

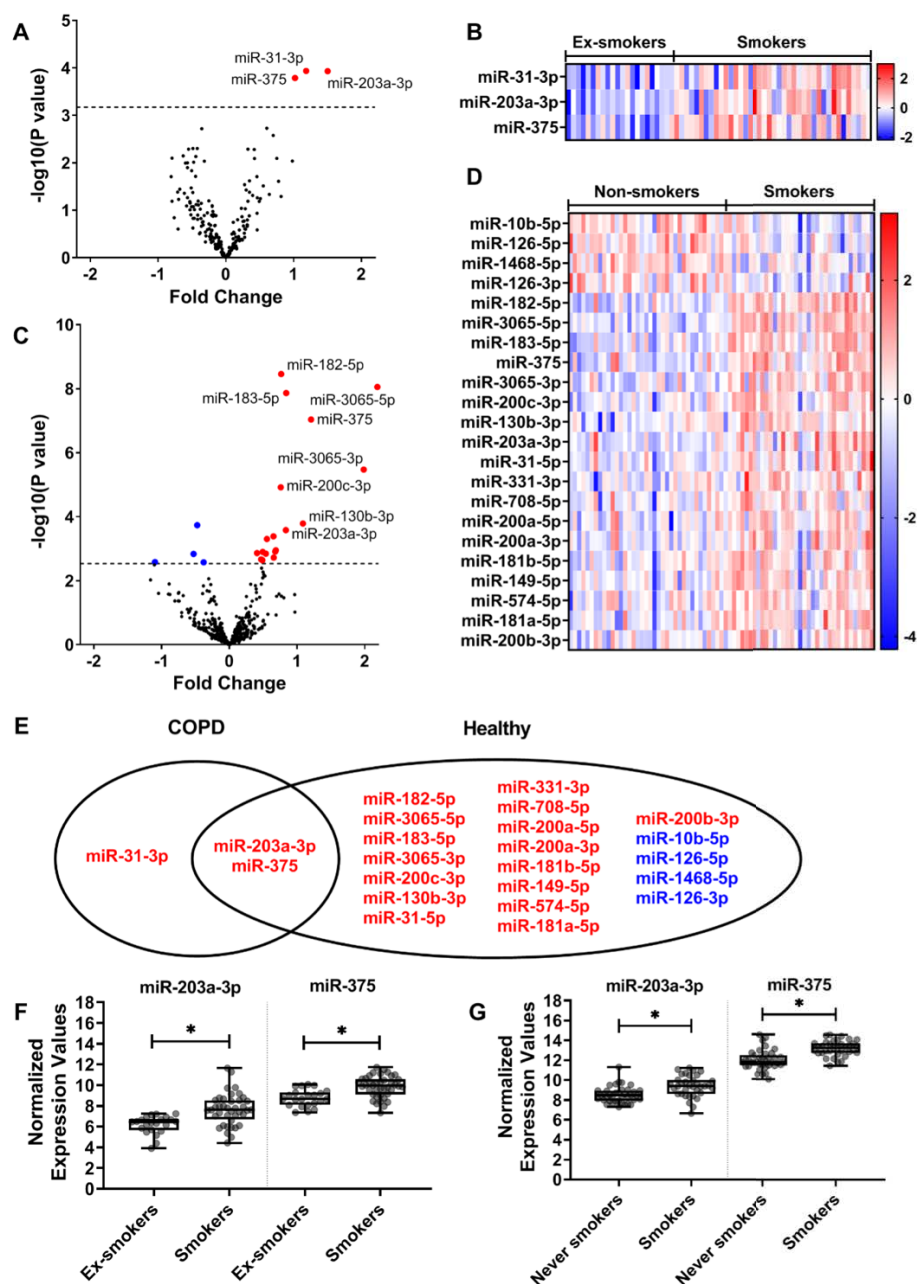


Figure 5.2) Differentially expressed microRNAs (miRNAs) when comparing smoking status in patients with COPD and respiratory healthy controls. Volcano plot (a) shows changes in expression levels of 220 miRNAs in patients with COPD (n = 62). Heat map (b) shows expression changes of the three significantly differentially expressed miRNAs in COPD patients. Volcano plot (c) shows changes in the expression of 395 miRNAs in respiratory healthy controls (n = 73). Heat map (d) shows expression changes of the 22 significantly differentially expressed miRNAs in respiratory healthy controls. Venn diagram (e) shows the overlap between significantly differentially expressed miRNAs in COPD patients compared to respiratory healthy controls. Boxplot (f) shows the normalized expression values of the miRNAs miR-203a-3p and miR-375 per smoking status group in the COPD patients, and boxplot (g) shows the normalized expression values for the same miRNAs in the respiratory healthy controls. Red indicates an increase in expression, whereas blue indicates a decrease in expression. A false discovery rate (FDR) cut-off of 0.05 was used.

We next repeated this analysis in respiratory healthy individuals. A total of 22 miRNAs were differentially expressed in the current- versus non-smokers (FDR < 0.05). Of these, 18 miRNAs

showed a higher expression level in current smokers than non-smokers, while four were lower expressed (Table 5.2). A volcano plot and heatmap are depicted in Figure 5.2C,D, respectively.

Two miRNAs, miR-375 and miR-203a-3p, were associated with current smoking in the same direction in both COPD patients and respiratory healthy controls (Figure 5.2E–G).

MiR-31-3p, the third smoking-associated miRNA in the COPD patients, was not present in the respiratory healthy small-RNA-seq dataset.

We validated our findings in an independent dataset of epithelial brushes. We confirmed higher miR-375 expression in current- than ex-smokers in COPD and non-COPD, but no effect of smoking status on miR-203a-3p or miR-31-3p. (Supplementary Table 5.4).

Associations between changes in miRNA expression and mRNA expression

To identify potential genes and pathways that are regulated by the three smoking-associated miRNAs, we used two different methods: (1) we investigated the global effects of the miRNAs by doing direct correlations with all expressed genes in the same biopsies, and (2) we investigated the direct effects by focusing on the negatively correlated predicted targets for each miRNA.

For the first analysis, we examined whether miR-375, miR-203a-3p and miR-31-3p affected the expression of mRNA transcripts in each study separately and then performed a meta-analysis for the miRNAs present in both studies (miR-375 and miR-203a-3p). Here, we identified 983 significant negatively correlated genes for miR-203a-3p (FDR < 0.05), and 2254 positively correlated genes. We found 2442 significantly negatively correlated genes for miR-375 and 865 positively correlated genes, and we found 1106 significant negative correlations for miR-31-3p and 1029 positive correlations (Supplementary Table 5.1). The top 3 most strongly correlated genes (*PDGFD*, *ARL4C* and *MBNL1*) of both miR-375 and miR-203a-3p are shown in Figure 5.3C–H. Gene functions are described in Table 5.3.

Table 5.3) Selective literature of predicted target genes that were negatively correlated to miR-203a-3p, miR-375 and miR-31-3p, whilst significant after our meta-analysis (miR-203a-3p and miR-375) or FDR significant (miR-31-3p) after correlation.

miRNA id	Gene symbol	r (COPD)	r (Healthy)	Meta-analysis FDR	Gene function
miR-375	<i>ARL4C</i>	-0.49	-0.52	1.45E-06	Involved in tumorigenesis in lung ^{31, 32}
miR-375	<i>MBNL1</i>	-0.50	-0.49	5.52E-06	Tumorigenesis via an isoform ³³ , regulates alternative splicing ³⁴
miR-375	<i>SLC16A2</i>	-0.30	-0.53	3.71E-05	Transportation of diiodothyronine, thyroxine and triiodothyronine ³⁵
miR-375	<i>RLF</i>	-0.35	-0.45	2.75E-04	Tumorigenesis ³⁶
miR-203a-3p	<i>PDGFD</i>	-0.43	-0.37	9.58E-04	Fibroblast proliferation and survival, associated with adrenal suppression ^{37, 38}
miR-375	<i>TCF12</i>	-0.34	-0.39	1.86E-03	Tumorigenesis in several cancers. Cell differentiation, repression of E-cadherin, cell development, differentiation of lymphocytes ³⁹⁻⁴³
miR-375	<i>WBP1L</i>	-0.25	-0.43	2.20E-03	Regulates CXCR4 signalling in leukocytes and alters B-cell development ⁴⁴
miR-375	<i>PDE5A</i>	-0.49	-0.29	3.13E-03	Regulates cyclic GMP, intercellular messengers that mediate the effects of extracellular signalling molecules. ^{45, 46} Regulates pulmonary hypertension.

miR-375	<i>APBB2</i>	-0.42	-0.32	3.72E-03	Adapter protein, signal transduction ^{47, 48}
miR-375	<i>ZNF385D</i>	-0.54	-0.24	4.87E-03	
miR-375	<i>CNIH4</i>	-0.39	-0.33	5.13E-03	Located in the secretory pathway, it promotes the exit of GPCRs ⁴⁹
miR-375	<i>MMD</i>	-0.45	-0.27	7.77E-03	Macrophages activation ⁵⁰ , highly expressed in non-small cell lung cancer ⁵¹
miR-375	<i>HAS2</i>	-0.40	-0.29	9.63E-03	Regulates cell adhesion, extracellular matrix formation, tumorigenesis ^{52, 53}
miR-375	<i>CXCL12</i>	-0.39	-0.29	1.03E-02	Chemoattractant for T-lymphocytes and monocytes ⁵⁴
miR-375	<i>LDHB</i>	-0.41	-0.27	1.35E-02	Involved in tumorigenesis ⁵⁵ , a subunit of lactate dehydrogenase
miR-375	<i>LST1</i>	-0.10	-0.39	2.94E-02	Inhibition of lymphocyte proliferation ⁵⁶
miR-375	<i>DTHD1</i>	-0.02	0.44	3.09E-02	Involved in apoptosis
miR-375	<i>TRAPPC6B</i>	-0.28	-0.29	3.73E-02	Vesicle transport ⁵⁷ , involved in the secretory pathway ⁵⁸
miR-375	<i>APBB1IP</i>	-0.33	-0.26	4.02E-02	Adapter protein, signal transduction ^{47, 48}
miR-203a-3p	<i>EBF3</i>	-0.58	-0.10	4.22E-02	B-cell differentiation, bone development, neurogenesis and tumour suppressor via cell cycle arrest and apoptosis ⁵⁹
FDR					
miR-31-3p	<i>NFATC2</i>	-0.58		2.74E-06	nuclear factor of activated T cells - translocates to the nucleus upon T cell receptor (TCR) - immune response ^{60, 61}
miR-31-3p	<i>PLCB1</i>	-0.50		8.61E-05	intracellular transduction of many extracellular signals
miR-31-3p	<i>PDE5A</i>	-0.48		1.40E-04	cAMP binding - smooth muscle relaxation in the cardiovascular system ^{62, 63}
miR-31-3p	<i>RXFP1</i>	-0.48		1.43E-04	
miR-31-3p	<i>DLC1</i>	-0.47		2.52E-04	GAP family proteins participate in signalling pathways that regulate cell processes involved in cytoskeletal changes ⁶⁴ . – angiogenesis ⁶⁵
miR-31-3p	<i>TOX</i>	-0.45		4.06E-04	HMG box DNA binding domain. - chromatin assembly, transcription and replication ⁶⁶ - T-cell development ⁶⁷ .
miR-31-3p	<i>C1QTNF7</i>	-0.45		4.54E-04	
miR-31-3p	<i>PIK3CG</i>	-0.42		1.15E-03	immune response, proliferation and survival ^{68, 69}
miR-31-3p	<i>IL15</i>	-0.40		1.98E-03	regulates T and natural killer cell activation and proliferation ⁷⁰
miR-31-3p	<i>DIRAS3</i>	-0.40		2.24E-03	Inhibits RAS/MAPK signaling ⁷¹
miR-31-3p	<i>WASF3</i>	-0.38		4.02E-03	transduce signals that involve changes in cell shape, motility or function ⁷²
miR-31-3p	<i>LGI2</i>	-0.37		4.53E-03	
miR-31-3p	<i>KCNIP2</i>	-0.37		4.54E-03	Voltage-gated potassium channels
miR-31-3p	<i>TMOD2</i>	-0.36		5.48E-03	actin regulatory protein ⁷²
miR-31-3p	<i>RAB3C</i>	-0.35		8.29E-03	

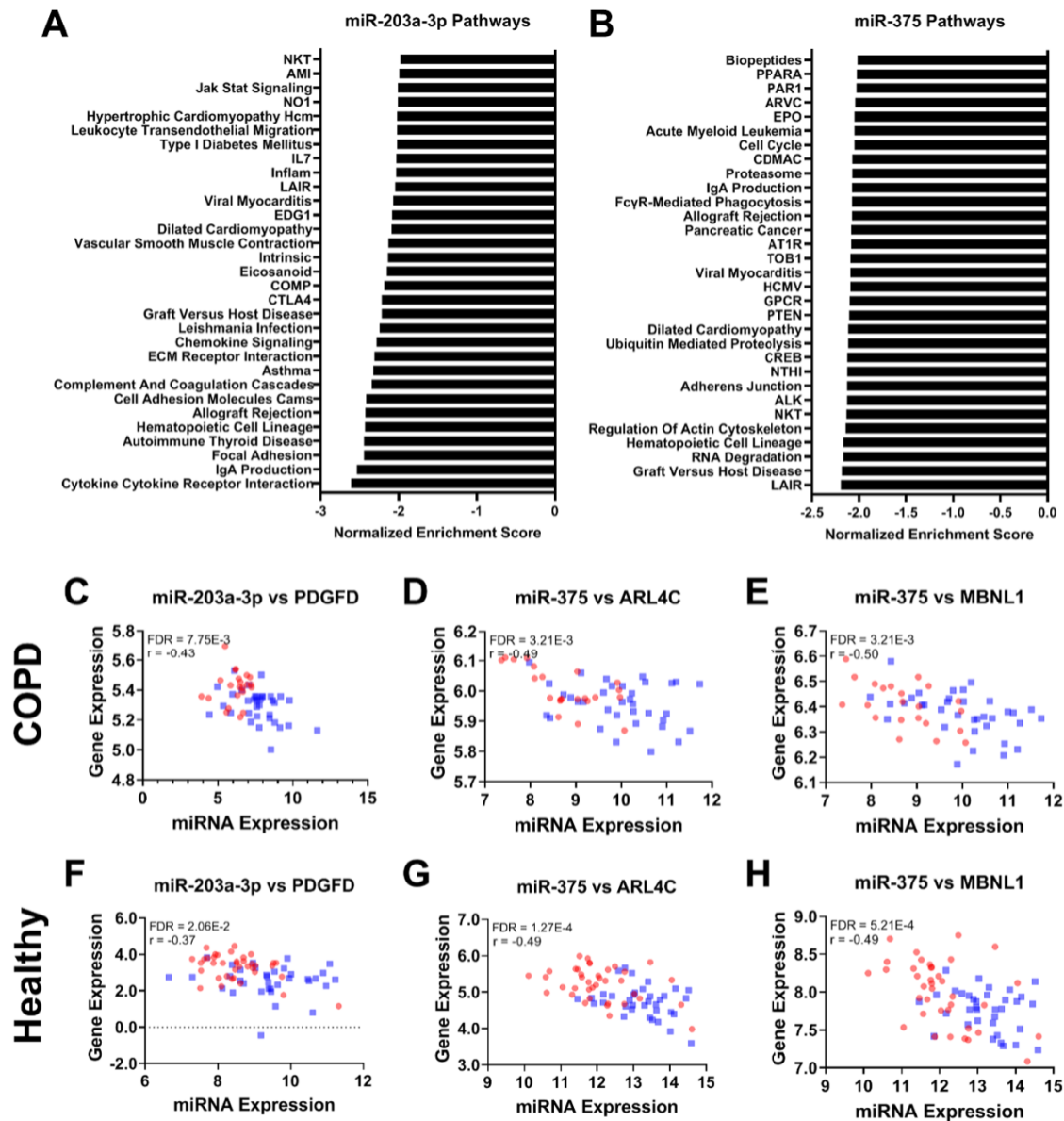


Figure 5.3) Significant pathway negative enrichment by miRNA-203a-3p and miR-375 and correlation of aforementioned miRNAs with genes in both COPD patients as well as respiratory healthy controls, excluding correlations that were considered not biologically relevant due to a low r-value. (A) Enriched pathways affected by miR-203a-3p in COPD patients. (B) The enriched pathways affected by miR-375 in COPD patients. Scatter plots (C–E) show the correlation between miRNA expression and gene expression levels in COPD patients for all significant negatively correlated predicted targets; miR-203a-3p and PDGFD, miR-375, and ARL4C, and miR-375 expression and MBNL1, respectively. Smoking participants are shown in blue, ex-, and non-smokers are shown in red. Scatter plots (F–H) show the correlation between miRNA expression and gene expression levels in respiratory healthy controls between miR-203a-3p and PDGFD, miR-375, and ARL4C, and miR-375 and MBNL1, respectively.

Correlation with predicted target genes

For the targeted analysis, we focused on negatively correlated predicted targets for miR-203a-3p, miR-375 and miR-31-3p. First, predicted targets were identified using miRNetap, and then the overlap was taken from the negatively correlated genes from the previous analysis.

Of the 983 genes negatively correlated with miR-203a-3p, two genes were also predicted targets; *PDGFD*, previously associated with fibroblast proliferation and survival, and *EBF3*, previously associated with B cell differentiation and cell survival through apoptosis and cell cycle arrest^{37, 38, 59}.

Of the 2442 genes negatively correlated to miR-375, 18 were also predicted targets; *ZNF385D*, *MBNL1*, *ARL4C*, *PDE5A*, *MMD*, *APBB2*, *LDHB*, *HAS2*, *CXCL12*, *CNIH4*, *APBB1IP*, *DTHD1*, *LST1*, *RLF*, *SLC16A2*, *TCF12*, *TRAPPC6B* and *WBP1L* (Figure 5.3C–H).

For miR-31-3p, we found that 14 of the 1106 negatively correlated genes were also predicted targets; *NFATC2*, *PLCB1*, *PDE5A*, *RXFP1*, *DLC1*, *TOX*, *C1QTNF7*, *PIK3CG*, *IL15*, *DIRAS3*, *WASF3*, *LGI2*, *KCNIP2* and *TMOD2*.

Pathways affected by miRNA expression that change with current smoking patients

To investigate the pathways influenced by changes in miRNA levels, we performed pathway analysis on (1) all genes of which the expression correlated to the identified miRNAs, for a global effect (Figure 5.3A,B), and (2) the significantly negatively correlated predicted targets of the miRNAs, for a direct effect of these miRNAs.

For miR-203a-3p, we found a higher expression associated with genes involved in cellular protection (xenobiotic and glutathione metabolism pathways), cellular repair (cell cycle, citrate cycle pathways), and protease pathway, as well as negative associations with pro-inflammatory pathways. In addition, the MAPK signalling pathway was negatively associated with miR-203a-3p associated genes in COPD patients (Table 5.4).

We found a global positive effect of genes associated with a higher expression of miR-375 on olfactory transduction, glycan biosynthesis and linoleic acid metabolism, and adverse effects on pro-inflammatory pathways in COPD patients, for example, via the cytokine-cytokine receptor integration pathway (Table 5.4).

A higher expression of miR-31-3p had positive associations with genes involved in pathways related to xenobiotic metabolism, cellular repair and linoleic acid metabolism, as well as negative associations with genes involved in pro-inflammatory pathways.

Table 5.4) Significant gene sets negatively affected by the identified miRNAs in COPD patients and asymptomatic participants. Only the top pathways with a family-wise error rate < 0.05 in the COPD patients are shown.

miRNA	Gene sets	COPD				Asymptomatic			
		NES	FWER P- value	Core- enriched genes	NES	FWER P- value	Core- enriched genes	Total genes	
miR-203a-3p	Cytokine-Cytokine Receptor Interaction	-2.58	0.00E+00	122	-2.78	0.00E+00	125	253	
	Intestinal Immune Network For IgA Production	-2.53	0.00E+00	36	-2.92	0.00E+00	26	45	
	Focal Adhesion	-2.46	0.00E+00	78	-2.27	8.00E-03	88	197	
	Systemic Lupus Erythematosus	-2.43	0.00E+00	27	-2.96	0.00E+00	31	50	
	Autoimmune Thyroid Disease	-2.43	0.00E+00	31	-2.71	0.00E+00	22	49	
	Hematopoietic Cell Lineage	-2.42	0.00E+00	51	-2.80	0.00E+00	49	83	
	Asthma	-2.42	0.00E+00	14	-2.54	0.00E+00	14	28	
	Cell Adhesion Molecules Cams	-2.41	0.00E+00	60	-2.58	0.00E+00	61	126	
	Allograft Rejection	-2.41	0.00E+00	26	-2.99	0.00E+00	23	35	
	Complement And Coagulation Cascades	-2.32	0.00E+00	26	-2.18	1.60E-02	36	67	
	ECM Receptor Interaction	-2.29	0.00E+00	40	-2.04	4.20E-02	43	83	
	Chemokine Signaling	-2.27	0.00E+00	72	-2.21	1.30E-02	53	179	
	Leishmania Infection	-2.25	0.00E+00	37	-2.84	0.00E+00	26	65	
	Ribosome	-3.05	0.00E+00	53	-2.72	0.00E+00	65	75	
	Leishmania Infection	-2.75	0.00E+00	35	-1.93	1.20E-02	38	65	
miR-375	Focal Adhesion	-2.74	0.00E+00	105	-2.04	0.00E+00	103	197	
	IL2rb	-2.71	0.00E+00	25	-2.20	0.00E+00	26	37	
	MET	-2.66	0.00E+00	25	-2.08	0.00E+00	22	33	
	B Cell Receptor Signaling	-2.62	0.00E+00	44	-1.84	1.03E-01	40	75	
	PDGF	-2.60	0.00E+00	20	-1.94	1.20E-02	23	28	
	GH	-2.58	0.00E+00	16	-2.13	0.00E+00	20	27	
	FCER1	-2.58	0.00E+00	25	-1.93	1.20E-02	29	39	
	GLEEVEc	-2.51	0.00E+00	18	-1.94	1.10E-02	15	23	
	Renal Cell Carcinoma	-2.51	0.00E+00	37	-2.03	1.00E-03	39	66	
	TCR	-2.49	0.00E+00	29	-1.98	3.00E-03	33	44	
	CTCF	-2.48	0.00E+00	18	-2.05	0.00E+00	15	24	
	Integrin	-2.46	0.00E+00	22	-2.06	0.00E+00	21	34	
	Hematopoietic Cell Lineage	-2.56	0.00E+00	44				83	
	Intestinal Immune Network For IgA Production	-2.56	0.00E+00	24				46	
	Cytokine-Cytokine Receptor Interaction	-2.55	0.00E+00	110				254	
miR-31-3p	Systemic Lupus Erythematosus	-2.46	0.00E+00	29				50	
	Focal Adhesion	-2.44	0.00E+00	74				197	
	Leishmania Infection	-2.43	0.00E+00	28				66	
	ECM Receptor Interaction	-2.39	0.00E+00	45				83	
	Cell Adhesion Molecules Cams	-2.31	0.00E+00	51				128	
	Chemokine Signaling Pathway	-2.28	0.00E+00	65				180	
	Autoimmune Thyroid Disease	-2.26	0.00E+00	35				49	
	Asthma	-2.26	0.00E+00	18				28	
	Allograft Rejection	-2.25	0.00E+00	28				35	
	LAIR Pathway	-2.24	0.00E+00	13				16	

NES: normalized enrichment score; FWER P-value: family-wise error rate p-value

Relation between cell-type composition and genes associated with miRNAs

Two methods were used to assess potential effects of cell-type composition on the expression of the miRNA-associated genes, i.e. GSVA based on single-cell gene signatures derived from a publically available dataset and correlation analysis with inflammatory cell counts assessed in adjacent biopsies of the same patients.

With the GSVA analysis, we found a significant negative correlation between miR-203a-3p expression and cell type composite scores of single-cell signatures for fibroblasts, luminal macrophages, endothelial cells, mast cells, smooth muscle cells, B-cells, submucosal cells and inflammatory dendritic cells. In contrast, we found a positive association of this miRNA with proliferating basal cells and club and goblet cells with current smoking in both COPD patients and respiratory healthy controls.

For miR-375, we found a negative correlation with cell type composite scores for endothelial cells, smooth muscle cells, mast cells, fibroblasts, basal cells and luminal macrophages., and a positive correlation with club and goblet cells, ionocytes and submucosal cells in both studies (Supplementary Table 5.2 and Supplementary Table 5.3). Activated endothelium, fibroblasts, luminal macrophages, mast cells and smooth muscle cells GSVA composite scores were negatively associated with miR-31-3p while proliferating basal cells and club and goblet cells were positively associated.

In addition, we correlated the miRNA expression profiles to cell counts of eosinophils, neutrophils, macrophages, mast cells, CD3 + , CD4 + or CD8 + T-cells, CD20 + B-cells, and percentage of PAS staining positive mucus in the COPD study. We found a negative correlation between eosinophils and miR-203a-3p expression and a positive correlation between miR-31-3p and epithelial cells, but no association with miR-375 (data not shown).

5.6 – Discussion

In the current study, we identified two miRNAs (miR-203a-3p and miR-375) that are significantly higher expressed with current smoking in COPD patients and respiratory healthy subjects, indicating a common effect of smoking regardless of disease status. Additionally, we found that miR-31-3p is associated with current smoking only in COPD patients.

We found that miR-375 is higher expressed in bronchial biopsies of current-smoking COPD patients and respiratory healthy controls, and we validated this finding in an independent dataset. We identified 18 predicted target genes that decrease with increased miR-375 expression. These included *TCF12*, *WBP1L*, *CXCL12*, *MMD*, and *LST1* that were previously associated with inflammation via T-cell precursor differentiation⁷³, altering B-cell development⁴⁴, chemotaxis⁵⁰, differentiation of monocytes to macrophages⁴⁶, or differentiation of other cells⁵². These predicted target genes were not enriched in one specific pathway. However, our global miRNA pathway analysis showed that genes associated with miR-375 are lower expressed in pro-inflammatory pathway activity, such as the IL2RB and B-cell receptor signalling pathway. We continued our investigation using cellular deconvolution and found that higher levels of miR-375 are associated with lower levels of basal cells, fibroblasts, mast cells, smooth muscle cells, endothelial cells and luminal macrophages and higher levels of secretory cells, submucosal cells and ionocytes. In contrast, no association was found with inflammatory cell populations as determined histologically. As we consistently find miR-375 being increased by smoke across all datasets and a positive correlation between miR-375 with secretory cells, this may indicate cell-specific expression in secretory cells as it is well established that the levels of secretory cells are higher in current smokers⁷⁴. The high number of genes and

pathways related to inflammation changed by miR-375, but the lack of correlation to gene signatures typical for inflammatory cells might indicate an increased activation of inflammatory cells caused by miR-375 rather than a change in cell numbers or composition.

We showed that miR-203a-3p is more highly expressed in bronchial biopsies of current- compared to ex-smokers with COPD and in current- versus non-smoking respiratory healthy subjects. Furthermore, two negatively correlated predicted targets of miR-203a-3, *EBF3* and *PDGFD*, were involved in cell cycle arrest and apoptosis³⁰ and fibroblast proliferation^{28, 29}, respectively. Therefore, the lower *EBF3* and *PDGFD* in response to smoking in association with miRNA-203a-3p by smoking potentially indicates a protective response and stimulation of repair as a response to cigarette smoke. In addition, we found that genes associated with xenobiotic metabolism were positively correlated with miR-203a-3p, suggesting that this miRNA facilitates the metabolism of harmful particles in cigarette smoke⁷⁵. Additionally, we found that genes involved in pro-inflammatory pathways (e.g. cytokine-cytokine receptor interaction, IgA production and chemokine signalling) were negatively correlated with miR-203a-3p, suggesting an anti-inflammatory effect for this miRNA. We associated these microRNAs with cell signatures to assess whether a change in cell populations may drive the smoking-induced increase in miR-203a-3p expression. We found higher smoking-induced expression of miR-203a-3p associated with lower proportions of fibroblasts, luminal macrophages, endothelial cells, mast cells, smooth muscle cells, B-cells, inflammatory dendritic cells, and submucosal cells and higher proportions of secretory cells (club and goblet cells) and proliferating basal cells. Therefore, it could be speculated that miR-203a-3p might play a role in the basal cell's differentiation towards secretory cells, leading to more mucus production in smokers or that this miRNA is selectively expressed in secretory cells. This, however, has now only been demonstrated in biopsies and requires further investigation in isolated cells to see the relationship of this miRNA with basal cell differentiation towards secretory cells, with and without smoke exposure.

MiR-31-3p was higher in current- compared to ex-smoking COPD patients, and this miRNA was not found in the respiratory healthy control dataset. We found 14 predicted target genes that decreased with an increase of miR-31-3p. Among these genes was *PDE5A*, which was involved in smooth muscle contraction and relaxation, as well as cell proliferation, cell signalling and included pro-inflammatory w Inhibition of *PDE5A* is involved in preventing tobacco smoke-induced emphysema⁷⁵, suggesting a protective role of this miRNA in current smokers with COPD as an attempt to limit the harmful effects of smoking⁷⁶. In a previous study, it was shown that *PDE5A* was decreased after smoke exposure in lung tissue of healthy individuals⁷⁶. The higher expression of miR-31-3p could suppress *PDE5A* expression even further in patients with COPD.

