Functional Genetics of Lung Function Associated Gene *GPR126*

By

Robert J Hall BSc, MRes

Student ID number: 4253757

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

November 2019



University of Nottingham

UK | CHINA | MALAYSIA

Abstract

GPR126 lies on chromosome 6q24 and variants in this region are reproducibly associated with lung function and COPD in genome wide association studies (GWAS). This association signal includes a non-synonymous single nucleotide polymorphism (SNP), rs17280293 (minor allele frequency = ~2.9%) which results in the amino acid substitution Ser123Gly within the GPR126 coding region. GPR126 is a G protein coupled receptor belonging to the adhesion family of receptors. The aim of the work described in this thesis was to define the role of GPR126 in the human lung and in pulmonary disease. My specific aims were to i) identify the GPR126 expression profile in the lung and in airway structural cells ii) determine potential causal variants within the association signal iii) define the effects of missense variant Ser123Gly on GPR126 function iv) determine the function of GPR126 signalling in human airway structural cells.

Linkage disequilibrium (LD) and expression quantitative trait loci (eQTL) analyses were used to determine potential causal variants within the 6q24 association signal. GPR126 mRNA expression in the lung, human bronchial epithelial cells (HBEC) and human airway smooth muscle cells (HASM) was established using qPCR and RNA-seq. GPR126 protein expression was characterised in lung tissue using immunohistochemistry. A recombinant CRE-luciferase reporter cell line was engineered to express the Ser123 and Gly123 variants of GPR126. These cell lines were treated with potential GPR126 agonists and cAMP driven gene expression was measured by luciferase activity. These agonists included the stachel peptide which is a short amino acid sequence in the N-terminal region of the receptor which can activate receptor signalling. Column chromatography was used to measure GPR126 driven cAMP accumulation in HBECs and HASMs in response to activation with the stachel peptide. To assess the downstream effects of GPR126 activation in human airway cells, HASMs were activated with the stachel peptide in the presence and absence of PKA inhibitor, H89, to identify the PKA dependency of this response. Differential gene expression was

analysed by RNA-seq. Pathway analysis was carried out with GSEA from the Broad Institute.

GPR126 mRNA is expressed in the lung, HBEC and HASMs and GPR126 protein expression was detected in human lung. The stachel peptide activates GPR126 in a recombinant cell system, although this signal was not genotype-dependent. Moreover, the stachel peptide also activates endogenous GPR126 in HASMs but not HBECs.

GPR126 activation of HASMs resulted in 514 and 520 differentially expressed genes at 4 and 24 hours post-treatment respectively (5% FDR). There was approximately 20% (61 upregulated and 121 downregulated) overlap between these two time points. This sustained response was partially cAMP-PKA dependent with ~40% of genes no longer differentially expressed with H89 pre-treatment at both time points. Genes upregulated by GPR126 activation are involved in inflammatory pathways (IL33 and ADAM8) and downregulated genes include CTGF and SERPINE1, which both have known roles in fibrosis. Gene set enrichment analysis revealed that beta defensin pathways were upregulated by GPR126 activation where pathways involved in cancer, cell cycle checkpoints and cell proliferation were downregulated.

GPR126 expression was identified in the human lung and airway structural cells. While a potentially causal variant was identified in Ser123Gly no potential effects of this mutation on GPR126 signalling were observed. GPR126 signals in human airway smooth muscle cells and the downstream effects of GPR126 activation were assessed by RNA-seq. HASMs respond to GPR126 agonist and this response has partial PKA dependency implicating other pathways driving gene expression differences we observed. Taken together, this work describes novel biology for GPR126 implicating a role for GPR 126 in regulating airway inflammation and airway remodelling.

Acknowledgments

Firstly, I would like to thank my supervisors, Professor Ian Hall and Professor Ian Sayers for giving me the opportunity to study for my PhD and for their continued support, advice and guidance throughout the duration of this project.

Thanks also goes to the entire Department of Respiratory Medicine for training, advice and friendship over the past 4 years. Thank you to Mike Portelli, Amanda Henry, Charlotte Billington and Suzanne Miller for time spent training me in many of the techniques used throughout this thesis as well as general advice. Thanks also goes to everyone in the office for making it a genuinely nice environment to come in to everyday.

I would also like to thank my parents for all their unwavering support (financial and otherwise!). Thank you for your patience while I was living at home after completing my undergraduate degree and supporting my move down to Nottingham (not to mention the wedding and house purchase!).

Finally, thank you to my wife, Hatty, for moving to Nottingham with me, agreeing to marry me, and for making home a haven from the PhD stress. Also thanks for stopping asking me when I was going to submit!

Clear eyes. Full hearts. Can’t lose

Table of Contents

Functional Genetics of Lung Function Associated Gene *GPR126*.....i

Abstract.....ii

Acknowledgments.....iv

Table of Contents.....vi

Abbreviations.....xiii

1 General Introduction.....1

1.1 Chronic Obstructive Pulmonary Disease.....3

1.1.1 COPD Pathology.....4

1.1.2 COPD Pathogenesis.....6

1.1.3 Lung Function and COPD diagnosis.....7

1.1.4 Heritability of Lung Function and COPD.....8

1.2 Identification of Genes for COPD and Lung Function.....10

1.2.1 Early Genetics – Linkage Studies and Candidate Gene Studies 10

1.2.2 Genome-Wide Association Studies (GWAS).....11

1.2.3 Chromosome 6q24 Association Signal.....23

1.3 GPR126/ ADGRG6.....24

1.3.1 G-Protein Coupled Receptors.....26

1.3.2 Adhesion G-Protein Coupled Receptors.....29

1.3.3 GPR126 Signals via cAMP and is Important for a Number of Developmental Pathways 31

1.3.4 Human traits associated with GPR126 variants 35

1.4 Hypothesis and Aims 36

2 General Materials and Methods 37

2.1 Mammalian cell culture methods 38

2.1.1 Culture of GloResponse™ CRE-luc2P HEK293 Cell Line 39

2.1.2 Culture of primary human airway cells 40

2.1.3 Trypsinisation and Passaging Cell Cultures 44

2.1.4 Cyopreservation of Mammalian Cells 44

2.1.5 Cell Counting using a Haemocytometer and Assessment of Cell Viability 45

2.1.6 Transfections using Fugene 6 45

2.1.7 Luciferase Assays 46

2.1.8 ^{3}H-cAMP accumulation assays by column chromatography 47

2.2 Molecular biology methods 50

2.2.1 Total RNA extraction from mammalian cell cultures 50

2.2.2 DNase I pre-treatment of total RNA 51

2.2.3 cDNA synthesis from total RNA 51

2.2.4 Polymerase Chain Reaction 53

2.2.5 Agarose Gel Electrophoresis 54

2.2.6 Quantitative-PCR (Taqman) 55

2.2.7 Plasmid Preparations 57

2.2.8 Nucleic Acid Quality Control 60

2.2.9 Sanger Sequencing 61

2.2.10 RNA sequencing 62

2.3 Protein Detection methods 64

2.3.1 Immunohistochemistry 64

2.4 Bioinformatics 66

2.4.1 Processing Sequencing data 66

2.5 Statistical analysis 70

2.5.1 Analyses of Variance 70

3 Bioinformatic Analysis of the 6q24 Association Signal and Expression Profiling of *GPR126* mRNA and Protein in Human Lung and Airway Cells .. 71

3.1 Introduction 72

3.1.1 Aims 78

3.2 Methods 79

3.2.1 Bioinformatic analysis of rs148274477 signal 79

3.2.2 GPR126 expression in human cells and tissue by quantitative-PCR 80

3.2.3 Identification of GPR126 Expression in Human Airway Cells by RNA Sequencing 81

3.2.4 GPR126 expression in GEO datasets 82

3.2.5 GPR126 Protein Expression in Human Lung Tissue 83

3.3 Results 85

3.3.1 Missense Variant Ser123Gly is in Linkage Disequilibrium with Sentinel SNP rs148274477 85

3.3.2 SNPs are not eQTLs for GPR126 or Other Genes 90

3.3.3 Ser123Gly (rs17280293) is Predicted to be Damaging to Protein Function. 90

3.3.4 GPR126 mRNA is expressed in the Lung and Airway Relevant Cells 92

3.3.5 GPR126 Expression in GEO RNA-seq Datasets 93

3.3.6 RNA Sequencing Quality Control Results 94

3.3.7 RNA-sequencing Data of GPR126 mRNA Expression in HASM and HBEC 94

3.3.8 Immunohistochemistry of GPR126 protein expression in lung tissue 104

3.4 Discussion 109

4 Investigation of GPR126 Ligand Specificity and Effects of Ser123Gly on cAMP Signalling in a Recombinant Cell System 114

4.1 Introduction 115

4.1.1 Aims 122

4.2 Methods 123

4.2.1 Plasmid Production and Preparation 123

4.2.2 Cell Biology Methods 130

4.2.3 Characterisation of transfected cells by qPCR 141

4.2.4 cAMP Accumulation Assay on pLV-GPR126 and pLV-EV Transfected Cells Treated with Stachel Peptide 141

4.2.5 Statistical Analyses 142

4.2.6 Column statistics 142

4.3 Results 144

4.3.1 Optimisation of the CRE-Luc2P Assay with Isoprenaline and Forskolin 144

4.3.2 Plasmid DNA Quality Control 145

4.3.3 GPR126 Signalling in CRE-luc2P HEK293 Cells Transiently eExpressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV. 148

4.3.4 Production of a Mixed Population of Cells Stably Expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV. 151

4.3.5 Establishing GPR126 Agonists by Luciferase Assays in a Mixed Population of Cells Expressing GPR126 154

4.5 Discussion 166

5 GPR126 signalling in primary human airway cells 176

5.1 Introduction 177

5.1.1 Aims 180

5.2 Methods 181

5.2.1 Culture of Human Airway Smooth Muscle (HASM) Cells for ^{3}H-cAMP Accumulation Assays and CRE-luciferase Transfections 181

5.2.2 Culture of Human Bronchial Epithelial Cells (HBEC) for ^{3}H-cAMP Accumulation assays. 181

5.2.3 cAMP assays using column chromatography 181

5.2.4 Preparation and transfection of the CRE-luciferase plasmid (p6CRE/luc) 182

5.2.5 CRE-luciferase assays in HBECs and HASMs 182

5.2.6 Statistical analyses 183

5.3 Results 184

5.3.1 Cyclic-AMP Accumulation in Primary Human Airway Cells Treated with the Stachel Peptide 184

5.3.2 CRE-luciferase transfections in primary human airway cells .. 188

5.3.3 Adhesion GPCR expression in primary human airway cells ... 190

5.4 Discussion 193

6 Analysis of GPR126 Downstream Signalling and Potential Functions using Transcriptomics 197

6.1 Introduction 198

6.2 Methods 201

6.2.1 Cell Biology Methods 201

6.2.2 Molecular Biology Methods 202

6.2.3 Analysis of RNA-seq data 203

6.3 Results 208

6.3.1 Quality Control of Samples for RNA-sequencing 208

6.3.2 FastQC Quality Control of Sequencing Data 210

6.3.3 GPR126 Expression Determined by Cuffnorm 210

6.3.4 Differential gene expression in HASMs in response to stachel 211

6.3.5 Pathway based analyses identifies a potential role for GPR126 in inflammation and control of cellular proliferation 230

6.3.6 The stachel mediated response is partly cAMP-PKA dependent 235

6.4 Discussion 241

7 General Discussion 250

7.1 Hypothesis and Aims 251

7.2 Main Findings 253

7.3 Study Limitations 256

7.4 Future Work 258

7.5 Final Conclusions 261

8 Appendices 263

8.1 List of reagents 264

8.1.1 General Reagents 264

8.1.2 Cell Culture Reagents 266

8.2 RNA-sequencing Quality control 267

8.3 Linux Scripts for RNA-sequencing Analysis 273

8.3.1 Processing Sequencing data 273

8.4 R scripts used for visualisation of RNA-seq data 277

8.4.1 Parsing files for GSEA 277

8.4.2 Volcano Plots in R studio 277

8.4.3 Heatmaps in R studio 279

References 280

Abbreviations

2-ME	2-mercaptoethanol
7TM	7 Transmembrane spanning
A1AT	Alpha 1 Antitrypsin
AC	adenylyl cyclase
aGPCR	Adhesion G Protein Coupled Receptor
AMC	Arthrogryposis Multiplex Congenita
ANOVA	Analysis of Variance
ASM	Airway smooth muscle
ATF1	Activation transcription factor 1
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
cAMP	Cyclic adenosine monophosphate
CAT	COPD assessment test
cDNA	Complementary DNA
CF	Cystic Fibrosis
CNGC	cyclic nucleotide gated ion channel
COPD	Chronic Obstructive Pulmonary Disease
CRE	cAMP response elements
CREB	cAMP response binding element
CREB	cAMP response element-binding protein
CREM	cAMP response element modulation
CTF	C-terminal fragment
CT	Cycle threshold
CUB	complement C1r-C1s, Uegf and Bmp1
DAB	3,3'-Diaminobenzidine
DEPC	Diethyl pyrocarbonate
DMEM	Dulbecco's Modified Eagles Medium

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNPT	Deoxynucleotide triphosphates
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPAC	Exchange protein directly activated by cAMP
eQTL	Expression quantitative trait loci
EV	Empty Vector
FCS	Foetal calf serum
FEV1	Forced expiratory volume in 1 second
FPKM	Fragments Per Kilobase Million
FVC	Forced vital capacity
GEO	Gene Expression Omnibus
GPCR	G protein coupled receptor
GPS	GPCR proteolysis site
GSEA	Gene set enrichment analysis
GWAS	Genome wide association studies
HASM	Human airway smooth muscle
HBEC	Human bronchial epithelial cell
HCL	Hydrochloric acid
HEK	Human embryonic kidney
LABA	Long acting beta agonist
LAMA	long-acting muscarinic antagonist
LD	Linkage disequilibrium
LOF	loss of function
MAF	Minor allele frequency
mMRC	Modified British Medical Research Council
NH4OH	Ammonium hydroxide
NTF	N-terminal fragment

PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PDAR	Pre developed assay reagent
PDE	Phosphodiesterase
PKA	Protein Kinase A
Poly-A	Polyadenylated
PPA	Posterior probability of association
PTX	Pentraxin
qPCR	quantitative PCR
RIN	RNA Integrity Number
RLU	Relative light units
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Reverse transcription
SABA	S1hort acting beta agonist
SFM	Serum free media
SNP	Single nucleotide polymorphism

1 General Introduction

Chronic obstructive pulmonary disease (COPD) is a world leading cause of morbidity and mortality with no cure currently available. Treatments can alleviate symptoms yet there is an unmet need for therapeutic interventions that can halt or reverse disease progression. The well-understood risk factors for disease development are environmental, with cigarette smoking among the leading causes of COPD. However, whilst there is a known genetic component of the disease the mechanisms underlying genetic effects are mostly unclear.

Genome wide association studies (GWAS) identify associations between genetic variants and human traits such as lung function or disease states like COPD. For lung function, there are hundreds of loci which have shown association to date. Understanding the genes underlying these associations in health and disease is important for understanding disease mechanisms and in identifying therapeutic targets. There is now a need to translate GWAS association signals to biological understanding of gene function in pulmonary health and disease.

GPR126 lies on chromosome 6q24, a region which harbours an intronic variant rs148274477 which has reproducibly been associated with FEV_1/FVC. *GPR126* codes for an adhesion g-protein coupled receptor and its role in the human lung is currently unknown. However it is known to be important for a number of developmental processes such as nerve development and cardiovascular development in mammals. This thesis aimed to explore the role of GPR126 in the lung and in structural airway cells and examines the potential downstream effects of GRP126 signalling in human airway cells.

1.1 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory disease which results in airflow obstruction in the lungs. Individuals with COPD suffer shortness of breath, persistent cough with more phlegm than usual, wheezing and frequent chest infections. These symptoms gradually increase in severity as the disease progresses over time and COPD patients experience exacerbations where symptoms get markedly worse for a period of time. Airflow obstruction in COPD is caused by irreversible and progressive damage to the airways usually presenting as either emphysema (damage to the alveoli) and/or chronic bronchitis (inflammation of the bronchial airways). [1, 2] Worldwide in 2015, COPD resulted in the death of approximately 3.2 million people while affecting 174.5 million individuals in total. [3]

Treatment for COPD can reduce symptoms and the frequency and severity of exacerbations and in doing so can increase exercise capacity and improve overall health. However, none of the current treatments for COPD can modify disease progression or prevent the progressive decline in lung function. [4] The main pharmaceutical interventions usually consist of inhaled steroids to reduce inflammation and bronchodilators which reduce bronchoconstriction. Inhaled corticosteroids work by reducing inflammation in the lung and have been shown to successfully reduce the rates of COPD exacerbation and reduce the rate of progressive airflow decline in COPD patients. [5, 6] There are three types of bronchodilators commonly used to treat COPD: β-agonists, anticholinergics and methylxanthines. Anticholinergics inhibit bronchoconstriction and mucus secretion. [7] They have been shown to improve symptoms and health status resulting in reduced exacerbations and related hospitalisations. [8] β_2-agonists, including short-acting (SABA) and long-acting (LABA) drugs are used to relax airway smooth muscle and reduce bronchoconstriction. Studies suggest that long-acting antimuscarinic antagonists (LAMA) have a greater impact on rate of exacerbations compared to LABAs. [9] Methylxanthines have been shown to improve lung function and breathlessness when used in combination with salmeterol and exhibit a bronchodilator effect in stable COPD. [8] Lifestyle changes such as smoking cessation, which is the main intervention for

preventing COPD, are also important. Patients with mild COPD who quit smoking have a higher lung function and higher survival rate compared to those who persist. [10] The lack of effective treatments for COPD highlights that further insight is needed into disease mechanisms in order to develop more effective therapies which can halt or reverse disease progression.