There were some limitations to the current study. As the miRNAs expression in the healthy group were conducted on different platforms, these miRNAs were not directly comparable. This may have led to the lack of expression of miRNA-31-3p in the healthy cohort. However, as the healthy cohort was conducted with small-RNA-Seq, it should have been picked up when present, and therefore we are confident in the current results. As a strength of the study, we were able to directly correlate the miRNA to matched expression data in the same subject allowing for the identification of direct miRNA-target gene interactions. Predicted targets for the miRNAs were identified using miRNAatap, which relies on algorithms and databases. However, as these are only predictions and not necessarily accurate, having the paired gene expression and miRNA expression needs additional confirmation of the associations.

In conclusion, we found two miRNAs, miR-203a-3p and miR-375, that are upregulated in bronchial biopsies of smokers compared to ex- and non-smokers in both COPD and respiratory healthy controls. These miRNAs play a role in the detoxification and inflammatory response to smoke response. In addition, we identified that miR-31-3p was upregulated only in current smokers versus ex-smokers with COPD. miR-31-3p might have a protective role via decreasing *PDE5A* expression and thus smoke-induced emphysema. We propose that our identified miRNAs are candidates for future studies aimed at unravelling which smoking-induced inflammatory mechanisms are mainly reactive and which are involved in COPD pathogenesis.

5.7 – Supplementary Material

Tables

Supplementary Table 5.1) Correlations of miR-203q-3p and miR-375 to genes in both the COPD and respiratory healthy dataset. Meta-analysis is performed between the datasets.

See <http://www.GitHub.com/vanNijnatten/PhD-Thesis> .

Supplementary Table 5.2) Significance of enrichment scores for single-cell signatures in named cell types.

Cell Type	COPD			Asymptomatic		
	P-value	T-Statistic	Standard Error	P-value	T-Statistic	Standard Error
Activated endothelium	6.64E-03	2.86	0.13	7.98E-01	0.26	0.11
All Basal cells	8.43E-02	1.76	0.08	2.79E-02	2.25	0.08
B cells	7.55E-03	2.79	0.13	5.47E-01	0.61	0.12
Proliferating Basal cells	4.60E-05	-4.49	0.12	2.30E-05	-4.54	0.12
Ciliated cells	1.64E-01	1.42	0.12	4.54E-01	0.75	0.10
Club and Goblet cells	5.96E-03	-2.93	0.09	2.53E-10	-7.41	0.08
Fibroblasts	1.38E-03	3.41	0.12	1.72E-01	1.38	0.13
Inflammatory DCs	8.82E-02	1.74	0.11	3.64E-01	-0.91	0.13
Ionocytes	1.64E-02	-2.52	0.13	2.44E-07	-5.79	0.10
Luminal macrophages	2.59E-02	2.30	0.15	6.99E-02	1.84	0.15
Mast cells	3.91E-03	3.03	0.14	6.75E-01	0.42	0.13
Neutrophils	3.65E-01	0.92	0.15	2.58E-01	1.14	0.15
Smooth muscle	8.84E-03	2.74	0.13	1.39E-02	2.52	0.13
Submucosal	5.05E-01	-0.67	0.14	8.27E-02	-1.76	0.13

Supplementary Table 5.3) Enrichment scores of single-cell signatures within named cell types correlated with expression values of miRNAs.

Cell Type	COPD			Asymptomatic		
	r	T-Statistic	FDR	r	T-Statistic	FDR
miR-203a-3p						
Activated endothelium	-0.53	-4.69	2.63E-04	-0.24	-2.07	7.38E-02
All Basal cells	-0.22	-1.69	1.50E-01	0.05	0.41	7.34E-01
B cells	-0.24	-1.80	1.37E-01	-0.34	-3.08	7.63E-03
Proliferating Basal cells	0.13	0.97	3.85E-01	0.48	4.57	2.82E-04
Ciliated cells	0.12	0.93	3.85E-01	-0.11	-0.97	4.25E-01
Club and Goblet cells	0.30	2.33	5.44E-02	0.43	4.02	9.98E-04
Fibroblasts	-0.49	-4.22	6.43E-04	-0.20	-1.75	1.32E-01
Inflammatory DCs	-0.18	-1.38	2.19E-01	-0.08	-0.67	5.89E-01
Ionocytes	0.20	1.49	1.98E-01	0.33	2.92	9.43E-03
Luminal macrophages	-0.37	-2.99	1.46E-02	-0.34	-3.04	7.63E-03
Mast cells	-0.33	-2.60	3.36E-02	-0.15	-1.25	3.02E-01
Neutrophils	-0.02	-0.16	8.73E-01	-0.38	-3.46	3.18E-03
Smooth muscle	-0.47	-3.94	1.08E-03	-0.38	-3.50	3.18E-03
Submucosal	-0.24	-1.83	1.37E-01	-0.02	-0.20	8.40E-01
miR-375						
Activated endothelium	-0.57	-5.18	4.62E-05	-0.14	-1.19	3.35E-01
All Basal cells	-0.30	-2.35	3.89E-02	-0.53	-5.28	1.85E-05
B cells	-0.21	-1.63	1.41E-01	-0.03	-0.29	7.71E-01
Proliferating Basal cells	0.11	0.84	4.69E-01	0.08	0.68	5.81E-01
Ciliated	0.04	0.28	8.38E-01	0.14	1.21	3.35E-01
Club and Goblet cells	0.49	4.13	4.30E-04	0.37	3.38	5.10E-03
Fibroblasts	-0.41	-3.32	3.75E-03	-0.35	-3.15	5.96E-03
Inflammatory DCs	-0.21	-1.62	1.41E-01	-0.08	-0.70	5.81E-01
Ionocytes	0.48	4.02	4.99E-04	0.37	3.31	5.10E-03
Luminal macrophages	-0.34	-2.64	2.16E-02	-0.06	-0.49	6.75E-01
Mast cells	-0.50	-4.27	3.62E-04	-0.35	-3.13	5.96E-03
Neutrophils	0.02	0.17	8.63E-01	-0.20	-1.69	1.67E-01
Smooth muscle	-0.51	-4.36	3.62E-04	-0.32	-2.84	1.18E-02
Submucosal	0.30	2.30	3.93E-02	0.46	4.40	2.58E-04
miR-31-3p						
Activated endothelium	-0.46	-3.81	1.25E-03			
All Basal cells	-0.06	-0.48	6.36E-01			
B cells	-0.27	-2.09	7.17E-02			
Proliferating Basal cells	0.37	2.99	1.17E-02			
Ciliated	-0.08	-0.57	6.16E-01			
Club and Goblet cells	0.33	2.58	2.53E-02			
Fibroblasts	-0.46	-3.85	1.25E-03			
Inflammatory DCs	-0.19	-1.44	1.96E-01			
Ionocytes	0.12	0.93	4.16E-01			
Luminal macrophages	-0.48	-4.07	1.25E-03			

Mast cells	-0.35	-2.73	1.98E-02
Neutrophils	-0.25	-1.95	8.85E-02
Smooth muscle	-0.46	-3.89	1.25E-03
Submucosal	-0.23	-1.72	1.27E-01

FDR: False discovery rate

Supplementary Table 5.4) Differential expression analysis of miRNAs in smokers compared to ex-smokers in an independent dataset.

	Smoking Status		COPD	
	T statistic	ANOVA	T statistic	ANOVA
miR-203	-0.55	5.87E-01	0.31	7.60E-01
miR-375	6.22	7.58E-08	1.70	9.55E-02
miR-31-3p	1.05	3.00E-01	1.34	1.86E-01

Supplementary Table 5.5) Differential expression analysis of miRNAs in smokers compared to ex-smokers in an independent dataset.

	Smoking Status COPD		Smoking Status Non-COPD	
	T statistic	ANOVA	T statistic	ANOVA
miR-203	-0.03	9.74E-01	1.05	3.03E-01
miR-375	-2.86	8.49E-03	-5.86	4.15E-06
miR-31-3p	-0.65	5.22E-01	0.97	3.39E-01

5.8 – References

1. Hoffmann, D. and I. Hoffmann, *Letters to the Editor - Tobacco smoke components*. Beiträge zur Tabakforschung International/Contributions to Tobacco Research, 1998. **18**(1): p. 49-52.
2. Hecht, S.S., *Cigarette smoking and lung cancer: chemical mechanisms and approaches to prevention*. Lancet Oncol, 2002. **3**(8): p. 461-9.
3. Govindan, R., et al., *Genomic landscape of non-small cell lung cancer in smokers and never-smokers*. Cell, 2012. **150**(6): p. 1121-34.
4. Klareskog, L., L. Padyukov, and L. Alfredsson, *Smoking as a trigger for inflammatory rheumatic diseases*. Curr Opin Rheumatol, 2007. **19**(1): p. 49-54.
5. Andres, S.A., et al., *Interaction between smoking history and gene expression levels impacts survival of breast cancer patients*. Breast Cancer Res Treat, 2015. **152**(3): p. 545-56.
6. Nedeljkovic, I., et al., *COPD GWAS variant at 19q13.2 in relation with DNA methylation and gene expression*. Hum Mol Genet, 2018. **27**(2): p. 396-405.
7. Perdomo, C., et al., *MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis*. Proc Natl Acad Sci U S A, 2013. **110**(47): p. 18946-51.
8. Schembri, F., et al., *MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium*. Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2319-24.
9. Hammond, S.M., et al., *An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells*. Nature, 2000. **404**(6775): p. 293-6.
10. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs*. Genome Res, 2009. **19**(1): p. 92-105.
11. Lim, L.P., et al., *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs*. Nature, 2005. **433**(7027): p. 769-73.
12. Podshivalova, K. and D.R. Salomon, *MicroRNA regulation of T-lymphocyte immunity: modulation of molecular networks responsible for T-cell activation, differentiation, and development*. Crit Rev Immunol, 2013. **33**(5): p. 435-76.
13. Hwang, H.W. and J.T. Mendell, *MicroRNAs in cell proliferation, cell death, and tumorigenesis*. Br J Cancer, 2006. **94**(6): p. 776-80.
14. Lacedonia, D., et al., *Expression profiling of miRNA-145 and miRNA-338 in serum and sputum of patients with COPD, asthma, and asthma-COPD overlap syndrome phenotype*. Int J Chron Obstruct Pulmon Dis, 2017. **12**: p. 1811-1817.
15. Ezzie, M.E., et al., *Gene expression networks in COPD: microRNA and mRNA regulation*. Thorax, 2012. **67**(2): p. 122-31.
16. Kim, W.J., et al., *Altered miRNA expression in lung tissues of patients with chronic obstructive pulmonary disease*. Molecular & Cellular Toxicology, 2017. **13**(2): p. 207-212.
17. Faiz, A., et al., *Effect of long-term corticosteroid treatment on microRNA and gene-expression profiles in COPD*. Eur Respir J, 2019. **53**(4).
18. Faiz, A., et al., *Profiling of healthy and asthmatic airway smooth muscle cells following interleukin-1beta treatment: a novel role for CCL20 in chronic mucus hypersecretion*. Eur Respir J, 2018. **52**(2).
19. Osei, E.T., et al., *miR-146a-5p plays an essential role in the aberrant epithelial-fibroblast cross-talk in COPD*. Eur Respir J, 2017. **49**(5).
20. Boudewijn, I.M., et al., *A Novel Role for Bronchial MicroRNAs and Long Noncoding RNAs in Asthma Remission*. Am J Respir Crit Care Med, 2020. **202**(4): p. 614-618.
21. Conickx, G., et al., *microRNA profiling in lung tissue and bronchoalveolar lavage of cigarette smoke-exposed mice and in COPD patients: a translational approach*. Sci Rep, 2017. **7**(1): p. 12871.
22. Tasena, H., et al., *microRNA-mRNA regulatory networks underlying chronic mucus hypersecretion in COPD*. Eur Respir J, 2018. **52**(3).
23. Lapperre, T.S., et al., *Smoking cessation and bronchial epithelial remodelling in COPD: a cross-sectional study*. Respir Res, 2007. **8**(1): p. 85.

24. Hoonhorst, S.J., et al., *Advanced glycation endproducts and their receptor in different body compartments in COPD*. Respir Res, 2016. **17**: p. 46.
25. Ong, J., et al., *Age-related gene and miRNA expression changes in airways of healthy individuals*. Sci Rep, 2019. **9**(1): p. 3765.
26. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts*. Genome Biol, 2014. **15**(2): p. R29.
27. Steiling, K., et al., *A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment*. Am J Respir Crit Care Med, 2013. **187**(9): p. 933-42.
28. Vieira Braga, F.A., et al., *A cellular census of human lungs identifies novel cell states in health and in asthma*. Nat Med, 2019. **25**(7): p. 1153-1163.
29. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. BMC Bioinformatics, 2013. **14**: p. 7.
30. van den Berge, M., et al., *Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity*. Thorax, 2014. **69**(1): p. 14-23.
31. Fujii, S., et al., *Arl4c expression in colorectal and lung cancers promotes tumorigenesis and may represent a novel therapeutic target*. Oncogene, 2015. **34**(37): p. 4834-44.
32. Kimura, K., et al., *ARL4C is associated with initiation and progression of lung adenocarcinoma and represents a therapeutic target*. Cancer Sci, 2020. **111**(3): p. 951-961.
33. Tabaglio, T., et al., *MBNL1 alternative splicing isoforms play opposing roles in cancer*. Life Sci Alliance, 2018. **1**(5): p. e201800157.
34. Goers, E.S., et al., *MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing*. Nucleic Acids Res, 2010. **38**(7): p. 2467-84.
35. Jansen, J., et al., *Functional analysis of monocarboxylate transporter 8 mutations identified in patients with X-linked psychomotor retardation and elevated serum triiodothyronine*. J Clin Endocrinol Metab, 2007. **92**(6): p. 2378-81.
36. Makela, T.P., et al., *Intrachromosomal rearrangements fusing L-myc and rlf in small-cell lung cancer*. Mol Cell Biol, 1991. **11**(8): p. 4015-21.
37. Warshamana, G.S., et al., *Dexamethasone activates expression of the PDGF-alpha receptor and induces lung fibroblast proliferation*. Am J Physiol, 1998. **274**(4): p. L499-507.
38. Hawcutt, D.B., et al., *Susceptibility to corticosteroid-induced adrenal suppression: a genome-wide association study*. The Lancet Respiratory Medicine, 2018. **6**(6): p. 442-450.
39. Chen, W.S., et al., *Secreted heat shock protein 90alpha (HSP90alpha) induces nuclear factor-kappaB-mediated TCF12 protein expression to down-regulate E-cadherin and to enhance colorectal cancer cell migration and invasion*. J Biol Chem, 2013. **288**(13): p. 9001-10.
40. Lee, C.C., et al., *TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer*. J Biol Chem, 2012. **287**(4): p. 2798-809.
41. L'Abbate, A., et al., *t(15;21) translocations leading to the concurrent downregulation of RUNX1 and its transcription factor partner genes SIN3A and TCF12 in myeloid disorders*. Mol Cancer, 2015. **14**: p. 211.
42. Uittenbogaard, M. and A. Chiaramello, *Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis*. Gene Expression Patterns, 2002. **1**(2): p. 115-121.
43. Mesman, S. and M.P. Smidt, *Tcf12 Is Involved in Early Cell-Fate Determination and Subset Specification of Midbrain Dopamine Neurons*. Front Mol Neurosci, 2017. **10**: p. 353.
44. Borna, S., et al., *Transmembrane adaptor protein WBP1L regulates CXCR4 signalling and murine haematopoiesis*. J Cell Mol Med, 2020. **24**(2): p. 1980-1992.
45. Schudt, C., et al., *Therapeutic potential of selective PDE inhibitors in asthma*. Pulm Pharmacol Ther, 1999. **12**(2): p. 123-9.

46. Movsesian, M.A., *Beta-adrenergic receptor agonists and cyclic nucleotide phosphodiesterase inhibitors: shifting the focus from inotropy to cyclic adenosine monophosphate*. Journal of the American College of Cardiology, 1999. **34**(2): p. 318-324.
47. Yang, Z., et al., *A dominant role for FE65 (APBB1) in nuclear signaling*. J Biol Chem, 2006. **281**(7): p. 4207-14.
48. Schettini, G., et al., *Phosphorylation of APP-CTF-AICD domains and interaction with adaptor proteins: signal transduction and/or transcriptional role--relevance for Alzheimer pathology*. J Neurochem, 2010. **115**(6): p. 1299-308.
49. Sauvageau, E., et al., *CNIH4 interacts with newly synthesized GPCR and controls their export from the endoplasmic reticulum*. Traffic, 2014. **15**(4): p. 383-400.
50. Liu, Q., et al., *Monocyte to macrophage differentiation-associated (MMD) positively regulates ERK and Akt activation and TNF-alpha and NO production in macrophages*. Mol Biol Rep, 2012. **39**(5): p. 5643-50.
51. Li, W. and F. He, *Monocyte to macrophage differentiation-associated (MMD) targeted by miR-140-5p regulates tumor growth in non-small cell lung cancer*. Biochem Biophys Res Commun, 2014. **450**(1): p. 844-50.
52. Koyama, H., et al., *Significance of tumor-associated stroma in promotion of intratumoral lymphangiogenesis: pivotal role of a hyaluronan-rich tumor microenvironment*. Am J Pathol, 2008. **172**(1): p. 179-93.
53. Bernert, B., H. Porsch, and P. Heldin, *Hyaluronan synthase 2 (HAS2) promotes breast cancer cell invasion by suppression of tissue metalloproteinase inhibitor 1 (TIMP-1)*. J Biol Chem, 2011. **286**(49): p. 42349-59.
54. Ghosh, S., et al., *Cannabinoid receptor CB2 modulates the CXCL12/CXCR4-mediated chemotaxis of T lymphocytes*. Mol Immunol, 2006. **43**(14): p. 2169-79.
55. McClelland, M.L., et al., *Lactate dehydrogenase B is required for the growth of KRAS-dependent lung adenocarcinomas*. Clin Cancer Res, 2013. **19**(4): p. 773-84.
56. Rollinger-Holzinger, I., et al., *LST1: a gene with extensive alternative splicing and immunomodulatory function*. J Immunol, 2000. **164**(6): p. 3169-76.
57. Kummel, D., et al., *The structure of the TRAPP subunit TPC6 suggests a model for a TRAPP subcomplex*. EMBO Rep, 2005. **6**(8): p. 787-93.
58. Kim, J.J., Z. Lipatova, and N. Segev, *TRAPP Complexes in Secretion and Autophagy*. Front Cell Dev Biol, 2016. **4**: p. 20.
59. Boller, S. and R. Grosschedl, *The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function*. Immunol Rev, 2014. **261**(1): p. 102-15.
60. Peng, S.L., et al., *NFATc1 and NFATc2 Together Control Both T and B Cell Activation and Differentiation*. Immunity, 2001. **14**(1): p. 13-20.
61. Rengarajan, J., B. Tang, and L.H. Glimcher, *NFATc2 and NFATc3 regulate T(H)2 differentiation and modulate TCR-responsiveness of naive T(H) cells*. Nat Immunol, 2002. **3**(1): p. 48-54.
62. Kass, D.A., H.C. Champion, and J.A. Beavo, *Phosphodiesterase type 5: expanding roles in cardiovascular regulation*. Circ Res, 2007. **101**(11): p. 1084-95.
63. Li, S., et al., *Phosphodiesterase-5a Knock-out Suppresses Inflammation by Down-Regulating Adhesion Molecules in Cardiac Rupture Following Myocardial Infarction*. J Cardiovasc Transl Res, 2021.
64. Lahoz, A. and A. Hall, *DLC1: a significant GAP in the cancer genome*. Genes Dev, 2008. **22**(13): p. 1724-30.
65. Shih, Y.P., S.Y. Yuan, and S.H. Lo, *Down-regulation of DLC1 in endothelial cells compromises the angiogenesis process*. Cancer Lett, 2017. **398**: p. 46-51.
66. Page, N., et al., *Persistence of self-reactive CD8+ T cells in the CNS requires TOX-dependent chromatin remodeling*. Nat Commun, 2021. **12**(1): p. 1009.
67. Aliahmad, P., et al., *TOX is required for development of the CD4 T cell lineage gene program*. J Immunol, 2011. **187**(11): p. 5931-40.

68. Kaneda, M.M., et al., *PI3Kgamma is a molecular switch that controls immune suppression*. Nature, 2016. **539**(7629): p. 437-442.
69. Sasaki, T., et al., *Function of PI3Kgamma in thymocyte development, T cell activation, and neutrophil migration*. Science, 2000. **287**(5455): p. 1040-6.
70. Nandagopal, N., et al., *The Critical Role of IL-15-PI3K-mTOR Pathway in Natural Killer Cell Effector Functions*. Front Immunol, 2014. **5**: p. 187.
71. Sutton, M.N., et al., *DIRAS3 (ARHI) Blocks RAS/MAPK Signaling by Binding Directly to RAS and Disrupting RAS Clusters*. Cell Rep, 2019. **29**(11): p. 3448-3459 e6.
72. Corral-Serrano, J.C., et al., *PCARE and WASF3 regulate ciliary F-actin assembly that is required for the initiation of photoreceptor outer segment disk formation*. Proc Natl Acad Sci U S A, 2020. **117**(18): p. 9922-9931.
73. Braunstein, M. and M.K. Anderson, *HEB-deficient T-cell precursors lose T-cell potential and adopt an alternative pathway of differentiation*. Mol Cell Biol, 2011. **31**(5): p. 971-82.
74. Cerveri, I. and V. Brusasco, *Revisited role for mucus hypersecretion in the pathogenesis of COPD*. Eur Respir Rev, 2010. **19**(116): p. 109-12.
75. Seimetz, M., et al., *Cigarette Smoke-Induced Emphysema and Pulmonary Hypertension Can Be Prevented by Phosphodiesterase 4 and 5 Inhibition in Mice*. PLoS One, 2015. **10**(6): p. e0129327.
76. Zuo, H., et al., *Cigarette smoke exposure alters phosphodiesterases in human structural lung cells*. Am J Physiol Lung Cell Mol Physiol, 2020. **318**(1): p. L59-L64.

Chapter 6 – Elucidating the association between DNA methylation and COPD in bronchial biopsies

Jos van Nijnatten, Alen Faiz, Wim Timens, Cornelis Joseph Vermeulen, Maaïke de Vries, Sandra Casas, Rosa Faner, Maarten van den Berge[†], Corry-Anke Brandsma[†]

Status: Not published

6.1 – Abstract

Introduction: COPD is a smoking-related disease characterised by various degrees of airway remodelling and emphysematous lung tissue destruction. COPD is associated with changes in gene expression and epigenetics, such as DNA methylation. Although many studies have found differences in DNA methylation in blood between COPD patients and healthy controls¹, few studies investigated how DNA methylation differs in the airway wall.

Aim: To investigate DNA methylation levels in the airway wall associated with the presence and severity of COPD.

Methods: We performed a differential methylation analysis comparing patients with COPD (n=90) and non-COPD (n=21) controls and associated DNA methylation levels with airflow obstruction and hyperinflation severity. eQTM analysis assessed how COPD-associated DNA methylation levels affect gene expression. eQTMs were checked for associations with the calculated cell type proportions in our RNA-Seq data.

Results: The presence of COPD was associated with six differentially methylated CpG sites compared to non-COPD controls, of which three were hypermethylated and three hypomethylated. Within COPD, 16 CpG sites were hypomethylated and nine hypermethylated, associated with more severe airflow obstruction. A total of 2274 CpG sites were hypomethylated and 1680 hypermethylated in association with hyperinflation, as reflected by a higher RV/TLC % pred. eQTM analysis identified 817 eQTMs for hyperinflation-associated CpGs and none for the COPD- and airflow obstruction-associated CpGs. Pathway analysis showed enrichment of hematopoietic cell lineage pathways and pathways involved in the immune system and (immune cell) signal transduction.

Conclusions: In conclusion, we show DNA methylation changes associated with the presence and severity of COPD. CpG sites hypomethylated in association with more severe hyperinflation in COPD were found to affect genes involved in immune regulation signal transduction.

Keywords: COPD methylation

6.2 – Introduction

Chronic obstructive pulmonary disease (COPD) is one of the major causes of death worldwide². It is characterised by various degrees of airway remodelling and emphysematous lung tissue destruction. Smoking is the main environmental risk factor of COPD. Previously, we and others have shown large transcriptional differences between COPD and non-COPD individuals in different compartments of the lungs^{3,4}. Interestingly, even within COPD, there is large heterogeneity within these transcriptional profiles, which predict the severity of the disease⁷ and the response to current therapies⁵. It is still unclear how these expression profiles are regulated; epigenetic regulation likely plays an important role. Epigenetic regulation in the form of DNA Methylation regulates gene expression by adding a methyl group to a cytosine base next to a guanine base⁶. These methylation sites can promote or suppress gene expression depending on whether they are positioned within a promoter or repression region of a gene. It is well established that active smoking has significant effects on the epigenome of the blood and airways^{7,8}. Although many studies have found differences in DNA methylation in the blood of COPD patients independent of smoking¹, few studies investigated the direct association between DNA methylation in the airway wall and the presence or severity of COPD. Whereas DNA methylation in blood has provided an important view on the epigenetic landscape of COPD, it did not provide information about methylation changes in the lung itself. At the same time, the airways form the first line of defence against smoke exposure. They are characterised by important pathological changes in COPD, including chronic inflammation, airway wall thickening, and mucus hypersecretion.

In the current study, we aimed to identify differences in DNA methylation in airway wall biopsies between current smokers with and without COPD and to assess how differences in DNA methylation are associated with the severity of airflow obstruction and hyperinflation. In addition, we investigated the impact of differential DNA methylation on gene expression in the airway wall to explain the possible role of differential DNA methylation in COPD pathology.

6.3 – Methods

Patient data

We used the previously described Groningen and Leiden Universities Corticosteroids in Obstructive Lung Disease (GLUCOLD⁹, $n_{\text{COPD}}=76$, ClinicalTrials.gov: NCT00158847) and Stop-Smoking¹⁰ ($n_{\text{COPD}}=14$, $n_{\text{non-COPD}}=21$) studies. In short, GLUCOLD included patients with COPD stage II or III patients without a history of asthma, were aged 45 to 75 years, current or former smokers with at least ten pack-years, and had not used ICS for at least six months prior to the study⁹. For Stop-Smoking, all subjects smoked and either had COPD or were asymptomatic with a post-bronchodilator FEV₁/FVC ratio higher than 70%. All patients were extensively characterised, i.e., lung function, blood tests, and questionnaire data. In addition, a bronchoscopy was performed, during which bronchial biopsies were obtained¹⁰. The local medical ethics committees approved the study, and all subjects gave their written informed consent.

DNA Methylation and RNA sequencing data processing

DNA was isolated from bronchial biopsies of all 90 patients with COPD and 21 non-COPD controls at the same time, carefully taking into account an even distribution across the batches with respect to the presence of COPD, sex, and smoking status. DNA methylation profiling was performed using the HumanMethylation850K BeadChip (Illumina Inc., San Diego, CA, USA) and gene expression profiling using RNA-Seq¹¹. The methods for DNA and RNA extraction, 850K array processing, and RNA-Seq have been described previously¹¹⁻¹⁵. Raw DNA methylation intensity values were processed using the Minfi package¹⁶, implemented in R. Samples and probes failing QC were removed, and raw beta

values normalised using the Dasen method as implemented in the WateRmelon package¹⁷. Raw RNA-Seq reads were aligned using STAR-Aligner and normalised to counts per million. Technical variation between the RNA-Seq datasets from both studies was eliminated using CombatSeq¹⁸.

Association of DNA methylation with COPD, airflow obstruction and hyperinflation

We performed differential methylation analysis comparing the patients with COPD and non-COPD controls using robust linear modelling using limma (version 3.54.2) adjusting for age and sex as potential confounders¹⁹ and only included current smokers in both groups. We used a robust linear model comparing the severity of airflow obstruction and hyperinflation as reflected by the FEV₁ % pred and RV/TLC % pred in smoking COPD patients only, adjusting for age and sex. A Benjamini-Hochberg False Discovery Rate (FDR) < 0.05 was maintained to define statistical significance.

Validation in blood and lung tissue

We performed a validation analysis of the differential methylation analyses in blood in independent datasets. The first dataset was the previously described LifeLines COPD&C DNA methylation dataset²⁰. We selected the current smokers for the differential methylation analysis comparing COPD (n=249) and non-COPD (n=350) participants and the association with FEV₁ % pred in participants with COPD only. We used a robust linear model using limma while adjusting for age, sex, the technical principal components explaining more than 5% of the variation, and the cellular populations of monocytes, eosinophils, granulocytes and lymphocytes.

The second dataset for the validation of the association of DNA methylation with COPD and airflow obstruction in COPD was performed in lung tissue samples patients previously described²¹. Included in this cohort were 28 current smoking patients with different levels of airflow limitation (FEV₁ % pred). DNA was isolated from lung tissue, and DNA methylation was assessed with the EPIC array (Illumina). The 'ChAMP' pipeline was used for quality control, covariates/batch adjustment, and the differential methylation analysis of the EPIC raw data. Batch effects (Sentrix Position and samples batch plate) were adjusted using ComBat²². CpG sites associated with FEV₁ % pred in lung tissue in COPD only were determined using a robust linear model using limma, and we corrected for age and sex²¹.

As none of the validation cohorts included information about hyperinflation, we could not validate our hyperinflation-associated CpG sites.