COPD is a heterogeneous disease with both environmental and genetic factors contributing to disease development and progression. Smoking tobacco is well known as the major risk factor for developing COPD, however only 10-20% of heavy smokers develop the disease. In developing countries, the burning of biomass is a major cause of COPD particularly in women who tend to cook over indoor stoves with little or no ventilation.[11, 12] While it is known environmental factors play an important role in COPD, it is also clear that there are populations who are genetically more susceptible to developing COPD than others.

1.1.1 COPD Pathology

A summary of the pathological changes seen in the COPD lung are presented in Table 1-1. COPD is characterised by irreversible airway obstruction and an abnormal inflammatory response in the lungs. This inflammatory response is a results of long term exposure to noxious particles and gasses like those in cigarette smoke. While an immune response to these particles is normal, the immune response in COPD individuals is abnormal and/or enhanced.[13] This inflammatory signature is characterised by increased numbers of neutrophils, macrophages and T-lymphocytes which the release of cytokines and immune mediators which contribute to disease progression. [13]

Proximal Airways	• Increased numbers of macrophages and CD8 T lymphocytes • Increased neutrophils and eosinophils during exacerbations. • Submucosal bronchial gland enlargement and goblet cell metaplasia • Cellular infiltrates (neutrophils and lymphocytes) of bronchial glands • Airway epithelial squamous metaplasia, ciliary dysfunction, hypertrophy of smooth muscle and connective tissue
Peripheral Airways	• Increased numbers of macrophages and T lymphocytes • Increased numbers of B lymphocytes, lymphoid follicles, and fibroblasts • Few neutrophils or eosinophils • Bronchiolitis at an early stage • Luminal and inflammatory exudates • Pathological extension of goblet cells and squamous metaplasia into peripheral airways • Peribronchial fibrosis and airway narrowing with progressive disease
Lung Parenchyma	• Increased numbers of macrophages and CD8 T lymphocytes • Alveolar wall destruction from loss of epithelial and endothelial cells • Development of emphysema • Microscopic emphysematous changes: dilatation and destruction of respiratory bronchioles; destruction of the whole acinus • Macroscopic emphysematous changes: Microscopic changes progress to bulla formation
Pulmonary Vasculature	• Increased numbers of macrophages and T lymphocytes • Intimal thickening, endothelial dysfunction • Hypertrophy of vascular smooth muscle, collagen deposition, destruction of capillary bed, development of pulmonary hypertension

Table 1-1 Pathological features observed in COPD lungs. [13, 14]

The two primary sub-categories of COPD are emphysema and chronic bronchitis. Individuals with chronic bronchitis have a persistent cough and increased sputum production, which is a result from an innate immune response to inhaled toxic particles. [14] This is characterised by inflammation in the epithelium of the airways and in mucus-producing glands resulting in increased mucus production, reduced mucociliary clearance and increased permeability of the epithelial barrier. [1, 14] Emphysema is the enlargement of the distal airspaces beyond the terminal bronchioles as a result of airway wall destruction. [15] This lung destruction reduces the maximal expiratory airflow by decreasing the elasticity of the lung, meaning that the force that drives air from the lung is diminished. [14]

1.1.2 COPD Pathogenesis

In COPD, inflammation leads to tissue destruction, impairment of defence mechanisms that would normally prevent or limit such destruction and disruption of repair mechanisms. The inflammatory and structural changes in the airways observed in COPD increase with disease severity and persist even after smoking cessation. [13]

The abnormal inflammatory response in COPD is characterised by increased numbers of neutrophils, macrophages and T-lymphocytes which release pro-inflammatory cytokines and mediators that contribute to the disease process.[13] This response can result in a number of pathological features include mucus hypersecretion, tissue destruction and disruption of repair and defence mechanisms resulting in chronic inflammation and fibrosis of the lung. [13] These changes result in an increased resistance to airflow in the airways of the lung, increase compliance of the lungs and air trapping. [13]

Aside from inflammation, two other processes are important in the pathogenesis of COPD, protease activity and oxidative stress. A shift towards increased production/ activity of proteases (or reduced production/ inactivation of antiproteases) also contributes to disease pathology. The major proteases involved in the COPD process are produced by neutrophils and macrophages

as well as matrix metalloproteases.[13] Antiproteases involved in COPD include α1 antitrypsin, deficiency of which is a known cause of emphysema.[13]

Cigarette smoke as well as reactive oxygen species released from inflammatory cells are sources of oxidants in the lung. Markers of oxidative stress are increased in COPD and further increased during exacerbations.[13] Oxidative stress contributes to COPD by a number of possible methods. It can lead to inactivation of anti-proteases resulting in an imbalance between protease and anti-protease activity, stimulation of mucous production and enhanced expression of pro-inflammatory mediators.[13] These pathogenic mechanisms in the lung result in physiological abnormalities including mucous hypersecretion, ciliary dysfunction, airflow obstruction, hyperinflation, gas exchange abnormalities and pulmonary hypertension as well as other systemic effects.[13]

Structural changes in the lung are often referred to as airway remodelling. [16] Tissue remodelling describes a number of changes in tissue structure resulting from processes such as cell proliferation, cell volume increase and modification of the composition of the extracellular matrix (ECM). [17]

1.1.3 Lung Function and COPD diagnosis

Lung function measurements are key to the diagnosis of COPD and are measured by spirometry. The most commonly used measures of lung function are forced expiratory volume in one second ($FEV_{1,}$), the volume of air that can forcibly be blown out in one second; forced vital capacity (FVC), the volume of air that can forcibly be blown out after full inspiration; and the FEV_1/FVC ratio.

Spirometry is used to define the severity of airflow limitation with a post-bronchodilator FEV_1/FVC ratio of <0.70 commonly considered diagnostic for COPD. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) system categorises airflow limitation into 4 stages from mild to very severe, shown in Figure 1-1. The GOLD guidelines uses a combined approach to categorise patients according to symptom assessment and previous history of exacerbations using the modified British Medical Research Council (mMRC) or COPD assessment test (CAT) scale. A summary is shown in Figure 1-1 with

group A being low risk (0-1 exacerbations per year, not requiring hospitalisation) to group D, which is high risk (≥2 exacerbations per year, or one or more requiring hospitalisation and more symptoms)

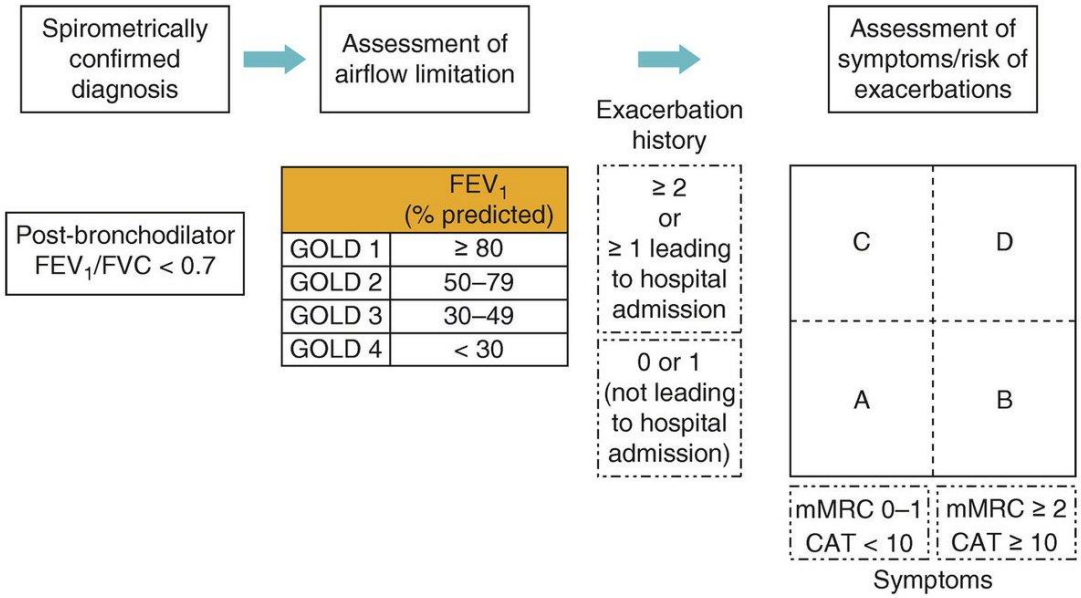


Figure 1-1 GOLD criteria for COPD diagnosis taking into account lung function measurements, symptom assessment and severity/ frequency of exacerbations. [18]

As lung function measures essentially define the severity of COPD they make ideal quantitative traits to use when studying the genetics of COPD in association studies. This approach, instead of a case/control approach, allows the use of greater numbers of individuals which provides increased statistical power to discover new genes with sufficient confidence.

1.1.4 Heritability of Lung Function and COPD

Twin studies in the 1980s estimated that heritability of FVC and FEV_1 could be as high as 90%, suggesting that variation in lung function is primarily due to genetic factors.[19, 20] More recent studies have estimated that heritability of lung function measures can range anywhere from 10%-80%.[21-25]

Studies to identify the genetic factors which may predispose an individual to developing COPD have shown considerable success with hundreds of genetic variants now identified as associated with lung function and COPD. Identification of genetic variants and disease genes adds to our understanding of disease mechanisms and is important for a number of reasons including; i) to identify risk alleles which may help with early diagnosis and possible disease prevention ii) the development of novel therapeutic interventions and iii) the identification of subgroups of patients who are most likely to gain benefit from particular therapies, an approach usually referred to as personalised or precision medicine. This information can also help to unravel the complex gene-environment interactions that lead to disease.

1.2 IDENTIFICATION OF GENES FOR COPD AND LUNG FUNCTION

The first gene identified to be associated with COPD was *SERPINA1* which codes for α1-antitrypsin (A1AT). A1AT deficiency, caused by mutations in A1AT, affects up to 1 in 5000 individuals in Europe and these individuals are more susceptible to developing emphysema.[26] This deficiency was first identified in 1963 by Laurell and Eriksson who described individuals who have approximately 10% of normal A1AT levels in their serum. While some individuals with A1AT deficiency have emphysema, only about 1-3% of COPD sufferers have the A1AT deficiency. [27] Therefore, despite the fairly well understood genetic and molecular causes of COPD in this small subset of patients, it is still largely unknown what the underlying disease mechanisms in the majority of sufferers are.

1.2.1 Early Genetics – Linkage Studies and Candidate Gene Studies

Even given recent success, identifying genes and variants involved in respiratory disease has been challenging. This is in part due to the heterogeneity in patient phenotypes and in part due to the likelihood that complex, common diseases are driven by a combination of common genetic variants each contributing small phenotypic effects combined with a significant environmental component. This is in contrast to monogenic diseases such as cystic fibrosis (CF) where a mutation in one gene results in the disease. This means that variation in phenotype is likely due to a combination of different genotypes making unravelling genetic causes of disease difficult.

Linkage analysis is a method of genetic mapping which allows the identification of genetic variants that co-segregate with disease through generations based on their chromosomal location. Linkage analysis studies are highly effective in identifying genetic variants causing monogenic disorders - CF, for example. However, this method has proved less useful for heterogeneous diseases like COPD due to the lack of sufficient power and the complex genetic basis of these diseases. In addition, due to the nature of COPD being a late-onset

disease, it is challenging to find many families with the disease in multiple generations. Despite these limitations, there were a number of studies which found regions of linkage with lung function and COPD phenotypes on chromosomes 2, 4, 6, 8, 12 and 19.[28-32]

Candidate gene studies are a hypothesis driven approach in which genes postulated to impact disease, due to known or hypothesised role in disease biology and/or chromosomal location, are investigated. For COPD, over 100 genetic studies have been carried out identifying a number of potentially important genes such as MMP12. [33] However, these studies have been limited by power and lack of replication of findings. [34]

These limitations ushered in the era of genome wide association studies (GWAS), which have been much more effective at identifying disease genes. The genotyping technology used in these studies allows the assaying of hundreds of thousands of single nucleotide polymorphisms (SNPs) in thousands of individuals, providing much more statistical power than previous genetic studies. However, this comes with caveat that much greater sample sizes are required to find loci of significance due to high statistical stringency and a need for replication of findings.

1.2.2 Genome-Wide Association Studies (GWAS)

In 2005, DeWan and colleagues performed a GWAS identifying a variant in the promoter region of *HTRA1* to be significantly associated with macular degeneration. [35] Since this first proof of concept study, GWAS have become highly effective for the discovery of disease genes across a large number of complex traits and diseases due to the rapid advances in genotyping technology and projects such as the Human Genome Project and the HapMap project. [36-39]

The basic concept of GWAS is that hundreds of thousands of common single nucleotide polymorphisms (SNPs) (generally >5% allele frequency) are tested for association with a specific trait, for example a lung function phenotype such as FEV_1 or a disease category such as asthma. A typical GWAS utilises a case-control design whereby the allele frequency of SNPs in the cases

(individuals with the trait or disease) is compared to the allele frequency of SNPs in the controls. However, lung function is usually studied as a quantitative trait and therefore is not generally studied in a case/ control design.

Individuals are genotyped on a SNP array, which is a chip-based microarray technique used to assay hundreds of thousands to millions of SNPs. Currently, there are two platforms commonly used for GWAS, which differ slightly in their methods to detect SNP variation in genomic DNA. Affymetrix uses an approach by which short DNA sequences are printed on a spot on the chip which recognises a specific allele by hybridisation. Alternatively, the Illumina platform uses a bead-based approach with slightly longer DNA sequences to detect specific alleles. [40]

A standard GWAS involves examining association of >500,000 SNPs spanning the entire genome and as well as the platform, the individual SNPs to be assayed is an important consideration as it is not currently feasible to assay all known SNPs in the human genome. In order to capture the maximum amount of information possible, panels are used that contain SNPs that tag haplotype blocks. A haplotype block is a group of SNPs that are statistically likely to be inherited together. Therefore, by choosing to include SNPs on the panel which tag haplotype blocks, it is possible to infer the genotype for many more SNPs than those which are actually on the panel. Other considerations for SNP selection generally are specific to the type of study. For example, the UK BiLEVE custom array was designed to 1) measure rare coding variants (variants which alter amino acid sequences); 2) provide a framework which would allow optimum imputation of non-genotyped common or low frequency variants (therefore increasing coverage) and 3) optimise coverage of genes and loci which have already been linked to respiratory health and/or disease (allows fine mapping of regions of interest). [41]

The allele frequency of each SNP is tested for association with the trait in question, requiring a stringent level of statistical significance due to the large number of tests. Typically $P<5x10^{-8}$ is considered significant to show genome-wide association. This stringent threshold is important to avoid the detection of

false positives, which is more likely when there are a large number of variants (and therefore hypotheses) being tested. A genome-wide scan of common variants is equivalent to approximately 1 million independent tests, hence this value represents Bonferroni correction. [42, 43] This level of stringency therefore requires large sample sizes for a signal to reach genome-wide significance. Replication of findings in an independent cohort is also standard practice to validate findings. Once SNPs are validated for association, further *in silico, in vitro* and *in vivo* studies required to explore the biological mechanisms underlying the association.

The advantage of GWAS is that they are mostly hypothesis free in the sense that associations are searched for over the entire genome. However, they are limited in that association signals are not directly informative with respect to the biology underlying the association signal and therefore further analysis and functional genetics is required to elucidate the causal variant(s) and mechanisms underlying the signal requiring further *in silico, in vitro* and *in vivo* studies. [44]

Early GWAS of lung function showed some success in identifying a number of loci associated with lung function and COPD. These loci include *HHIP* at the 4q31 locus, *CHRNA3/5* at 15q25 and *FAM13A* at 4q22. [45-47]

1.2.2.1 Meta-Analyses and Imputation

While early GWAS showed great success in identifying novel loci associated with lung function and COPD these studies made clear that greater statistical power and consequently, greater population sizes, would be required to identify genetic variation associated with modest effect sizes. As a result of these studies, meta-analyses were used to analyse the results of independent GWAS together thereby increasing the statistical power for novel loci discovery.

Essential for meta-analyses, imputation describes a method whereby genetic variants which were not directly genotyped can be predicted with a degree of confidence. This is done using reference genomes which are now widely available thanks to studies such as the 1000 genomes, 10,000 genomes and

the HapMap projects. [48] Imputation gives increased power because the reference panel is more likely to contain the causal variant than the original array while also facilitating fine-mapping of the region by giving a high-resolution overview of an association signal over a locus. In addition, imputation allows GWAS typed with different arrays to be combined to the reference panel, which is essential to combine independent GWAS for meta-analyses.