Association of DNA methylation with Gene expression

To find the effects of DNA methylation on gene expression, we performed a cis-eQTM analysis (< 250.000bp) in samples with both methylation and RNA-Seq data, separately for COPD-associated CpGs and CpGs associated with disease severity. The five principal components explaining 95% of the variation in the control probes of the 850K array were used as covariates to correct for technical variation in all analyses (Figure 6.1).

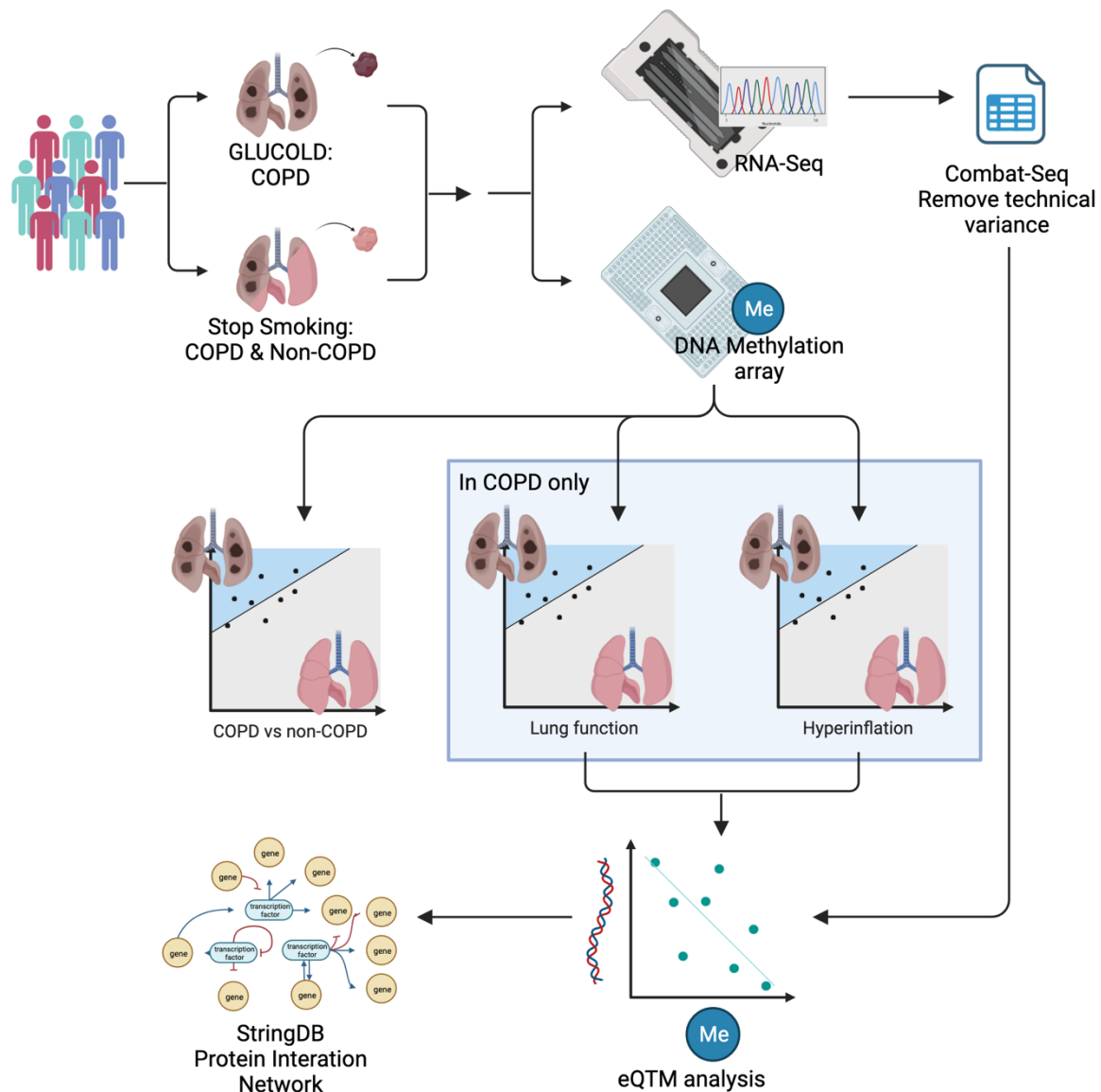


Figure 6.1) Flow chart of the statistical methods.

Cellular deconvolution

To estimate cell proportions for cell types in the bronchial biopsies, we performed RNA-seq-based cellular deconvolution with CIBERSORT as described previously¹¹, using airway wall cell matrices of B-cell lineage, ionocytes, mast cells, arterial endothelial cells, ciliated cells, goblet cells, cells of T-cell lineage, monocytes, submucosal secretory cells, basal cells, dendritic cells, basal cycling cells, alveolar macrophages, other macrophages, and fibroblasts²³. To determine if the shifts in DNA methylation levels were due to shifts in cell populations, we associated the eQTMs with cell proportions if more than 25% of samples had a cell proportion greater than zero. M-values of DNA-methylation sites were Spearman correlated to cell proportions, corrected for multiple testing over the CpG sites, and we considered an FDR < 0.05 significant.

Pathway analysis

To check the effects of the differentially methylated sites on a biological level, we performed a pathway analysis of genes affected by DNA methylation levels associated with RV/TLC % pred. To

this end, we used String-DB to create the network and then hide the disconnected nodes. We performed an enrichment analysis on the overall network to determine the biological pathways, biological processes, molecular functions and cellular components using Gene Ontology, KEGG and Reactome.

6.4 – Results

Patient demographics

The clinical characteristics of the included COPD (n=90) patients and non-COPD (n=21) controls are displayed in Table 6.1. COPD patients were older than non-COPD controls, with an average age of 61.3 ± 7.3 versus 50.2 ± 3.7 years (mean \pm SD, p-value $7.54E-10$). Participants with COPD had a median FEV₁/FVC % of 50.7 (32.2-69.9), a median FEV₁ % pred of 71.5 (38.0-102.9) and a median RV/TLC % pred of 125.6 (81.0-179.0) (median, IQR), whereas non-COPD controls had a median FEV₁/FVC % of 79.9 (70.3-87.6), a median FEV₁ % pred of 92.7% (81.2-149.4) and an RV/TLC % pred of 89.0 (71.0-131.0). Additionally, there was a higher percentage of male participants in COPD compared to non-COPD controls (83.3%, p-value $5.65E-03$).

Table 6.1) COPD and non-COPD patient demographics. FEV₁ is Forced Expiratory Volume in one second, FVC is Forced Vital Capacity, RV is Residual Volume, and TLC is Total Lung Capacity. ¹ Mean (SD), ² Median (range)

	Non-COPD (n=21)	COPD (n=90)
Age (Years) ¹	50.2 (3.7)	61.3 (7.3)*
Male (%)	52.4	83.3*
Packyears ²	25.2 (13.4-41.5)	41.8 (13.0-181.6)
FEV ₁ % pred ²	92.7 (81.2-149.4)	71.5 (38.0-102.9)*
FEV ₁ /FVC post bronchodilator ²	79.9 (70.3-87.6)	50.7 (32.2-69.9)*
RV/TLC % pred ³	89.0 (71.0-131.0)	125.6 (81.0-179.0)*
Blood Eosinophils (E10 ⁶ cell/ml) ²	0.1 (0.0-0.4)	0.2 (0.0-1.8)
Blood Eosinophils (%) ²	1.3 (0.6-4.9)	2.2 (0.5-15.6)

Differentially methylated sites associated with the presence and severity of COPD.

We identified six CpG sites differentially methylated between COPD patients and non-COPD controls (Figure 6.2A&B), of which three were hypermethylated (cg23800199, cg05748572, cg16560488) and three hypomethylated (cg10730398, cg01124573, cg12081267) in COPD. Five of the six CpG sites were found in the intronic regions of protein-coding genes (*DIRC3*, *GNG7*, *JMJD1C*, *ADGRL3* and *TMEM131*, respectively) (Figure 6.2). Of interest, cg01124573 is located within an intronic region of *ADGRL3*, a gene previously found to be highly expressed in asthma²⁴.

Next, we investigated how DNA methylation is associated with the severity of COPD within the COPD subset of patients. Here, we investigated two clinical measures of lung function: FEV₁ % pred and RV/TLC % pred. This analysis identified 25 CpG sites associated with the severity of airflow obstruction as reflected by the level of FEV₁ % pred (Figure 6.3A&B), with 16 DNA methylation sites less methylated with increased, and 9 DNA methylation sites more methylated with a decrease in FEV₁ % pred (Figure 6.3C). Nine of the CpG sites were in an intronic region of a gene (*ATAMTS18*, *ENSG00000267098*, *SPATA16* *CCDC57*, *ENSG00000289526*, *PAAF1*, *ENSG00000231918*, *CENPU*, and *NDUFA5*), six CpG sites were in the exonic region of genes (*RAB4A*, *HDAC4*, *ZNF776*, *CTCF*, *MSH4*, and *SPATA45*), and nine more CpG sites were located outside of genes (Table 6.2). Lastly, we found

3954 CpG sites associated with more hyperinflation, as reflected by a higher RV/TLC % pred (FDR < 0.05, Figure 6.3D&E). Of these, 2274 were hypomethylated, and 1680 CpG sites were hypermethylated with an increase in hyperinflation. Of the top 20 CpGs, five were in an exonic region (TSBP1-AS1, KIRREL3, MUC5AC, HEY2-AS1, and PI4KA), six were in an intronic region of a gene (CCDC102B, FSIP2, RGMB, FZD8, CBFA2T3, and DTNB), and 8 were close to a gene (ADGRA3, PRKCQ, LINC02247, FIBIN, LINC02376, PHETA1, NUA1, PARD3B, and LILRB2, Figure 6.3F). Of interest were PRKCQ, a gene that encodes for a protein kinase, and MUC5AC, a gene that encodes for mucin and has previously been linked with mucus hypersecretion in COPD²⁵.

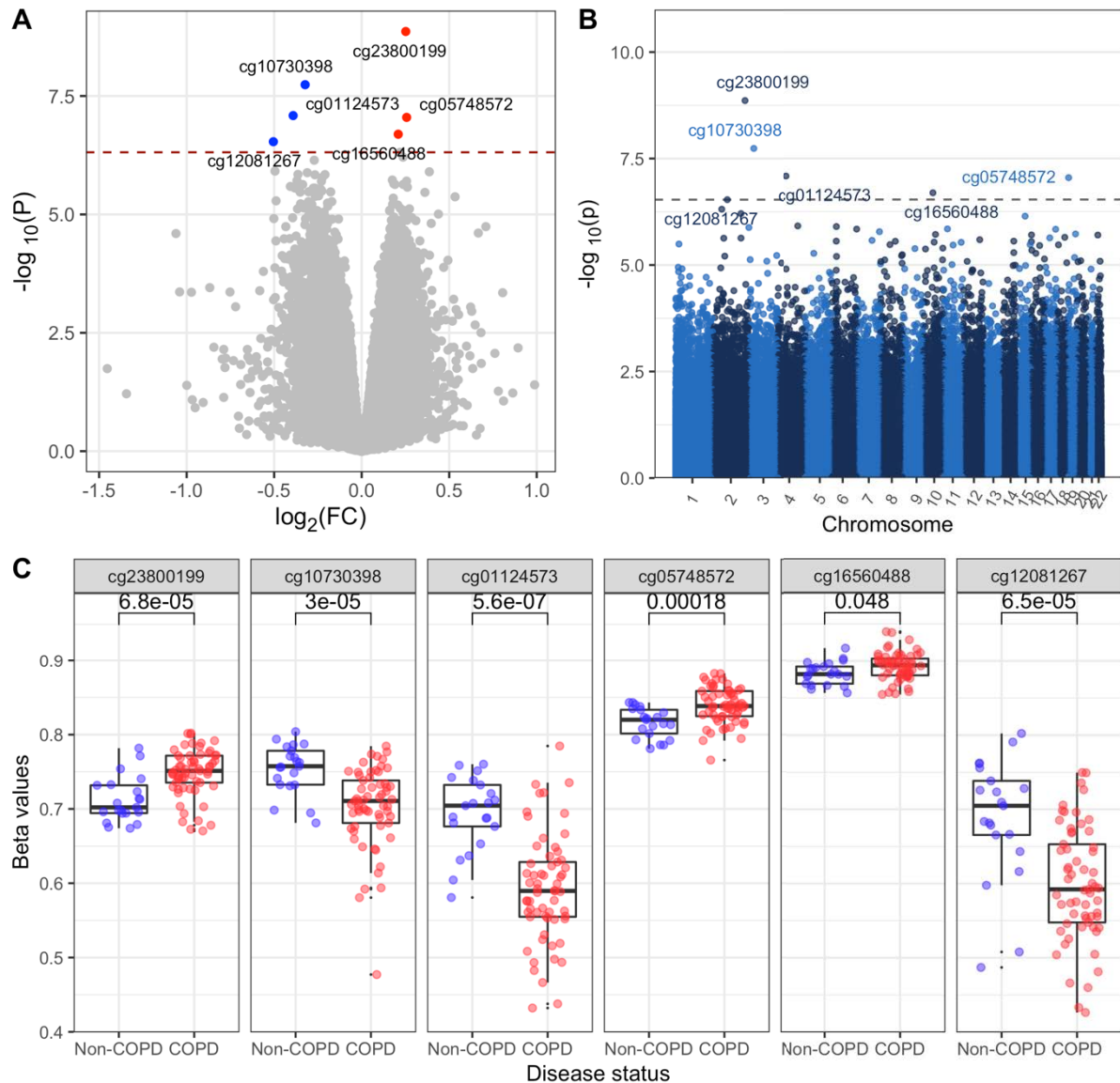


Figure 6.2) Differential Methylation comparing COPD and non-COPD. [A] Volcano plot showing the fold change and significance of the change in methylation in COPD, red are hypermethylated CpG sites, and hypomethylated CpG sites are in blue. [B] Manhattan plot showing the position and significance of the change in methylation in COPD. [C] The six differentially methylated CpG sites associated with COPD stratified by disease status.

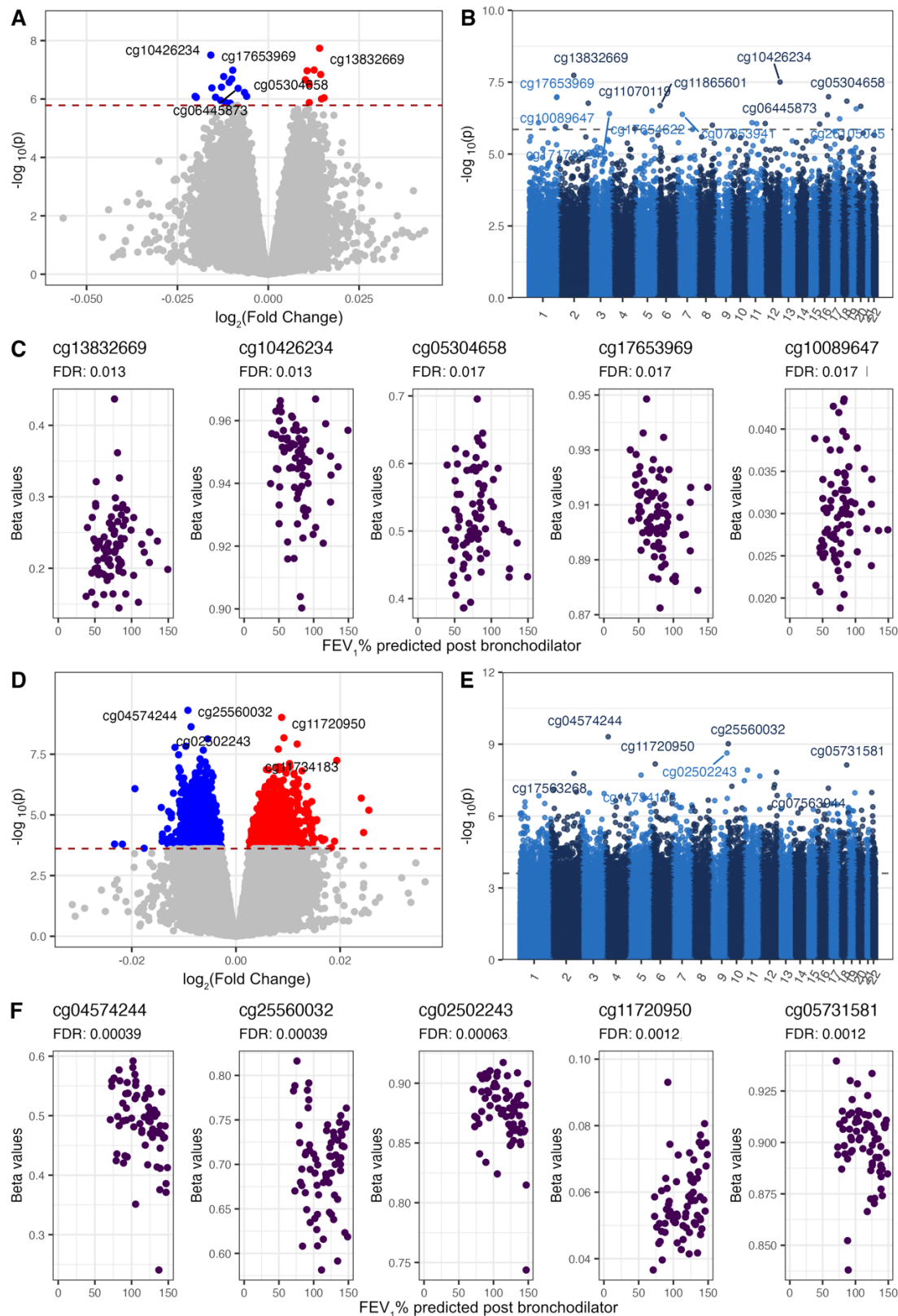


Figure 6.3) Differential Methylation associated with FEV₁ % pred (A-C) and RV/TLC % pred (D-F). Volcano plots [A] and [D] show the fold change and significance of the change associated with lung function and hyperinflation, respectively; in red are hypermethylated CpG sites, and hypomethylated CpG sites are in blue. Manhattan plots [B] and [E] show the position and significance of the change in methylation associated with lung function and hyperinflation, respectively. [C] and [F] show the associations of the top CpG sites with lung function and hyperinflation, respectively.

Table 6.2) Top differentially methylated sites associated with COPD, lung function and hyperinflation. Only the top 20 DNA methylation sites are shown. FDR is the false discovery rate, and FC is the Fold Change.

Associated with	CpG id	Log ₂ (FC)	P-value	FDR
COPD	cg23800199	0.25	1.37E-09	1.11E-03
COPD	cg10730398	-0.32	1.83E-08	7.37E-03
COPD	cg01124573	-0.39	8.20E-08	1.80E-02
COPD	cg05748572	0.26	8.94E-08	1.80E-02
COPD	cg16560488	0.21	2.03E-07	3.27E-02
COPD	cg12081267	-0.50	2.92E-07	3.92E-02
FEV ₁ % pred	cg13832669	0.01	1.83E-08	1.26E-02
FEV ₁ % pred	cg10426234	-0.02	3.13E-08	1.26E-02
FEV ₁ % pred	cg05304658	0.01	1.02E-07	1.74E-02
FEV ₁ % pred	cg17653969	-0.01	1.03E-07	1.74E-02
FEV ₁ % pred	cg10089647	0.01	1.08E-07	1.74E-02
FEV ₁ % pred	cg25969765	0.01	1.45E-07	1.94E-02
FEV ₁ % pred	cg11070119	-0.01	1.70E-07	1.96E-02
FEV ₁ % pred	cg11865601	-0.01	2.06E-07	1.96E-02
FEV ₁ % pred	cg19438514	0.01	2.19E-07	1.96E-02
FEV ₁ % pred	cg23480676	-0.01	2.69E-07	2.17E-02
FEV ₁ % pred	cg17654622	0.01	3.14E-07	2.30E-02
FEV ₁ % pred	cg17179924	-0.01	3.93E-07	2.48E-02
FEV ₁ % pred	cg06445873	-0.01	4.30E-07	2.48E-02
FEV ₁ % pred	cg07853941	-0.02	4.17E-07	2.48E-02
FEV ₁ % pred	cg26105045	-0.01	5.97E-07	3.21E-02
FEV ₁ % pred	cg23237611	-0.02	8.12E-07	3.66E-02
FEV ₁ % pred	cg20809822	-0.01	8.10E-07	3.66E-02
FEV ₁ % pred	cg00313758	-0.02	8.77E-07	3.66E-02
FEV ₁ % pred	cg13944507	-0.01	8.69E-07	3.66E-02
FEV ₁ % pred	cg07494218	0.02	9.08E-07	3.66E-02
RV/TLC % pred	cg04574244	-0.01	4.89E-10	3.87E-04
RV/TLC % pred	cg25560032	0.01	9.61E-10	3.87E-04
RV/TLC % pred	cg02502243	-0.01	2.33E-09	6.27E-04
RV/TLC % pred	cg11720950	0.01	6.68E-09	1.19E-03
RV/TLC % pred	cg05731581	-0.01	7.36E-09	1.19E-03
RV/TLC % pred	cg07042914	0.01	1.20E-08	1.62E-03
RV/TLC % pred	cg18205376	-0.01	1.48E-08	1.67E-03
RV/TLC % pred	cg17563268	-0.01	1.65E-08	1.67E-03
RV/TLC % pred	cg11734183	0.01	1.95E-08	1.73E-03
RV/TLC % pred	cg13174293	-0.01	2.15E-08	1.73E-03
RV/TLC % pred	cg09690118	-0.01	3.33E-08	2.44E-03
RV/TLC % pred	cg17456749	-0.01	4.91E-08	3.30E-03
RV/TLC % pred	cg25964032	0.02	5.72E-08	3.55E-03
RV/TLC % pred	cg00262225	-0.01	6.58E-08	3.72E-03
RV/TLC % pred	cg07563944	-0.01	6.91E-08	3.72E-03
RV/TLC % pred	cg00379635	0.01	7.71E-08	3.89E-03
RV/TLC % pred	cg17187439	-0.01	9.94E-08	4.19E-03
RV/TLC % pred	cg14776114	0.01	1.03E-07	4.19E-03
RV/TLC % pred	cg06806678	0.01	1.04E-07	4.19E-03
RV/TLC % pred	cg25422753	-0.01	1.08E-07	4.19E-03

Validation of or results in two independent datasets

We set out to validate our findings in blood and lung tissue in two independent datasets. The six CpG sites differentially methylated in COPD compared to non-COPD did not validate in either of the cohorts. We then set out to validate the differentially methylated CpG sites associated with FEV₁ % pred. Unfortunately, no validation was found in either blood or lung tissue (Supplementary Table 6.1).

Association between DNA methylation and gene expression

To investigate the function of the CpG sites that we identified in the previous analyses, we related the methylation levels of these sites to gene expression of genes in proximity. For the six COPD-associated CpG sites, no significant eQTM were identified.

For the 25 CpGs associated with FEV₁ % pred in COPD, we found two significant eQTM for one CpG site. Cg06445873 showed hypermethylation with a lower FEV₁ % pred and a lower expression of *HSD11B2* and *TPPP3* (Table 6.3). For the 3954 CpG sites that were associated with RV/TLC % pred, we found 817 eQTM, of which two of the top CpGs, cg11734183 and cg13421038, were hypermethylated in association with higher RV/TLC % pred and lower expression of the genes *RGMB* and *SFSWAP*, respectively (Figure 4D, Table 6.3).

Table 6.3) Top 25 eQTM associated with lung function and hyperinflation. FDR is the false discovery rate, and FC is the Fold Change.

Associated with	Differential Methylation			eQTM		
	CpG id	Log ₂ (FC)	FDR	Gene	Statistic	FDR (eQTM)
FEV ₁ % pred	cg06445873	-0.01	2.48E-02	HSD11B2	-2.99	4.77E-02
FEV ₁ % pred	cg06445873	-0.01	2.48E-02	TPPP3	-2.95	4.77E-02
RV/TLC % pred	cg25273081	0.00	3.41E-02	CD163L1	6.21	1.42E-06
RV/TLC % pred	cg00604356	-0.01	4.57E-02	PIK3CG	-5.55	2.91E-06
RV/TLC % pred	cg05997203	-0.01	4.90E-02	SNX31	5.49	4.83E-06
RV/TLC % pred	cg07369741	0.01	3.92E-02	CLEC7A	5.41	1.12E-05
RV/TLC % pred	cg14096569	0.00	4.94E-02	HLA-DPA1	5.55	2.07E-05
RV/TLC % pred	cg25273081	0.00	3.41E-02	C1RL	5.20	2.69E-05
RV/TLC % pred	cg09914480	-0.01	1.34E-02	SAA2	-5.28	2.78E-05
RV/TLC % pred	cg26236507	0.01	4.60E-02	PHYHD1	-5.33	3.91E-05
RV/TLC % pred	cg18696576	-0.01	1.87E-02	NUDT3	5.01	4.00E-05
RV/TLC % pred	cg08248202	-0.01	3.92E-02	LCP2	-4.79	7.04E-05
RV/TLC % pred	cg08248202	-0.01	3.92E-02	DOCK2	-4.53	8.79E-05
RV/TLC % pred	cg05000029	-0.00	4.55E-02	CFI	-4.78	8.84E-05
RV/TLC % pred	cg09574909	-0.01	3.88E-02	PTPRC	-4.46	8.89E-05
RV/TLC % pred	cg17139403	-0.01	1.18E-02	IRAK3	-4.73	1.04E-04
RV/TLC % pred	cg21986821	0.01	3.62E-02	TNFSF14	4.80	1.22E-04
RV/TLC % pred	cg25273081	0.00	3.41E-02	C1R	4.63	1.30E-04
RV/TLC % pred	cg27613455	-0.01	4.32E-02	IKZF3	-4.90	1.56E-04
RV/TLC % pred	cg06844181	-0.01	2.77E-02	ARHGEF19	4.80	1.78E-04
RV/TLC % pred	cg14236470	-0.00	4.34E-02	LGR6	4.74	2.38E-04
RV/TLC % pred	cg17232592	0.01	3.06E-02	GBP4	4.57	2.44E-04
RV/TLC % pred	cg05468843	-0.01	2.20E-02	AMICA1	-4.63	2.77E-04
RV/TLC % pred	cg08498833	0.01	4.72E-02	SLA	4.40	3.78E-04
RV/TLC % pred	cg25384897	-0.01	4.34E-02	HLA-DPA1	-4.62	4.04E-04

Association of methylation differences and eQTMs with cell types

To investigate if methylation differences could be related to altered cell proportions, we performed cellular deconvolution on bronchial biopsies. Goblet cells, cycling basal cells, ciliated cells, and fibroblasts had the highest predicted cell proportions in the airway wall biopsies, each having more than 2% of the total predicted cell population and therefore were included in the subsequent analyses (Figure 6.4A). First, we assessed whether cell proportions differed between COPD and non-COPD and then whether they were associated with FEV₁ % pred and RV/TLC % pred within COPD. Goblet cells, cycling basal cells, and ciliated cells showed a significant difference in cell proportion between COPD patients and non-COPD controls, being lower, higher, and lower in COPD, respectively (Figure 6.4B). A lower FEV₁ % pred was associated with lower ciliated and goblet cell proportions, whereas a higher RV/TLC % pred was associated with lower goblet cell proportions (Supplementary Table 6.2). We then associated the cell types with the gene expression levels of our eQTMs. We found that 56 out of 723 eQTMs were associated with cell types, possibly indicating an association of these eQTMs with a shift in cell populations (Figure 6.4C, Supplementary Table 6.3).

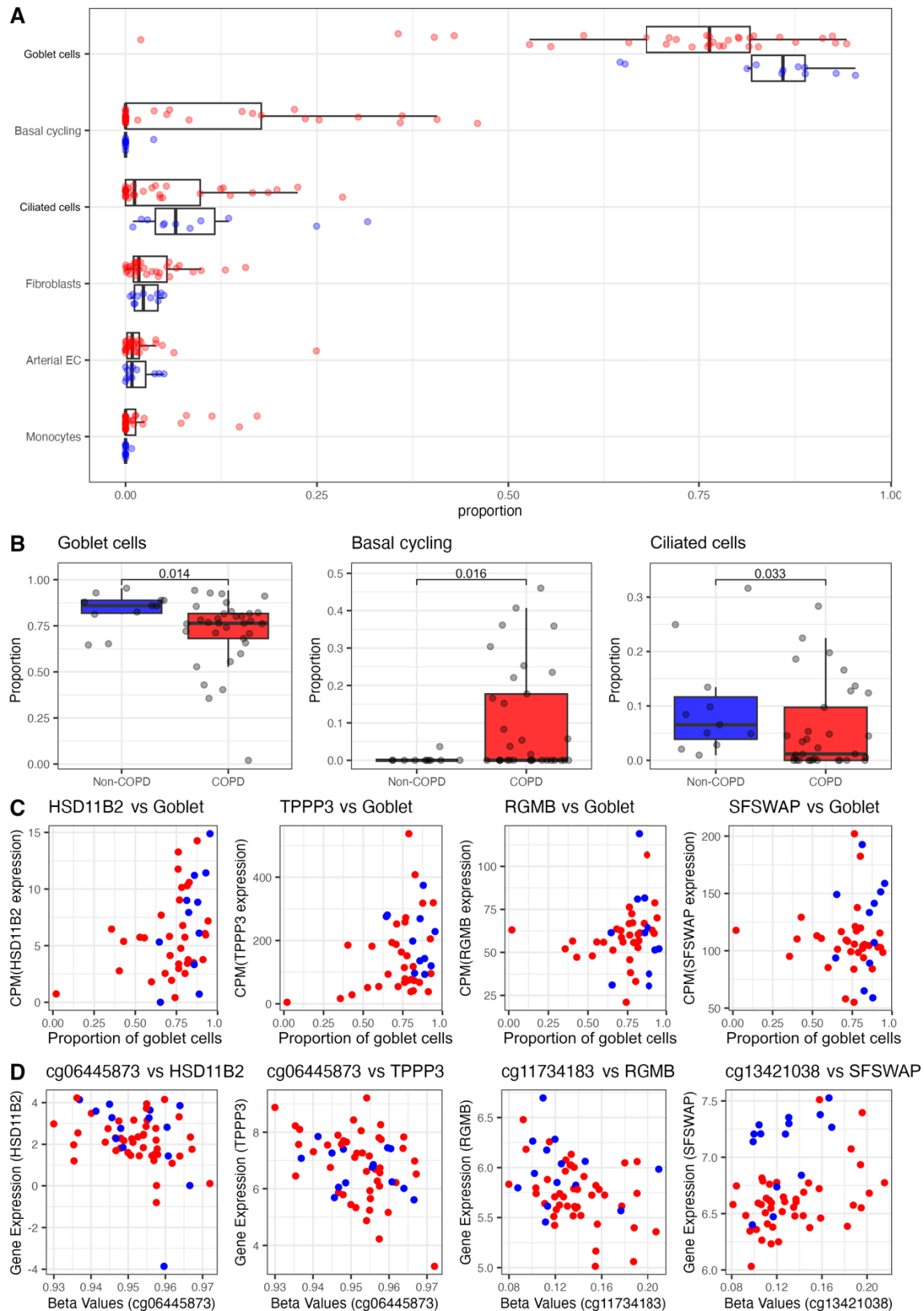


Figure 6.4) [A] Cell proportions calculated by cellular deconvolution stratified by COPD status for the cell types representing more than 1.5% of the total cell population. [B] Box plots for the three most abundant cell types, stratified by COPD status. [C] Scatter plots for the top eQTM associated with lung function and hyperinflation versus the goblet cell proportions. [D] Scatter plots showing the top two eQTM associated with lung function and hyperinflation.

Pathway analysis

To find the biological effects of the eQTM associated with RV/TLC % pred, we created a protein interaction network of the eQTM that were lower expressed with higher DNA methylation values. We found an enrichment of the hematopoietic cell lineage pathway as well as pathways involved in the immune system (Th1 and Th2 cell differentiation, type I diabetes mellitus, Th17 cell differentiation, cytokine-cytokine receptor interaction). Additionally, we found several pathways and cellular processes involved in (immune cell) signalling signal transduction (cytokine signalling, TCR signalling, and cell communication, signalling, and signal transduction) (Figure 6.5 and Table 6.4).

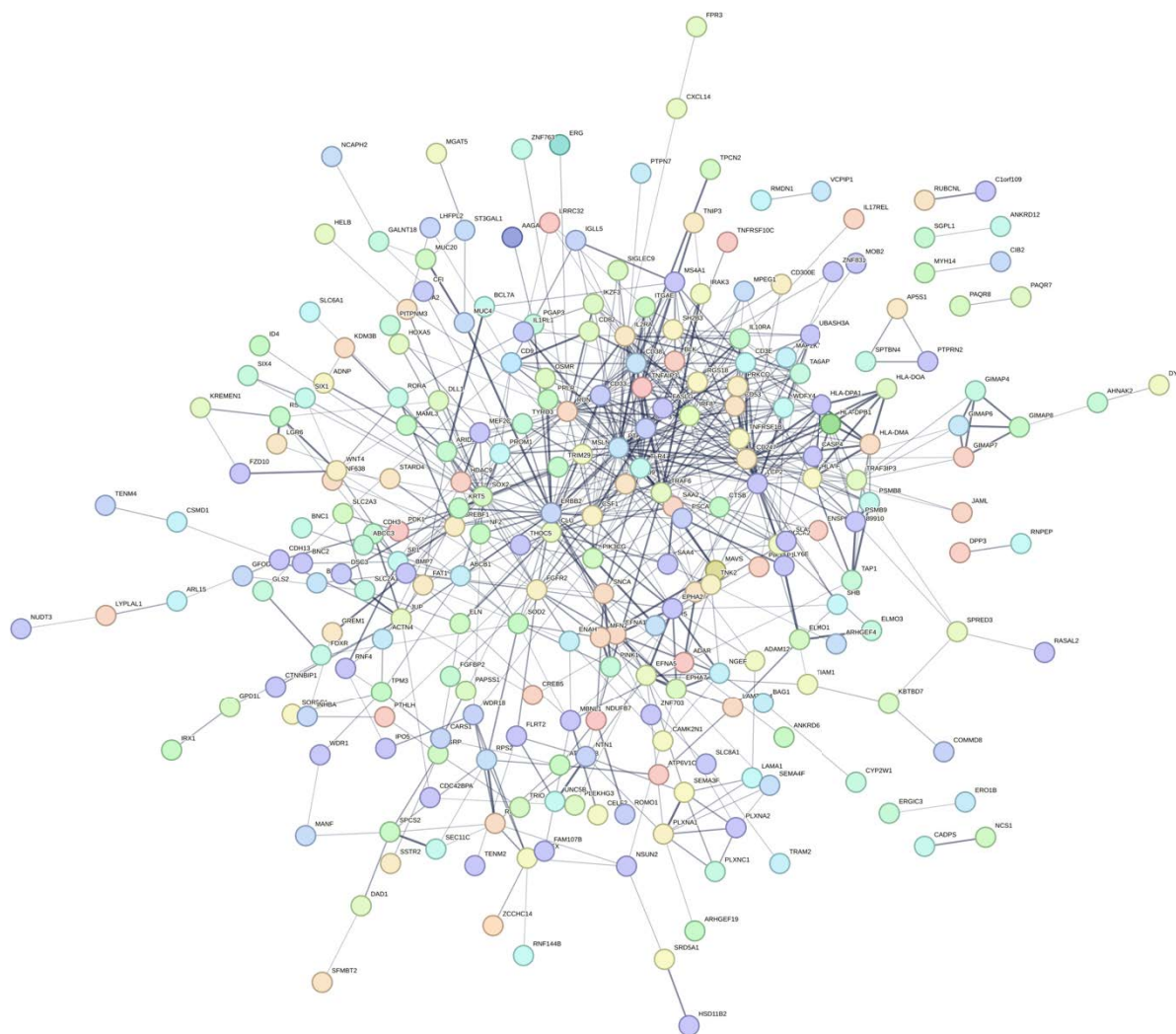


Figure 6.5) Protein interaction network using String-DB of all eQTM associated with RV/TLC % pred.

Table 6.4) Enriched pathways associated with eQTM and methylation in hyperinflation. Tested databanks were KEGG, Reactome and Gene Ontology (Molecular functions, Cellular Components and Biological Processes). The top 25 enriched pathways are shown. FDR is the false discovery rate, and FC is the Fold Change.

Group	ID	Term description	Strength	FDR
Process	GO:0007166	Cell surface receptor signaling pathway	0.38	1.98E-10
Process	GO:0007154	Cell communication	0.22	3.63E-08
Process	GO:0023052	Signaling	0.22	3.63E-08
Process	GO:0007165	Signal transduction	0.22	1.51E-07
Process	GO:0030155	Regulation of cell adhesion	0.5	1.51E-07
Process	GO:0051239	Regulation of multicellular organismal process	0.3	1.51E-07
Process	GO:0002376	Immune system process	0.33	1.88E-07
Process	GO:0050896	Response to stimulus	0.16	2.46E-07
Process	GO:0048583	Regulation of response to stimulus	0.24	4.67E-07
Process	GO:0048518	Positive regulation of biological process	0.18	1.02E-06
Process	GO:0009605	Response to external stimulus	0.3	1.22E-06
Component	GO:0009986	Cell surface	0.45	1.39E-06
Process	GO:0048522	Positive regulation of cellular process	0.19	1.45E-06
Process	GO:0051716	Cellular response to stimulus	0.17	1.45E-06
Process	GO:2000026	Regulation of multicellular organismal development	0.37	1.92E-06
Process	GO:0042127	Regulation of cell population proliferation	0.34	2.29E-06
Process	GO:0045321	Leukocyte activation	0.53	2.29E-06
Process	GO:0050793	Regulation of developmental process	0.29	2.29E-06
Process	GO:0048534	Hematopoietic or lymphoid organ development	0.48	2.45E-06
Process	GO:0002520	Immune system development	0.47	3.54E-06
Process	GO:0030097	Hemopoiesis	0.49	3.68E-06
Process	GO:0050865	Regulation of cell activation	0.49	3.92E-06
Process	GO:0002694	Regulation of leukocyte activation	0.51	4.67E-06
Process	GO:0045595	Regulation of cell differentiation	0.34	7.58E-06
Process	GO:0007411	Axon guidance	0.69	1.23E-05

6.5 – Discussion

In the current study, we show that changes in DNA methylation are associated with the presence of COPD and the severity of the disease, as reflected by the level of airflow obstruction and hyperinflation. COPD-associated CpG sites were mostly independent of cellular composition in the airway wall and were found to regulate genes reflecting biologically relevant inflammatory pathways. Our findings provide evidence of epigenetic dysregulation in the airway wall in COPD that is present to a greater extent with more advanced disease.

We found 6 CpG sites to be differentially methylated when comparing patients with COPD to non-COPD controls. Unfortunately, we were unable to replicate our findings due to the uniqueness of our study with available genome-wide DNA methylation and gene expression data in bronchoscopically derived airway wall biopsies from a large cohort of subjects with and without COPD. When investigating our COPD-associated CpG sites in blood and lung tissue in two independent datasets, we did not find any of the CpG sites associated with the presence of COPD or severity of airflow obstruction, suggesting the COPD-associated changes in the airway wall did not extend to the other compartments, i.e. blood or lung tissue. Five of the CpG sites we observed to be differentially methylated in COPD were located in intronic regions. Interestingly, we found cg01124573 to be hypomethylated in COPD, a CpG site located near *ADGRL3*, a gene involved in the contraction of airway smooth muscle²⁶. Further, the CpG site cg16560488 was hypermethylated in COPD and located in *JMJD1C*, a histone H3K9 demethylase gene²⁷. This finding is of interest as a previous

GWAS study showed the SNP rs7899503, for *JMJD1C*, to be associated with the level of FEV₁ in a multi-ethnic meta-analysis²⁸. These findings may suggest a role for histone modification, often a result of cigarette smoke exposure and high oxidant levels²⁹, to play a role in COPD pathogenesis.

We found several DNA methylation changes associated with COPD severity as reflected by the severity of airflow obstruction or hyperinflation. We found cg06445873 hypermethylation to be associated with a lower FEV₁ % pred, as well as lower *HSD11B2* gene expression. *HSD11B2* enzymatic activity was previously found to convert cortisol into inactive cortisone through increasing 11-beta-dehydrogenase activity^{30, 31}. This activity is not specific to cortisol but can influence the inactivation of inhaled corticosteroids used by COPD patients as well³², though evidence for this metabolic activity of synthetic-corticosteroids is conflicting³³. Possibly, a higher efficacy of ICS due to hypermethylation of cg06445873 could contribute to less airflow obstruction in COPD patients.

Additionally, *SFSWAP*, hypermethylated at cg13421038 and less expressed with higher RV/TLC % pred, regulates splicing and has been associated with regeneration and cellular differentiation³⁴. The latter is of interest given the chronic damage and repair responses contributing to COPD pathogenesis. Possibly, differential regulation of splicing due to *SFSWAP* hypermethylation could impact the functionality of genes and proteins involved in COPD pathology.

Finally, we found genes associated with differential methylation in association with more severe hyperinflation to be involved in biological pathways relevant to COPD, such as the regulation of immune responses and cell signalling.

Cellular deconvolution revealed the presence and severity of COPD to be associated with lower predicted goblet cell proportions. This was an unexpected finding, as goblet cells have been shown to be increased in COPD. As cellular deconvolution is a prediction based on transcriptomics data and comprises predicted cell type proportions that are dependent on each other, it is important to consider changes in other cell types as well. Possibly, the shift in goblet cell proportion we observed does not reflect lower absolute numbers in COPD but rather may be related to a relative increase in basal cycling cells in response to the repair of damaged epithelium in COPD. Extensive histological assessment would be needed to confirm the actual changes in these cell type proportions in relation to COPD.

There are some limitations in the current study. The current analysis only focuses on smokers. This was done to prevent the influence of smoking, which is known to have a substantial effect on methylation. Additionally, one of the major points was the fact that although bronchial biopsies collected for methylation and gene expression were collected on the same day, they were serial biopsies collected from adjacent locations within the bronchial tree. This could potentially affect the cellular composition and, subsequently, the concordance between methylation and gene expression, potentially limiting our ability to identify eQTMs. Nevertheless, we did find many eQTMs for the CpG sites associated with FEV₁ % pred and RV/TLC % pred. Lastly, the validation was performed in a blood and a lung tissue cohort, which may explain why we could not validate our findings. Moreover, the LifeLines (i.e. blood) cohort is a large cohort based on the general population. In this cohort, there are less severe cases of COPD, which may affect the statistical power of this analysis.

In conclusion, we described how altered DNA methylation may contribute to the presence and severity of COPD. We identified CpG sites that regulate the expression of genes involved in airway smooth muscle contraction, corticosteroid metabolism, and regeneration and cellular differentiation. We also identified hyperinflation-associated CpG sites that regulate genes involved

in the immune system and signalling signal transduction pathways, all important processes involved in COPD, thus supporting a role for differential methylation in airway changes in COPD.

6.6 – Funding

Funded by the PPP-allowance by Health Holland (LSH) and the UMCG. This collaboration project is co-financed by the Ministry of Economic Affairs and Climate Policy by means of the PPP-allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships

6.7 – Supplementary Material

Tables

Supplementary Table 6.1) Validation of the CpG sites associated with COPD and lung function in the Lifelines and COPD-Sp cohorts.

Associated with	CpG id	GLUCOLD, Stop-Smoking		LifeLines Blood		COPD-Sp Blood	
		FDR	Log ₂ (FC)	FDR	Estimate	FDR	Log ₂ (FC)
COPD	cg23800199	1.11E-03	0.25	NA	NA	NA	NA
COPD	cg10730398	7.37E-03	-0.32	1.00E+00	599.00	NA	NA
COPD	cg01124573	1.80E-02	-0.39	NA	NA	3.63E-01	0.01
COPD	cg05748572	1.80E-02	0.26	NA	NA	NA	NA
COPD	cg16560488	3.27E-02	0.21	NA	NA	NA	NA
COPD	cg12081267	3.92E-02	-0.50	1.00E+00	599.00	NA	NA
FEV ₁ % Pred	cg13832669	1.26E-02	0.01	5.98E-01	0.00	NA	NA
FEV ₁ % Pred	cg10426234	1.26E-02	-0.02	3.36E-01	-0.00	8.37E-01	-0.00
FEV ₁ % Pred	cg05304658	1.74E-02	0.01	NA	NA	2.35E-01	-0.00
FEV ₁ % Pred	cg17653969	1.74E-02	-0.01	5.74E-01	0.00	9.29E-01	-0.00
FEV ₁ % Pred	cg10089647	1.74E-02	0.01	5.74E-01	-0.00	4.50E-01	0.00
FEV ₁ % Pred	cg25969765	1.94E-02	0.01	NA	NA	NA	NA
FEV ₁ % Pred	cg11070119	1.96E-02	-0.01	5.98E-01	-0.00	9.45E-01	0.00
FEV ₁ % Pred	cg11865601	1.96E-02	-0.01	5.74E-01	-0.00	8.11E-01	0.00
FEV ₁ % Pred	cg19438514	1.96E-02	0.01	5.74E-01	-0.00	8.47E-01	0.00
FEV ₁ % Pred	cg23480676	2.17E-02	-0.01	3.36E-01	0.00	4.78E-01	0.00
FEV ₁ % Pred	cg17654622	2.30E-02	0.01	NA	NA	2.24E-01	0.00
FEV ₁ % Pred	cg17179924	2.48E-02	-0.01	5.98E-01	-0.00	8.82E-01	0.00
FEV ₁ % Pred	cg06445873	2.48E-02	-0.01	5.74E-01	0.00	9.63E-01	-0.00
FEV ₁ % Pred	cg07853941	2.48E-02	-0.02	NA	NA	NA	NA
FEV ₁ % Pred	cg26105045	3.21E-02	-0.01	5.74E-01	-0.00	NA	NA
FEV ₁ % Pred	cg23237611	3.66E-02	-0.02	6.57E-01	-0.00	7.41E-01	0.00
FEV ₁ % Pred	cg20809822	3.66E-02	-0.01	NA	NA	NA	NA
FEV ₁ % Pred	cg00313758	3.66E-02	-0.02	NA	NA	3.66E-02	-0.01
FEV ₁ % Pred	cg13944507	3.66E-02	-0.01	5.74E-01	-0.00	3.45E-01	0.00
FEV ₁ % Pred	cg07494218	3.66E-02	0.02	5.74E-01	0.00	7.84E-01	-0.00
FEV ₁ % Pred	cg05261028	3.81E-02	0.01	5.74E-01	-0.00	5.28E-01	0.00
FEV ₁ % Pred	cg04219507	4.08E-02	-0.01	NA	NA	6.82E-01	-0.00
FEV ₁ % Pred	cg10166610	4.46E-02	-0.01	NA	NA	2.28E-01	0.00
FEV ₁ % Pred	cg18930361	4.46E-02	0.01	NA	NA	7.02E-01	0.00
FEV ₁ % Pred	cg23531488	4.50E-02	-0.01	3.36E-01	-0.00	9.20E-01	-0.00

Supplementary Table 6.2) Associations between lung function measurements and cell proportions.

Cell type	Spearman correlation test	
	FEV ₁ % pred	RV/TLC % pred
Arterial EC	4.87E-01	3.48E-01
Basal	4.18E-01	1.90E-01

Cell type	Spearman correlation test	
	FEV ₁ % pred	RV/TLC % pred
Basal cycling	3.40E-02	3.70E-01
Ciliated	6.78E-02	1.03E-01
Dendritic cells	9.11E-01	4.24E-01
Fibroblasts	8.94E-01	7.00E-01
Goblet	1.44E-03	2.35E-02
Monocytes	2.77E-02	2.59E-02
Submucosal Secretory	2.15E-01	8.70E-02

Supplementary Table 6.3) Associations of the eQTM (associated with hyperinflation) with cell proportions.

CpG id	Gene	Differential Methylation		eQTM		Gene-Cell type spearman association	
		Log ₂ (FC)	FDR	Statistic	FDR (eQTM)	Rho	Genewise FDR
cg13718141	SHB	-0.01	4.97E-03	-2.77	1.94E-02	0.44	2.50E-03
cg22022910	MED16	0.01	7.04E-03	-4.07	1.09E-03	0.49	6.71E-04
cg22022910	STK11	0.01	7.04E-03	-3.09	1.28E-02	0.45	2.29E-03
cg05203855	MAP2K7	-0.00	1.58E-02	3.04	4.01E-02	0.52	3.33E-04
cg26187339	DLK2	0.01	1.55E-02	-3.17	3.33E-02	0.50	5.20E-04
cg00021532	ADCY9	0.01	1.64E-02	-4.07	1.93E-03	0.50	6.10E-04
cg21082130	PTPRN2	-0.01	1.69E-02	2.35	2.26E-02	0.64	3.24E-06
cg27232494	RAI1	0.01	2.30E-02	-2.97	1.49E-02	0.50	5.90E-04
cg16664394	SKIV2L2	-0.01	2.32E-02	3.38	1.37E-02	-0.49	6.87E-04
cg23533824	ZNF446	0.01	2.30E-02	-2.70	3.57E-02	0.49	6.99E-04
cg15015194	PINK1	-0.00	2.51E-02	3.44	8.05E-03	0.46	1.90E-03
cg10420435	SIPA1L3	-0.01	2.53E-02	3.00	3.78E-02	0.43	3.41E-03
cg19471591	FCGRT	0.01	2.55E-02	-3.52	1.66E-02	0.62	8.62E-06
cg19471591	PTOV1	0.01	2.55E-02	-2.72	4.76E-02	0.47	1.27E-03
cg05726450	NTN1	-0.01	2.63E-02	3.08	1.32E-02	0.50	5.38E-04
cg16968650	KLHL21	0.01	2.69E-02	-3.18	1.48E-02	0.48	1.05E-03
cg11905574	PLXNA1	-0.01	2.79E-02	3.10	6.29E-03	0.54	1.34E-04
cg27263395	SPPL2B	0.00	2.88E-02	-3.46	1.62E-02	0.45	2.35E-03
cg05956890	SP1	-0.00	2.95E-02	-3.34	2.46E-02	-0.51	4.03E-04
cg21079272	RAI1	-0.01	2.99E-02	2.62	2.70E-02	0.50	5.90E-04
cg05409612	PRDM11	-0.01	2.96E-02	2.67	2.03E-02	0.46	1.79E-03
cg07919775	MED16	0.01	3.06E-02	-2.82	3.18E-02	0.49	6.71E-04
cg15962136	CD82	-0.01	3.06E-02	2.37	3.20E-02	0.52	3.03E-04
cg11685767	IBTK	-0.00	3.38E-02	-2.28	2.68E-02	-0.44	3.10E-03
cg14030728	TMEM8B	0.00	3.39E-02	-3.30	3.19E-02	0.54	1.40E-04
cg24781461	MEGF6	0.00	3.44E-02	-3.83	1.93E-03	0.59	2.89E-05
cg24781461	ARHGEF16	0.00	3.44E-02	-3.26	7.23E-03	0.51	3.78E-04
cg24251800	NGFR	0.01	3.56E-02	-2.49	4.83E-02	0.43	3.84E-03
cg02110048	EPT1	0.01	3.56E-02	-2.68	4.97E-02	-0.47	1.21E-03
cg21986821	C3	0.01	3.62E-02	3.19	1.10E-02	0.46	1.57E-03
cg27531927	IPO5	-0.01	3.74E-02	3.60	1.43E-03	-0.45	2.40E-03
cg26181864	SEMA4F	-0.01	3.68E-02	-3.31	1.90E-02	0.43	3.66E-03
cg12009719	ACACB	0.01	3.74E-02	-2.92	3.06E-02	0.45	2.30E-03
cg17282301	THRA	0.01	3.79E-02	-3.06	2.88E-02	0.55	1.11E-04
cg13363708	MOB2	-0.00	3.88E-02	2.95	4.33E-02	0.44	2.67E-03
cg26358677	C3	0.01	3.95E-02	2.88	2.59E-02	0.46	1.57E-03
cg18711369	PGAP3	0.01	3.95E-02	-2.99	1.98E-02	0.52	2.95E-04
cg15983585	ZBTB42	0.01	4.03E-02	-2.76	3.98E-02	0.43	3.96E-03
cg02943188	RPUSD1	0.00	4.10E-02	-2.79	4.80E-02	0.51	4.43E-04

CpG id	Gene	Differential Methylation		eQTM		Gene-Cell type spearman association	
		Log ₂ (FC)	FDR	Statistic	FDR (eQTM)	Rho	Genewise FDR
cg24311376	DOK7	0.01	4.17E-02	-3.17	7.04E-03	0.48	9.45E-04
cg17237830	SPTBN4	-0.00	4.23E-02	3.59	1.16E-02	0.44	2.62E-03
cg07478324	POMT1	0.00	4.20E-02	-3.24	1.05E-02	0.48	8.62E-04
cg24430642	PANX2	0.01	4.27E-02	-4.40	9.66E-04	0.45	2.02E-03
cg24430642	ALG12	0.01	4.27E-02	-2.98	1.44E-02	0.51	3.49E-04
cg24430642	TUBGCP6	0.01	4.27E-02	-2.96	1.44E-02	0.50	6.13E-04
cg24430642	TRABD	0.01	4.27E-02	-2.78	1.94E-02	0.52	3.00E-04
cg24430642	MAPK11	0.01	4.27E-02	-2.66	2.08E-02	0.53	2.40E-04
cg05480110	TNK2	-0.01	4.27E-02	3.25	1.01E-02	0.55	1.03E-04
cg07905442	ZNF703	-0.01	4.33E-02	3.07	3.02E-02	0.50	5.50E-04
cg27613455	PGAP3	-0.01	4.32E-02	2.99	3.35E-02	0.52	2.95E-04
cg25841279	SCAND1	0.01	4.44E-02	-3.02	1.36E-02	0.45	2.35E-03
cg08498542	FRMD4B	-0.01	4.44E-02	2.99	2.98E-02	-0.49	6.57E-04
cg13457900	TNFRSF1B	-0.01	4.60E-02	-3.12	1.45E-02	0.43	3.22E-03
cg26236507	ZER1	0.01	4.60E-02	-3.01	3.63E-02	0.45	2.41E-03
cg06618764	LFNG	0.00	4.94E-02	-4.31	5.09E-04	0.51	4.70E-04
cg06618764	BRAT1	0.00	4.94E-02	-2.59	4.38E-02	0.45	2.31E-03

6.8 – References

1. Machin, M., et al., *Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies*. BMC Pulm Med, 2017. **17**(1): p. 54.
2. Lopez-Campos, J.L., W. Tan, and J.B. Soriano, *Global burden of COPD*. Respirology, 2016. **21**(1): p. 14-23.
3. van Nijnatten, J., et al., *A bronchial gene signature specific for severe COPD that is retained in the nose*. ERJ Open Research, 2023.
4. Boudewijn, I.M., et al., *Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression*. Respiratory research, 2017. **18**(1): p. 1-10.
5. Faiz, A., et al., *Th2 high and mast cell gene signatures are associated with corticosteroid sensitivity in COPD*. thorax, 2023. **78**(4): p. 335-343.
6. Ziller, M.J., et al., *Charting a dynamic DNA methylation landscape of the human genome*. Nature, 2013. **500**(7463): p. 477-81.
7. Joehanes, R., et al., *Epigenetic Signatures of Cigarette Smoking*. Circ Cardiovasc Genet, 2016. **9**(5): p. 436-447.
8. Sundar, I.K., et al., *DNA methylation profiling in peripheral lung tissues of smokers and patients with COPD*. Clin Epigenetics, 2017. **9**: p. 38.
9. Lapperre, T.S., et al., *Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: a randomized trial*. Ann Intern Med, 2009. **151**(8): p. 517-27.
10. Willemse, B.W., et al., *Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers*. Eur Respir J, 2005. **26**(5): p. 835-45.
11. Rathnayake, S.N.H., et al., *Smoking induces shifts in cellular composition and transcriptome within the bronchial mucus barrier*. Respirology, 2023. **28**(2): p. 132-142.
12. Rathnayake, S.N.H., et al., *Longitudinal Effects of 1-Year Smoking Cessation on Human Bronchial Epithelial Transcriptome*. Chest, 2023. **164**(1): p. 85-89.
13. Tasena, H., et al., *microRNA-mRNA regulatory networks underlying chronic mucus hypersecretion in COPD*. Eur Respir J, 2018. **52**(3).
14. Slob, E.M.A., et al., *Association of bronchial steroid inducible methylation quantitative trait loci with asthma and chronic obstructive pulmonary disease treatment response*. Clin Transl Allergy, 2022. **12**(8): p. e12173.
15. Vermeulen, C.J., et al., *Differential DNA methylation in bronchial biopsies between persistent asthma and asthma in remission*. Eur Respir J, 2020. **55**(2).
16. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. Bioinformatics, 2014. **30**(10): p. 1363-9.
17. Pidsley, R., et al., *A data-driven approach to preprocessing Illumina 450K methylation array data*. BMC Genomics, 2013. **14**: p. 293.
18. Zhang, Y., G. Parmigiani, and W.E. Johnson, *ComBat-seq: batch effect adjustment for RNA-seq count data*. NAR Genom Bioinform, 2020. **2**(3): p. lqaa078.
19. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
20. de Vries, M., et al., *From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function*. Respir Res, 2018. **19**(1): p. 212.
21. Casas-Recasens, S., et al., *Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity*. Am J Respir Crit Care Med, 2021. **203**(1): p. 129-134.
22. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments*. Bioinformatics, 2012. **28**(6): p. 882-3.
23. Aliee, H., et al., *Determinants of expression of SARS-CoV-2 entry-related genes in upper and lower airways*. Allergy, 2022. **77**(2): p. 690-694.