1.2.2.2 GWAS for Lung Function

In 2010, the SpiroMeta and Charge consortium published back to back papers describing meta-analyses of GWAS on FEV_1 and FEV_1/FVC. [49, 50] These two studies had significantly greater numbers of subjects than previously used with the SpiroMeta study using 20,288 individuals for discovery analysis. This was followed up by a meta-analysis of 32,184 genotyped individuals with *in silico* association data from the Charge consortium's 21,092 individuals and an additional 883 individuals from the Health 2000 study. [50] The SpiroMeta study replicated the association locus at 4q31 (*HHIP*) and identified five additional novel loci associated with either FEV_1 or FEV_1/FVC in or near *GSTCD, HTR4, AGER, THSD4*. [50] The Charge meta-analyses included 20,890 participants of European ancestry in the discovery phase and 20,228 individuals from the SpiroMeta consortium in the replication cohort. [49] This study identified another 9 loci associated with lung function. *HHIP, GPR126, ADAM19, AGER-PPT2, FAM13A, PTCH1, PID1* and *HTR4* were associated with FEV_1/FVC while the one locus was associated with FEV_1 was *INTS12-GSTCD-NPNT*, replicating results from the SprioMeta study. [49, 50]

Since these first meta-analyses of GWAS of lung function, the scale of the studies has increased significantly with respect to both the number of subjects and the number of markers and this has facilitated the growth in the number of signals found. For example, following up these studies in 2011, a meta-analyses of 23 lung function GWAS from the SpiroMeta and CHARGE consortiums was published which used a total of 48,201 individuals. The result was the identification of 16 additional novel loci for lung function including 12 novel loci for FEV_1/FVC, 3 novel loci for FEV_1 and 1 locus for both traits. [51] In

addition to the novel loci, the study provided further evidence for ten previously identified loci. [51]

In more recent years, meta-analyses using SpiroMeta and CHARGE GWAS was performed by Loth and colleagues in 2014, this time focused to forced vital capacity (FVC). This analysis used 52,253 individuals from 26 studies, following up the top associations in 32,917 individuals. They identified six novel loci associated with FVC at *EFEMP1, BMP6, MIR129-2-HSD17B12, PRDM11, WWOX* and *KCNJ2* as well as association with known signals at *GSTCD* and *PTCH1*. [52] While most studies have focused to cross sectional lung function, a study of longitudinal change in adult lung function provided suggestive evidence for association at novel loci *IL16/STARD5/TMC3* on chromosome 15 and *ME3* on chromosome 11. [53]

A further sixteen novel signals were identified in a 2015 SpiroMeta paper in or near *ENSA, RNU5F1, KCNS3, AK097794, ASTN2, LHX3, CCDC91, TBX3, TRIP11, RIN3, TEKT5, LTBP4, MN1* and *AP1S2* including novel signals in known loci *GPR126* and *NPNT*. [54] A study around the same time in the UK Biobank cohort identified six novel genome-wide significant signals of association with extremes of FEV_1. This study selected individuals from the middle and extremes of the FEV_1 distribution among heavy smokers and never smokers. These samples were selected from the UK Biobank cohort with 50,008 samples selected in total. Samples were taken from both heavy smoker and never smoker groups: 10,002 with low FEV_1, 10,000 with average FEV_1 and 5,002 with high FEV_1 were chosen from both of the never smoker and heavy smoker groups. [41] Signals of association were found at 4 novel loci; *KANSL1, TSEN54, TET2* and *RBM19/TBX5* and two novel signals from previously reported loci *NPNT* and *HLA-DQB1/HLA-DQA2*. [41]

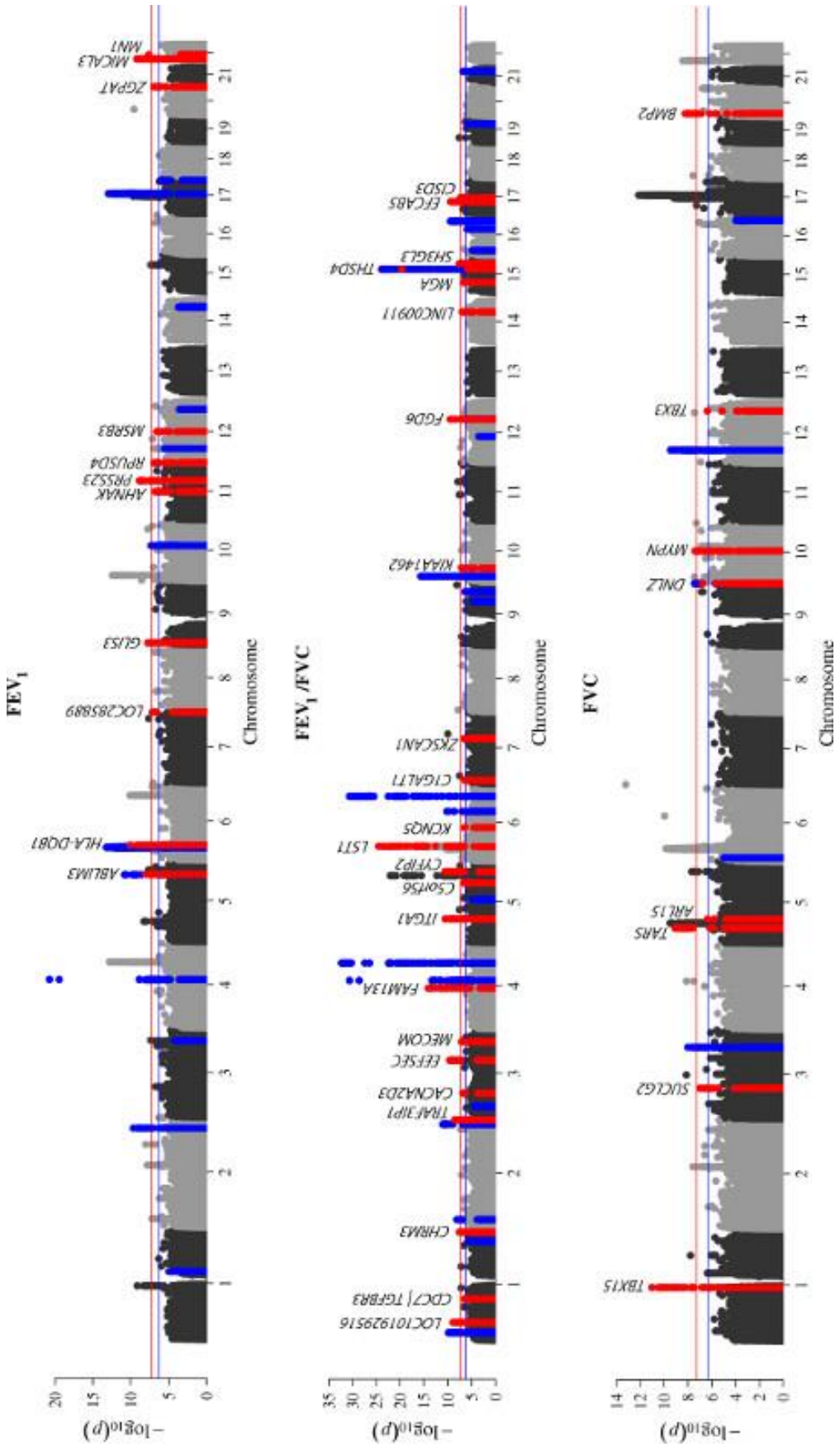

FEV1
FEV1/FVC
FVC
Chromosome
−log10(p)
ZGPAT
MKLN1
MSRB3
RPUSD4
PRSS23
AHNAK
GLIS3
LOC285889
HLA-DQB1
ABLIM3
LST1
ITGA1
FAM13A
MECOM
EFSEC
CACNA2D3
TMEM38B
CHRM3
LOC101929516
CDC7
TGFBR3
LINC00911
MGA
SH3GL3
THSD4
EFCAB5
CISD3
FGD6
KIAA1462
ZKSCAN1
CTGALT1
KCNQ5
CYP7B1
C5orf56
TBX15
SUCLG2
TARS
AKL15
ANPN
ONLZ
TBX3
BMP2

Figure 1-2 Genome-wide association results for FEV_1 (top), FEV_1/FVC (middle) and FVC (bottom).[55] *Previously reported signals are highlighted in dark blue and novel signals are coloured in red. The red and blue lines correspond to the genome-wide significance level ($P=5x10^{-8}$, -log10P=7.3) and the threshold used to select signals for follow up in stage 2 ($P=5x10^{-7}$, -log10P=6.3) respectively. Labels show the nearest gene to the novel sentinel variants. Figure reproduced from Wain LV et al (2017) Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets.* Nat Genet. *2017 Mar; 49(3):416-425.*

In 2017, a large scale lung function GWAS almost doubled the number of signals that have been associated with lung function by identifying 43 new signals, increasing the number of independent signals to 97 (Figure 1-2) [55] There was enhancement in genes for epigenetic regulation pathways, elastic fibres and development. [55] Additionally, 51 of the variants identified showed nominal association with COPD ($p<0.05$) and 30 showed associated reaching Bonferroni correction threshold for 95 tests ($p<5.26x10^{-4}$). [55]

eQTL analysis, using the top variant or a proxy at each of the 97 signals, identified 234 genes with potentially causal effects on lung function. These genes were enriched in elastic fibre pathways and in signalling pathways regulated by the Hedgehog family, which play important roles in early development. This list of potential causal genes was further narrowed down to 68 high priority genes which either harboured a deleterious variant or based on a more stringent criteria for gene expression co-localisation ($r^2 \geq 0.9$). These genes were enriched for gene ontology terms such as SH3 domain binding, GTPase binding, actin binding and fibroblast migration. [55] In order to identify potential drug targets among this list of candidate genes, it was identified that the protein product of 7/234 genes were already targets for approved drugs, or drugs in development including *CHRM3*, *SLC6A4* and *CRHR1*. [55]

More recently, this number of independent loci associated with lung function (FEV_1, FVC and the ratio) has reached 279 with 139 novel signals discovered

by Shrine et al (2019).[56] Of the novel signals, 131 were associated with at least two lung function traits. To identify the putative causal genes for each signal, deleterious variants and variants which had eQTL effects within each credible set were identified. This revealed 107 potentially causative genes with 43 being implicated by a novel signal and 32 from a signal from previous studies.[56] This study identified new genes and pathways which may be involved in airflow obstruction such as cilia development, TGF-β signalling, ECM organisation and elastic fibres.[56] Moreover, in combination these variants strongly predict COPD in independent populations

1.2.2.3 Overlap in Genes Associated with Lung Function and COPD

As one would expect, many of these signals have also shown an association with COPD (Figure 1-3). [41, 45, 47, 49-52, 54, 55, 57-60] A number of studies have succeeded in finding disease specific SNPs associated with COPD phenotypes including emphysema, COPD exacerbations, chronic bronchitis, airway responsiveness and severity of disease, e.g. AGER SNP rs2070600 and emphysema and gas trapping. [61] Significant association has been found for a number of loci already associated with lung function with COPD diagnosis including *FAM13A, TNS1, INTS12/GSTCD/NPNT, HTR4, HHIP, AGER, TSHD4, CHRNA5/CHRNA3* and *ADAM19*. [47, 55, 58-60, 62-66]

A 2017 GWAS using 15,256 COPD cases and 47,936 controls identified 22 loci at genome-wide significance which included 13 new COPD associations, nine of which have previously been associated with lung function. [57] The remaining four loci (*EEFSEC*, *DSP*, *MTCL1*, and *SFTPD*) were novel. [57] Also in 2017, Wain and colleagues analysed 95 of 97 known independent signals associated with lung function for replication with association with COPD. Of the 95, 51 showed nominal association ($P<0.05$) and 30 showed association with COPD susceptibility. [55] This finding that there is an overlap between genetic variants associated with lung function and COPD provides confidence that 1) the locus is a true association and 2) it is important for COPD. By knowing that a variant contributes to lung function and a specific COPD phenotype we can begin to discover the biological mechanisms that underlie the association with disease.

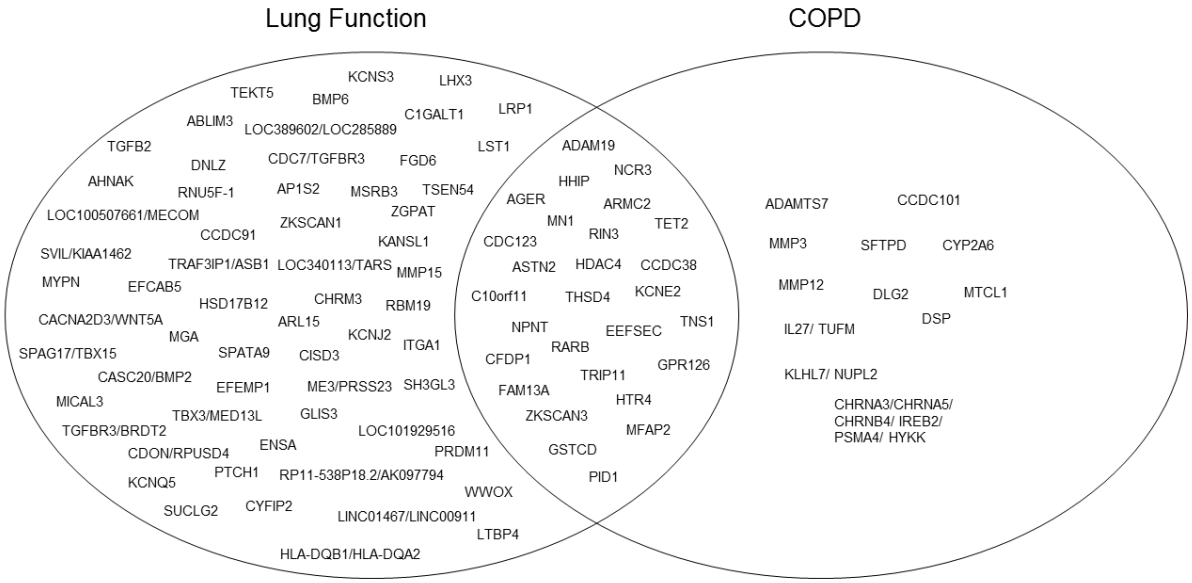


Figure 1-3 Reported genes associated with lung function and COPD from Wain et al (2017).[55] The diagram shows the genes which have association signals for lung function and COPD and the genes which have signals for both. The signals displayed are those which showed highest levels of association, while additional signals which showed less significant association are not shown. In addition to this overlap, the genetic risk score based on 95 lung function variants is highly predictive of COPD risk. [43]

A 2019 GWAS in 35,735 COPD cases and 222,076 controls identified 82 loci associated with COPD ($p < 5 \times 10^{-8}$), 47 of which were previously associated with COPD and/or lung function.[67] Of the 35 new loci, 9 were also recently described in the Shrine et al (2019) paper as signals associated with lung function. Gene set enrichment analysis identified pathways involved in development, including lung specific pathways such as lung morphogenesis and alveolar development and others such as Wnt receptor, MAPK-ERK and nerve growth factor signalling pathways.[67]

1.2.2.4 Translational Approaches for Follow-up Analyses

GWAS have identified many signals relating to respiratory disease, the challenge now is translating these findings into biological insight. There are a

number of considerations to take into account when selecting an association locus for follow-up analysis. Depending on the variant in question, the types of analyses will vary significantly. In brief, a few considerations are; whether the SNP is coding or non-coding, expression quantitative trait loci (eQTL) effects, linkage disequilibrium (LD) with the sentinel SNP. Due to the availability of datasets such as that of the ENCODE (Encyclopaedia of DNA Elements) project, it is possible to identify the likelihood of variants occurring in functional elements in the genome such as promoters, transcription factor binding sites and regions of chromatin and histone modification.

Coding region variants are obvious targets to analyse as they can result in altered amino acid sequences and therefore altered protein structure and function. In addition, coding region variants make up a disproportionate contribution to heritability given that the exome (coding region of the genome) makes up only 1.5% of the genome. [68] Therefore, coding region variants have an enhanced probability of being causal compared to those in non-coding regions. As a result, coding region variants are often assumed to be causal, an assumption that may be inaccurate. [68] However, it has recently been suggested that coding region variants previously thought to be casual may not be. Mahajan et al (2018) assessed coding variants associations with disease (type 2 diabetes) by fine mapping association signals using a meta-analysis from 24 GWAS. Of the 37 regions fine-mapped, they constructed credible variant sets which collectively accounted for 99% of posterior probability of association (PPA). Within each signal they then calculated the proportion of PPA attributable to coding region variants. Only 2 signals had coding variants which accounted for ≥80% PPA whereas other signals showed weak support for causative coding region variants. In large these findings were due to the fact that coding variants were in high LD r2>0.9 with large numbers of non-coding variants so the PPA was distributed across many sites. [68]

The vast majority of variants, approximately 93%, found through GWAS lie in non-coding regions [68, 69] and non-coding region variants that lie in promoter or enhancer regions can have eQTL effects. eQTLs are variants which alter the mRNA expression of a gene. Interestingly, SNPs identified as regulating

mRNA levels in lung tissue overlap with SNPs identified to be associated with lung function demonstrating a clear enrichment of cis and trans eQTL in the associated SNPs. [70] LD is a term to describe the statistical likelihood that SNPs are inherited together. So, it is possible or even likely that the sentinel SNP identified through GWAS is not the variant driving the association and it is actually driven by a SNP in high LD with the sentinel SNP.

For the purposes of this introduction I will use two examples to demonstrate how GWAS hits are translated into information that can be used to inform on the underlying mechanism of the association signal.

1.2.2.5 Coding Region Variant Analysis

The Advanced Glycation End Product Receptor (*AGER)*, encoding the RAGE protein, contains coding region variant rs2070600 which has been associated with FEV_1/FVC in GWAS. [49, 50] The T allele of this variant results in a substitution in the amino acid sequence of RAGE from a glycine at position 82 to a serine. [49, 50] Further analyses also showed the variant to be associated with COPD, emphysema and gas trapping and lung cancer, highlighting the need for further analysis of the role of the gene and variant Ser82 in COPD. [62, 63, 71, 72]

Further study of the variant identified that RAGE is expressed in control and COPD lung as well as the foetal lung. [73] A functional effect of Ser82 was also identified as it was found to be an eQTL for soluble RAGE in serum of UK smokers and in a recombinant cell system. [73] sRAGE serum or plasma levels in COPD patients is lower than in controls. This led to the hypothesis that lower levels of sRAGE, which itself acts as a decoy to limit inflammation, actually may lead to greater RAGE driven inflammation in COPD. [73] While the relationship between genetics, lung function, sRAGE levels and COPD is not fully defined, these findings provide evidence that the Ser82 variant, discovered through GWAS, is important for RAGE function in COPD and began to explain the underlying biological processes behind this association.

1.2.2.6 Non-Coding Region Variants

The majority of SNPs identified through GWAS localise outside of protein coding regions, indicating that regulation of gene expression is likely to be important in disease processes. [69] As an example, a number of SNPs near the *HHIP* gene on chromosome 4q31 have been reproducibly associated with pulmonary function and COPD. [46, 49, 50] *HHIP* codes for a protein which is likely to be an important regulator in a number of pathways including those important for lung development.

HHIP mRNA and protein expression is reduced in COPD lung tissues. Two SNPs, from the haplotype associated with COPD risk, rs6537296 and rs1542725 lie within a *HHIP* enhancer region. [74] These SNPs have been shown to modulate expression of *HHIP,* suggesting that reduced expression of *HHIP* may be involved in the pathogenesis of COPD. [74] To determine the biological mechanism underlying the influence of *HHIP* on COPD susceptibility a knock-down cell model was used. [75] Silencing of the gene resulted in differential expression of nearly 300 genes in the bronchial epithelial cell line Beas-2B. [75] Interestingly, there was enrichment for variants associated with COPD and seven of these genes were found to be differentially expressed in human COPD and control lung tissues. [75] Additional analyses found enrichment for extracellular matrix genes and genes involved in cell growth. [75] The functionality of the specific SNPs in the long range enhancer was also demonstrated using promoter-reporter approaches (66). The *HHIP* case demonstrates how SNPs in a non-coding region can begin to reveal mechanisms underlying disease process and highlight key pathways and genes for further study.

While the last decade or so has been promising in terms of number of genes and variants associated with respiratory disease, there is still considerable work to do to translate these findings into clinical practise. In the future, translation of GWAS signals will be further facilitated by improved annotation of the human genome and improved eQTL datasets, for example in the lung and in structural airway cells, as well as advances in molecular and cell biology techniques such as CRISPR/Cas9 gene editing to ultimately identify causal

variants and mechanisms leading to new understanding and opportunities for therapies.

1.2.3 Chromosome 6q24 Association Signal

The 6q24 locus is one that has reproducibly been associated with lung function measurements in a number of GWAS dating back to 2010. [49, 51, 54] More recently, Soler Artigas et al identified a novel signal over *GPR126* from rs148274477 which showed association with FEV_1/FVC, [54] a signal which has since been replicated. [55, 56]

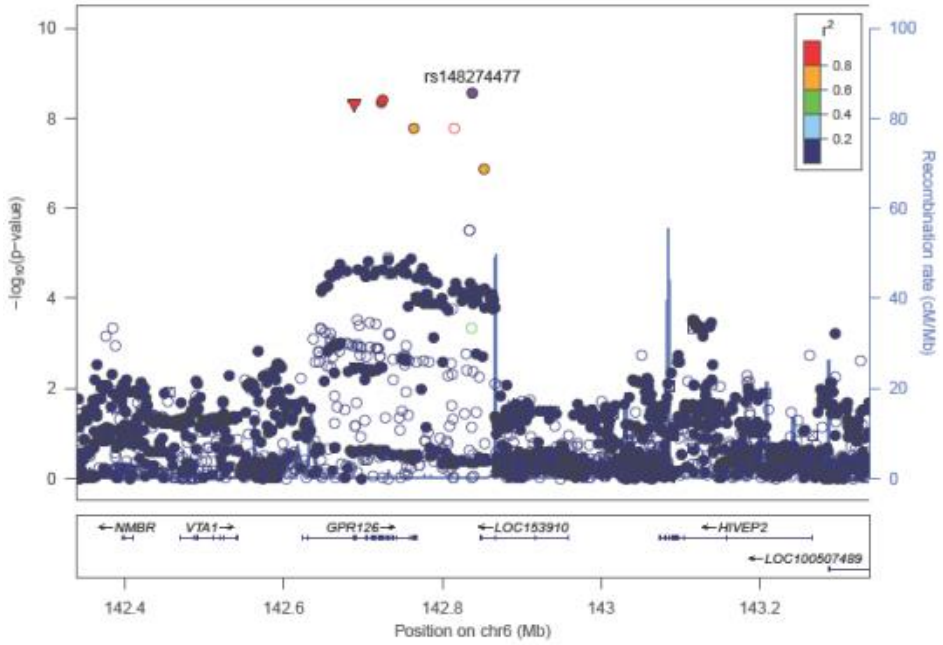


Figure 1-4 Region plot showing SNPs associated with FEV1/FVC in the GPR126 locus in Stage 1 analysis. The x axis represents the chromosomal location and is annotated with gene names and the y axis shows the p-value, with more significantly associated SNPs higher up the axis. The key SNP (rs148274477) in this region is highlighted by the blue circle ($p=9.6 \times 10^{-26}$ after combined stage 1 and stage 2 analysis). Figure reproduced from Soler Artigas et al (2015) Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nat Commun. 2015; 6: 8658.

The region plot in Figure 1-4 shows signals of association over chromosome 6 and *GPR126.* Each individual point represents one SNP. The strongest signal in this region originates from intergenic SNP rs148274477 ($P=9.6 \times 10^{-26}$), which has a minor allele frequency (MAF) of 2.4%. The common (C) allele (frequency = 97%) is associated with a reduction in FEV_1/FVC (beta = -0.162). More recently, the 6q24 signal has since been replicated for FEV_1/FVC ($p=4.48 \times 10^{-17}$) with the rare allele in this instance, T, showing an association with increased FEV_1/FVC (beta = 0.175). In addition, the C allele has shown association with COPD risk (odds ratio = 1.21, $p=1.77 \times 10^{-6}$) by Wain et al (2017).[55]

Interestingly, this SNP is in high LD with a missense variant ($r^2=0.9262$), resulting in an amino acid change. This variant, rs17280293, results in a glycine replacing serine at position 123 of the amino acid sequence (Ser123Gly) of GPR126. Furthermore, a recent GWAS identified rs17280293 to be associated with reduced FEV_1/FVC with the common allele (A) presenting the risk of reduced lung function (beta = -0.1803).[56] An independent signal from an intronic SNP in *GPR126* (rs9399401 (T/C), MAF=0.28, $p=1.81 \times 10^{-19}$) has also been recently associated with COPD with the T allele being the risk allele (odds ratio = 1.17).[57]

1.3 GPR126/ADGRG6

GPR126 (also known as *ADGRG6*) is alternatively spliced into four 'full-length' protein coding transcripts containing 24 to 26 exons (Figure 1-5). These 4 transcripts code for 4 protein isoforms which are denoted as alpha 1 (ADGRG6-201), alpha 2 (ADGRG6-202), beta 1 (ADGRG6-204) and beta 2 (ADGRG6-203) coding for proteins 1221, 1193, 1250 and 1222 amino acids in length respectively.[76] The alpha 2 and beta 2 transcripts lack exon 7 while exon 26 is skipped in the beta 1 and beta 2 transcripts.

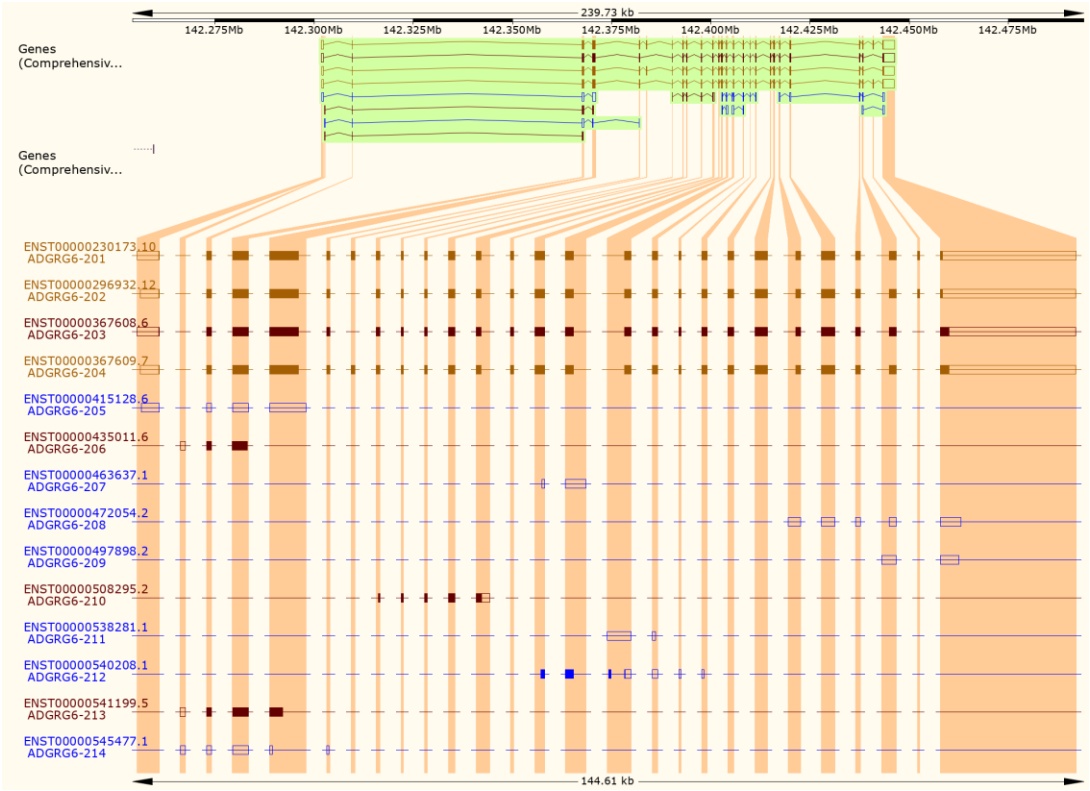


Figure 1-5 GPR126 isoforms as reported by Ensembl.[76] There are four "full-length" protein coding transcripts. ADGRG6-201 is the alpha 1 protein which is 1221 amino acids long. ADGRG6-203 is the beta 2 transcript consisting of 1222 amino acids. ADGRG6-204 and ADGRG-202 are the beta 1 and alpha 2 transcripts of 1250 and 1193 amino acids respectively.

GPR126 codes for an adhesion G-protein coupled receptor (aGPCR). Like all members of the GPCR family of proteins, GPR126 has 7 transmembrane spanning domains as well as adhesion domains in its N-terminal, which are the CUB (complement C1r-C1s, Uegf and Bmp1) and pentraxin (PTX) domains, typical for this sub-class of GPCRs. The G-protein coupled receptor auto-proteolysis-inducing (GAIN) domain containing the GPCR proteolytic site (GPS) lies downstream of the adhesion domains and is important for receptor function (Figure 1-6). Cleavage of GPR126 at the GPS results in a 70kDa fragment (NTF) and the membrane bound 35kDA fragment (CTF). [77]

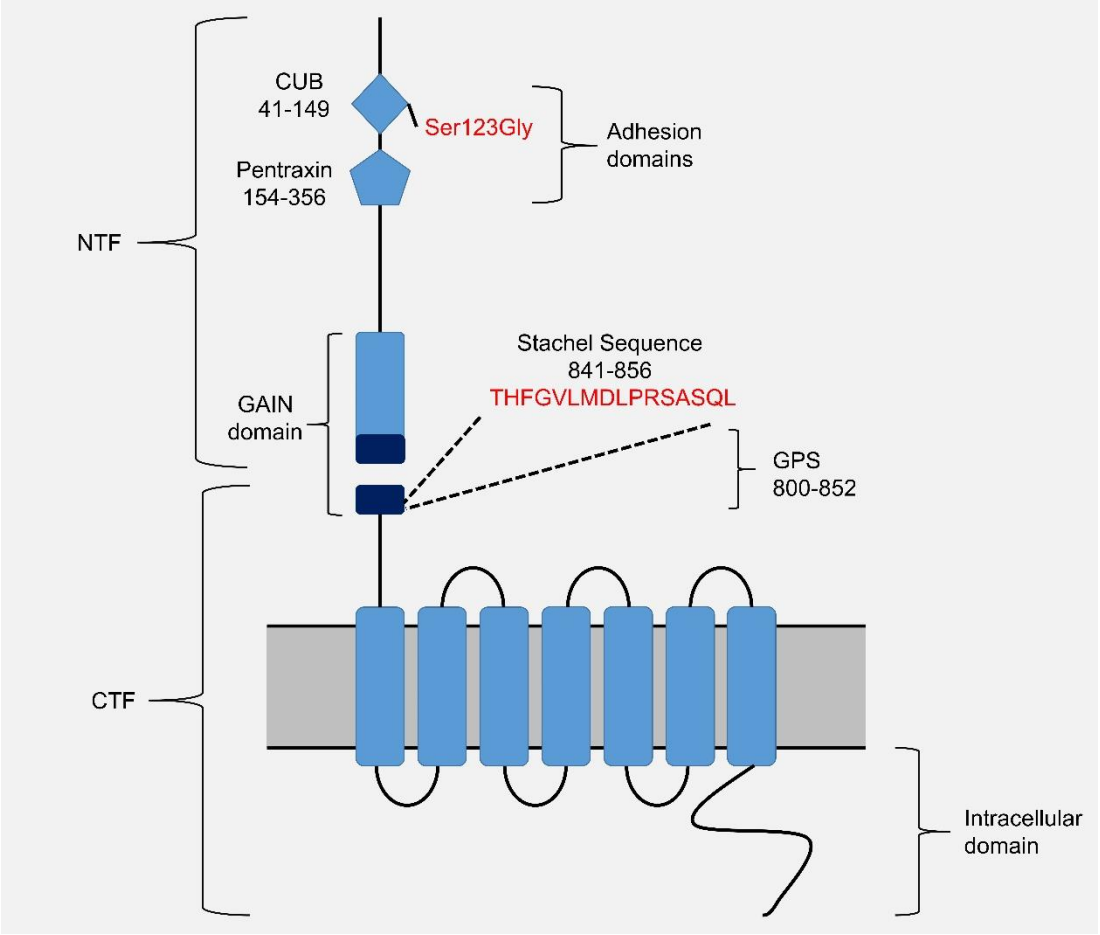


Figure 1-6 The GPR126 structure. The N-terminal fragment, which is cleaved at the GPS motif contains pentraxin (PTX) and CUB domains. Ser123, a single nucleotide polymorphism associated with lung function, lies within the CUB domain, a domain thought to be important for binding collagen type IV. The tethered agonist – the stachel sequence – lies C terminal to the GPS motif and is important in receptor activation by interacting with the 7TM domains following NTF cleavage. NTF = N-terminal fragment. CTF = C-terminal fragment.

1.3.1 G-Protein Coupled Receptors

G protein coupled receptors (GPCRs) are membrane spanning receptors that contain an extracellular domain (N-terminal), seven transmembrane (7TM) spanning domains and an intracellular domain (C-terminal) that interacts with heterotrimeric G proteins to transduce extracellular signals into intracellular signalling cascades resulting in a cellular response. GPCRs are a superfamily

of proteins that collectively can detect a diverse array of extracellular signals including but not limited to neurotransmitters, hormones, odorants, tastants, and light. [78] Because of this, and the various downstream pathways they signal through, GPCRs are of major therapeutic interest with 475 drugs, approximately 34% of all drugs approved by the US Food and Drug Administration (FDA), targeting 108 unique GPCRs. [79] In vertebrates, GPCRs are traditionally categorised into one of five families based on structure; rhodopsin (family A), secretin (family B), glutamate (family C), adhesion and Frizzled/Taste2. [80] Many GPCRs including GPR126 signal via PKA mediated pathways. [81, 82]

1.3.1.1 Cyclic-AMP Signalling

Cyclic adenosine monophosphate (cAMP) is a second messenger important for intracellular signal transduction and is involved in numerous cellular processes (Figure 1-7). cAMP is synthesised from ATP by adenylyl cyclases (AC) and broken down to 5' AMP by phosphodiesterases (PDEs) and therefore the intracellular levels of cAMP are determined by the activity of these two groups of enzymes. AC activity is stimulated by interaction with the alpha subunit of G_s proteins. $G_{\alpha s}$ subunits are coupled with a variety of membrane receptors in heterotrimeric complexes with β and γ subunits from which they dissociate after ligand binding and receptor activation. cAMP has three main targets in the cell, protein kinase A (PKA), exchange protein directly activated by cAMP (EPAC) and ion-gated channels.