24. Donovan, C., et al., *Abnormal Expression Of Latrophilin's Contributes To Increased Airway Smooth Muscle Mass And Contraction In Asthmatic Airways*, in *C30. AIRWAY HYPERRESPONSIVENESS: DO WE HAVE A COMPREHENSIVE EXPLANATION?* p. A4012-A4012.
25. Song, J., et al., *Targeted epigenetic editing of SPDEF reduces mucus production in lung epithelial cells*. *Am J Physiol Lung Cell Mol Physiol*, 2017. **312**(3): p. L334-L347.
26. Faiz, A., et al., *Latrophilin receptors: novel bronchodilator targets in asthma*. *Thorax*, 2017. **72**(1): p. 74-82.
27. Nakajima, R., H. Okano, and T. Noce, *JMJD1C Exhibits Multiple Functions in Epigenetic Regulation during Spermatogenesis*. *PLoS One*, 2016. **11**(9): p. e0163466.
28. Wyss, A.B., et al., *Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function*. *Nat Commun*, 2018. **9**(1): p. 2976.
29. Sundar, I.K., H. Yao, and I. Rahman, *Oxidative stress and chromatin remodeling in chronic obstructive pulmonary disease and smoking-related diseases*. *Antioxid Redox Signal*, 2013. **18**(15): p. 1956-71.
30. Tomlinson, J.W. and P.M. Stewart, *Cortisol metabolism and the role of 11beta-hydroxysteroid dehydrogenase*. *Best Pract Res Clin Endocrinol Metab*, 2001. **15**(1): p. 61-78.
31. Orsida, B.E., Z.S. Krozowski, and E.H. Walters, *Clinical relevance of airway 11 β -hydroxysteroid dehydrogenase type II enzyme in asthma*. *American journal of respiratory and critical care medicine*, 2002. **165**(7): p. 1010-1014.
32. Feinstein, M.B. and R.P. Schleimer, *Regulation of the action of hydrocortisone in airway epithelial cells by 11beta-hydroxysteroid dehydrogenase*. *Am J Respir Cell Mol Biol*, 1999. **21**(3): p. 403-8.
33. Suzuki, S., et al., *Dexamethasone upregulates 11beta-hydroxysteroid dehydrogenase type 2 in BEAS-2B cells*. *Am J Respir Crit Care Med*, 2003. **167**(9): p. 1244-9.
34. Moayed, Y., et al., *The candidate splicing factor Sfsmap regulates growth and patterning of inner ear sensory organs*. *PLoS Genet*, 2014. **10**(1): p. e1004055.

Chapter 7 – ICS use in COPD changes DNA methylation and targets FKBP5 methylation and expression.

Jos van Nijnatten, Kathy Phung, Maarten van den Berge, Corry-Anke Brandsma, Wim Timens[†], Alen Faiz[†]

Status: Not published

7.1 – Abstract

Introduction: Chronic inflammation is a key hallmark in COPD, leading to airway damage and lung tissue destruction. Inhaled Corticosteroids (ICS) have broad anti-inflammatory effects, but their overall efficacy in COPD is limited, with only a subset responding favourably, and the remaining COPD patients being insensitive to corticosteroids. The mechanism underlying corticosteroid responsiveness or unresponsiveness is not fully understood. Epigenetics, such as DNA methylation changes, may play a role in this context, but this has not been properly investigated yet.

Aims: To assess how the use of ICS in COPD changes the level of DNA methylation in bronchial biopsies.

Methods: We utilised data from the Groningen and Leiden Universities Corticosteroids in Obstructive Lung Disease (GLUCOLD) study, which included moderate-to-severe COPD patients, of whom 26 were treated with fluticasone with or without added salmeterol for 30 months, 10 with placebo for 30 months and 15 with fluticasone for 6 months followed by placebo for the remaining 24 months of the study. Bronchial biopsies were obtained at baseline, and after 6 and 30 months of treatment, genome-wide DNA methylation (EPIC 850K array) and gene expression profiling were performed. Differential methylation analysis was conducted using a linear model, and gene set variation analysis (GSVA) and differential methylation analysis were used to assess the persistence of methylation changes after ICS withdrawal. We used a glucocorticoid receptor (GR) ChIP-Seq dataset to cross-reference if the differentially methylated sites could directly affect GR binding. Finally, we performed epigenetic editing in cultured epithelial cells to confirm the direct effects of a change in methylation in one CpG site that was both significantly affected by ICS and located in a GR binding site.

Results: Analysis of bronchial biopsy samples revealed 114,098 differentially methylated CpG sites after six months of ICS therapy, with a significant proportion of the top sites being hypomethylated. GSVA and differential methylation analysis indicated that 93.6% of methylation changes persist beyond the cessation of ICS therapy, suggesting long-term epigenetic effects of ICS treatment in COPD patients. 907 (0.79%) of the differentially methylated sites were in GR binding sites, including cg2341608 in the FKBP5 gene, a co-chaperone protein and known repressor of GR. Hypermethylation after epigenetic editing of cg2341608 resulted in lower expression of FKBP5, confirming the direct effect of DNA methylation at this site on the expression of FKBP5 in lung epithelial cells.

Conclusions: ICS treatment in COPD patients leads to widespread changes in DNA methylation, with a notable persistence of these changes even after treatment withdrawal. These findings highlight the long-term epigenetic modifications resulting from ICS therapy and provide insights into the mechanisms of corticosteroid insensitivity.

Keywords: COPD, Inhaled corticosteroids (ICS), Glucocorticoid receptor (GR), Corticosteroid insensitivity, Smoking-related COPD, Transcriptional response

7.2 – Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death globally¹⁻³, and the disease burden is expected to rise due to a growing and ageing population².

Chronic inflammation, one of the main pathophysiological changes in COPD, is triggered by smoking, the key risk factor for developing COPD. Prolonged smoking leads to repeated injury of the airway wall, which irritates and causes damage to airway epithelium. This then leads to the release of pro-inflammatory cytokines by the airway epithelium and smooth muscle cells^{4, 5}, activation of the innate immune and adaptive immune system. Immune cells, such as neutrophils and macrophages, can damage the lung's integrity by releasing proteases and reactive oxygen species (ROS) to clear the airway and parenchyma from the toxic particles in smoke. This can cause cell death, goblet cell hyperplasia, and mucus hypersecretion, obstructing the airways^{6, 7}.

Inhaled Corticosteroids (ICS) are known for their broad anti-inflammatory effects. Although they are the cornerstone of treatment in asthma, their role in COPD is less clear, with a considerable proportion of patients being corticosteroid insensitive. The underlying mechanisms of the corticosteroid insensitivity in COPD remain to be elucidated. Epigenetic regulation in the form of DNA methylation of key transcription factors, including the GR, has been proposed to play a role. The glucocorticoid receptor (GR) is a key regulator of the anti-inflammatory function of corticosteroids. The GR plays a dual role: 1) it promotes the expression of anti-inflammatory genes, including IL10, and 2) it inhibits the effect of pro-inflammatory transcription factors, the most important one being NF- κ B.

The DNA methylation profile in the peripheral blood of patients with COPD has been shown to differ from that of non-COPD controls⁸. These COPD-associated DNA methylation changes include locations where the GR binds to the DNA, possibly interfering with corticosteroids' efficacy. Although these findings help to understand the factors involved in developing corticosteroid insensitivity, studies have mainly investigated blood.

In this study, we aimed to shed light on the mechanisms behind corticosteroid insensitivity, potentially leading to more effective treatments for COPD. To this end, we investigated changes in DNA methylation in bronchial biopsies after treatment with ICS in COPD patients. The effect of these ICS-induced methylation changes on gene expression was subsequently assessed *in vitro*.

7.3 – Methods

Patient data

We used available DNA methylation and gene expression data of bronchial biopsies derived from 89 COPD patients included in the Groningen and Leiden Universities study on the role of Corticosteroids in Obstructive Lung Disease (GLUCOLDClinicalTrials.gov: NCT00158847). Briefly, GLUCOLD included subjects with COPD stage II or III without a history of asthma, aged 45 to 75 years, who were current or former smokers with at least ten pack-years and had not used ICS for at least six months prior to the study⁹. Subjects were randomly allocated to one of the following treatment arms: (1) fluticasone 500ug with or without added salmeterol b.i.d for 30 months, (2) fluticasone b.i.d for six months followed by placebo b.i.d for 24 months, or (3) placebo b.i.d for 30 months. Subjects were included if treatment adherence remained 70% or above. All subjects were extensively characterised, i.e., lung function, blood tests, and questionnaire data. In addition, a bronchoscopy was performed, during which bronchial biopsies were obtained at 0, 6 and 30 months¹⁰. Gene expression and methylation profiling were performed using the Affymetrix Human Gene 1.0 ST Array and Infinium Human

Methylation 850k Bead Chip Array, respectively¹¹. The local medical ethics committees approved the study, and all subjects gave written informed consent.

Statistical analysis

All analyses were done in R version v4.2.2. To identify differentially methylated CpG sites following ICS therapy, we used a linear model using limma (version 3.54.1), correcting for age, sex, smoking status and the principal components explaining 95% of the variability of the control probes, and paired samples were taken into account using blocking to compare each patient with their own sample pre- and post-treatment. We adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR), with a value of < 0.05 considered statistically significant. The statistical methods are shown in Figure 7.1.

We calculated the GSVA enrichment score (GSVA version 1.48.3) of the CpG sites differentially methylated after six months of treatment. Methylation sites that were hyper (increased) and hypo (decreased) methylated were analysed separately. Statistical comparisons for the GSVA analysis between treatment groups were conducted with unpaired t-tests, while comparisons between timepoints within treatment groups were completed with paired t-tests.

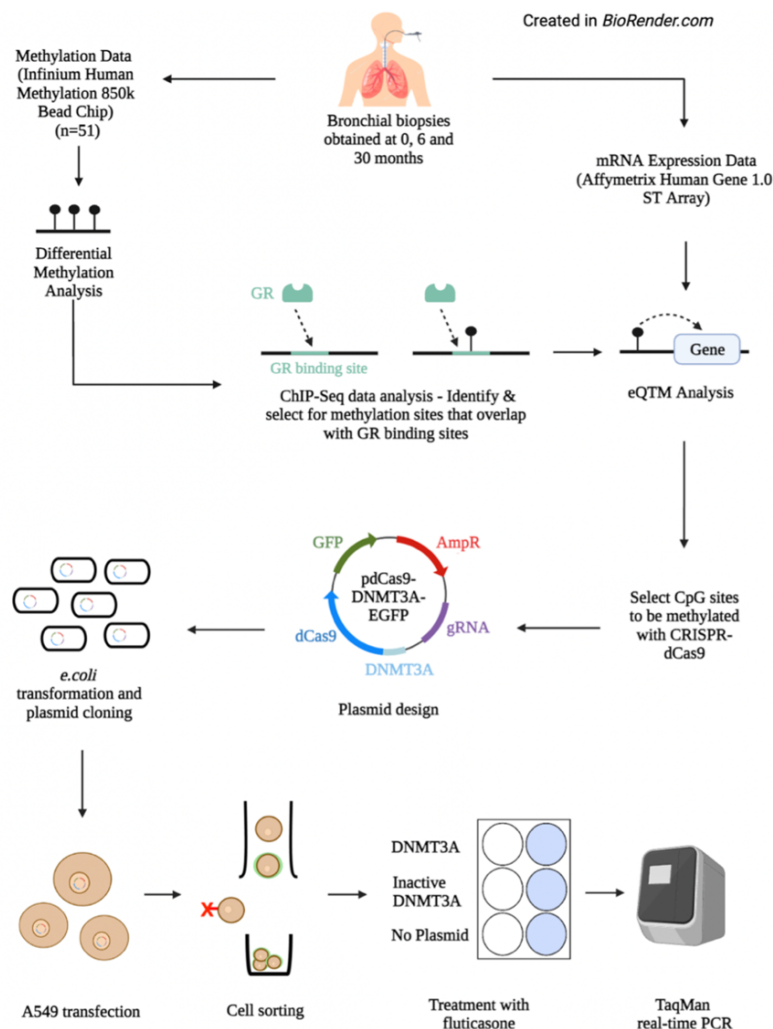


Figure 7.1) Outline of study methods. In the Groningen and Leiden Universities study of Corticosteroids in Obstructive Lung Disease (GLUCOLD), bronchial biopsies from moderate-to-severe COPD patients after 0,6 and 30 months of treatment. Differential DNA methylation analyses were performed and combined with ChIP-Seq and gene expression data¹¹. The part of plasmid design a549 transfection, etc, is not explained.

Chromatin Immunoprecipitation followed by DNA sequencing (ChIP-Seq) analysis.

We used a publicly available ChIP-Seq dataset on A549 lung carcinoma epithelial cells treated with dexamethasone or ethanol as a control to select methylation sites that are directly targeted by corticosteroids using ChIP-Seq¹².

Association of DNA methylation with gene expression and pathway analysis

To assess the downstream ICS-induced changes in DNA methylation on the expression of genes, we performed a cis-eQTM analysis (< 50.000bp) using Matrix EQTL (version 2.3), correcting for age, sex, and smoking status. This analysis was performed separately for all ICS-associated CpG sites and for those located in GR binding sites. We performed pathway analysis using g:Profiler on the eQTMs associated with GR receptor binding and separately on all genes associated with all differentially expressed DNA methylation sites.

CpG Site Selection and CRISPR Guide Design

As methylation editing via CRISPR-dCas9-DNMT3A peaks at about 25–30 bps downstream from the targeted PAM sequence¹³, we designed various CRISPR guides with a PAM sequence (NGG) within a section of ~150 bps flanking the target CpG site using the Guide Design function in Benchling (Table 7.1). CRISPR guides 15-35 bps apart on both sides from the target CpG sites, and the NGG sequences of the CRISPR guides were selected for experimental validation of the validity of the eQTMs of methylations sites found to be altered by ICS.

Table 7.1) Guide RNA sequences. Guide RNA sequences for targeted methylation with CRISPR-dCas9-DNMT3A were designed using the Guide Design function in Benchling. Guides with PAM sequences 25-30 bps from the targeted CpG site or as close as possible to this range were selected. PAM: Protospacer adjacent motif, NGG: PAM sequence 5'-NGG-3', where N is any nucleotide, FKBP5: FK506 binding protein 5, PDK4: Pyruvate dehydrogenase lipoamide kinase isozyme 4. Distance to CpG site: Distance between PAM (NGG) sequence and targeted CpG

Target CpG Site	Target Gene	Guide RNA Sequence	On Target Score	Distance to CpG site
cg2341608	FKBP5	aaacTGTGTTTCATATGGCCTCTGC CACCGCAGAGGCCATATGAAACACA	70.30%	17 bps
cg2341608	FKBP5	aaacCTGTCCAGTCCTACACTGGC CACCGCCAGTGTAGGACTGGACAG	72.90%	34 bps
cg2645126	PDK4	aaacCAGGCCAGCTGCTTCTCTCTC CACCGAGGAAGAAGCAGCTGGCCTG	62.80%	27 bps

Plasmid Design and Transformation

Guide RNA oligonucleotides designed in Benchling were purchased from Integrated DNA Technologies, using PX458 as the selected vector backbone. The gRNAs were inserted into pdCas9-DNMT3A-EGFP vectors through restriction enzyme digestion with BbsI and T7 DNA Ligase, with each plasmid containing genes for the expression of dCas9, DNMT3A, enhanced green fluorescent protein (EGFP) and ampicillin-resistance (AmpR). pdCas9-DNMT3A-EGFP (ANV) vectors containing DNMT3A with inactive catalytic domains were used to create control plasmids, with cells expected to remain unedited. The ligation reaction was treated with PlasmidSafe exonuclease to remove unwanted recombination products. The plasmids were transformed into Escherichia coli using 42°C heat shock, then grown in ampicillin-positive broth to select for plasmid-containing bacteria. Plasmid DNA was subsequently isolated with the ISOLATE II Bioline Plasmid Mini Kit and used to transfect A549 cells.

Cell Transfection, Sorting and Corticosteroid Treatment and Real-Time PCR

In preparation for transfection, A549 cells were seeded in a 12-well plate at 80,000 cells per well. At 50% confluence, the A549 cells were transfected with the plasmid DNA using Roche X-tremeGENE HP DNA Transfection Reagent, supplemented with Opti-MEM. For each gene, one well was transfected with plasmids containing active DNMT3A (*experimental*), a second well with plasmids containing inactivate DNMT3A (*plasmid control*) and a third well was not transfected (*reagent control*). After 4 hours in Opti-MEM, cells were grown in culture media for 48 hours to prepare for cell sorting. For each condition (experimental, plasmid control and reagent control), all GFP-positive cells were sorted using the BD Biosciences FACSymphonyS6 Cell Sorter. Flow cytometry parameters were configured using the FACSDiva 9.1.2 software. The sorted cells for each condition were then split across two wells of a 24-well plate. Cells were grown to 100% confluence, then washed with Hanks and quiesced for 24 hours. Out of the two wells for each condition, one was treated with 10nM fluticasone, and one was treated with 10nM DMSO as a control. After 8 hours, RNA was extracted using the Bioline Isolate RNA II kit and 500 ng was used to synthesise cDNA. We used a combination of equal concentrations oligo (dT) primers and random hexamer primers from the Bioline Tetro cDNA Synthesis Kit. RNA was extracted using the Bioline Isolate RNA II kit, and cDNA was synthesised using a combination of equal concentrations of oligo (dT) primers and random hexamer primers (Bioline Tetro cDNA Synthesis Kit). Duplex Taqman real-time PCR was performed to quantify gene expression (ThermoFisher Scientific) (Details in supplementary methods).

7.4 – Results

Clinical Characteristics

Good quality DNA methylation and clinical data were available from 51 patients included in GLUCOLD, 26 undergoing treatment with fluticasone with or without salmeterol for 30 months (n=26), 15 undergoing treatment with ICS followed by placebo for 24 months, 10 undergoing treatment with placebo. Clinical characteristics are presented in Table 7.2.

Table 7.2) Patients demographics. ¹ Mean (SD), ² Median (range), ³ Post-bronchodilator

	Fluticasone 30 months (n=26)	Fluticasone 6 months (n=15)	Placebo (n=10)
Age (Years) ¹	61.1 (6.5)	61.5 (7.9)	60.9 (7.4)
Male (%)	96.2	80.0	90.0
Current Smoking (%)	60.7	53.3	60.0
Packyears ²	48.5 (15.8 - 67.4)	40.8 (19.5 - 82.7)	40.8 (26.1 - 54.5)
FEV ₁ % pred ^{2, 3}	64.6 (40.8 - 76.4)	67.2 (51.9 - 77.7)	61.0 (46.1 - 72.1)
FEV ₁ /FVC [#] % ²	66.1 (43.1 - 86.3)	71.4 (54.6 - 91.5)	63.1 (49.3 - 82.2)
RV/TLC [#] (%) of predicted ²	128.5 (91.6 - 159.0)	125.6 (82.7 - 150.4)	123.3 (97.4 - 139.0)

Changes in DNA methylation after 6 months of ICS-containing treatment

We identified 114,098 differentially methylated CpG sites following six months of ICS therapy (FDR < 0.05) (Figure 7.2A&B), compared to baseline. 42% of these were hypomethylated, while the remainder were hypermethylated. The top 20 differentially methylated sites are shown in Table 7.3, and the top two CpG sites are shown in Figure 7.2C&D. We next conducted an additional analysis to investigate whether ICS-induced CpG sites may have been due to cellular composition. To this end, we adjusted for changes in plasma cells, mast cells, and CD3⁺, CD4⁺ and CD8⁺ T cells as previously

measured in the biopsies¹⁴. This analysis showed that 88% of the hypomethylated CpG sites remain hypomethylated.

Changes in methylation persist even after 24 months of withdrawal from ICS.

We next investigated whether ICS-induced DNA methylation changes would revert to normal after treatment withdrawal. To this end, we conducted a gene set variation analysis (GSVA) as well as a CpG site-per-site analysis. We defined our two groups of interests as i) all CpG sites that were differentially hypermethylated with six months of ICS therapy and ii) all CpG sites that were differentially hypomethylated with six months of ICS therapy (FDR < 0.05). Using GSVA, we found that the CpG sites which were differentially methylated after 6 months of ICS therapy remained significantly methylated in the same direction after 24 months of withdrawal compared to baseline ($p = 0.0242$ and $p = 0.0179$ for hyper-methylated and hypo-methylated sites, respectively), suggesting overall that methylation changes are not immediately reversed but rather persists even after 24 months of treatment withdrawal (Figure 7.2E&F).

We next performed a CpG site-per-site analysis to assess if and to what extent the methylation of CpGs affected by ICS would revert towards their baseline methylation state after ICS was switched to placebo after 24 months. We identified 7371 differentially methylated CpG sites that were changed ($p < 0.05$). Of these, 3532 were hypermethylated, 3712 were hypomethylated in the opposite direction, and 70 and 57 were in the same direction, respectively, leaving 106,854 (93.6) CpG sites unchanged.

ChIP-Seq identified 907 CpG sites in GR binding sites.

Next, we filtered for the differentially methylated CpG sites positioned in GR binding sites, narrowing down the CpG sites most relevant to corticosteroid function and theoretically conferring direct effects on downstream gene expression. This narrowed down the number of differentially methylated CpG sites to 907 CpG sites located within GR binding sites. Of these, 73% were hypomethylated. The top 20 of these 907 CpG sites are displayed in Table 7.4.

Differentially methylated GR-binding CpG sites were associated with nearby genes.

Next, we investigated whether differentially methylated CpG sites affect gene expression. First, a genome-wide expression quantitative trait methylation (eQTM) analysis was conducted to identify associations between all 114,098 differentially methylated CpG sites and local genes within 50,000 bps. The top 15 associations are shown in Supplementary Table 7.1. We narrowed our analyses to only include the GR-specific differentially methylated sites and found 128 eQTMs (FDR < 0.05). The top 15 associations are shown in Table 7.5. Among them was cg23416081, hypomethylated after treatment with ICS in association with higher expression of FKBP5, a gene well-known to increase after treatment with corticosteroids, as shown in several studies^{15, 16}. Figure 7.3 shows its position to FKBP5 and the GR binding site, and Figure 7.4A/B shows the change in expression and methylation with ICS. Figure 7.4C provides a graphical representation of these results.

Pathways associated with eQTM genes

Gene Ontology enrichment analysis of the 128 eQTM genes was additionally performed using *g:Profiler* to gain insight into the biological pathways related to these genes. All pathways that were identified are shown in Table 7.6. The top pathways were related to protein binding, cytosol and cytoplasm, which are not particularly relevant in the context of this study. The nuclear receptors meta-pathway was identified, possibly induced by GR interactions at these genes.

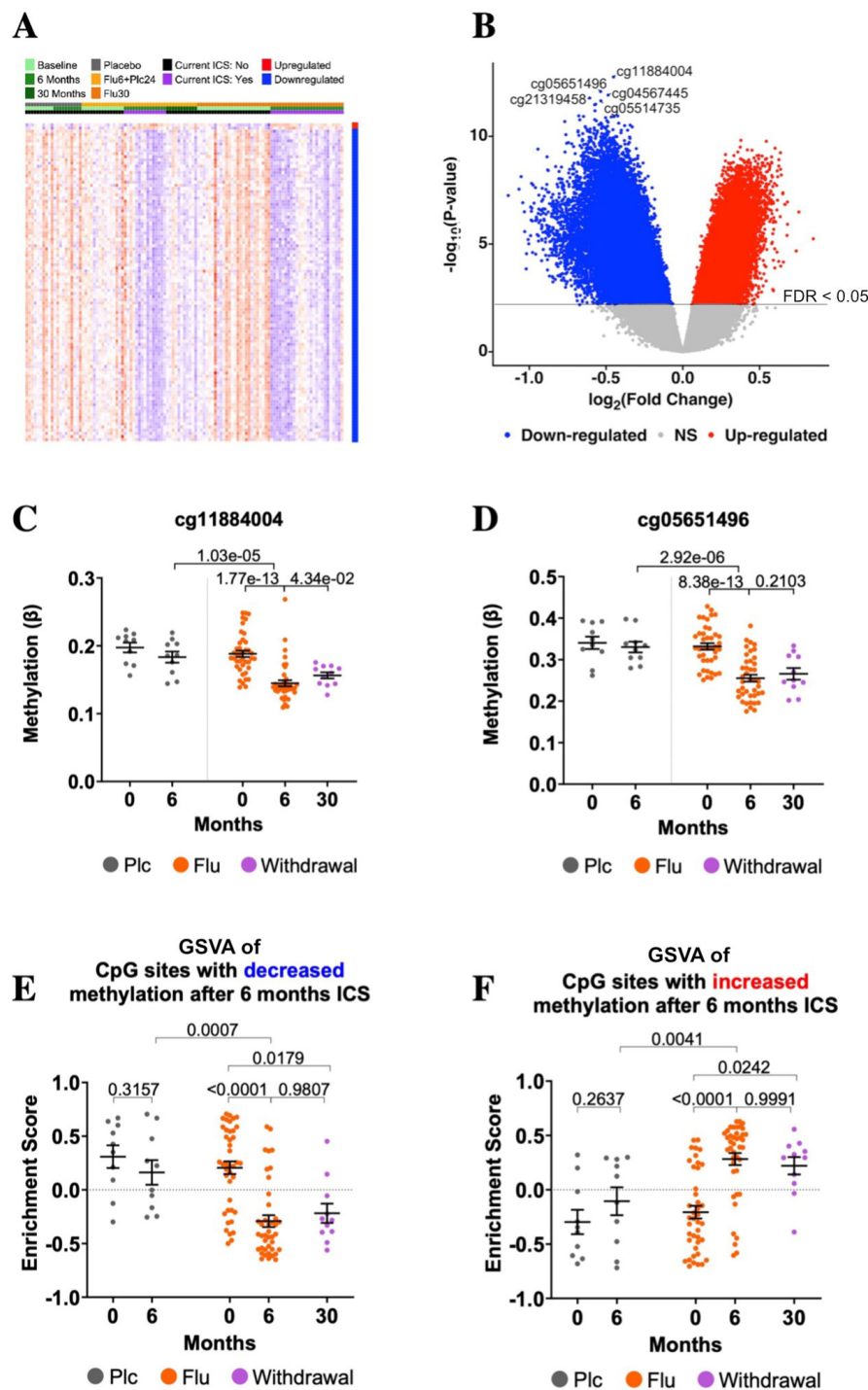


Figure 7.2) Differential methylation results. Methylation data was collected from bronchial biopsies of smoking-naïve, mild-to moderate- COPD patients in the GLUCOLD study. A) Heatmap of top 100 differentially methylated sites. Flu6+Flu24: Fluticasone treatment for six months followed by 24 months of withdrawal on placebo. Flu30: Fluticasone treatment for 30 months (Note: methylation data was not collected at the 30-month timepoint). B) Volcano plot of differentially methylated CpG sites between baseline and six months Flu. NS: Not significant. C) and D) Top 2 differentially methylated sites. Methylation significantly decreased with six months of ICS treatment. Withdrawal: Patients were treated with Flu for six months, then withdrawn with a placebo for 24 months. E) CpG sites that were differentially hypomethylated with six months of ICS treatment were decreased at 30 months, after 24 months after treatment withdrawal F) CpG sites that were differentially hypermethylated with six months of ICS treatment were increased at 30 months, after 24 months after treatment withdrawal. Mean \pm SEM is displayed. $P < 0.05$ was considered significant for paired and unpaired t-tests. Plc: Placebo, Flu: Fluticasone.

Table 7.3) Top 20 differentially methylated CpG sites after six months of treatment with fluticasone (n=51). Log₂(FC); log fold change, FDR: False Discovery Rate.

Differentially Methylated CpG sites			
CpG Site	Log ₂ (FC)	P-Value	FDR
cg11884004	-0.451	1.77E-13	1.43E-07
cg05651496	-0.535	8.38E-13	3.28E-07
cg04567445	-0.486	1.22E-12	3.28E-07
cg21319458	-0.607	1.67E-12	3.37E-07
cg05514735	-0.534	2.26E-12	3.65E-07
cg01172150	-0.568	3.24E-12	3.94E-07
cg19858017	-0.574	3.42E-12	3.94E-07
cg24727480	-0.572	6.54E-12	6.59E-07
cg14419424	-0.678	1.12E-11	7.83E-07
cg07867876	-0.424	1.13E-11	7.83E-07
cg15598244	-0.647	1.20E-11	7.83E-07
cg08886600	-0.448	1.21E-11	7.83E-07
cg03891590	-0.466	1.26E-11	7.83E-07
cg18941817	-0.503	1.37E-11	7.90E-07
cg16276850	-0.574	1.68E-11	8.49E-07
cg15010714	-0.671	1.71E-11	8.49E-07
cg26876444	-0.494	1.90E-11	8.49E-07
cg01740660	-0.564	1.90E-11	8.49E-07
cg12349858	-0.769	2.07E-11	8.79E-07
cg24888168	-0.487	2.21E-11	8.92E-07

Table 7.4) Top 20 differentially methylated CpG sites after six months of treatment with fluticasone (n=51), filtered for those in GR binding sites. Log₂(FC); log fold change, FDR: False Discovery Rate

Differentially Methylated CpG sites in GR binding sites			
CpG Site	Log ₂ (FC)	P-Value	FDR
cg24812349	-0.611	1.10E-10	1.22E-06
cg24326398	-0.497	1.99E-10	1.41E-06
cg12582959	-0.501	2.48E-10	1.51E-06
cg24190730	-0.529	3.51E-10	1.78E-06
cg03834031	-0.459	3.99E-10	1.87E-06
cg07730301	-0.348	4.71E-10	2.06E-06
cg07730301	-0.348	4.71E-10	2.06E-06
cg06422467	-0.558	6.06E-10	2.30E-06
cg13195254	-0.571	6.61E-10	2.40E-06
cg13954667	-0.387	7.04E-10	2.46E-06
cg01987202	-0.470	7.66E-10	2.57E-06
cg20000342	-0.463	9.33E-10	2.73E-06
cg26350671	-0.436	1.03E-09	2.89E-06
cg20274944	-0.431	1.04E-09	2.89E-06
cg20228862	-0.394	1.24E-09	3.08E-06
cg02675394	-0.470	1.71E-09	3.36E-06
cg08925954	-0.510	1.77E-09	3.38E-06
cg00976121	-0.393	1.91E-09	3.45E-06
cg24417971	-0.546	2.89E-09	4.15E-06
cg05059480	-0.416	3.41E-09	4.45E-06

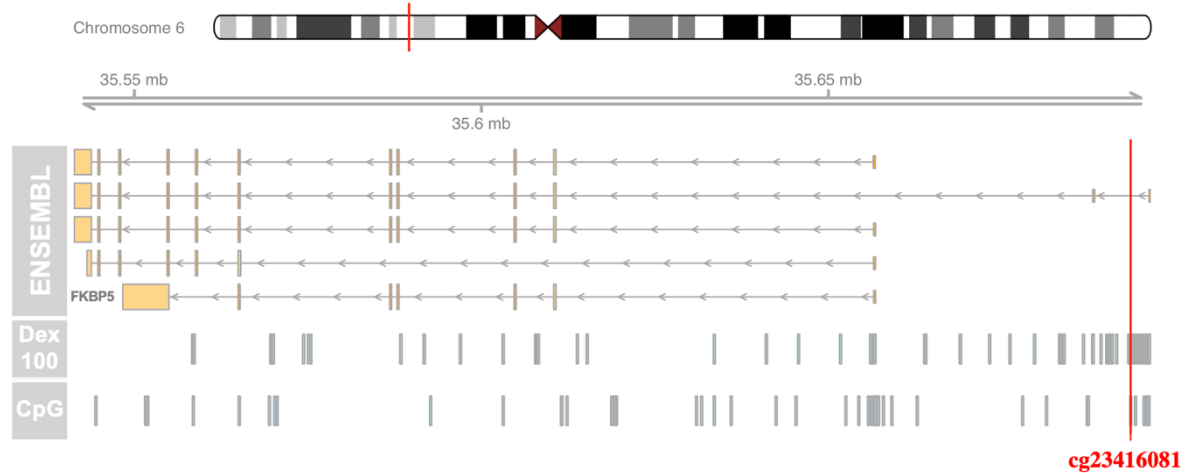


Figure 7.3) cg23416081 located in a GR binding site. Publicly available ChIP-Seq data on GR binding sites in dexamethasone-treated (Dex) A549 cells is displayed for the genomic region surrounding cg23416081. The 'ENSEMBL' track shows FKBP5 transcripts, and the 'Dex 100' track shows GR binding regions identified in cells treated with 100 nM Dex. The vertical red line indicates the location of cg23416081, which was identified as differentially methylated with ICS therapy. This analysis was completed on all differentially methylated CpG sites to select only those located within GR binding sites.