PKA is made up of two regulatory subunits and two catalytic subunits. It is activated by cAMP binding to regulatory subunits, which results in dissociation from catalytic subunits. PKA modulates cAMP response by phosphorylating other components of the cAMP pathway such as receptors, ACs and PDEs. Phosphorylation of these proteins results in either an enhancement of the cAMP signal or in downregulation of the pathway. PKA is involved in many different signalling pathways including the modulation of other protein kinases, regulation of intracellular calcium concentration and transcriptional regulation. Regulation of transcription by PKA is mainly achieved by direct phosphorylation of transcription factors. Responses to intracellular cAMP

increase are typically via the cAMP response binding element (CREB), cAMP response element modulation (CREM) and activation transcription factor 1 (ATF1). [83]

cAMP can activate intracellular pathways by binding to and modulation the function of cyclic nucleotide-gated ion channels. These channels are relatively nonselective and conduct cations such as Ca^{2+}, which stimulates CaM and CaM-dependent kinases which in turn modulate cAMP production by regulation of ACs and PDEs.[84] cAMP can also directly modulate the mitogen-activated protein kinase (MAPK) by binding to and activating EPACs, which are GTP exchange-protein for the GTPase Rap1 and Rap2. [85, 86] Epac proteins bind to cAMP with high affinity and activate the Ras superfamily small GTPases Rap1 and Rap2. [86]

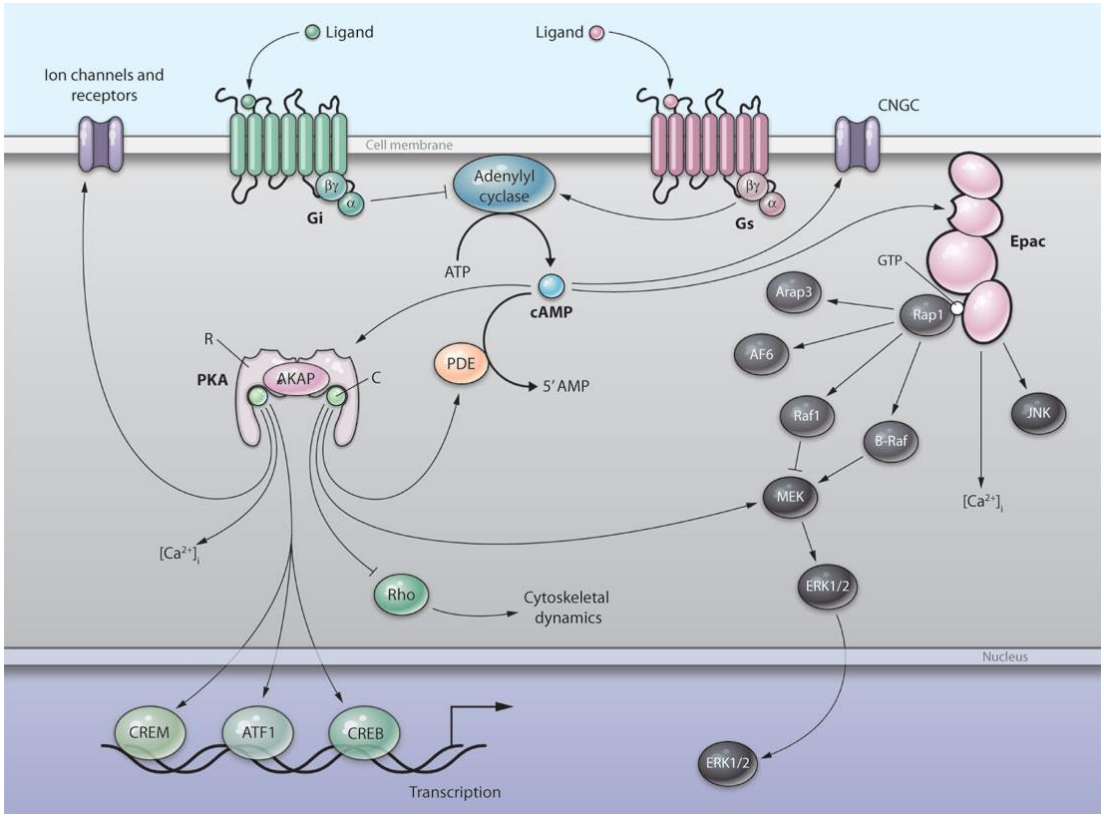


Figure 1-7 cAMP signalling pathways. Image from Murray (2008) Pharmacological PKA Inhibition: All May Not Be What It Seems. Sci. Signal. *03 Jun 2008: Vol. 1, Issue 22, pp. re4* [87]

1.3.2 Adhesion G-Protein Coupled Receptors

GPR126 belongs to the adhesion GPCRs (aGPCR) sub-family of GPCRs. This family has 33 members, defined by the presence of their large N-terminal regions. [88, 89] These extracellular regions contain a number of conserved domains that are important for adhesion interactions such as epidermal growth factor-like repeats, thromobospondin-like repeats and cadherin-like repeats. [80, 88, 90] While ligands have been identified for some aGPCRs, most members of this protein family are still considered to be orphan receptors i.e. they currently have no identifiable ligand. Despite this, aGPCRs have been shown to signal through traditional GPCR pathways via G proteins and so far $G_{\alpha 12/13}$, $G_{\alpha o}$, $G_{\alpha s}$, $G_{\alpha i}$ and $G_{\alpha q}$ have all been shown to be involved in aGPCR signalling pathways (reviewed in [78]).

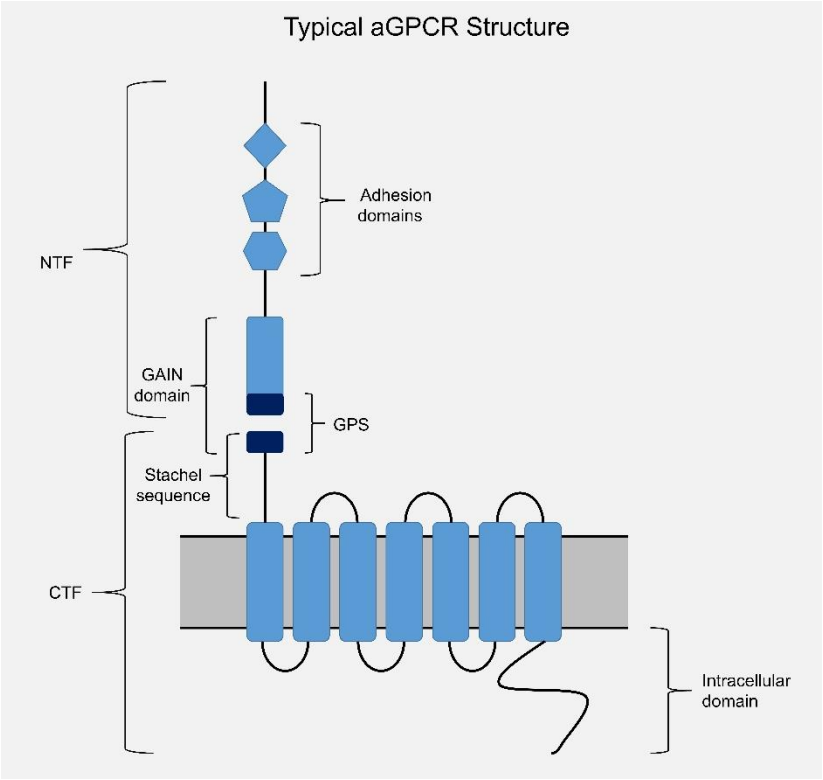


Figure 1-8. The typical structure of an adhesion-GPCR. This family of receptors contains the characteristic 7 transmembrane spanning domains of a GPCR. They also contain long N-terminal domains which contain domains important for adhesive functions and the GAIN domain, which contains the autoproteolytic site, the GPS motif.[91]

1.3.2.1 Adhesion-GPCR activation

A common feature for all aGPCRs is the GPCR autoproteolysis-inducing (GAIN) domain which contains the GPCR proteolysis site (GPS) motif (Figure 1-8). [92, 93] Cleavage at this site results in two fragments, the cleaved N-terminal fragment (NTF) and the still membrane bound C-terminal fragment (CTF). Proteolytic cleavage at this site is an important modification in the processing of aGCPRs and even after cleavage at the GPS motif, this transmembrane domain can remain non-covalently associated to the 7TM region. [78, 94, 95] Interestingly, whilst the NTF is still associated with the 7TM it exerts an inhibitory effect on receptor signalling and it is hypothesised that interaction with a potential ligand would result in removal or rearrangement of the NTF thereby removing the inhibitory effect. [78]

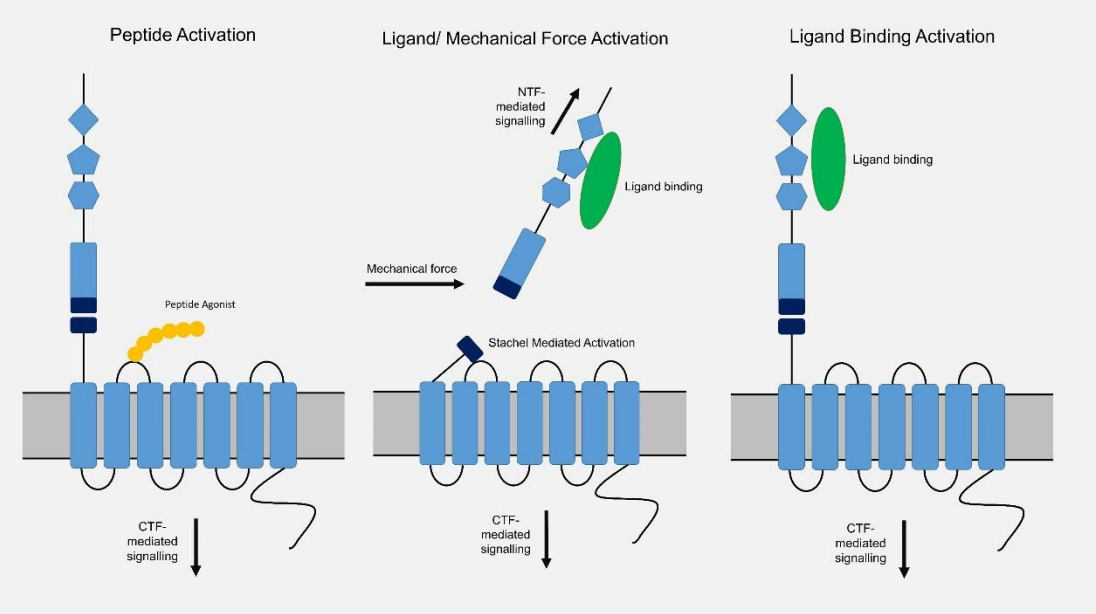


Figure 1-9. Reported mechanisms of aGPCR activation. aGPCRs contain a tethered agonist sequence – termed the stachel sequence - which can activate receptor signalling after cleavage of the N-terminal. Synthetic peptides derived from this sequence can activate receptor sequencing.[96] Additionally, mechanical forces in tandem with ligand binding is thought to result in cleavage of the NTF at the GPS site resulting in stachel-mediated receptor activation.[97] aGPCRs are also thought to activated by ligand binding to the adhesive domains.[98]

The cleavage of the NTF is important for aspects of aGPCR signalling. While the binding of ligands to the N-terminal can activate aGPCR signalling, receptors with removed NTFs are actually constitutively active, reinforcing the idea that the NTF exerts an inhibitory effect. [78, 96] More recent evidence suggests that this might be due to an activating sequence (termed the stachel sequence) within the GAIN domain, C terminal to the GPS motif (Figure 1-9). [96, 99] This tethered agonist is thought to be exposed when the two β-sheets surrounding it are removed along with the NTF at the GPS motif whereby it can then activate the 7TM domain resulting in downstream signalling. [96] The signalling modalities of these receptors is further complicated by the fact that the NTF fragment can have distinct signalling properties from the CTF. [78, 97] Together, this research suggests a number of roles for the NTF; i) cell adhesion ii) binding of ligands leading to traditional GPCR signalling iii) inhibition of signalling and iv) additional signalling mechanisms as a secreted ligand. [78]

1.3.3 GPR126 Signals via cAMP and is Important for a Number of Developmental Pathways

The role of *GPR126* in the human lung is currently unknown and therefore the mechanisms by which *GPR126* is associated with lung function is not understood. However, it is known that *GPR126* is differentially expressed throughout human lung development, suggesting that it may be important for normal lung development.[100] In addition, it has been shown that *GPR126* is an essential part of signalling pathways vital for a number of developmental processes with the majority of what is known about *GPR126* function coming from studies in zebrafish and mouse models.

Early studies into GPR126 signalling showed that the receptor is key for the correct development of peripheral nerves in both mouse and zebrafish.[101, 102] *Oct6* and *Krox20* are two genes that are necessary for Schwann cells to initiate myelination.[101] Monk and colleagues identified a mutation, *st49,* within GPR126 causing a premature stop codon N-terminal to the GPS in zebrafish with abnormally myelinated axons.[101] Cultured Schwann cells from Zebrafish with the *GPR126*^{st49} knockout mutation fail to express *Oct6* and *Krox20.* [101]

When treated with forskolin (used to raise cAMP levels via activation of adenyl cyclase) these mutant cells showed a reversal of the *GPR126* knockout phenotype and the myelination of peripheral nerves was restored, indicating *GPR126* regulates this process via cAMP signalling and regulation of *Oct6* and *Krox20* expression.[101] This function of GPR126 is conserved in mammals and Schwann cells in GPR126 knockout mice also arrest at the promyelinating stage and additional studies in mice showed that GPR126 is essential for many aspects of peripheral nerve development including defects in axonal sorting, the formation of mature non-myelinating Schwann cells and in the organisation of the perineurium. [82, 101-103] In addition to the nerve phenotype, GPR126 knockout mice have limb defects and the majority of homozygous mutant mice do not survive until birth and most of those that do survive die before weaning. [102] A follow-up study in 2013 showed that cAMP and protein kinase A (PKA) elevation rescues the myelin defect in dorsal root ganglia from mutant mice.[104] They also identified that GPR126 couples to G_{as} and G_{ai} to modulate intracellular cAMP levels. [104]

Aside from peripheral nerve development, GPR126 has been shown to be vital for angiogenesis by regulating VEGF signalling. GPR126 is enriched in endothelial cells and through the cAMP-PKA-CREB signalling pathway, regulates endothelial cell proliferation, migration and tube formation in the process by inducing the expression of *GAT2* and *STAT5* which in turn regulate the expression of *VEGFR2.* [81]

GPR126 is highly expressed in the otic vesicle and has a role in regulating semi-circular canal morphogenesis in both zebrafish and mouse.[105] During inner ear development, the semi-circular canals are formed from the migration and fusion of epithelium in the otic vesicle. [106] *GPR126* controls the adhesion and outgrowth of canal projections by regulating the expression of genes coding for components of the extracellular matrix (ECM) through cAMP. In addition, ECM gene expression is highly disrupted in zebrafish GPR126 knockout. [105]

1.3.3.1 GPR126 Signalling is Domain Dependent and Organ Specific

Cleavage at the GPS domain can result in two signalling pathways one of which is through the CTF via associated G-proteins as described above and the other is via the cleaved NTF presumably as a soluble ligand. [97, 104] *GPR126* knockdown in mice is lethal due to cardiovascular failure with this loss in viability coinciding with the *GPR126* expression pattern in mouse development. [98] Interestingly, the NTF can act independently of the CTF, and is essential and sufficient for aspects of cardiac development including energy metabolism, trabeculation and contractility. [104]

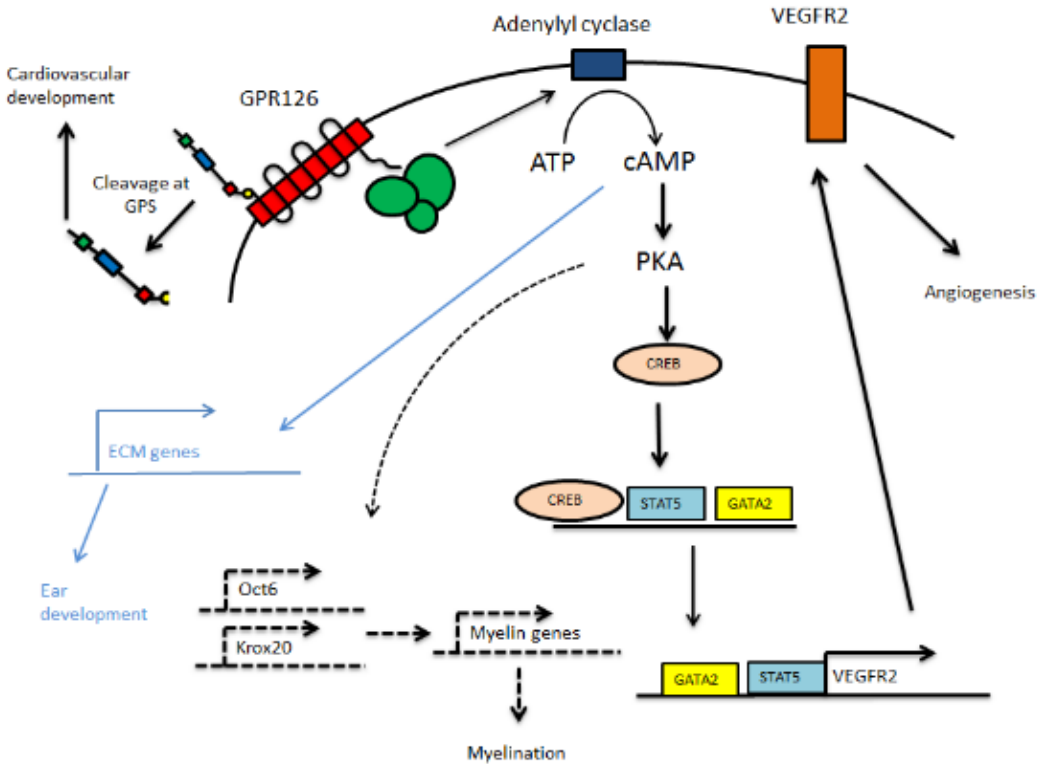


Figure 1-10 The signalling pathways of GPR126 are organ and domain specific. GPR126 can signal via G protein interaction or via the cleaved N-terminal fragment

In addition, the NTF and CTF have independent roles in peripheral nerve development. [97] The CTF promotes peripheral nerve wrapping through cAMP mediated pathways as previous described whereas the NTF is essential and

sufficient for radial sorting of peripheral nerves and acts independently of cAMP regulated pathways. [97]

1.3.3.2 GPR126 Potential Ligands

GPR126's roles in nerve are mediated by the ligand Laminin-211 which binds to the NTF and can have both inhibitory and activating effects on GPR126 signalling. [97] Interestingly, collagen type IV, a major component of basement membranes, has been identified as an activating ligand for GPR126 in rodent Schwann cells and HEK293 cell lines overexpressing the zebrafish clone.[98] This interaction is thought to involve the CUB and PTX domains in the N-terminal domain and initiate cAMP signalling pathways. [98] However, the human GPR126 receptor has shown only a modest, and not statistically significant response to collagen type IV when overexpressed in a recombinant cell system. [97]

Both collagen type IV and laminin 211 are key components of the basement membrane (BM). [107, 108] Basement membranes are thin sheets of ECM that lie beneath endothelial and epithelial cells and surround smooth muscle, fat and peripheral nerve cells. [108] This membrane serves as a scaffold for cells which directs cell proliferation, migration, differentiation and apoptosis. The main components of BMs are collagen type IV, nidogen, heperan sulphated proteoglycans and laminins. In the lung, BMs are associated with the airway and alveolar epithelium, endothelium, bronchial and vascular smooth muscle cells, nerve cells and visceral pleura and are therefore an important aspect of the air-blood barrier.[108]

Laminins are heterotrimeric glycoproteins integral to BMs that are made up of α, β and γ chains. Laminins are effectors of tissue morphogenesis and vital for the formation of multiple organs during development including the lung.[109, 110] Specifically, the laminin α2 chain, the α chain in laminin 211, and laminin 211 have an important role in smooth muscle cell differentiation. [111]

Collagen type IV is a major component of epithelial and endothelial BMs and is essential for its stability.[107] Collagen type IV and the BM have important roles in the formation of the epithelium and vasculature and alveolar

myofibroblast development by induction of proliferation, differentiation and migration. Two collagen type 4 chains, COL4A1 and COL4A2, are upregulated in chick lung development during the formation of the avian blood-gas barrier. [112] In mouse, mutations in murine COL4 genes affected the balance between lung epithelial progenitors and differentiated cells. COL4A1 mutants have disrupted myofibroblast proliferation, differential and migration suggesting that collagen type IV and the BM play fundamental roles in the coordination of alveolar morphogenesis. [112]

Similar to other aGPCRs. GPR126 signalling activity is mediated by a tethered agonist, called the stachel peptide. Studies show that deletion of the extracellular domain activates signalling in GPR126 suggesting that the extracellular domain is constitutively inhibiting signalling activity – in keeping with what is known about other aGPCRs. Activation of the receptor could be caused by extracellular molecules causing structural changes in the extracellular domain or mechanical cleavage of NTF may expose the agonising sequence. [96]

1.3.4 Human traits associated with GPR126 variants

Polymorphisms and mutations in *GPR126* have been shown to be associated with human phenotypes aside from lung function phenotypes. A number of polymorphisms have been identified to be associated with stature, specifically trunk height. [113, 114] In addition, SNPs in the gene have been associated with adolescent idiopathic scoliosis, which is the sideways curving of the spine in late childhood/adolescence [115, 116]. Homozygous loss of function (LOF) mutations have been shown to cause arthrogryposis multiplex congenita (AMC). [117] AMC is a heterogeneous condition that is characterised by the presence of congenital non progressive contractures in more than one major joint. [117] These skeletal abnormalities are in keeping with the abnormal limb defects seen in mutant mice. [102]

1.4 HYPOTHESIS AND AIMS

The hypothesis underlying this project was that genetic variants in GPR126 are the driving causal variants underlying the association signal on chromosome 6q24 with FEV_1/FVC. I hypothesised that that altered GPR126 expression or function, possibly due to the Ser123Gly variant, leads to an altered lung phenotype.

In addition, due to known GPR126 and aGPCR biology, I hypothesised that GPR126 may have a role in the process of airway remodelling or lung development based on its known interactions with important components of the basement membrane, collagen type IV and laminin-211.

The overarching aim of this project is to identify the driving mechanisms underlying the association signal at GPR126 for lung function.

The primary aims are as follows:

- To identify a GPR126 expression profile in the lung and in airway structural cells
- To determine potential causal variants within the association signal
- To define the effects of missense variant Ser123Gly on GPR126 function
- To determine the capacity of primary human airway cells to signal via GPR126
- To determine the downstream effects of GPR126 activation in human airway smooth muscle cells

2 General Materials and Methods

A list of all reagents and antibodies used in this thesis, including manufacturers and catalogue numbers, can be found in the appendices.

2.1 MAMMALIAN CELL CULTURE METHODS

Cell culture is the method of growing cells under controlled artificial conditions. The cultures grown in this thesis were proliferated at 37°C and 5% CO_2 humidity in incubators. Only adherent cells were used in this project. These cells grow in a monolayer requiring a surface to attach to in order to grow. Most adherent cells grow in a monolayer until reaching confluence (the cells are occupying all the physical space), at which point cell growth will arrest. At this point cells need to be sub-cultured in new flasks or plates (passaged) in order to proliferate further.

Cell lines are cells which have acquired the ability to proliferate indefinitely by avoiding cellular senescence allowing them to be used for prolong periods of experimentation *in vitro*. In this project, a HEK293 cell line was used for recombinant overexpression of GPR126. The HEK293 cell line is derived from human embryonic kidney cells grown in tissue culture and was immortalised due to the introduction of adenovirus 5 DNA. [118] Primary cells are cultures derived directly from tissue and can be maintained *in vitro* but only for a limited time. While these cells are useful as they are more physiologically relevant to cells *in vivo,* their limited life-span limits the time in which they can be used for experimentation.

The following section outlines the cell culture methods used in this thesis. Firstly, a CRE-*luc2P* HEK293 cell line was used for recombinant expression of GPR126 variants and functional studies on receptor activity. And secondly, human primary airway cells were used for total RNA extraction for expression profiling and RNA-sequencing experiments as well as endogenous GPR126 functional studies.

2.1.1 Culture of GloResponse™ CRE-luc2P HEK293 Cell Line

The GloResponse™ CRE-luc2P HEK293 cells (Promega, E8500) is a HEK293 cell line stably expressing the firefly luciferase gene (*luc2P)* under the control of a minimal HSV thymidine kinase promoter with multiple cyclic-AMP response elements (CREs). The *luc2P* gene is positioned downstream of the cAMP response elements (CRE) which are activated upon cAMP signalling cascades. The CRE is the DNA response element for the cAMP response element-binding protein (CREB). In response to intracellular cAMP elevation, CREB relocates to the nucleus of the cell and binds to CRE, promoting gene transcription. The location of the *luc2P* gene downstream of a CRE ensures that luciferase is produced upon activation of the CRE hence making this cell line ideal to study G_s- and $G_{i/o}$-coupled GPCRs, which signal through cAMP and the cAMP response element (CRE).[119, 120]

Two vials of CRE-*luc2P* HEK293 cells in freezing medium were delivered on dry ice. One vial was placed in liquid nitrogen storage while the other was immediately propagated into a T75 flask. The cells were grown in Dulbecco's Modified Eagles Medium with 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate (DMEM) with the addition of 10% FCS with 50µg/ml hygromycin B. Table 2-1 shows the media compositions used for culturing this cell line. Cells were experimented on in serum-free media (SFM) unless otherwise stated. Cells were routinely passaged in T75 flasks in 15mLs media.

Growth Medium	90% DMEM 10% FCS
Growth Medium + hygromycin B	90% DMEM 10% FCS 50µg/ml hygromycin B
Selection medium for cells transfected with pLV vectors	90% DMEM 10% FCS 50µg/ml hygromycin B 0.05µg/ml Puromycin
Freezing Medium	75% DMEM 15% FCS 10% DMSO
Assay Medium	100% DMEM

Table 2-1 Cell culture media for culturing GloResponse CRE-luc2P HEK293 cells.

2.1.2 Culture of primary human airway cells

2.1.2.1 Human Airway Smooth Muscle Cells

Human airway smooth muscle (HASM) cells were obtained from lung tissue donors by Dr Charlotte Billington and Dhruma Thakker using the explant model outlined below. Human lung tissue was obtained from the Nottingham Research Biorepository or the Papworth Hospital Tissue Bank from patients undergoing lung resection surgery. Details of these tissue donors can be found in Table 2-2. Written consent was obtained from all patients and the study was conducted under the approval of the research Tissue Banks (ethics reference 10/H1008/72 and 08/H0304/56+5).

Donor Number	D1515	D063	D023	D006	D007	D00529
Study Number	54	54	54	61	61	61
Tissue source	Royal Papworth	Royal Papworth	Royal Papworth	Nottingham Biobank	Nottingham Biobank	Nottingham Biobank
Donor Information	Age	86	64	59	63	53	57
	Sex	F	M	F	M	M	M
	Smoker	Ex-smoker	Heavy smoker	Ex-smoker	Ex-smoker	Ex-smoker	Ex-smoker
	Disease Status/ reason for operation	Lung cancer; upper lobectomy	Squamous Cell Carcinoma of Right Lung	Right Upper Lobe Squamous Cell Carcinoma	COPD; Double Lung Transplant	COPD; Double Lung Transplant	Pulmonary Fibrosis

Table 2-2 HASM donor information

Small sections of airways were dissected from the surrounding tissue and then finely chopped with scissors. The resulting small explants were smeared across the bottom of one well of a 6 well plate. Each well contained two or three explants. These were allowed to dry briefly to ensure adherence and then DMEM (plus 10% FCS and 5% PSF) was added. The plates were then kept in the incubator and media was changed every few days. After a week or

two, cells would start to grow out from the explants. Once the well was confluent it was trypsinised (p0) and the contents transferred to a T25 (p1) and then later on to a T75 at which point cells were named p2. They were then frozen back or used in experiments.[121]

HASMs were cultured in DMEM+10% FCS. They were generally passaged in T75 flasks until being passaged into assay specific culture vessels. During experimentation, HASMs were cultured in serum-free media (100% DMEM).

2.1.2.2 Human Bronchial Epithelial Cells

HBECs were purchased from Lonza. Table 2-3 shows the demographics of the donors of these cells as well as the cell properties and the results of the safety testing.

Donor		1	2	3	4	5	6	7	8	9	10	11
	Lot number	7F3158	7F3206	195307	124040	355687	369144	494599	328324	382850	543643	485960
Donor Information	Age	56	50	19	43	38	43	19	18	60	57	38
	Sex	M	M	M	M	M	M	F	F	F	F	M
	Ethnicity	C	C	C	C	C	C	C	C	C	C	C
	Smoker	Y	Y	N	N	N	N	N	N	N	N	N
Cell properties	Viability (%)	70	75	91	71	75	87	83	85	82	87	83
	Seeding efficiency (%)	31	45	66	60	68	41	61	99	49	30	45
	Doubling time (hours)	24.5	21	16	20	23	18	28	31	19	23	35
Virus testing	HIV	N	N	N	N	N	N	N	N	N	N	N
	HBV	N	N	N	N	N	N	N	N	N	N	N
	HCV	N	N	N	N	N	N	N	N	N	N	N
Safety tests	Sterility test	N	N	N	N	N	N	N	N	N	N	N
	Mycoplasma test	N	N	N	N	N	N	N	N	N	N	N

Table 2-3 Characteristics of Lonza donors and properties of the HBECs cultured from these donors. C=Caucasian, N=negative.

HBECs were maintained in PneumaCult-Ex Basal Medium (490mls PneumaCult-Ex Basal Medium plus 10mls PneumaCult-Ex 50X Supplement and 500µl 200X hydrocortisone stock solution) with the addition of 5mls antibiotic, antimycotic solution containing penicillin, streptomycin and fungizone (PSF). [122]

HBECs frozen in liquid nitrogen stocks at a concentration of 250,000 cells/mL were thawed and seeded to T75 cell culture flasks in complete medium. Media was replaced the day after seeding to remove DMSO. Cells were initially expanded in a T75 cell culture flask with media replacements every 2-3 days to replenish supplements. Cells were grown to ~80% confluency at which point cells were passaged to new cell culture vessels for experimentation or frozen down.

2.1.3 Trypsinisation and Passaging Cell Cultures

Cells are passaged as they reach confluence to prevent contact inhibition and cell senescence. This involves disassociating the adherent cells from the cell culture dish and culturing in a new vessel at a lower concentration of cells to promote growth. To remove cells from the culture dishes through trypsinisation. Media was removed from cells, cultures were briefly washed with PBS and trypsin/EDTA solution was added for ~5 minutes until cells had dissociated from the flask surface. For HBECs, this reaction was stopped by adding the same volume of trypsin inhibitor. For HEKs and HASMs, ~10mL media was added. Cells were centrifuged at 200g and resuspended in the appropriate culture media and seeded to the required culture vessel.

2.1.4 Cryopreservation of Mammalian Cells

Cryopreservation of cells was performed in order to store stocks of both cell lines and primary cell lines. Freezing cells means that cell expansion can be halted and cells can be maintained at a lower passage number (and therefore maintain cell viability) – particularly important for primary cells – for future use. Cells were frozen to a temperature of -180°C in a cryogenic tank filled with liquid nitrogen after one night frozen at -80°C in an isopropanol bath.

All cells were frozen in normal culture media with 10% DMSO, which is a cryoprotective agent. This slows the rate of cooling which prevents cell dehydration and ice crystal formation which would otherwise cause structural damage to the cells.

2.1.5 Cell Counting using a Haemocytometer and Assessment of Cell Viability

Cell density was determined using an improved Neubauer haemocytometer. A 10µl cell suspension was added under the coverslip of the haemocytometer and the cells were visualised using an inverted light microscope. Cells were counted in the four corner squares of the haemocytometer and the average was taken. The average value was then multiplied by 10^4 to give number of cells per mL as the volume of each corner square is $0.1mm^3$.

Cell viability was established using Trypan Blue. Trypan Blue is a stain which does not react with the cell unless it has a damaged membrane. Viable cells do not take up the stain whereas non-viable cells do. An Eppendorf was set up with 50µl of cell suspension and 50µl of Trypan Blue was added. The solution was mixed and left to incubate for 5 minutes. Cells were viewed and counted as explained above. Cells stained blue were not included in the viable cell count.

2.1.6 Transfections using Fugene 6

In this project CRE-luc2P HEK293 cells were transfected with the pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV in order to generate recombinant cell lines overexpressing GPR126. In addition, primary human airway cells were transfected with a CRE-luciferase plasmid in order to study cAMP-driven gene transcription via luciferase assays.

Cell cultures (passage 2 for all cell types) were grown to ~80% confluency in a T75 flask. Cells were then split into individual wells of the appropriate culture vessel. The cells were left to incubate at 37°C with 5% CO_2 until ~70% confluent, usually 3-4 days post-thawing.

All transfections of mammalian cell cultures were carried out using the lipid-based transfection reagent Fugene ® 6 (Promega). [123] In this method, the liposome and DNA complex moves into the cell by endocytosis or fusion with the plasma membrane. Transfections were carried out as per the manufacturer's instructions. The Fugene ® 6 transfection reagent was diluted in normal culture media followed by a 5 minute incubation. Plasmid DNA was then added to the required ratio of Fugene:DNA and incubated with the Fugene ® 6/ media for 15 minutes to allow the plasmid DNA to complex with the transfection reagent. The complex was then added to cells and left for 24 or 48 hours at normal cell culture conditions.

2.1.7 Luciferase Assays

Luciferase is a common choice as a reporter for measuring signalling events and gene expression as the functional luciferase enzyme is immediately produced upon translation. This makes luciferase reporters ideal for high throughput assays as they are rapid and sensitive. The luciferase (*luc2P)* gene catalyses the mono-oxygenation of beetle luciferin using ATP as a cofactor - a reaction which produces bioluminescence (Figure 2-1) which can be quantified by measuring luminescence. [119] The *luc2P* gene is a destabilised variant of the luciferase gene with a protein half-life of approximately 1 hour due to the attached PEST sequence. These sequences are peptide sequences which are rich in proline (P), glutamic acid (E), serine (S) and threonine (T) and act as signal peptides for protein degradation.

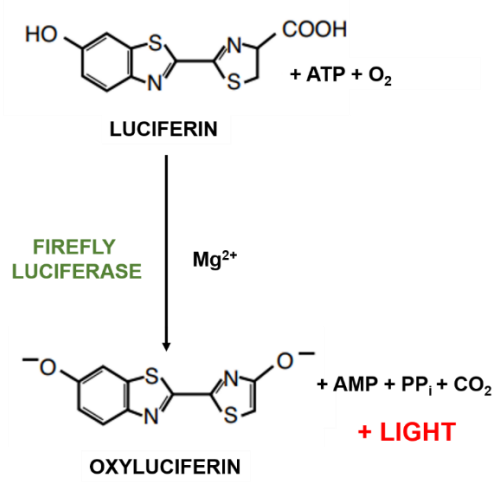


Figure 2-1. Firefly luciferase catalyses the production of oxyluciferin from luciferin, a reaction which produces light.

The following assays were all performed using this basic protocol except where stated otherwise. Cells were thawed from liquid nitrogen stock cryogenic vials, seeded to T75 flasks and cultured in complete media.

Cells were treated with experiment specific agonists for the required time before the Bright-Glo™ Luciferase Assay System (Promega) was used to lyse cells and produce the luciferase reaction. Figure 2-1 displays the process of the mono-oxygenation of luciferin which is catalysed by luciferase in the presence of Mg2+, ATP and molecular oxygen. The provided Bright-Glo buffer and Bright-Glo substrate were mixed and 100µl of the solution was added to each well. Plates were incubated at room temperature for at least 2 minutes before luminescence was measured on a Flexstation 3 microplate reader (Molecular Devices) for 0.5 seconds/well.

2.1.8 ^{3}H-cAMP accumulation assays by column chromatography

The basis of column chromatography is to use radioactively labelled adenine to detect cAMP production by cells. Cells are incubated with tritiated adenine which is used by the cells to produce more cAMP as required. The underlying principle behind this method is that cells which are treated with a condition which stimulates cAMP production (e.g isoprenaline or forskolin) will use more

of the radioactively labelled adenine to produce more radioactively labelled cAMP.[124] Cells are washed to remove excess tritiated adenine and cells are lysed with hydrochloric acid (HCL). The cell lysis therefore contains the tritiated compounds from the cell. The tritiated cAMP is then separated from other adenine derivatives with column chromatography.

The Dowex column uses a dowex cation exchange resin, which allows negatively charged cellular components such as AMP, ATP and ADP pass through. cAMP is acidic but does get absorbed into the column which is washed off Dowex columns onto Alumina columns with water. Alumina is neutral and binds cAMP less effectively than remaining adenine containing compounds and therefore cAMP can be washed off the columns with imidazole and collected. [125]

Cells were grown to 80% confluence in a T75 flask after being thawed from liquid nitrogen stocks. The day before assaying, cells were seeded to wells of a 48 well plate in culture medium. A ^{3}H adenine working solution was produced by diluting 100µl stock adenine in 5ml serum free media. On the day of experimentation cells were removed from incubation and loaded with 50µl ^{3}H adenine per well. Plates were left for 2 hours in 37°C with 5% CO_2 in order to incorporate the adenine. Media containing ^{3}H adenine was aspirated from the cells and cell cultures were washed with serum free media.