Table 7.5) Top 15 eQTM identified from a genome-wide eQTM analysis, including the 907 differentially methylated CpG sites located within GR binding sites. FDR; False Discovery Rate

Expression Quantitative Trait Methylation (eQTM) Analysis focused on GR binding sites				
Gene	CpG site	Beta	P-value	FDR
PDK4	cg09436502	-12.667	5.52E-07	5.52E-07
FKBP5	cg23416081	-5.233	3.80E-06	3.80E-06
DDIT4	cg10642599	-5.094	3.24E-05	9.73E-05
PDK4	cg04498739	-11.903	3.29E-04	3.29E-04
PNRC1	cg09501509	-1.759	3.87E-04	3.87E-04
UBE2L6	cg11554335	2.275	3.94E-04	3.94E-04
EMC3	cg13363829	-2.618	6.02E-04	7.76E-04
JAGN1	cg13363829	-2.083	7.76E-04	7.76E-04
ZMYND8	cg00152127	-2.458	8.21E-04	8.21E-04
NDUFA3	cg05944623	7.141	8.40E-04	8.40E-04
ZMYND8	cg25541209	-2.237	8.79E-04	8.79E-04
ELL2	cg23792274	11.610	1.17E-03	1.17E-03
MTHFD1L	cg03792573	1.415	1.29E-03	1.29E-03
AHNAK	cg18825531	-2.139	1.31E-03	1.31E-03
MTHFD1L	cg01510688	1.282	1.42E-03	1.42E-03

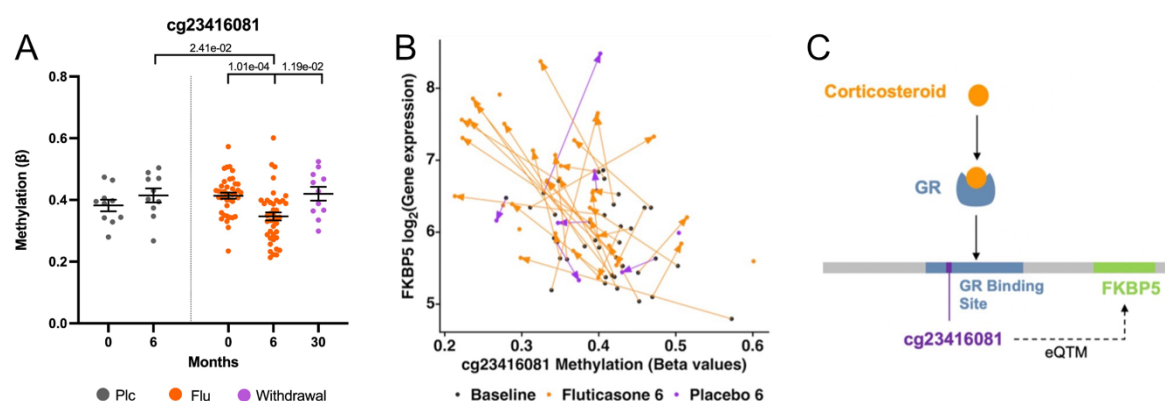


Figure 7.4) The second highest eQTM is cg23416081 and FKBP5. A) Methylation (beta values) of cg23416081 between treatment groups and visits. Methylation significantly decreased with six months of ICS treatment. Withdrawal: Patients were treated with Flu for six months, then withdrawn with a placebo for 24 months. B) Correlation plot showing the change in cg23416081 methylation and FKBP5 expression after six months of ICS therapy (n=44). Paired samples between 0 and 6 months are connected by lines, with arrows indicating the direction of change over time, in orange the paired samples of fluticasone treated patients are shown, in purple the paired samples of placebo treated patients are shown. Black circles indicate the baseline measurement C) Diagram depicting the differentially methylated site cg23416081 and its location within a GR binding site, as well as correlation to FKBP5 gene, which is located within 50,000 bps of the CpG site. GR: Glucocorticoid receptor. eQTM: Expression Quantitative Trait Methylation.

Table 7.6) Pathway analysis of eQTM genes. Pathway analysis was conducted on the 128 eQTM genes using g:Profiler. All results are shown. GO: Gene Ontology, MF: Molecular Function, CC: Cellular Component, WP: WikiPathways, FDR: False Discovery Rate.

Source	Term ID	Term Name	FDR
GO:MF	GO:0005515	Protein binding	6.54E-06
GO:CC	GO:0005829	Cytosol	4.84E-05
GO:CC	GO:0005737	Cytoplasm	1.35E-04
GO:CC	GO:0005925	Focal adhesion	5.08E-03
GO:CC	GO:0005900	Oncostatin-M receptor complex	5.53E-03
GO:CC	GO:0030055	Cell-substrate junction	6.21E-03
WP	WP:WP2882	Nuclear receptors meta-pathway	1.41E-02
GO:MF	GO:0004924	Oncostatin-M receptor activity	4.80E-02

Targeted methylation of cg23416081 in A549 cells with CRISPR-dCas9-DNMT3A led to a decrease in FKBP5 expression.

We then looked to establish a causative relationship between cg23416081 methylation and the induction of FKBP5 expression during corticosteroid treatment. We conducted this experimentally using CRISPR-dCas9-DNMT3A, a novel epigenetic tool that allowed us to introduce DNA methylation at target CpG sites. A549 cells were transfected with plasmids containing two gRNA sequences, an active DNMT3A or inactive DNMT3A. The distance between the targeted PAM sequence and cg23416081 was 34 bps (*gRNA-34*) or 17 bps (*gRNA-17*) (Supplementary Figure 7.1). We found that methylation with DNMT3A plasmids, led by *gRNA-34*, induced a 34.7% decrease in FKBP5 expression after stimulation with fluticasone compared to inactive DNMT3A (n=3, p=0.0180) (Figure 7.5B), whereas we did not identify a difference in FKBP5 expression when using *gRNA-17* (Figure 7.5A). For cells transfected with inactive DNMT3A plasmids, treatment with fluticasone led to an increase in FKBP5 expression versus without treatment (p = 0.0231 for *gRNA-34* and p=0.0596 for *gRNA-17*). In addition, we targeted another eQTM, cg2645126 and PDK4, with CRISPR-dCas9-DNMT3A. We

designed one gRNA that targeted a PAM sequence 27 bps away from cg2645126, but no changes in expression were observed upon fluticasone stimulation (Figure 7.5C).

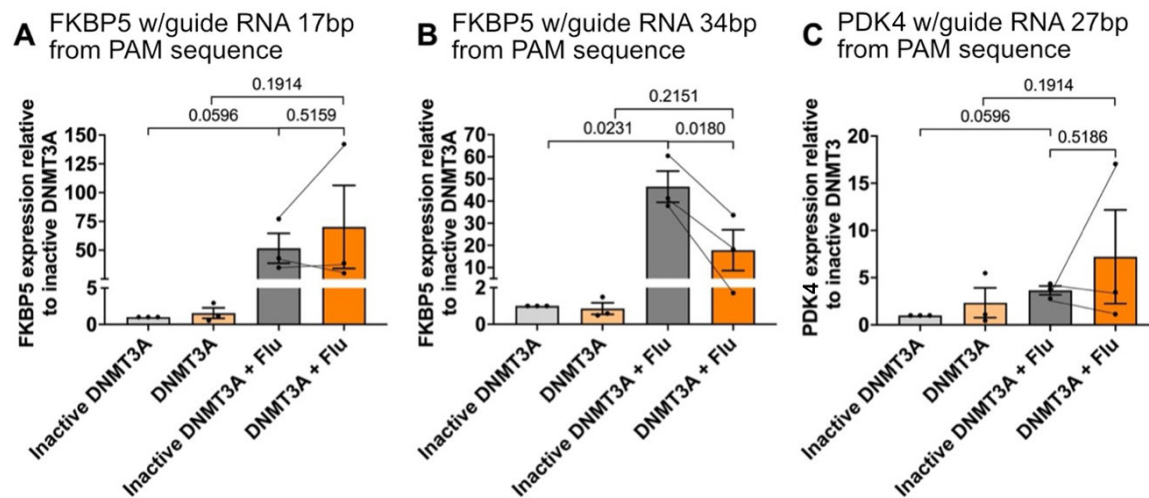


Figure 7.5) A) and B) FKBP5 expression following targeted methylation of cg23416081 (n=3) with a distance of 17 bps and 34 bps from the PAM sequence, respectively, C) PDK4 expression following targeted methylation of cg2645126 (n=2) with a distance of 27 bps from the PAM sequence, using CRISPR-dCas9-DNMT3A. FKBP5 and PDK expression were measured in transfected A549 cells treated with 10 nM fluticasone. DNMT3A: Cells transfected with pdCas9-DNMT3A-EGFP plasmids. Inactive DNMT3A: Cells transfected with pdCas9-DNMT3A-EGFP (ANV) plasmids (containing a catalytically inactive DNMT3A), Flut: Fluticasone.

7.5 – Discussion

In the current study, we show that treatment with inhaled corticosteroids strongly affects DNA methylation in the airway wall of patients with COPD, with the majority persisting even after 24 months of ICS cessation. When focusing on CpG sites located within GR binding sites, we identified several CpG sites likely to directly affect gene expression upon ICS-induced methylation changes. For example, we identified cg23416081, which was strongly hypomethylated by ICS in association with higher expression of FKBP5, a gene that inhibits binding and activation of activated GR to its response element (GRE). Using CRISPR-dCas9, we confirmed that hypomethylation of this CpG increases FKBP5 expression. Our findings indicate that ICS treatment can have long-lasting effects on DNA methylation, including on CpG sites within the GRE, directly inhibiting the effects of ICS, which can potentially explain at least part of corticosteroid insensitivity.

Our findings show that ICS extensively changes airway wall DNA methylation. In current literature, methylation profiling has generally shown corticosteroid treatment to be associated with hypomethylation and hypermethylation in similar proportions¹⁷⁻¹⁹, which is also reflected in our study overall. It is difficult to determine whether these methylation sites reflect changes in cell composition or are dynamic changes within cells.

Interestingly, the majority of the sites that were differentially methylated after six months of ICS treatment retained their methylation state, even after ICS withdrawal for two years. There is a general understanding that while the methylation status of most CpG sites is stable, approximately 20% are transient and driven by both internal and environmental stimuli²⁰. One in-vitro study on arsenic treatment of HeLa cells demonstrated that not all arsenic-induced methylation changes were reversed after ten days of treatment withdrawal, and changes in CpG islands were particularly more permanent²¹. Our study similarly showed that overall methylation levels do not revert towards

baseline levels after withdrawing ICS treatment, showing for the first time that at least part of the epigenetic effects of corticosteroids may be longer-lasting. Additionally, previous studies on GLUCOLD data showed that the discontinuation of fluticasone therapy at six months led to increased numbers of bronchial CD3⁺ cells (120%), plasma cells (118%) and mast cells (218%) at 30 months compared to during treatment^{14, 22}. This shift in inflammatory cells could result in a different DNA methylation pattern. Despite an increase in inflammatory cells that should increase methylation sites associated with inflammatory cells that were decreased with ICS, we did not see many of the CpG sites return to baseline. Additionally, correcting for these inflammatory cell types in our differential methylation analyses comparing baseline versus 6 months of ICS usage did not have a big impact.

Differential methylation has shown differences in the methylome of COPD patients compared to non-COPD controls²³⁻²⁵. This is believed to influence the activity of genes involved in COPD pathology^{23, 26}. However, differential methylation has not been demonstrated in the context of inhaled corticosteroids. We focussed on differentially methylated CpG sites located within GR binding sites, because this offers insight into the direct effects of ICS on DNA methylation. In theory, CpG methylation in promoter regions can influence gene expression in several ways. The presence of a methyl group within a binding motif can actively impede transcription factor binding. Our data shows that roughly 70% of the differential methylation occurring in the GR binding sites was hypomethylation. Previously, statistical models and *in vitro* and *in-vivo* studies have confirmed strong negative correlations between methylation and transcription factor binding. Areas with low methylation, particularly in promoters, are reliable predictors of transcription factor binding^{27, 28}. These results indicate that the hypomethylation we observed of CpG sites in GR binding sites likely contributes to enhanced GR activity during ICS treatment, leading to increased gene expression.

One of the top associations was between cg23416081 and FKBP5, which was hypomethylated by ICS treatment, resulting in increased FKBP5 expression and reversed after ICS withdrawal. We verified this finding in an *in-vitro* model using epigenetic editing, showing hypermethylation of this site resulted in lower expression of FKBP5. Increased production of FKBP5 creates an intracellular negative feedback loop, where FKBP5 reduces the affinity of the GR for corticosteroids, preventing its translocation to the nucleus²⁹. Therefore, higher FKBP5 expression may be linked to corticosteroid insensitivity.

Pathway analysis of the eQTM genes using *g:Profiler* did identify the oncostatin-M- receptor complex pathway as one of the top enriched pathways. Although less significant ($p = 5.53E-03$), the oncostatin-M- receptor complex pathway may be of interest as it has been strongly linked to the pathogenesis of chronic inflammatory disease, particularly inflammatory bowel disease^{30, 31}. A recent *in-vivo* mouse study demonstrated that higher oncostatin-M- receptor complex expression was linked to lung inflammation, damage to lung structure and thicker alveolar walls³². These findings support a link between corticosteroid-induced methylation changes and inflammatory responses.

There are several limitations to our findings. We minimised the effect of cell composition by selecting only the differentially methylated sites located within GR binding regions, as this is known to be directly relevant to corticosteroid function. Additionally, we referenced ChIP-Seq data on direct GR binding sites; however, this excludes situations where corticosteroids tether to NF- κ B to indirectly confer transcriptional effects. Therefore, it is possible that by excluding NF- κ B binding sites, we excluded some additional CpG sites and genes relevant to our study. Our eQTM analysis showed that cg23416081 and FKBP5 had the second-highest correlation. We identified another

differentially methylated CpG site associated with FKBP5, cg23416081, but it was situated too far from our target site to be unintentionally methylated (>100kb away).

In conclusion, ICS treatment affects DNA methylation across the genome, with sustained effects after treatment withdrawal. Some of these sites were situated within GR binding sites and associated with local genes, suggesting that ICS-induced methylation changes can directly influence gene expression and possibly explain mechanisms underlying ICS insensitivity. We experimentally validated one of the top associations, cg23416081 and FKBP5, in an *in-vitro* lung epithelial cell model. Overall, this study demonstrates that DNA methylation is long-term altered by inhaled corticosteroid treatment and influences transcriptional processes to drive changes in gene expression.

7.6 – Funding

Funded by the PPP-allowance by Health Holland (LSH) and the UMCG. This collaboration project is co-financed by the Ministry of Economic Affairs and Climate Policy by means of the PPP-allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships.

7.7 – Supplementary Material

Methods

A549 Cell Culture

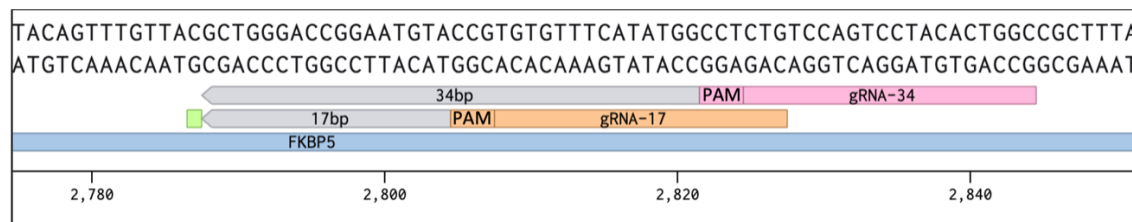
A549 cells were cultured in growth media comprised of RPMI, supplemented with 10% Foetal Bovine Serum (FBS) and 1% penicillin/streptomycin. Cells were incubated at 37 °C in a 5% CO₂/ 95% O₂ atmosphere and passaged at 85% - 90% confluence until ready for transfection. Hank's Balanced Salt Solution (HBSS) was used to wash cells before trypsinisation with trypsin-EDTA and neutralising with growth media.

Real-Time PCR

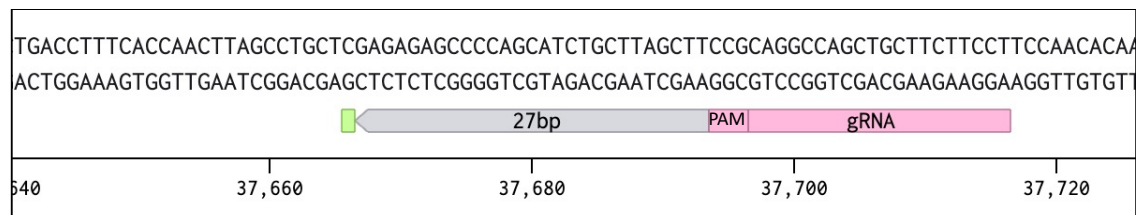
Gene expression was quantified following ThermoFisher Scientific protocols for duplex Taqman real-time PCR assays. 500 ng of cDNA was used for each reaction, with 20X FAM-labelled primers for the gene targets, 20X VIC-labelled primers for the 18S rRNA endogenous control and 2X Gene Expression Master Mix for a total reaction volume of 10 µL. All reactions were performed in triplicates in 96-well plates and run on the BioRad CFX96 Touch Real-Time PCR Detection System. Conditions for amplifications were: 50 °C for 2 mins, 95 °C for 10 mins, then 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Relative quantification of gene expression was then performed using the delta-delta Ct method, with 18S rRNA as the reference gene. Expression across samples was calculated relative to the expression of those with the inactive DNMT3A plasmid.

Figures

A FKBP5



B PDK4



Supplementary Figure 7.1) Design of gRNA sequences on Benchling. **A)** Green square shows location of cg23416081 and associated gene was FKBP5. gRNA sequences were designed to target PAM sequences at two different distances from cg23416081, 17 bps and 34 bps. **B)** Green square shows location of cg2645126 and associated gene was PDK4. gRNA sequence was designed to target a PAM sequence 27 bps away from cg2645126.

Tables

Supplementary Table 7.1) Top 15 eQTMS identified from a genome-wide eQTM analysis, including all 114,098 differentially methylated CpG sites. FDR; False Discovery Rate.

Genome wide Expression Quantitative Trait Methylation (eQTM) Analysis				
Gene	CpG site	Beta	P-value	FDR
THNSL2	cg26963093	-4.100	1.08E-11	1.08E-11
ITGAE	cg13868654	12.962	6.18E-08	6.18E-08
TRIM31	cg19006429	14.248	4.53E-08	2.26E-07
TRIM31	cg05853632	8.493	4.69E-08	2.34E-07
EPSTI1	cg09843907	-5.879	1.38E-07	2.76E-07
SLC44A5	cg09694258	4.778	3.79E-07	3.79E-07
GLYATL2	cg27230697	6.252	4.51E-07	4.51E-07
PDK4	cg09436502	-12.667	5.52E-07	5.52E-07
THNSL2	cg05124136	6.860	5.69E-07	5.69E-07
ITGAE	cg19585196	11.441	9.61E-07	9.61E-07
ZBTB16	cg25744613	-11.769	1.24E-06	1.24E-06
ETV7	cg18728732	-3.183	1.40E-06	1.40E-06
TRIM31	cg25247859	4.507	2.02E-06	2.02E-06
ELMO1	cg21086690	-5.466	2.48E-06	2.48E-06
ITGAE	cg08396824	-7.422	1.59E-06	3.18E-06

7.8 – References

1. Organisation, W.H. *Chronic obstructive pulmonary disease (COPD)*. 2023 [28 Oct 2023].
2. Safiri, S., et al., *Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019*. BMJ, 2022. **378**: p. e069679.
3. Adeloye, D., et al., *Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis*. Lancet Respir Med, 2022. **10**(5): p. 447-458.
4. Comer, D.M., et al., *Airway epithelial cell apoptosis and inflammation in COPD, smokers and nonsmokers*. Eur Respir J, 2013. **41**(5): p. 1058-67.
5. Schuliga, M., *NF-kappaB Signaling in Chronic Inflammatory Airway Disease*. Biomolecules, 2015. **5**(3): p. 1266-83.
6. Jing, Y., et al., *NOTCH3 contributes to rhinovirus-induced goblet cell hyperplasia in COPD airway epithelial cells*. Thorax, 2019. **74**(1): p. 18-32.
7. Prescott, E., P. Lange, and J. Vestbo, *Chronic mucus hypersecretion in COPD and death from pulmonary infection*. Eur Respir J, 1995. **8**(8): p. 1333-8.
8. Machin, M., et al., *Systematic review of lung function and COPD with peripheral blood DNA methylation in population based studies*. BMC Pulm Med, 2017. **17**(1): p. 54.
9. Lapperre, T.S., et al., *Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: a randomized trial*. Ann Intern Med, 2009. **151**(8): p. 517-27.
10. Willemse, B.W., et al., *Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers*. Eur Respir J, 2005. **26**(5): p. 835-45.
11. van den Berge, M., et al., *Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity*. Thorax, 2014. **69**(1): p. 14-23.
12. Gertz, J., et al., *Distinct properties of cell-type-specific and shared transcription factor binding sites*. Mol Cell, 2013. **52**(1): p. 25-36.
13. Vojta, A., et al., *Repurposing the CRISPR-Cas9 system for targeted DNA methylation*. Nucleic Acids Research, 2016. **44**(12): p. 5615-5628.
14. Lapperre, T.S., et al., *Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: a randomized trial*. Ann Intern Med, 2009. **151**(8): p. 517-27.
15. Bancos, I., et al., *Evaluation of FKBP5 as a cortisol activity biomarker in patients with ACTH-dependent Cushing syndrome*. J Clin Transl Endocrinol, 2021. **24**: p. 100256.
16. Faiz, A., et al., *FKBP5 a candidate for corticosteroid insensitivity in COPD*. European Respiratory Journal, 2016. **48**(suppl 60): p. OA1779.
17. Wang, A.L., et al., *DNA methylation is associated with inhaled corticosteroid response in persistent childhood asthmatics*. Clin Exp Allergy, 2019. **49**(9): p. 1225-1234.
18. Braun, P.R., et al., *Genome-wide DNA methylation investigation of glucocorticoid exposure within buccal samples*. Psychiatry and Clinical Neurosciences, 2019. **73**(6): p. 323-330.
19. Wang, L., et al., *Genome-wide analysis of DNA methylation in endometriosis using Illumina Human Methylation 450 K BeadChips*. Molecular Reproduction and Development, 2019. **86**(5): p. 491-501.
20. Hernando-Herraez, I., et al., *DNA Methylation: Insights into Human Evolution*. PLoS Genet, 2015. **11**(12): p. e1005661.
21. Eckstein, M., M. Rea, and Y.N. Fondufe-Mittendorf, *Transient and permanent changes in DNA methylation patterns in inorganic arsenic-mediated epithelial-to-mesenchymal transition*. Toxicol Appl Pharmacol, 2017. **331**: p. 6-17.
22. Lisette, I.Z.K., et al., *Airway inflammation in COPD after long-term withdrawal of inhaled corticosteroids*. European Respiratory Journal, 2017. **49**(1): p. 1600839.

23. Vucic, E.A., et al., *DNA methylation is globally disrupted and associated with expression changes in chronic obstructive pulmonary disease small airways*. Am J Respir Cell Mol Biol, 2014. **50**(5): p. 912-22.
24. Casas-Recasens, S., et al., *Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity*. American Journal of Respiratory and Critical Care Medicine, 2020. **203**(1): p. 129-134.
25. Morrow, J.D., et al., *Human Lung DNA Methylation Quantitative Trait Loci Colocalize with Chronic Obstructive Pulmonary Disease Genome-Wide Association Loci*. American Journal of Respiratory and Critical Care Medicine, 2018. **197**(10): p. 1275-1284.
26. Clifford, R.L., et al., *Altered DNA methylation is associated with aberrant gene expression in parenchymal but not airway fibroblasts isolated from individuals with COPD*. Clinical Epigenetics, 2018. **10**(1): p. 32.
27. Medvedeva, Y.A., et al., *Effects of cytosine methylation on transcription factor binding sites*. BMC Genomics, 2014. **15**(1): p. 119.
28. Héberlé, É. and A.F. Bardet, *Sensitivity of transcription factors to DNA methylation*. Essays Biochem, 2019. **63**(6): p. 727-741.
29. Fries, G.R., N.C. Gassen, and T. Rein, *The FKBP51 Glucocorticoid Receptor Co-Chaperone: Regulation, Function, and Implications in Health and Disease*. Int J Mol Sci, 2017. **18**(12).
30. West, N.R., et al., *Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease*. Nature Medicine, 2017. **23**(5): p. 579-589.
31. Verstockt, S., et al., *Oncostatin M Is a Biomarker of Diagnosis, Worse Disease Prognosis, and Therapeutic Nonresponse in Inflammatory Bowel Disease*. Inflammatory Bowel Diseases, 2021. **27**(10): p. 1564-1575.
32. MacDonald, K., et al., *Type I Interferon Signaling is Required for Oncostatin-M Driven Inflammatory Responses in Mouse Lung*. Journal of Interferon & Cytokine Research, 2022. **42**(11): p. 568-579.

Chapter 8 – Summary

8.1 – English Summary

This thesis focuses on Chronic Obstructive Pulmonary Disease (COPD), a chronic inflammatory condition resulting from inhaling toxic particles and gases. We hypothesised that severe COPD and severe early-onset disease represent unique phenotypes rather than stages within COPD. Therefore, this thesis aimed to find characteristics of COPD in general and severe (early-onset) disease, explicitly using a multi-omics approach. More precisely, we aimed to understand the transcriptomic and epigenetic signatures in this respect. Additionally, this thesis examines the potential replication of these signatures in the nose and explores the impact of the patients' smoking statuses and inhaled corticosteroids (ICS) use.

In **Chapter 2**, we investigated severe COPD's bronchial epithelial gene expression profile. We found a unique gene signature that separates severe COPD from mild to moderate and non-COPD cases. This signature reflected abnormalities in epithelial repair, impaired fibroblast function and decreased angiogenesis, supporting the hypothesis of distinct pathology in severe COPD. Furthermore, genes differentially expressed in bronchial brushings of severe COPD subjects were also observed in association with disease in nasal brushings, indicating the potential of nasal brushes as a proxy for the lung and supporting the one-airway hypothesis.

Chapter 3 shifted focus to peripheral lung tissue to determine if severe early-onset (SEO-)COPD exhibits a distinct gene expression profile compared to non-COPD controls and those with common COPD. We found a differential gene expression profile that showed enrichment of pathways involved in B-cell-mediated adaptive immunity, chemotaxis, and extracellular matrix (ECM) organisation along with higher expression of the B-cell gene CD79A and the ECM gene FBLN1 in SEO-COPD lung tissue. CD79A-positive inflammatory infiltrate numbers were increased in SEO-COPD lung sections. Using the GWAS catalogue list of COPD and lung function-associated genetic variants, we found that 15 GWAS SNPs were associated with the expression of SEO-COPD-associated genes, including FBLN1 and FBN1. In comparison, genome-wide cis-eQTL analysis identified six additional SNPs associated with the expression of genes found to be dysregulated in SEO-COPD.