Cells were treated with the required agonists as detailed in each specific chapter. Plates were left to incubate for the appropriate amount of time and stopped by adding 50µl of concentrated hydrochloric acid. Plates were taped together and frozen at -20°C until columns were run.

2.1.8.1 Pouring the Columns

This assay uses 2 sets of columns; 1) Bio-Rad Dowex AG 50 W x4 columns and 2) Alumina columns which are prepared as outlined below.

The Dowex columns are made by mixing the Dowex with distilled water to produce a slurry (roughly 50:50 volume). This slurry is allowed to settle for ~1 hour before the water is poured off – this washes the Dowex. This process is

repeated several times until the water becomes clear and the very fine Dowex, which will otherwise clog up the column filters, is washed out. The washed Dowex is then resuspended in distilled water (again, a 50:50 volume) and 2.4mls is added to each column which, when the water has dripped through, results in 1.2mls Dowex in each column.

The Alumina columns are made by putting 0.6g Alumina in each column.

2.1.8.2 Daily Column Preparation

Before the cell contents in run through the columns a number of steps are taken to prepare the columns. For the Dowex columns, 10ml 1M HCl is added to each column and allowed to drip through before 10ml distilled water is added to each column and allowed to drip though twice. To the Alumina columns, 10ml 0.1M imidazole is added to each column and allowed to drip through twice.

2.1.8.3 Running cAMP columns

In short, 500µl of diluted ^{14}C-cAMP solution (100µl Stock 14C-cAMP mixed with 10ml water) was further diluted in 10ml H_2O and 100µl ^{14}C-cAMP was added to each well. The whole well contents were transferred to the corresponding Dowex column and were allowed to drip through. 3ml H_2O was added to each column and allowed to drip through. Dowex columns were placed on top of alumina columns and 6mls H_2O was added to dowex columns which was allowed to drip through both sets of columns. Dowex columns were removed and the alumina columns were placed over scintillation vials. 5ml 0.1M imidazole was applied and allowed to drip through the columns into the scintillation vials. 10ml scintillation fluid was added to each vial and vials were counted on a Packard Scintillation counter using a ^{3}H/^{14}C dual count program for 5 minutes per vial. Three standards were made up containing 100 100µl stock ^{14}C-cAMP + 5ml 0.1M imidazole + 10ml scintillation fluid.

2.2 MOLECULAR BIOLOGY METHODS

2.2.1 Total RNA extraction from mammalian cell cultures

RNA isolation was carried out using the GenElute™ Mammalian Total RNA Miniprep Kit as per the manufacturer's instruction (Sigma).[126] The protocol utilizes spin-column technology to bind RNA from cell lysates onto a column which can then be eluted. What follows is a brief outline of the steps involved and full details can be found in the manufacturer's instructions.

2.2.1.1 Cell lysis and harvest

Mammalian cell cultures were grown as detailed in section 0. For adherent cells, standard growth medium was removed and cells were briefly washed with PBS. Cell lysis buffer was made by the addition of 10µl 2-mercaptoethanol (2-ME) to 1mL lysis solution (both reagents are provided in the kit). The addition of 2-ME is required to inactive RNases and therefore prevent RNA degradation. For RNA extractions, cells were typically grown in 3 wells of a 6 well plate and pooled together after cell lysis. Lysis buffer was applied directly to the cells (typically 200µl per well of a 6 well plate or 600µl per T75 flask) and a cell scraper was used to disrupt the cells. Cell lysates were frozen at -80°C until required.

2.2.1.2 Lysate filtration and sample preparation

Cell lysates were thawed on ice and the whole lysate was applied to the filtration column which removes cellular debris and shears DNA. Columns were centrifuged and filtrate containing the RNA is mixed with equal volume of 70% ethanol.

2.2.1.3 DNase treatment and column binding

The lysate-ethanol mixture is applied to a binding column and spun down. The columns were then treated with DNase from the On-Column DNase 1 digestion set. The columns were washed with the wash buffers provided in the kit and

DNA was eluted in 50µl elution solution. Eluted RNA was assessed for purity and quantity with a before proceeding with further analysis and experimentation (see section 2.2.8.1).

2.2.2 DNase I pre-treatment of total RNA

RNA extracted from mammalian cells (section 2.2.1) was treated with DNase I to remove genomic DNA contamination prior to cDNA synthesis. This is important as genomic DNA contamination can result in false positive gene expression, which could be detected during PCR assays. The DNase reactions were carried out using the reagents outlined in Table 2-4, all of which were provided with the kit with the exception of DEPC-treated water.

Component	Volume
RNA (total 1µg)	xµl
DEPC treated water	8-xµl
10x DNase reaction buffer	1µl
DNase I	1µl
Total volume	10µl

Table 2-4 DNase I reaction set up

These reagents were mixed and left to incubate for 15 minutes at room temperature. Following this, 1µl EDTA was added and the reaction was incubated for 10 minutes at 65°C to inactivate the DNase I enzyme.

2.2.3 cDNA synthesis from total RNA

Single stranded complimentary DNA (cDNA) is single stranded DNA that is synthesised from mRNA templates by the process of reverse transcription (RT). cDNA was synthesised from DNase I treated total RNA extracted from mammalian cells (section 2.2.2). cDNA synthesis was carried out using random hexamers (6 base pair oligonucleotides of random sequence) provided in the Superscript First Strand Synthesis System for RT-PCT kit

(Invitrogen). This kit uses SuperScript® II reverse transcriptase to catalyse the cDNA synthesis reaction. The cDNA samples generated were used for expression profiling by PCR. What follows is a brief description of the protocol provided by Invitrogen.

For each positive cDNA synthesis reaction (RT positive) a RT negative reaction was set up in which no superscript enzyme was added. This is used as a control for contamination. There should be no cDNA detected in these samples as there is no enzyme. The RT positive and RT negative samples were set up as detailed in Table 2-5. A negative reaction was also set up containing no RNA template.

Component	Volume
DNase treated RNA	5.5µl
DEPC treated water	2.5µl
dNTPs	1µl
Random hexamers	1µl
Total volume	10µl

Table 2-5 cDNA synthesis reaction 1

The reaction mixtures were incubated at 65°C for 5 minutes and placed on ice for 1minute before the following reagents were added.

Component	Volume
10x RT buffer	2µl
25mM MgCl2	4µl
0.1M DTT	2µl
40 units/µl RNaseOUT™ Recombinant Ribonuclease Inhibitor	1µl
SuperScript™ II RT enzyme or water	1µl

Table 2-6 cDNA synthesis reaction 2

The tubes were incubated for 2 minutes at 25°C. To the RT positive reactions, 1µl of SuperScript™ II RT enzyme was added. Tubes were incubated for 10 minutes at 25°C followed by 50 minutes at 42°C and 15 minutes at 70°C. Samples were kept on ice and then treated with 1µl *E.coli* RNaseH for 20 minutes at 37°C in order to degrade the template RNA. The resulting cDNA samples (RT positive and RT negative) were checked for a successful reaction and genomic DNA contamination by PCR using GAPDH primers and gel electrophoresis (sections 2.2.4 and 2.2.5). All cDNA samples were stored at -20°C.

To assess whether cDNA synthesis reactions were successful, a PCR was carried out using primers for the housekeeping gene GAPDH. These primer sequences are shown in Table 2-7.

Primer name	Primer Sequence	Location
GAPDH Forward	AGGTCATCCATGACAACTTTGG	6645523 - 6645544
GAPDH Reverse	GCTTCACCACCTTCTTGATGTC	6646000 - 6646021

Table 2-7 GAPDH primer sequences used in PCR reactions to test successful cDNA synthesis reactions.

2.2.4 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a technique used to amplify a specific sequence of DNA to generate large copy numbers of that specific DNA sequence which is useful for isolation of a specific DNA sequence for cloning or sequencing and quantification of a specific sequence within a sample.[127] PCR relies on thermal cycling which exposes the reactants to repeated cycles of heating and cooling which is optimum for the different stages of the PCR. These stages are denaturation, annealing, extension and final extension. The reaction includes the target DNA, DNA polymerase, two primers (one forward, one reverse) specific to the target sequence, deoxynucleotide triphosphates (dNTPs), buffer and magnesium. The initial step of the reaction denatures double stranded DNA. The following cycles (typically 30-40 cycles) are a

sequence of denaturation, primer annealing and primer extension followed by the final extension. [128]

PCR was carried out using a Techne® PrimeG thermal cycler using the reagents shown in Table 2-8 with the thermocycler set to create the following conditions:

Initial denaturation at 94°C for 3 minutes followed by 35 cycles of:

- 94°C for 45 seconds
- 60°C for 45 seconds
- 72°C for 1.5 minutes

A final cycle at 72°C for 7 minutes at which point samples are kept at 4°C.

Reagent	Volume (µl)
Sterile H_2O (To make up to 25 µl total)	16.75
10x PCR reaction buffer	2.5
MgCl2 (50mM stock)	0.75
dNTPs (5mM stock)	1
5' Primer (10µM)	1
3' Primer (10µM)	1
Taq DNA polymerase	1
Target DNA	1
Total	25

Table 2-8 Reaction components used for PCR mastermix

2.2.5 Agarose Gel Electrophoresis

PCR products were analysed using gel electrophoresis. Gel electrophoresis is a process used to separate DNA fragments based on their size. [129] DNA, being negatively charged, migrates towards a positive charge and bands are

separated on the basis that smaller bands are able to migrate through the gel more quickly.

1% agarose gels were made up by dissolving agarose powder in 1X TAE buffer by heating with the addition of ethidium bromide to a final concentration of 0.5µg/mL. Samples were loaded with using 6x Orange G loading dye. Gels were run in 1X TAE buffer at approximately 45 volts for an hour. DNA bands were visualised under UV light using a Syngene GeneGenius Gel Doc system and GeneSnap software (Syngene).

2.2.6 Quantitative-PCR (Taqman)

Quantitative PCR (qPCR) is a method used to quantify mRNA levels which is used to measure gene activity. [130] The TaqMan® method was used for all qPCR experiments in this thesis. This method uses a primer pair to amplify a cDNA template but in the TaqMan® method, a fluorescent oligonucleotide probe is used to bind a specific sequence between the two primers. The probe is labelled with a fluorescent dye at the 5' end and a quencher dye at the 3' end. The method relies on the 5' to 3' exonuclease activity of Taq DNA polymerase, which cleaves the quencher dye during PCR amplification resulting in accumulation of fluorescence, which is used to quantify gene expression. [131]

Reagent	Volume per reaction (µl)
TaqMan® Gene Expression Master Mix	10
DEPC water	8
PDAR	1
cDNA (1µg)	1
Total reaction volume	20

Table 2-9 qPCR reaction mixture

Reactions were run in triplicate alongside a housekeeper gene, to normalise the expression data. In this thesis, 18s was used as this reference gene. The

GPR126 Taqman™ assay details are shown in Table 2-7. The cycle threshold (Ct) value is the cycle number at which the generated fluorescence crosses a significance threshold (i.e. a level at which the fluorescence exceeds the background level). The lower the C_t value, the greater the quantity of template. Relative gene expression was calculated using the formula - Relative mRNA expression = $2^{-\Delta CT}$. [132]

Interrogated Sequence	Translated Protein	Isoform	Exon Boundary	Assay Location	Amplicon Length
NM_001032394.2	NP_001027566.1	a2	18 - 19	2997	80
NM_001032395.2	NP_001027567.1	b2	18 - 19	2997	80
NM_020455.5	NP_065188.4	a1	19 - 20	3081	80
NM_198569.2	NP_940971.1	b1	19 - 20	3081	80

Table 2-10 GPR126 Taqman™ assay information

Each qPCR reaction was run on a Stratagene qPCR machine using the parameters outlined in Table 2-11. Data was collected for ROX (reference), FAM (gene of interest) and HEX (housekeeper) and analysed using the MxPro software. ROX was set as the reference dye and a background threshold level of fluorescence (dRN) was set for all dyes by examination of amplification plots. Cycle thresholds (CT) for individual samples were determined by this fluorescence threshold.

Temperature	Time	Number of Cycles
50°C	2 minutes	1
95°C	10 minutes	1
95°C	15 seconds	40
60°C	1 minute	

Table 2-11 qPCR cycling parameters

Two housekeeping genes were used to normalise GPR126 expression (Figure 2-2). The idea of a housekeeping gene is to normalise the values obtained

from the gene of interest to a gene which's expression level should not alter under different experimental conditions. GAPDH and 18s are two commonly used housekeeping genes. Both housekeepers showed fairly consistent CT values between the different cell lines however there is less variation between the samples with the 18s housekeeper. Therefore, the 18s PDAR was used as the housekeeper of choice for the remaining experiments.

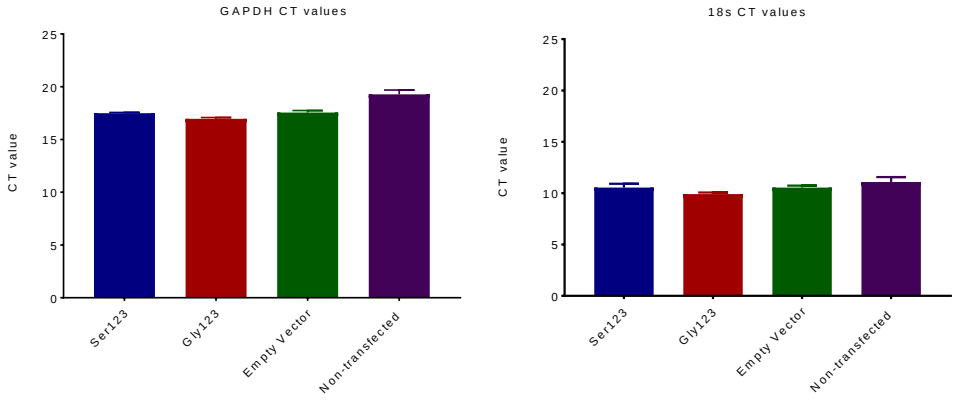


Figure 2-2 CT values for the two housekeeper genes used, GAPDH and 18s. The lower the CT values, the higher the abundance of transcripts.

2.2.7 Plasmid Preparations

In this thesis pLV-GPR126Ser123, pLV-GPR126Gly123 and pLV-EV plasmids were designed and purchased from Vector Builder. The plasmids were delivered in stab culture in cloning host stbl3, a chemically competent strain of *E.coli*. The frozen bacterial stocks were stored at -80°C until needed. The p6CRE/luc plasmid was produced as previously described[133] and propagated from pre-existing glycerol stocks.

Bacterial cells containing plasmids stored in glycerol stocks or stab culture require culturing to replicate the plasmid DNA they contain. The plasmids used in this thesis all contained an ampicillin resistance gene which allowed antibiotic selection of bacterial cells which possessed the plasmid DNA. Following bacterial cell culture and selection, plasmid DNA is isolated from the bacteria and verified by sequencing and restriction digest before use.

2.2.7.1 Bacterial cell culture

Bacterial cells were grown on LB agar plates or in LB broth. LB agar is prepared from a powder in Duran bottles in a 400ml volume and autoclaved. Once cooled, ampicillin was added to the agar to a final concentration of 100µg/ml. Plates were poured and stored at 4°C. Prior to inoculation, plates were pre-warmed to 37°C. A sterile inoculation loop was used dipped into the glycerol stocks/ stab culture and streaked onto the plate. Plates were left to incubate for ~16 hours at 37°C. Plates were checked for colony growth and stored at 4°C for ~8 hours until a colony was picked for inoculation of LB broth. LB broth is also made from a powder stock in 500ml volumes in Duran bottles, immediately autoclaved and stored at room temperature. One colony was picked from each culture with an inoculation loop to inoculate 5ml LB broths which were left to incubate overnight at 37°C with shaking at ~200RPM. From this step, inoculated broths were used for minipreps of DNA (section 2.2.7.2) or used to inoculate larger broths for maxipreps (section 2.2.7.3). For maxipreps, 100mL LB broth was inoculated with 100µl of the 5ml starter broth in a conical flask and left to incubate overnight at 37°C with shaking at 200RPM.

2.2.7.2 Purification of Plasmid DNA (small scale) - Monarch® Plasmid Miniprep

This method utilises the standard method for the purification of plasmid DNA. In brief these standard steps are as follows. The process begins with cell resuspension, alkaline lysis and neutralization followed by binding of the DNA to a column by a silica matrix under high salt conditions. Wash buffers are used to remove salts, proteins, RNA and other cellular components before DNA is eluted.

Before the miniprep was run, 24ml of 100% ethanol was added to 6ml of Monarch Plasmid Wash Buffer 2. Throughout the protocol all centrifugation steps were carried out at 16,000g (~13,000 RPM).

The first step of the mini prep was to pellet 1.5ml bacterial culture by centrifugation for 30 seconds which was then resuspended in 200µl Plasmid Resuspension Buffer. Cells were then lysed with 200µl Plasmid Lysis Buffer and inverted until thoroughly mixed – this turns the solution a uniform dark pink colour – and the solution was incubated for 1 minute. The solution was neutralised with the addition of 400µl Plasmid Neutralisation Buffer and inverted until fully mixed, indicated by a uniformly yellow solution and the formation of a precipitate. The neutralised solution was left to incubate for 2 minutes. Lysate was clarified by centrifugation for 5 minutes and the supernatant was transferred to the spin column and centrifuged for 1 minute. The spin column is placed back into a collection tube and wash buffers are applied to the column. Plasmid Wash Buffer 1 removes RNA, protein and endotoxin and 200µl was applied and centrifuged for 1 minute. The same was done with 400µl Plasmid Wash Buffer 2. Plasmid DNA was eluted in 50µl DNA Elution Buffer.