Chapter 4 investigated whether miRNA profiling in peripheral lung tissue can distinguish between SEO-COPD, common COPD, and non-COPD controls. Here, we identified a differential miRNA signature in SEO-COPD of three miRNAs (miR-202-5p, miR-193b-3p and miR-331-3p). MiR-193b-3p is involved in regulating extracellular vehicles, miR-331-3p may be a new regulator of ECM dysregulation in SEO-COPD pathology, and miR-202-5p may be involved in post-translational effects.

In **Chapter 5**, we investigated the expression of the miRNA profile in relation to smoking status in both COPD and respiratory healthy controls in airway wall biopsies. We found two miRNAs, miR-203a-3p and miR-375, more expressed in current smokers' bronchial biopsies than ex- and never-smokers in COPD and respiratory healthy controls. These miRNAs play a role in the detoxification and inflammatory response to smoke. In addition, we identified that miR-31-3p was upregulated only in current smokers with COPD and not in asymptomatic smokers compared to ex-/never smokers. MiR-31-3p might have a protective role via decreasing *PDE5A* expression in the airways. This study provides candidate miRNAs for future studies to understand how smoking-induced inflammation influences epigenetics and how they may play a role in COPD pathogenesis.

In **Chapter 6**, we identified CpG sites associated with the presence of COPD and the severity of the disease. With the presence of COPD, we identified six DNA methylation sites in five genes involved in

smooth muscle contraction, corticosteroid metabolism and regeneration, and cellular differentiation. We also performed an expression quantitative trait methylation (eQTM) analysis, which did not show a correlation between the expression of these genes and methylation levels. We identified 25 and 3954 DNA methylation sites associated with disease severity as measured by FEV₁ % pred and RV/TLC % pred, respectively. For these CpG sites, we did find significant eQTMs (two for FEV₁ % pred and 817 for RV/TLC % pred) related to genes that play roles in the immune system and signalling signal transduction pathways, all important processes involved in COPD. These findings show that DNA methylation changes are associated with both the presence and severity of COPD and may be involved in COPD development.

In **Chapter 7**, we investigated the longitudinal changes in DNA methylation in bronchial biopsies from COPD patients during a period of ICS treatment and following treatment withdrawal. In this chapter, we found that ICS treatment of moderate-to-severe COPD patients was associated with differential methylation at CpG sites across the genome, with sustained effects after treatment withdrawal (93.6%). We focused on a proportion (0.79%) of these methylated sites positioned within GR binding sites and associated with local gene expression to address methylation changes that may directly influence the mechanism of action of corticosteroids and dictate the expression of pro- and anti-inflammatory genes. A large proportion of the methylation sites were found to be outside of GR binding sites. Their roles might be to indirectly regulate anti-inflammatory genes by positioning them in their promoter, while others positioned in gene deserts might have functions yet to be explored. Using epigenetic editing, we experimentally confirmed one of the top eQTM associations, i.e. cg23416081 and FKBP5, in an in-vitro lung epithelial cell model. This study demonstrates that DNA methylation is altered by inhaled corticosteroid treatment and influences transcriptional processes to drive changes in gene expression associated with the response of corticosteroids. In addition, we found that a large proportion of methylation sites altered by corticosteroids are retained after corticosteroid withdrawal.

A summary of all the main findings of our studies is presented in Figure 8.1.

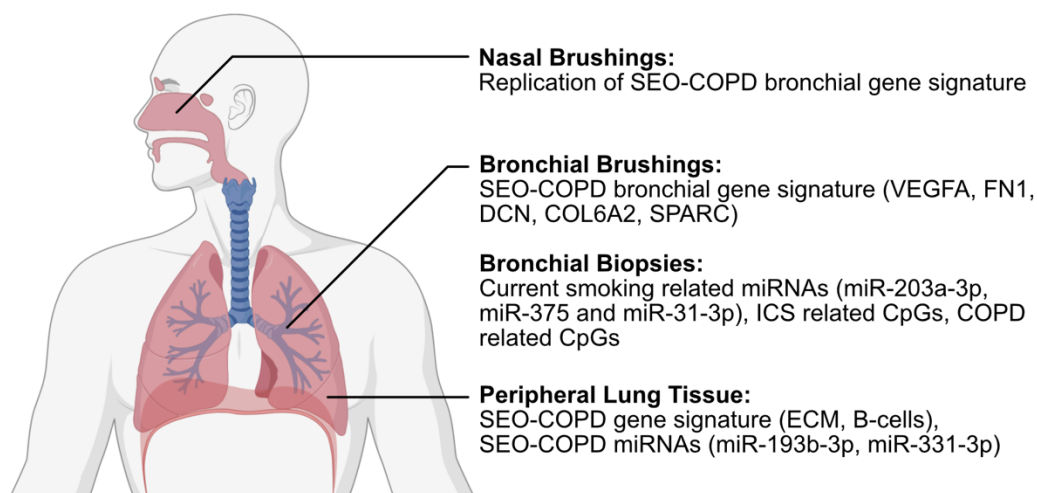


Figure 8.1) Summary of the main results per chapter.

8.2 – Nederlandse Samenvatting (Dutch Summary)

Dit proefschrift richt zich op de chronische longziekte COPD, ook wel Chronisch Obstructief Longlijden genoemd. Deze ziekte wordt gekenmerkt door een chronische ontstekingsreactie in de longen die het gevolg is van het inademen van giftige deeltjes en gassen, waaronder sigarettenrook. COPD is een ziekte dat sterk varieert tussen individuen, met verschillende verschijningsvormen en ernst-stadia van de ziekte.

Hoofdstuk 1 omschrijft het ziektebeeld van COPD, de werking van de ziekte en de huidige problemen met het onderscheiden en behandelen van de verschillende COPD-subgroepen. Het doel van het onderzoek beschreven in dit proefschrift was om nieuwe en specifiekere kenmerken van COPD te ontdekken die de verschillende verschijningsvormen en stadia beter van elkaar kunnen onderscheiden. Hiermee kunnen we mogelijk in de toekomst nieuwe behandelingen ontwikkelen voor deze subgroepen van COPD-patiënten. We stelden de hypothese dat ernstig COPD wat begint op oudere leeftijd, en ernstig COPD wat begint op relatieve jonge leeftijd, unieke subgroepen vertegenwoordigen binnen COPD met elk hun onderscheidende kenmerken.

Om deze kenmerken te ontdekken hebben we gebruik gemaakt van een zogenaamde “multi-omics” benadering. Dit is een onderzoeksmethode waarbij we grootschalig kijken naar de moleculaire kenmerken (o.a. DNA, RNA en eiwitten) in het weefselmateriaal van veel verschillende patiënten met of zonder COPD. De specifieke moleculaire kenmerken die we vinden in de COPD-patiënten worden dan in verband gebracht met de specifieke ziektekenmerken van COPD. We richten ons hierbij op 3 moleculaire kenmerken: genexpressie, miRNA expressie en DNA methylering. Genexpressie zegt iets over de hoeveelheid van een bepaald gen dat wordt afgeschreven van ons DNA, dus over de activiteit van bepaalde genen. Actieve genen worden vaak vertaald, via mRNA, naar eiwitten. miRNAs kunnen een negatief effect hebben op gen expressie door, met behulp van een speciaal eiwit complex, te binden aan complementaire mRNA sequenties en deze te af te breken. Hierdoor zeggen miRNAs iets over de genregulatie, ofwel de mate waarin de expressie van genen vertaald kan worden naar eiwitten. DNA methylering is het fenomeen waar een deel van het DNA-molecuul wordt aangepast door er een methyl groep aan toe te voegen. Dit heeft effect op het mechanisme dat DNA vertaalt naar mRNA. Daarom zegt mate van DNA methylering iets over in hoeverre de genexpressie wordt gestimuleerd of geremd. Deze moleculaire kenmerken onderzoeken wij in verschillende compartimenten van het ademhalingssysteem, waaronder de neus, de luchtwegen en het longweefsel. Daarnaast kijken we in dit het onderzoek naar de invloed van de rookstatus en het COPD-medicijngebruik van patiënten op de moleculaire kenmerken.

In **Hoofdstuk 2** onderzoeken we het genexpressieprofiel in de luchtwegen van patiënten met ernstig COPD en vergelijken we dat met gezonde controles en patiënten met mild tot matig COPD. We vonden een uniek genexpressieprofiel dat ernstig COPD onderscheidt van zowel mild tot matig COPD en gezonde controles. Dit profiel weerspiegelt afwijkingen in luchtwegherstel, verminderde functie van belangrijke structurele cellen in de long (fibroblasten), en verminderde vorming van nieuwe bloedvaten. Deze bevindingen ondersteunen de hypothese dat ernstig COPD zich onderscheidt van mild-matig COPD en laat zien dat andere ziekteprocessen een rol spelen. Bovendien vonden we dat een groot deel van de genen die kenmerkend waren voor ernstig COPD in de luchtwegen ook terug te vinden waren in de neus. Deze studie suggereert dat ernstige COPD een ander ziektebeeld is dan mild tot matige COPD, en dat de neus potentieel gebruikt kan worden als minder invasieve alternatief voor monsterafname in COPD- en longonderzoek.

In **Hoofdstuk 3** wordt de focus verlegd naar perifeer longweefsel om te bepalen of ernstig COPD op jonge leeftijd, ook wel “severe early-onset” (SEO)COPD genaamd, een verschillend

genexpressieprofiel vertoont in vergelijking met controles zonder COPD en patiënten met mild-matig COPD. We vonden een specifiek genexpressieprofiel dat verrijking laat zien van genen die betrokken zijn bij een adaptieve immuunrespons gemedieerd door B-cellen, de ontstekingsreactie en de organisatie van de extracellulaire matrix (ECM). Deze verrijking gaat samen met een hogere expressie van het B-cel gen CD79A en het ECM-gen FBLN1 in SEO-COPD longweefsel. Daarnaast tonen we op eiwit niveau aan dat het aantal CD79A-positieve ontstekingsinfiltraten verhoogd is in SEO-COPD longweefsel. Ook vonden we een verband tussen de expressie van de SEO-COPD specifieke genen en specifieke genetische varianten in het DNA, wat mogelijk wijst op een rol voor genetische vatbaarheid voor het ontwikkelen van SEO-COPD. Samenvattend, we vinden een genexpressieprofiel dat gelinkt kan worden aan o.a. het immuunrespons, genetische risicofactoren, en veranderingen in eiwit expressie.

In **Hoofdstuk 4** onderzoeken we de expressie van miRNAs in perifeer longweefsel van SEO-COPD in vergelijking met controles zonder COPD en patiënten met mild-matig COPD. Hier hebben we een miRNA-profiel geïdentificeerd in SEO-COPD van drie miRNAs: miR-202-5p, miR-193b-3p en miR-331-3p, die we kunnen koppelen aan genexpressie. MiR-193b-3p is betrokken bij communicatie tussen cellen door middel van uitgescheiden blaasjes met moleculen. MiR-331-3p is mogelijk betrokken bij de regulatie van de ECM-veranderingen in SEO-COPD, die we in hoofdstuk 3 aantoonen. Tot slot is miR-202-5p mogelijk betrokken bij de veranderingen van eiwitstructuren. Dit onderzoek geeft inzicht in hoe deze drie miRNAs betrokken zijn in SEO-COPD via genexpressie.

In **Hoofdstuk 5** onderzoeken we de expressie van miRNAs in de luchtwegen in relatie tot rookstatus in zowel COPD als gezonde controles. We vonden twee miRNA's, miR-203a-3p en miR-375, die meer tot expressie komen in de luchtwegen van huidige rokers dan bij ex- en nooit-rokers in COPD en gezonde controles. Door miRNA aan genexpressie te koppelen, vonden we dat deze miRNA's een rol spelen in het ontgiften van de long na, en de ontstekingsreactie op, rook. Daarnaast hebben we ontdekt dat miR-31-3p alleen in COPD-patiënten hoger tot expressie komt bij huidige rokers en niet in 'gezonde'. MiR-31-3p kan een beschermende rol spelen in COPD-rokers via het verminderen van de PDE5A-expressie in de luchtwegen. Deze studie levert kandidaat miRNA's voor toekomstig onderzoek om te begrijpen hoe door roken veroorzaakte ontsteking de epigenetica beïnvloedt en hoe deze een rol kunnen spelen in de pathogenese van COPD.

In **Hoofdstuk 6** identificeren we veranderingen in DNA methylatie die geassocieerd zijn met de aanwezigheid van COPD en de ernst van de ziekte. Met de aanwezigheid van COPD identificeren we op zes plaatsen in het DNA, verandering in methylering. Vijf van de zes plaatsen zijn in genen die betrokken zijn bij gladde spiercontractie, cellulaire differentiatie, corticosteroïd metabolisme en regeneratie. Daarnaast hebben we de associatie tussen de methylatie veranderingen in het DNA en de expressie van de genen die gelegen zijn in dat deel van het DNA onderzocht. We identificeerden 819 methylatie-gen expressie correlaties voor longfunctie (FEV₁ % pred) en residuele longvolume als fractie van het totale longvolume (RV/TLC % pred), welke beide belangrijke longfunctie parameters zijn voor de ernst van COPD. De gevonden methylatie-gen correlaties zijn gerelateerd aan genen die een rol spelen in het immuunsysteem en signaaltransductieroutes, belangrijke processen die betrokken zijn bij chronische ontsteking en longverbouwing in COPD. Deze studie toont aan dat DNA-methylatie veranderingen geassocieerd zijn met zowel de aanwezigheid als de ernst van COPD, en mogelijk betrokken zijn bij de ontwikkeling van COPD.

In **Hoofdstuk 7** onderzoeken we de invloed van langdurige behandeling met inhalatiecorticosteroïden (ICS, COPD-medicijnen) en het effect van staken van deze behandeling op de veranderingen in DNA-methylatie in de luchtwegen van COPD-patiënten. We vonden dat ICS-

behandeling geassocieerd is met veranderingen in DNA methylatie in het hele genoom, met effecten die aanhouden na het staken van de behandeling (93,6%). Het overgrote deel van de methylatie veranderingen bevindt zich buiten de gebieden waar de ICS kunnen binden. Dit zou mogelijk een rol kunnen spelen in de indirecte regulatie van ontstekingsremmende genen. We hebben ons vervolgens gericht op een klein deel (0,79%) van de gemethyleerde plaatsen in het DNA die zich bevinden in gebieden waar de ICS kunnen binden en daarmee een direct effect kunnen bewerkstelligen. Met behulp van functionele experimenten bevestigen we de correlatie in long cellen tussen één van de meer gemethyleerde plekken en ICS, namelijk methylatie plek cg23416081 en het betrokken gen FKBP5. Dit onderzoek toont aan dat DNA-methylatie verandert door behandeling met inhalatiecorticosteroïden en dat dit vervolgens de genexpressie beïnvloedt die geassocieerd is met de respons op corticosteroïden. We hebben gevonden dat een groot deel van de veranderingen in DNA methylatie na ICS-behandeling blijvend zijn en dus mogelijk consequenties hebben op langere termijn.

Figuur 8.1 geeft een overzicht van de belangrijkste bevindingen van de onderzoeken die beschreven zijn in dit proefschrift.

Chapter 9 – General Discussion

Clustering together disease severities and sub-phenotypes is common in most transcriptional analyses focussing on COPD. This has led to several issues with understanding this complex disease¹. It is now well-established that the major pathologies in COPD², chronic bronchitis, small airways disease and emphysema only scratch the surface of the heterogeneity of this disease³, also because these can manifest in various combinations and severities in individual patients. Recent studies have attempted to classify COPD based on several variables, including genetics, sex, environmental exposures, premature birth, severity of airflow obstruction and rate of lung function decline⁴⁻⁸. Despite this, there are still large gaps in knowledge of which pathways contribute to the development and progression of disease and specific disease subgroups.

9.1 – Severe (early-onset) COPD is associated with differences in gene expression involved in extracellular matrix composition

Chapters 2, 3 and 4 defined a transcriptomic and epigenetic signature of severe COPD and severe early-onset SEO-COPD in the airways (chapter 2) and peripheral lung tissue (chapters 3 and 4). In **Chapter 2**, we found severe COPD to be associated transcriptionally with a lower expression of ECM-related genes in bronchial brushings. From our findings, it is clear that a transcriptional decrease in key components of the ECM may play a role in large airways in severe COPD. One explanation for the lower expression of ECM-related genes in brushings could be the cell types responsible for ECM production, as brushes mainly contain epithelial cells and a lower proportion of stromal cells, known to be the main ECM producers. It can also be noted that the decrease in gene expression observed in our study may be a negative feedback loop where an over-abundant protein can negatively regulate its own expression. Hence, future studies should be done to determine if these transcriptional effects in bronchial brushings are also present on the protein level in serial airway tissue sections, comparing ECM deposition in the bronchial epithelial cells and in the directly underlying subepithelial area. In **Chapter 3**, we identified a lung gene expression profile for SEO-COPD in peripheral lung tissue, with 21 genetic variants related to the SEO-COPD gene signature indicating a genetic pre-disposition in SEO-COPD. The profile harboured an increased B-cell immunity and ECM organisation signature, supporting a role for autoimmunity and aberrant tissue repair in the SEO-COPD phenotype. Specifically, the ECM signature indicated increased ECM fibre formation via FBLN1, FBN1, EMILIN1 and others. This could be a sign of attempted lung tissue repair. However, as in peripheral lung tissue with COPD, emphysema is present, resulting in structural tissue loss. This may explain why the attempted repair was unsuccessful and cannot be countered. Additionally, we identified genetic risk factors in FBLN1 and FBN1 for airway obstruction, suggesting a potential genetic factor underlying ECM-related gene expression differences in SEO-COPD and associations between genetic variants (SNPs) and gene expression of genes involved in inflammation. These genetic variants can have direct and indirect effects on gene expression. Future studies are now needed to confirm and further clarify the relationship between the identified variants and their effects on gene expression *in vitro*. However, we should realise that *in vitro* models cannot fully recapitulate the *in vivo* situation. In general, there is loss of repair in parenchyma and excessive repair leading to fibrosis in the bronchi. It is hard to include cell-to-cell interactions *in vitro* that likely play a role, and examination can show mixed results. Additionally, some SNPs have structural effects or effects on alternative gene splicing. A systems biology approach can be undertaken to determine if any of the identified genetic variants have structural impact or can be associated to gene splicing. These associations then should be associated to single-cell data to determine what types of cells express the identified genetic variants, which, lastly, then can be followed up with *in-vitro* validation.

Chapter 4 found that miR-331-3p was involved in ECM organisation in severe early onset (SEO-) COPD in lung tissue. Here, we found that lower miR-331-3p expression was associated with more COL6A1, COL6A2, ELN, CSPG4 and NTN1 in peripheral lung tissue, all these genes being involved in the ECM. However, with more severe COPD, there is increased emphysema and loss of ECM in the peripheral lung. Again, the increased expression of genes involved in ECM organisation, where there is increased destruction of lung tissue, might indicate a failed repair attempt. It should be noted here that bronchial brushings are sampled from the large airways, which in COPD are characterised by thickening of the airway wall due to increased ECM deposition, whereas peripheral lung tissue is sampled from parenchymal areas in the lung characterised by alveolar destruction and overall loss of ECM in COPD. Hence, as also discussed above, the decreased ECM gene expression profile in bronchial brushes may result from a negative feedback loop induced by increased ECM protein deposition, while the increased ECM expression in peripheral lung tissue may reflect an attempt at tissue repair. This could explain the seemingly opposite processes occurring in the airways and parenchyma in COPD lungs. based on our findings, we think that miR-331-3p is an important player in ECM dysregulation and tissue repair, specifically in the peripheral lung in SEO-COPD, via the regulation of ECM molecules such as COL6A1, COL6A2 and ELN. Future studies should validate our results in an *in-vitro* model and see if targeting miR-331-3p can restore ECM homeostasis in SEO-COPD parenchyma or if tissue repair will be further impaired. Additionally, future studies could combine DNA methylation and other epigenetic data to check if the identified genes affected by miR-331-3p are also differentially methylated.

Chapter 2 additionally showed protein interactions in severe COPD between the ECM (FN1, COL6A2, SPARC), collagen crosslinkers (DCN) and VEGFA. We found lower expression levels of genes involved in ECM composition, with lower transcriptional levels of genes involved in ECM formation and lower levels of VEGFA. We discussed that loss of VEGFA leads to endothelial cell apoptosis and is associated with emphysema⁹⁻¹³, possibly due to reduced blood flow or damage in small capillaries that likely leads to local destruction of alveolar septa. Our findings suggested a role for VEGFA in the development of severe COPD beyond its role in the alveoli, as shown by previous studies¹¹, in the bronchus and the nose. This role of decreased VEGFA in the bronchus and its effect on blood vessel development in the bronchus should be investigated in future studies. One suggestion could be to investigate endothelial apoptosis in the bronchus, and another could be to investigate the endothelium – ECM interactions further. Specifically, we could investigate the VEGFA protein levels in the bronchial epithelial cells.

Clearly, differences in ECM gene expression are present in severe COPD and SEO-COPD. The direction of the differences is different across tissue types and locations sampled. Several studies have investigated the role of ECM gene expression and proteins in COPD^{14, 15}. It would now be interesting to determine why ECM gene expression is lower in the epithelium of the large airways, in contrast to its increased expression in peripheral lung tissue.

9.2 – Airways epigenetic response to smoking, ICS treatment, and an association with COPD severity

Epigenetics are marks on the DNA that do not modify the DNA sequence and that can change and adapt during an individual's lifetime. These epigenetic marks can directly or indirectly interfere with transcription factor binding and, by doing so, suppress or activate gene transcription. Types of epigenetic markers are DNA methylation, adding a methyl group to cytosines in the DNA, and miRNA expression that causes post-transcriptional regulation^{16, 17}. It is known that environmental factors,

such as tobacco smoking and inhaled corticosteroids (ICS), influence the epigenome, including DNA methylation^{18, 19}.

Recent studies have attempted to study the effects of smoking and ICS in COPD on epigenetics and how this subsequently leads to changes in gene expression²⁰⁻²³. The effects of smoking have mostly been studied by focusing on epigenetic changes in blood^{20, 23}. However, blood does not necessarily reflect the epigenetic state in lung tissue or airway epithelium, providing limited insight into smoking effects relevant to lung diseases.

In **Chapter 5**, we defined the effects of active smoking on miRNA expression in bronchial biopsies of individuals with and without COPD. Additionally, we linked miRNAs affected by smoking to transcriptomic changes to investigate their functional effects. Here, we found that smoking-related miRNA expression changes were associated with transcriptional changes involved in inflammation. Specifically, miR-203a-3p, miR-375 and miR-31-3p were higher expressed in actively smoking individuals with and without COPD compared to never or ex-smokers and linked to decreased expression of pro-inflammatory genes. However, miR-31-3p was not affected by smoking in the respiratory healthy cohort.

For miR-375, we identified 18 predicted target genes that were lower expressed with higher miR-375 expression. One of these genes was CXCL12, a ligand for CXCR4 (fusin) and CXCR7 (atypical chemokine receptor 3 (ACKR3)). CXCR4 controls the transduction of different downstream signalling pathways that are involved in cell survival, proliferation, and migration²⁴⁻²⁶. Previous studies suggest that CXCR7 is associated with signalling similar to CXCR4^{27, 28}. CXCL12 has been described in the lungs to suppress neutrophil apoptosis, enhance neutrophilia and contribute to neovascularization²⁹⁻³¹. Though we did not find a difference in receptor expression, less ligand may indicate worse cell survival, more apoptosis, and increased neutrophil apoptosis. A high number of genes and pathways related to inflammation were inversely correlated with miR-375 expression, but there was a lack of correlation with gene signatures typical for inflammatory cells. Therefore, we suggested that miR-375 likely causes more activation of inflammatory cells rather than a change in cell numbers or composition. To further study this, we could use CRISPR/cas9 in a smoking mouse model to knock down miR-375 to investigate if miR-375 has an influence on neutrophilic cell survival, as well as to assess whether miR-375 has an influence on the activation of other inflammatory cells.

In **Chapter 5**, we also showed that miR-203a-3p, higher expressed in current smokers, was negatively associated with the expression of EBF3 and PDGFD. These two genes are involved in cell cycle arrest, apoptosis, and fibroblast proliferation, respectively. Additionally, miR-203a-3p was associated with xenobiotic response and genes involved in inflammatory pathways. This indicates an important role for miR-203a-3p in response to the toxic particles in smoke. An *in vitro* model using stimulation with smoke extract to specifically inhibit and induce miR-203a-3p expression could be used to validate our results in a future study, which could then also determine the direct interaction between miR-203a-3p and its target genes EBF3 and PDGFD. Another interesting observation was that when comparing current to never-smoking healthy controls, we observed 20 differentially expressed miRNAs, which were not different when compared to current to ex-smoking COPD patients. Partly, this could be related to the difference in the groups that we compared, i.e. current versus ex-smoking in COPD and current vs never smoking in control, but it may also indicate a different response to smoking in healthy controls, which may be related to protective mechanisms, and hence of interest for further research.

In **Chapter 6**, in a mild to moderate COPD population, we found the level of bronchial obstruction and hyperinflation to be associated with differences in DNA methylation in bronchial biopsies in COPD. These differences were associated with the enrichment of inflammatory response pathways and cell signalling pathways. We also found that hyperinflation was associated with hypermethylation and lower expression of HSD11B2, a gene involved in cortisol and corticosteroid metabolism^{32, 33}, and SFSWAP, a regulatory gene for alternative splicing³⁴. It is worth investigating the direct effects and mechanisms behind HSD11B2 to see if HSD11B2 and other hydroxysteroid dehydrogenases are involved in corticosteroid insensitivity, regardless of ICS usage. Importantly, we focussed on finding a DNA methylation profile in bronchial biopsies that was different in COPD compared to non-COPD controls. We found a few sites differentially methylated with the presence of COPD and no direct association with gene expression. However, the identified CpG sites were located in intronic regions of genes. One of the CpG sites was in ADGRL3, a gene involved in the contraction of the airway smooth muscle³⁵. The lack of associations between DNA methylation and gene expression, the normal route by which DNA methylation exerts its effects, is notable. However, it is known that DNA methylation does have a role in alternative splicing as well³⁶, and a recent study from our group showed splicing differences in mild to severe COPD of bronchial brushings³⁷. The different methylation patterns of SFSWAP associated with hyperinflation could also influence this. Hence, future studies are needed to further clarify the role of alternative splicing in COPD and its relation to DNA methylation.

In **Chapter 7**, we focussed on changes in DNA methylation in bronchial biopsies after ICS treatment and the effects of ICS withdrawal. We identified differential methylation sites associated with ICS usage in GR sites and gene expression. Specifically, we found methylation sites associated with decreased expression of 38 genes and methylation sites associated with increased expression of 90 genes. Interestingly, among the methylation sites that were associated with decreased gene expression, we found associations involved in ECM regulation and the PI3K-Akt pathway. For example, COL6A3 was associated with the hypermethylated site cg21346966, decreasing gene expression. This site is also located in a GR binding site. COL6A3 is another member of the collagen family of genes and encodes for part of the type VI collagen, which has mechanical roles and cytoprotective functions via the inhibition of apoptosis and oxidative damage³⁸. Other studies found increased expression of COL6A3 in peripheral lung tissue in COPD, probably in response to emphysema to repair itself³⁹. We showed for the first time a link between ICS usage, DNA methylation and the expression of this gene, and other ECM genes, an interesting topic to investigate further. Additionally, we identified hypomethylation of cg23416081, a methylation site in a GR binding site near FKBP5, associated with increased FKBP5 expression, which we confirmed via epigenetic editing in an *in vitro* model. This results in the activation of a feedback loop that blocks the activation of the GR pathway (Figure 9.1).

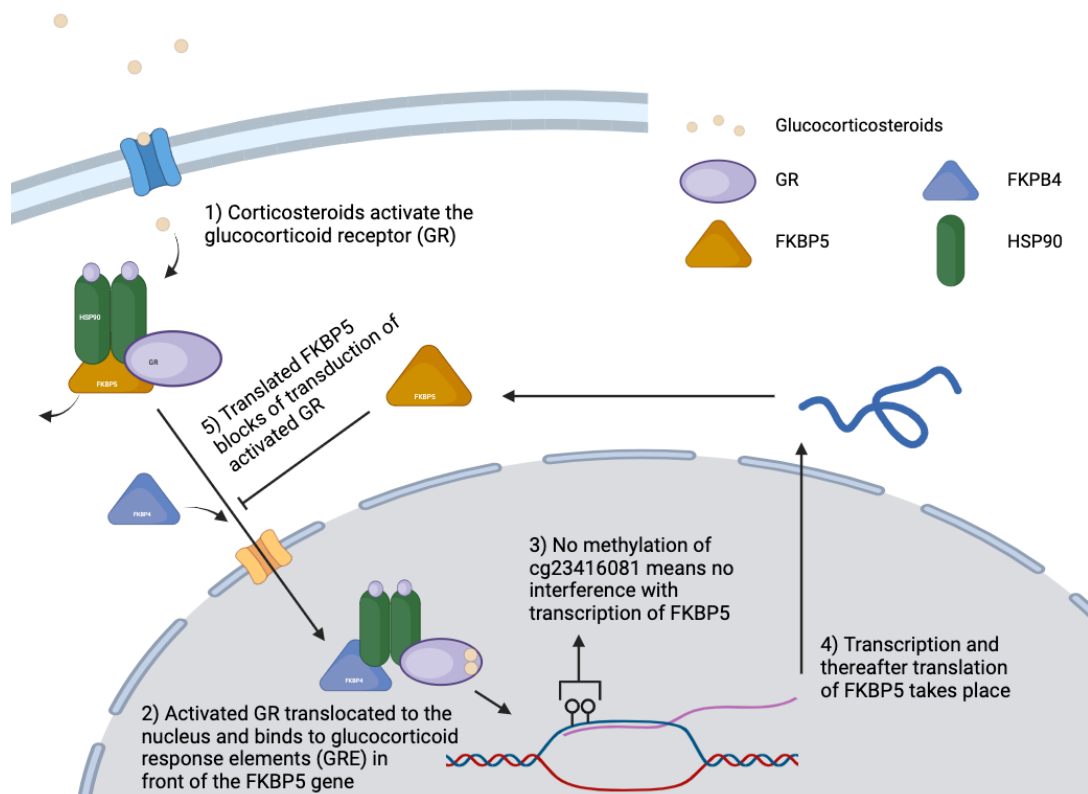


Figure 9.1) The activated feedback loop after hypomethylation of cg23416081 utilises FKBP5 to block transduction of activated GR.

In a longitudinal study, we showed this hypomethylation for the first time in the lungs after ICS. Severe COPD patients often use ICS for a prolonged time. However, most patients do not respond to this treatment. The upregulation of FKBP5 due to hypomethylation in bronchial epithelium could explain why some COPD patients respond less to ICS. Moreover, it is known that oxidative stress is a major factor in the pathogenesis of COPD, and previous research has shown oxidative stress to promote corticosteroid insensitivity^{40, 41}. Therefore, it is worth investigating if prolonged cellular stress, senescence or smoke could introduce methylation of FKBP5 and affect ICS response. Additionally, GR activation is involved in the trans-repression of the NF- κ B pathway, which may activate the adaptive and innate immune response⁴². In COPD, an amplified inflammatory response contributes to airway obstruction and emphysema. ICS are used in part to suppress this inflammatory response, and trans-repression of the NF- κ B pathway is the mechanism by which this occurs. Studying the secondary effects of corticosteroids on the NF- κ B pathway could provide valuable insights in how this trans-repression occurs. Therefore, future investigations into ICS activity and GR binding could be extended to include secondary activation of the NF- κ B pathway, verify the findings using the same epigenetic editing methodology as we used, and verify the epigenetic editing and the gene expression levels of the identified gene association *in vitro*. This could uncover additional CpG sites of importance that affect GR action.

Lastly, in **Chapter 6**, we found hyperinflation to be associated with differences in DNA methylation in COPD. Of these methylation sites associated with hyperinflation, 817 were associated with expression levels of genes, i.e. eQTMs. Of the methylation sites associated with hyperinflation, 147 sites changed following ICS treatment, as described in **Chapter 7**. These 147 methylation sites were mainly involved in oxidative stress and the inflammatory response via associations with gene expression. Little is known about the molecular mechanisms that contribute to or affect

hyperinflation, a result of a variety of complex changes in different lung compartments and an important aspect of COPD pathology. Further research into these overlapping sites and identified molecular mechanisms could highlight how ICS may affect hyperinflation and lung function in COPD. One of these DNA methylation sites associated with hyperinflation (cg26236507) was also located in GR binding sites of not well-defined genes: PHYHD1 and ZER1. The identified associations between CpG sites that are in GR binding sites and affected by ICS and associated with hyperinflation can point to ICS mechanisms that could have an unknown positive effect on hyperinflation and emphysema. It would be interesting to investigate the function of these genes, possibly via CRISPR/Cas9, as this may provide further insights into their role in COPD and potential therapeutic targets. Additionally, investigating the downstream effects of ICS treatment on oxidative stress and inflammatory pathways could help elucidate the mechanisms underlying the observed changes in DNA methylation.

9.3 – The nose as a proxy of the lower airways

Research in COPD is highly influenced by the availability of suitable samples and tissues that allow for an understanding of the disease. One of the applications of research is to monitor but also to subtype or classify diseases for personalised treatment. However, sampling of the lung is an invasive, uncomfortable and expensive procedure. Therefore, there is a need for less invasive samples that reflect the lower airways.

The one airway hypothesis originally stated that the upper and lower airways respond in the same way to external stimuli and features of disease⁴³. Our group has previously shown that nasal epithelium can distinguish both asthma and COPD from healthy individuals^{44, 45}. Furthermore, our group has previously shown that smoke exposure gives an overlapping transcriptional profile in nasal and bronchial brushes⁴⁵. A sufficiently shared transcriptional profile between the nasal and bronchial brushes suggests that the nasal epithelium may be a surrogate for studying respiratory diseases. However, no research has been performed to determine whether a shared transcriptional response in COPD exists between the bronchial and the nasal epithelium.

In **Chapter 2**, we showed that our unique severe COPD gene signature obtained from bronchial brushes was partially replicated in the nose. In the nose, we found a clear replication of the genes less expressed in severe COPD but no replication for genes more expressed in COPD. Hereafter, we performed a meta-analysis on two nasal brush cohorts to identify which genes specifically could be replicated in the nose. We found that 83 of our 219 severe COPD genes in the bronchial epithelium were differentially expressed in the nasal epithelium in the same direction. The 41 genes more expressed in severe COPD were not involved in specific pathways. The 42 genes less expressed in severe COPD were involved in ECM and immune system-related pathways via complement component C1. Again, VEGFA, FN1, SPARC, DCN were present, indicating a signal for a changed ECM response in the nose. Future research could build on this research, further profile the nasal and bronchial epithelium overlap, also on protein level, and investigate a possible uncovered underlying mechanism. It could also be interesting to focus on the effect of decreased expression of the identified ECM and angiogenesis genes on the bronchial and nasal epithelium.

9.4 – Impact of this research on patient

These findings have several implications for the future management of COPD. First, the identification of a unique transcriptional and epigenetic profile for severe COPD, including key genes and miRNAs, could aid in the development of more precise diagnostic tools. These molecular markers could be followed-up as biomarkers allowing for early detection and risk stratification, potentially identifying individuals at higher risk for developing severe disease at earlier stages. Additionally, these severe-

COPD specific markers provide valuable novel insight in underlying disease pathology, which could aid in discovery of new therapeutic targets that could be used for more personalized treatment approaches in COPD.

Second, our findings on the effects of ICS treatment on epigenetics, including the identification of specific methylation changes and gene expression alterations, suggest that ICS responsiveness in COPD patients could be better predicted by profiling DNA methylation patterns. This could enable clinicians to tailor corticosteroid therapy more effectively, minimizing the risk of unnecessary treatments and improving patient outcomes. Moreover, identifying patients who are less likely to respond to ICS could guide the development of alternative treatments or adjunct therapies aimed at restoring corticosteroid sensitivity.

Finally, the relationship between smoking, epigenetics, and COPD severity highlights the potential for preventive strategies based on epigenetic markers. For instance, understanding how smoking-induced DNA methylation affects gene expression in the airways could lead to the development of biomarkers that predict susceptibility to COPD, even before clinical symptoms appear. This could open new avenues for early intervention, potentially reducing the incidence of COPD in high-risk populations.

Overall, these advancements could ultimately lead to more effective prevention, earlier diagnosis, and personalized treatment strategies for patients with COPD, improving quality of life and reducing the burden of this disease.

9.5 – Concluding remarks

In summary, this thesis focussed on characteristics of severe COPD, the effects of inhaled corticosteroids (ICS) and smoking on epigenetics, and has highlighted several important findings. First, severe COPD appears to be a unique subtype of COPD with its own gene and miRNA expression profile. Furthermore, this stretches to DNA methylation, which was found to be different between COPD and non-COPD and especially related to COPD severity. Finally, we identified a considerable influence of corticosteroids on bronchial epithelium, which was largely irreversible. Future studies should investigate whether the characteristic genetic (expression) and epigenetic changes found in this thesis could predict COPD's development in its early stages and whether the epigenetic markers affected by ICS can be used to predict corticosteroid responsiveness.

9.6 – References

1. Polverino, F., A. Sam, and S. Guerra, *COPD: To Be or Not to Be, That is the Question*. Am J Med, 2019. **132**(11): p. 1271-1278.
2. Barnes, P.J., *Small airway fibrosis in COPD*. Int J Biochem Cell Biol, 2019. **116**: p. 105598.
3. Xu, X., et al., *The Heterogeneity of Inflammatory Response and Emphysema in Chronic Obstructive Pulmonary Disease*. Front Physiol, 2021. **12**: p. 783396.
4. Castaldi, P.J., et al., *Machine Learning Characterization of COPD Subtypes: Insights From the COPD Gene Study*. Chest, 2020. **157**(5): p. 1147-1157.
5. Ziyatdinov, A., et al., *Identifying COPD subtypes using multi-trait genetics*. medRxiv, 2023.
6. Young, K.A., et al., *Subtypes of COPD Have Unique Distributions and Differential Risk of Mortality*. Chronic Obstr Pulm Dis, 2019. **6**(5): p. 400-413.
7. Park, J., et al., *Subtyping COPD by Using Visual and Quantitative CT Imaging Features*. Chest, 2020. **157**(1): p. 47-60.
8. Corlateanu, A., et al., *"Chronic obstructive pulmonary disease and phenotypes: a state-of-the-art."*. Pulmonology, 2020. **26**(2): p. 95-100.
9. Soltani, A., et al., *Inhaled corticosteroid normalizes some but not all airway vascular remodeling in COPD*. Int J Chron Obstruct Pulmon Dis, 2016. **11**: p. 2359-2367.
10. Kasahara, Y., et al., *Endothelial cell death and decreased expression of vascular endothelial growth factor and vascular endothelial growth factor receptor 2 in emphysema*. Am J Respir Crit Care Med, 2001. **163**(3 Pt 1): p. 737-44.
11. Kawamoto, T., et al., *Evaluation of the severity of small airways obstruction and alveolar destruction in chronic obstructive pulmonary disease*. Respir Med, 2018. **141**: p. 159-164.
12. Kasahara, Y., et al., *Inhibition of VEGF receptors causes lung cell apoptosis and emphysema*. J Clin Invest, 2000. **106**(11): p. 1311-9.
13. Breen, E.C., et al., *Impaired pulmonary defense against Pseudomonas aeruginosa in VEGF gene inactivated mouse lung*. J Cell Physiol, 2013. **228**(2): p. 371-9.
14. Karakioulaki, M., E. Papakonstantinou, and D. Stolz, *Extracellular matrix remodelling in COPD*. Eur Respir Rev, 2020. **29**(158).
15. Annoni, R., et al., *Extracellular matrix composition in COPD*. Eur Respir J, 2012. **40**(6): p. 1362-73.
16. Greenberg, M.V.C. and D. Bourc'his, *The diverse roles of DNA methylation in mammalian development and disease*. Nat Rev Mol Cell Biol, 2019. **20**(10): p. 590-607.
17. Szymczak, I., J. Wiecefinska, and R. Pawliczak, *Molecular Background of miRNA Role in Asthma and COPD: An Updated Insight*. Biomed Res Int, 2016. **2016**: p. 7802521.
18. Szyf, M., *The early-life social environment and DNA methylation*. Clin Genet, 2012. **81**(4): p. 341-9.
19. Hou, L., D. Wang, and A. Baccarelli, *Environmental chemicals and microRNAs*. Mutat Res, 2011. **714**(1-2): p. 105-12.
20. Casas-Recasens, S., et al., *Lung DNA Methylation in Chronic Obstructive Pulmonary Disease: Relationship with Smoking Status and Airflow Limitation Severity*. Am J Respir Crit Care Med, 2021. **203**(1): p. 129-134.
21. Campisi, M., et al., *DNA Methylation-Based Age Prediction and Telomere Length Reveal an Accelerated Aging in Induced Sputum Cells Compared to Blood Leukocytes: A Pilot Study in COPD Patients*. Front Med (Lausanne), 2021. **8**: p. 690312.
22. Morrow, J.D., et al., *DNA Methylation Is Predictive of Mortality in Current and Former Smokers*. Am J Respir Crit Care Med, 2020. **201**(9): p. 1099-1109.
23. de Vries, M., et al., *From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function*. Respir Res, 2018. **19**(1): p. 212.
24. Marchesi, F., et al., *Increased survival, proliferation, and migration in metastatic human pancreatic tumor cells expressing functional CXCR4*. Cancer research, 2004. **64**(22): p. 8420-8427.

25. Yang, M., et al., *Impact of CXCR4 and CXCR7 knockout by CRISPR/Cas9 on the function of triple-negative breast cancer cells*. *OncoTargets and therapy*, 2019. **12**: p. 3849.
26. Kijima, T., et al., *Regulation of cellular proliferation, cytoskeletal function, and signal transduction through CXCR4 and c-Kit in small cell lung cancer cells*. *Cancer Research*, 2002. **62**(21): p. 6304-6311.
27. Jin, J., W.C. Zhao, and F. Yuan, *CXCR7/CXCR4/CXCL12 axis regulates the proliferation, migration, survival and tube formation of choroid-retinal endothelial cells*. *Ophthalmic Res*, 2013. **50**(1): p. 6-12.
28. Ma, D.M., D.X. Luo, and J. Zhang, *SDF-1/CXCR7 axis regulates the proliferation, invasion, adhesion, and angiogenesis of gastric cancer cells*. *World J Surg Oncol*, 2016. **14**(1): p. 256.
29. Atluri, P. and Y.J. Woo, *Pro-angiogenic cytokines as cardiovascular therapeutics: assessing the potential*. *BioDrugs*, 2008. **22**(4): p. 209-22.
30. Nagasawa, T., et al., *Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1*. *Nature*, 1996. **382**(6592): p. 635-8.
31. Isles, H.M., et al., *The CXCL12/CXCR4 Signaling Axis Retains Neutrophils at Inflammatory Sites in Zebrafish*. *Front Immunol*, 2019. **10**: p. 1784.
32. Tomlinson, J.W. and P.M. Stewart, *Cortisol metabolism and the role of 11beta-hydroxysteroid dehydrogenase*. *Best Pract Res Clin Endocrinol Metab*, 2001. **15**(1): p. 61-78.
33. Orsida, B.E., Z.S. Krozowski, and E.H. Walters, *Clinical relevance of airway 11 β -hydroxysteroid dehydrogenase type II enzyme in asthma*. *American journal of respiratory and critical care medicine*, 2002. **165**(7): p. 1010-1014.
34. Moayed, Y., et al., *The candidate splicing factor Sfsmap regulates growth and patterning of inner ear sensory organs*. *PLoS Genet*, 2014. **10**(1): p. e1004055.
35. Faiz, A., et al., *Latrophilin receptors: novel bronchodilator targets in asthma*. *Thorax*, 2017. **72**(1): p. 74-82.
36. Lev Maor, G., A. Yearim, and G. Ast, *The alternative role of DNA methylation in splicing regulation*. *Trends Genet*, 2015. **31**(5): p. 274-80.
37. Khalek, D., et al., *Alternative Splicing Is a Major Factor Shaping Transcriptome Diversity in Mild and Severe COPD*. *Am J Respir Cell Mol Biol*, 2024.
38. Di Martino, A., et al., *Collagen VI in the Musculoskeletal System*. *Int J Mol Sci*, 2023. **24**(6).
39. Francis, S.M., et al., *Expression profiling identifies genes involved in emphysema severity*. *Respir Res*, 2009. **10**(1): p. 81.
40. Kirkham, P.A. and P.J. Barnes, *Oxidative stress in COPD*. *Chest*, 2013. **144**(1): p. 266-273.
41. Lewis, B.W., et al., *Oxidative Stress Promotes Corticosteroid Insensitivity in Asthma and COPD*. *Antioxidants (Basel)*, 2021. **10**(9).
42. Neumann, M. and M. Naumann, *Beyond IkappaBs: alternative regulation of NF-kappaB activity*. *FASEB J*, 2007. **21**(11): p. 2642-54.
43. Steiling, K., et al., *The field of tissue injury in the lung and airway*. *Cancer Prev Res (Phila)*, 2008. **1**(6): p. 396-403.
44. Boudewijn, I.M., et al., *Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression*. *Respir Res*, 2017. **18**(1): p. 213.
45. Imkamp, K., et al., *Nasal epithelium as a proxy for bronchial epithelium for smoking-induced gene expression and expression Quantitative Trait Loci*. *J Allergy Clin Immunol*, 2018. **142**(1): p. 314-317 e15.

Chapter 10 – Appendices

10.1 – Acknowledgements

The past six years have been quite the ride. In 2019 I sold my apartment to start a PhD journey in Sydney and Groningen. I lost both my grandparents on my mother's side, needed to learn to live and work in a 12m² bedroom, and struggled financially as I would never like to again. In the past six years, I have lived in my apartment or with my parents in Breda, in a room in Groningen, in a room or with friends in Sydney, and with friends in Kincumber. I have visited Barcelona, Milan, and Christchurch for conferences, and Medellin, Queenstown, Melbourne, Hobart, the entire Australian East Coast, Stockholm, Coimbra and Paris. On top of that, I met hundreds of people, befriended tens of them, and will hopefully remember all of these friends throughout my life. However, the most important product of my past six years has been this thesis, the papers within it, and the teachings it entails.

The research and writing within this thesis would never have been possible without the encouragement and constant support of all my supervisors, both in the Netherlands and in Australia.

I would like to thank **Dr Alen Faiz**, who invited me to do this PhD program. I met you in 2010, when you were still a PhD student yourself, and hijacked the project that I applied to. Throughout the years, we kept in touch, especially when you came over to do a post-doc in Groningen. It is because you left Groningen that the possibility arose for me to do my PhD. Throughout the years you have been a great friend, and a great guide and help during my PhD. You taught me most of the bioinformatics I used in this project and were instrumental in writing this thesis. Thank you for your unconditional support and enthusiasm.

I am deeply grateful to **Prof Maarten van den Berge**. We first met during Alen and Anna's wedding in 2018 and had a quick chat during the reception. Then, when I came back to the Netherlands in 2019, you were there to support me and my project. That became a lot harder during the following months when COVID-19 took over the normal way of living around the world. But even then, you were there for me, and the project, as much as possible. On top of that, we worked a lot on my writing skills as well. You gave me a writing framework that I have been using since. Your expertise as a pulmonary clinician, was essential for this thesis as without it I would probably just have thought about the statistics instead. Therefore, your expertise is something I still greatly value. Thank you.

I extend my sincere gratitude to **Dr Corry-Anke Brandsma**. I first met you during one of the first days at the UMCG, and I remember feeling at ease right away. You kept me thinking about the basic research, or lab stuff, all the time. Even though I have worked in some laboratories, I have somewhat unlearned how to think like a laboratory analyst or researcher before my PhD. I would have loved to have done some of the experiments in the lab, but the circumstances were not optimal for that. Thank you for your emotional and academic support, during COVID and thereafter, and especially your academic support at the end or after my PhD. Yeah, I needed that.

My sincere thanks go to **Prof Wim Timens**. Though we did not converse much outside of the monthly meetings, your expertise has been invaluable. In the beginning, I was intimidated and did not understand many of the things you spoke about. I hope I showed you some growth in that and other regards, even if I did not in other ways. I admire your broad range of knowledge, not only in pathology, pulmonary diseases, and research, but also in technical subjects like bioinformatics, data science, and programming.

I would like to express my gratitude to **Dist Prof Brian Oliver**. You were my BAsC supervisor from 2010-2011, and again during my PhD. At the beginning of my PhD project, we had regular monthly meetings. However, you assessed correctly that having five supervisors was a bit much and put yourself more in the background. Even so, I could still count on your expertise and guidance, and even get some career advice. I could talk about my frustrations with the PhD process, and with UTS. Your optimism and groundedness are inspiring. Thank you.

Due to several factors, my PhD process has been more difficult than normal PhD projects. COVID had a profound impact on all our lives, disrupting not just our research but also our personal lives and well-being. Additionally, I faced challenges when my Australian stipend came to an unexpected end, creating significant financial strain. Without the support of several individuals, I would have been forced to cut my PhD journey short.

I am truly grateful to **Pooja** and **Vikram**. Pooja, when I reached out to you to explain the situation, your immediate response was nothing short of remarkable. You didn't hesitate for a second before inviting me into your home. Pooja and Vikram, you not only provided me with a roof over my head, but you also gave me a sense of security and belonging. You offered me a room to live in, a family to share meals with, and a peaceful environment where I could refocus and find the strength to continue. Your kindness gave me the mental and emotional space I needed to regain my balance during what was an incredibly difficult time. Additionally, it was incredibly inspiring to see you both caring for **Aarna**. Aarna, whenever you read this, it was incredible to have you around as well. It was always fun to play with you, and so joyful to see you grow. You have amazing parents, kind, wise, hard-working and inspiring. I am sure you will have an incredible bond with them.

I am also extremely grateful to **Esin** and **Bulent**. You too gave me a roof over my head, and a bed to sleep in. Esin, I enjoyed every meal you made, and every movie recommendation you gave, and am thankful for everything you did even if I did not see it. Bulent, thank you for geeking out about motorcycles, watching movies together, discussing politics and other topics. Together, you two gave me a place to call home in a beautiful environment. Hopefully, your new home is in an environment as lush as the Kincumber mountain.

I would also like to thank **Anna**, and once again, **Alen**. During the first half of my PhD, you both kindly opened your home to me as a friend. Though that time was brief and has long passed, I remember it fondly - cozy, welcoming, and always full of fun. Of course, winter brought a chill, as the rooms were often left open to air out the apartment. When I was in the Netherlands, you both started a new and exciting chapter of your life. **Lachlan** and **Charlotte** are now a beautiful addition to your family, and I have no doubt that you're both already amazing and loving parents. They are lucky to have you. You also introduced me to the inner workings of research life in a way that was both an inspiration and eye-opening. Your balance of family and academia is something I greatly admire. Thank you for being great friends.

During my PhD, I became friends with some amazing people, at both UTS and the UMCG. All of my dear new friends are a reason to consider this project a success, and any struggles are worth it.

Pooja, you were extremely open and slightly dark but very supportive to many people during your PhD. I am not sure where you got this from, but it is inspiring and a superpower. I had a great time travelling with you and the family and would do that again any day. Thank you for listening, giving advice and being an amazing friend.

Senani, you always smile, even when you are troubled. You always are busy, and slightly insecure, but always a good friend. People love you, and I think for a good reason. We both explored Barcelona, and I tried to help you out with your adventure in the Netherlands. You have an eventful life, and I hope all will be well and stable soon. Thank you for always being there, and sharing your life. I will value your friendship forever.

Lisa, you were so young when you first set foot into the office. I missed you too when I was in the Netherlands. You are a blast to have around, and an incredibly reliable friend and coworker. We went on a trip to Tasmania, and I would have loved to join you in New Zealand this past Christmas, knowing how immaculate you are in your trip planning. It is fun to play and converse with you, even though there is about 16.600km between Breda and Sydney. Thanks for being the glue to any and all friendships and staying in touch.

Neus, it was always fun to have you around. You were very insecure, but I think you were in the right place in our office and as a PhD student. You are smart, incredibly talented, emotional, an entrepreneur, and a busy bee. You just got to be a lecturer at your UTS, and with the incredible amount of students liking you, I think that it's the right place for you. I am not sure why you are uncertain or insecure about what you do or know, but I am sure you will be more than all right. I am grateful for sharing your wisdom and being a friend.

Tayyaba, you are kind but a little biased toward guys. I am glad that you have found yourself a job and stability in the city you want to live in. It is always fun listening to you, especially whenever it is nighttime in Sydney and daytime in the Netherlands. You wanted to do something with bioinformatics, and I sincerely hope that you will find a way to do that. I appreciate your kindness and the fun we had (and likely will have again). Keep on chatting at night!

Fred, during the first half a year of my PhD, we were able to chat bioinformatics, your project, mine, keyboards, programming, data science, and more. It was fun to have you around and would love to work with you again sometime in the future.

Niek, you are my paranimph and for a good reason. In the Netherlands, you were one of the first people I met and I stayed in contact with you throughout that COVID and when I left for Sydney. You are incredible at what you do, and immaculate about detail, and I am sure you are going to accomplish great things in your future. I appreciate you very much!

René, it is always fun to have you around. Thanks for showing me around Groningen, for getting me my drone back from the drench, and for being a friend throughout the years. I am grateful for having you as a friend, someone to talk to about bioinformatics, data science, programming, games and more. But also for exploring the rest of the Netherlands.

Rolf, you are a great data scientist, bioinformatician, programmer, and father by now. I am proud to have you working at my former employer, to make sure you were secure in your job. I know you are more knowledgeable about data and some algorithms. Thanks for being a friend!

Alejandra, you started out as an MD student and trainee with Machteld, but COVID changed your year in the Netherlands entirely. You could not work in the lab, as these got closed for non-essential personnel, and decided to do some bioinformatics. I have been incredibly honoured to be there for you during this time, and it has been a blast seeing you grow. I am sure you could teach me a thing or two about being a medical doctor. You are a beautiful soul, which shows in you helping others. Thank you for being my friend.

Tatiana, you are my second paranimph and I understand you have been one some times before. This means that you are appreciated by your peers, not only by me. Thank you for being there for me. I hope you will have a good time finishing your project, I am sure you will. You can do anything!

Rashad, you are as grounded as I am, which is quite refreshing. We had some good conversations about science, religion, academic culture, and more. I am grateful for all you have shown me, and I am sure you will be a great teacher to many, someday soon, even if I am not sure if that is as an academic or in another way. **Tashfia** can be proud of you, if not then I am. Keep being sceptical, and question everything!

Saima, I value your friendship, the chats we had, and the moments we shared. I loved helping you, going out with you, and exploring the city with you. I would do it all over. I should have been more open to you, even if I could not. I can see you being a beautiful mother sometime soon. I am thankful for the time we shared.

Sumaiya, I have found you to be an incredibly reliable friend and colleague, and appreciate you very much. It is inspirational to see you handling it all, and doing well. Thank you for introducing me to **Ayman** (wise, fun, and very helpful!), and the kids, and thank you for the short holiday in Portugal. I appreciate you and Ayman very much. You are almost at the end of your PhD journey, you can do it!

Fia, you nerd. I love your natural enthusiasm, and ability to understand a lot of different technologies. You were the most helpful person to most people in the group, even to me. I think you deserved the end-of-year awards you got in 2022 and 2024, though I am pretty sure it made me jealous. I am glad you have found your place at UTS and found yourself 'the one'. Thanks for listening and just being there.

Hao, it's incredibly fun to have you around in the office. You are incredibly talented and smart, and I am sure will have a good academic future for yourself. On top of that, it is great fun to have you around as our local bard. Thanks for being such a good sport in the office and a gaming buddy 16.600km away from home.

Elif, It was always great to be out with you. You were a rock in the surf, a person to count on. We could share our frustration about the PhD process, especially due to your own PhD hardship. I wish your PhD journey was easier than it has been. Regardless, I think you will be fine. You are so much stronger than I will ever be and are bound to persevere. Whatever comes your way, you can do it. I believe in you. I am deeply grateful to have you in my life and for the things you have done for me.

Ridhima, I know you will do great things, find your way up the academic ladder, and become a great researcher, and probably a professor. I know I have failed you, and hurt you by not communicating well. However, I appreciate your friendship so very much. These past two years would have been much worse without you. Know that I think and worry about you all the time.

Rochelle, dear, you find it hard to communicate over WhatsApp. Even so, you are a very kind person and friend, and very likeminded. You know you are beautiful, and I would love for you to find your place wherever you go. Let's try to catch up from time to time, ok?

We had a good time going out, exploring the city, other cities, nature, and going to festivals. I hope to stay in contact with you for the rest of my life, even if time and distance will prove to be a challenge. You are all amazing people.

Many thanks to the professors, researchers, lecturers, post-docs and collaborators in both Sydney and Groningen, as well as across the globe. Specifically, I would like to thank Prof Gerard Koppelman, Prof Peter Horvatovich, Prof Victor Guryev, Dr Corneel Vermeulen, Dr Judith Vonk, Dr Maaïke de Vries, and Dr Machteld Hylkema for your invaluable insights, guidance, and critical assertion of my work.

Thank you to the many people in our office, and my fellow PhD students in Sydney and Groningen. Specifically, I would like to thank Ben Ditz, Ben (at UTS), Daniela Gaio, Giang, Glenna, Hong, Inna, Iris, Luis, Marjam Roffel, Penny Dalla, Roderick de Hister, Roy Woldhuis, and Tessa Gilett.

I also have to thank the amazing people within the RBMB and RCMB groups. Many thanks to Antonia, Diren, Kyle, Manev, Margo, Mitchell, and Sobia, as well as Andra, Andy, Dia, Meng Wang, Min Feng, Patrick, Razia, Samaneh, and Xu Bai.

Special thanks to Alejandra Cardona Estepa, Douwe Metselaar, Kathy Phung, René Bults and Rinze-Pieter Jonker. I could not have asked for any better students. Kathy, special thanks for all your hard work on the DNA methylation project, I would have loved to help you more. Ask questions, bother me or your fellow group members. If we can, we will make time for you. So, Always.

Thank you for being my friends, Carl, Charles, Dale, Gido, Gustavo, Matt, Nadine, and Xandra. I should write a lot more about what every one of you means to me, but I think you all know.

Lastly, I would like to thank my family.

Thank you, Marieke, Anne, Thijs, Dennis, Johan, Eline, Brian, Jayden, Milan and Yinthe. Even though I was on the other side of the globe, and hard to keep in touch, we did and found new ways to do so. Thank you, mom and dad, for everything.