2.2.7.3 Purification of Plasmid DNA (large scale) - Qiagen Plasmid Endofree Maxi Kit

To generate greater amounts of plasmid DNA, the Qiagen Plasmid Endofree Maxi Kit was used which follows the same principles as the miniprep with DNA bound to the column, washed to remove contaminants and then eluted. The provided RNase A solution was added to buffer P1 and stored at 4°C. To produce with 70% endotoxin-free ethanol, 40mL 100% ethanol was added to the provided endotoxin-free water.

Bacterial cultures were prepared as previously explained. 100mL culture broths were poured into centrifuge bottles and the cells were centrifuged at 6000g for 15 minutes at 4°C using a Beckman Refrigerated Avanti J-301 centrifuge. The pellet was resuspended in 10mL buffer P1 by vortexing and mixed with the addition of 10mL buffer P2. The solutions were left to incubate for 5 minutes at room temperature and the QIAfilter cartridge was prepared. A further 10mL chilled P3 buffer was added to the lysate and mixed by inversion. The lysate was immediately loaded into the QIAfilter cartridge and incubated

for 10 minutes at room temperature. Following this the cell lysate was filtered by removing the cap of the cartridge and gently inserting the plunger to elute the filtered lysate. 2.5mL buffer ER was added to the filtered lysate and mixed by inversion – the mixture was then placed on ice for 30 minutes. A QIAGEN-tip 500 was equilibrated by applying 10mL buffer QBT and left to drain completely. The filtered lysate was then added to the QIAGEN-tip before the tip was washed with 20 x 30mL buffer QC. With the tip in a clean centrifuge tube, 15mL Buffer QN was added to elute the DNA. 10.5mL isopropanol was added to the elution solution and the mixture was centrifuged at 15,000g for 60 minutes at 4°C to precipitate DNA. The supernatant was decanted. To wash the DNA pellet, 5mL endotoxin free 70% ethanol was added and centrifuged at 15,000g for 10 minutes at 4°C. The ethanol supernatant was removed carefully and the pellet was left to dry for 5 minutes. The DNA was then re-dissolved in 600µl endotoxin free buffer TE.

2.2.8 Nucleic Acid Quality Control

To assess DNA, RNA and plasmid DNA, a number of steps were taken to confirm quality and quantity of nucleic acid material before it was used for any application. For cDNA synthesis, the RNA concentrations are normalised to ensure the same amount of cDNA (500ng) is synthesised per reaction. For transfections, the amount and concentration of the plasmid DNA is required to be of a certain standard and are generally normalised to reduce variation between different plasmid transfections. RNA is required to be of a standard before applications such as RNA-sequencing are carried out.

2.2.8.1 NanoDrop Quantification of Nucleic Acids

All RNA, cDNA and plasmid DNA purified was assessed using a Nanodrop ND-2000 Full spectrum Spectrophotometer (Thermo Fisher Scientific). This piece of equipment allows the rapid and easy quantification and assessment of the purity of samples. The machine measures the absorbance values of the sample at different wavelengths.

The 260/280 and 260/230 ratios are used to determine the 'purity' of the samples with a 260/280 value of ~1.8 and ~2 for DNA and RNA respectively being considered pure and free from protein and salts contamination. A 260/230 ratio above the respective 260/280 ratio is commonly considered pure and expected values lie between 2-2.0.

2.2.8.2 RNA Integrity Number

In addition to assessing RNA sample quality by NanoDrop, the RNA integrity Number (RIN) was calculated for each RNA sample used for RNA-sequencing (section 2.2.10). RIN analysis was carried out to assess the suitability of samples for RNA-sequencing with samples with a RIN score >8 generally considered optimal for sequencing. [134] This analysis was carried out using by the Post Genomics Facility, University of Nottingham using the Agilent 2100 Bioanalyser with the Agilent RNA 6000 Nano Kit. This kit includes Agilent RNA 6000 Nano gel matrix and RNA 6000 Nano dye.

The following protocol was performed by the Post Genomics Facility. To prepare the gel-dye mix, the RNA matrix gel was centrifuged at 1500 rpm for 10 minutes at room temperature. The Nano dye was vortexed and spun down. RNA dye was added to the matrix gel at a ratio of 1µl loading dye to 65µl matrix gel. The solution was vortexed and centrifuged at 13000rpm for 10 minutes. 9µl RNA matrix gel was loaded into the specific well of a RNA chip which was placed in a chop priming station. The priming station was closed for 30 seconds. A further 9µl of the RNA matrix gel was added into the same well. Following this, 5µl of the RNA 6000 Nano marker was loaded into the sample and ladder wells and 1µl of ladder or RNA sample was loaded into the specific wells. The chip was vortexed at 2400 rpm for 1 minute and run on the Agilent 2100 Bioanalyser. The output from this process is an electropherogram of the ladder and sample results as well as RIN and RNA concentration.

2.2.9 Sanger Sequencing

Sanger sequencing, also known as chain termination sequencing, is a method for determining the nucleotide sequence of DNA.[135] This method of

sequencing denatures the target DNA and anneals it to an oligonucleotide primer, which causes 5' to 3' elongation. DNA fragments of varying lengths are produced which are terminated with by combining primers, nucleotides (dATP, dCTP, dGTP and dTTP), dideoxynucleotides (ddATP, ddCTP, ddGTP and ddTTP), DNA polymerase and the DNA sequence of interest. The elongation is terminated by the addition of a labelled dideoxynucleotide which are inhibitors of DNA polymerase. This results in varying fragments of DNA all with a labelled (e.g. fluorescent) dideoxynucleotide at the end. DNA sequences are then separated by length using gel electrophoresis and each labelled nucleotide is excited by a laser and as each base is tagged with a different label, the light emitted can be attributed to the correct base. The laser generates a chromatogram showing fluorescent peaks for each nucleotide.

The sequencing reaction was carried out by The DNA Sequencing Facility (School of Life Sciences, University of Nottingham) utilising dye terminator chemistry. The facility use the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). The facility were provided with 5µl DNA plasmid (100ng/µl) and 3µl of primer (10pmole). Sequence analysis was carried out using a 3130xl ABI PRISM Genetic Analyzer (Life Technologies). Analysis of the sequence data was facilitated using GeneScreen (University of Leeds). (136)

2.2.10 RNA sequencing

RNA sequencing (RNA-seq) uses next generation sequencing (NGS) to analyse RNA from biological samples. NGS is a massively parallel high throughput sequencing method that allows rapid sequencing of base pairs in DNA or RNA samples. RNA-seq is now the preferred methodology for the study of gene expression. In brief, short cDNA fragments are prepared from the RNA samples and loaded onto glass flow cells where they are amplified into an array of millions of clones each of which contains a unique sequence to be determined. The cDNA templates are read nucleotide by nucleotide by measuring the fluorescence of each cluster after the incorporation of labelled nucleotides. After each cycle, a light source excites the fluorescent nucleotide

which emits a fluorescent signal (known as base-calling). After each cycle the fluorescent molecules are removed and the process is repeated. The number of cycles determines the length of the read.

2.2.10.1 Library Preparation and Sequencing

Library preparation involves converting the RNA population into a library of cDNA fragments with adaptor sequences attached to both ends. Library preparation and sequencing was carried out by Oxford Genomics Centre. Their library preparation begins with the enrichment for mature polyadenylated (poly-A) transcripts by annealing to oligo-(dT) magnetic beads. This method has the benefit that the prepared libraries and therefore the resulting data predominantly relates to protein coding transcripts.

Paired end sequencing was then carried out by Oxford Genomics Centre using an Illumina HiSeq4000 generating 25-30 million paired end reads for each sample.

2.3 PROTEIN DETECTION METHODS

2.3.1 Immunohistochemistry

Immunohistochemistry utilises immunological and biochemical techniques to selectively image antigens, and therefore proteins, by using labelled antibodies which specifically bind their target antigens. Visualisation of the antibody-antigen interaction is achieved in a number of ways, however the most common cases, the antibody is conjugated to an enzyme, such as peroxidase, which catalyses a colour-producing reaction. Other methods of visualisation may involve immunofluorescence. Reagents for immunohistochemistry were from the Dako REAL EnVision Detection System Peroxidase/DAB, Rabbit/Mouse.

Paraffin wax was removed by washing slides in histoclear, 100% ethanol and 95% ethanol each for 2x 10minutes. Slides were then washed in running water. For the antigen retrieval step, 750mL tap water was placed into a steamer and a sufficient volume of sodium citrate buffer to cover the slides was placed into an empty box which was placed in the steamer. Slides were treated in the sodium citrate buffer for 10 minutes and left to cool in the steamer for a further 20 minutes.

A humidified area was prepared for the immunohistochemistry by using wet paper towels. A hydrophobic pen was used to draw around sections on the slides before 100µl peroxidase blocking solution was added to cover the section, which was left on for 5 minutes. Slides were washed for 5 minutes in Phosphate-buffered saline (PBS) prior to the adding of 100µl of primary antibody (Anti-GPR126 antibody produced in rabbit – Sigma Aldrich HPA017346) which was diluted 1/50, 1/100, 1/250 and 1/1000 in antibody diluent. Slides were incubated for 1 hour. Slides were washed with PBS for 5 minutes and the secondary antibody was applied to cover the section (100µl). Slides were incubated for 30 minutes after which the slides were washed for 5 minutes in PBS. 3,3'-Diaminobenzidine (DAB) solution was made up by mixing 1ml DAB buffer solution and 20µl DAB provided in the Dako kit. 100µl of the

DAB solution was applied to cover the specimen and left to incubate for 5 minutes.

Slides were counterstained in Mayers haematoxylin for 10 seconds before being washed in running tap water. Slides were washed in 95% ethanol, 100% ethanol and histoclear each for 2x 5minutes. Slides were left for mounting. Once dry, slides were mounted using Vectamount and left to set. Slides were examined using microscopy.

2.4 BIOINFORMATICS

2.4.1 Processing Sequencing data

The resulting sequencing data was provided by Oxford Genomics Centre in FASTQ format. In this format, each nucleotide is associated to an encoded quality number corresponding to a Phred score (Q) which is a measure of the accuracy of the base call. This is therefore a measure of the quality of the sequencing and allows us to identify and trim regions of low quality. (137) FASTQ files were provided for both forward and reverse reads of each sample and corresponding to the multiple lanes of sequencing data. Processing of the RNA-seq data was performed using a linux operating system with the commands provided. (138)

Full commands and explanations of the tools used can be found in the appendices.

2.4.1.1 Concatenation of FastQ files

The first step of processing these samples is to concatenate the FASTQ files from each of the sequencing lanes. This combines the sequencing data from each of the multiple lanes of sequencing for each sample.

2.4.1.2 Adapter and Quality-Trimming

The adaptor sequences are trimmed from the concatenated files using the Scythe programme (https://github.com/vsbuffalo/scythe) and further quality trimming is achieved using the Sickle programme (https://github.com/najoshi/sickle) which trims reads with a Q score <20.

2.4.1.3 FASTQC Analysis of Data

The final step in quality control is the FASTQC analysis which generates graphical outputs to show the quality of sequence data. An example of these visual outputs is shown below in the appendices.

2.4.1.4 Alignment of reads using TopHat2

TopHat2 aligns the sequence reads to the genome of choice using a high-throughput short read aligner called Bowtie (http://bowtie-bio.sourceforge.net). In this project, TopHat2[139] was used to align RNA-seq reads to the human assembly GRCh37, the output of which is a single .bam file for each RNA sample which is used for downstream analysis. The data processing steps using the TopHat and Cufflinks pipeline is shown in Figure 2-3.

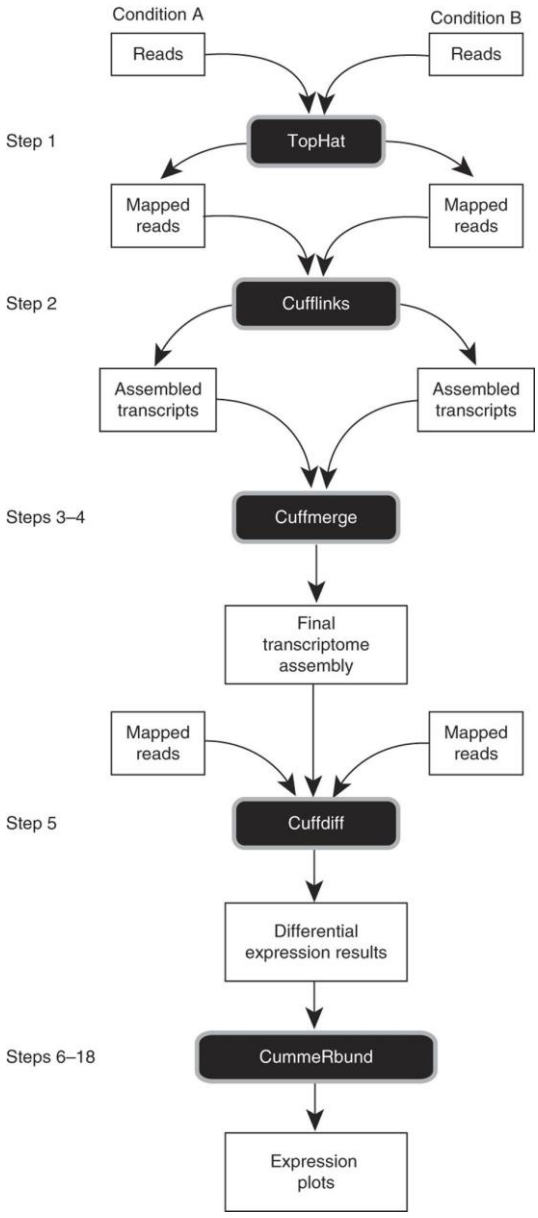


Figure 2-3 Summary of the Cufflinks tools used for RNA-seq analysis and the pipeline of data processing. [(139)]

The trimmed fastq files enter the analysis and the output is .bam and .bed files, the latter of which contains identified insertions, deletions and splice junctions.

2.4.1.5 Transcript Assembly using Cufflinks and Cuffmerge

Cufflinks is used to assemble the transcript by assigning genes to the TopHat analysis files. This creates transcript .gtf files which list the FPKM (Fragments Per Kilobase Million) values for genes and detected isoforms for each sample.

Cuffmerge is then used to merge all the transcript files into a single transcriptome assembly. This creates a merged.gtf for the total population of samples.

2.4.1.6 Gene and Transcript Expression (Cuffnorm)

In order to determine the individual expression levels of both gene and isoform expression we run Cuffnorm. This requires the merged.gtf file created by Cuffmerge and a user generated sample sheet directing the programme to the .bam files generated by TopHat.

2.4.1.7 Differential gene expression analysis using CuffDiff

Cuffdiff is used to identify differentially expressed genes and isoforms between different samples or different conditions based on their mean expression value across identified biological replicates.

2.5 STATISTICAL ANALYSIS

2.5.1 Analyses of Variance

Statistical tests were used to identify variance in the means (parametric) or medians (non-parametric) between two or more groups. For the comparison of two groups, a Student's t-test (parametric) was used for normally distributed data and a Mann-Whitney test (non-parametric) was used for data that was not normally distributed.

For analysis of variance across 3 or more groups a one way ANOVA (parametric) or a Kruskal-Wallis test (non-parametric) was performed. For paired data, a Friedman test was performed. Post-hoc tests were carried out in order to assess the effect of each group on the variation. The Tukey's test compares all columns to each other, where the Dunn's test compares groups to a selected control group.

A two-way ANOVA is a similar test to the one-way ANOVA in that the comparison is between multiple groups. However, a two-way ANOVA allows the testing of variance between groups where there are two ways of classifying the data, e.g. time and dose. This test can determine the main contributory effect of each independent variable and whether there is an interaction between independent variables. A post-hoc test was performed to identify variance within independent variables. A Dunnett's test was used to compare rows to control rows within independent variables.

All statistical analysis was carried out using GraphPad Prism 7.04.

3 Bioinformatic Analysis of the 6q24 Association Signal and Expression Profiling of *GPR126* mRNA and Protein in Human Lung and Airway Cells

3.1 INTRODUCTION

The 6q24 locus is one that has reproducibly been associated with lung function measurements in a number of GWAS since 2010.[49, 51, 54] In 2015, Soler Artigas et al identified a novel signal at this locus which showed association with FEV_1/FVC (Figure 1-4).[54] To summarise, the signal with the most significant p-value from this region ($P=9.6 \times 10^{-26}$) originates from intergenic SNP rs148274477 (C/T) near *GPR126.* This variant has a minor allele frequency (MAF) of 0.03 in the 1000 genomes European and Asian populations, 0.01 in African populations and 0.06 in American populations. The C allele (frequency = 0.97) is associated with a reduction in FEV_1/FVC (beta = -0.162). Furthermore, a recent GWAS identified rs17280293 (r^2=0.7 with rs148274477) to be associated with reduced FEV_1/FVC with the common allele (A) presenting the risk of reduced lung function (beta = -0.1803).[56] The variant lies within the GPR126 coding region and causes missense mutation Ser123Gly.

More recently, the 6q24 signal has since been replicated for FEV_1/FVC ($p=4.48 \times 10^{-17}$) with the rare allele in this instance, T, showing an association with increased FEV_1/FVC (beta = 0.175). In addition, the C (common) allele has shown association with COPD risk (odds ratio = 1.21, $p=1.77 \times 10^{-6}$) by Wain et al (2017).[55] This region of association is shown in the region plot in Figure 3-1. An independent signal from an intronic SNP in *GPR126* (rs9399401 (T/C), MAF=0.28, $p=1.81 \times 10^{-19}$) has also been recently associated with COPD with the T allele being the risk allele (odds ratio = 1.17).[57]

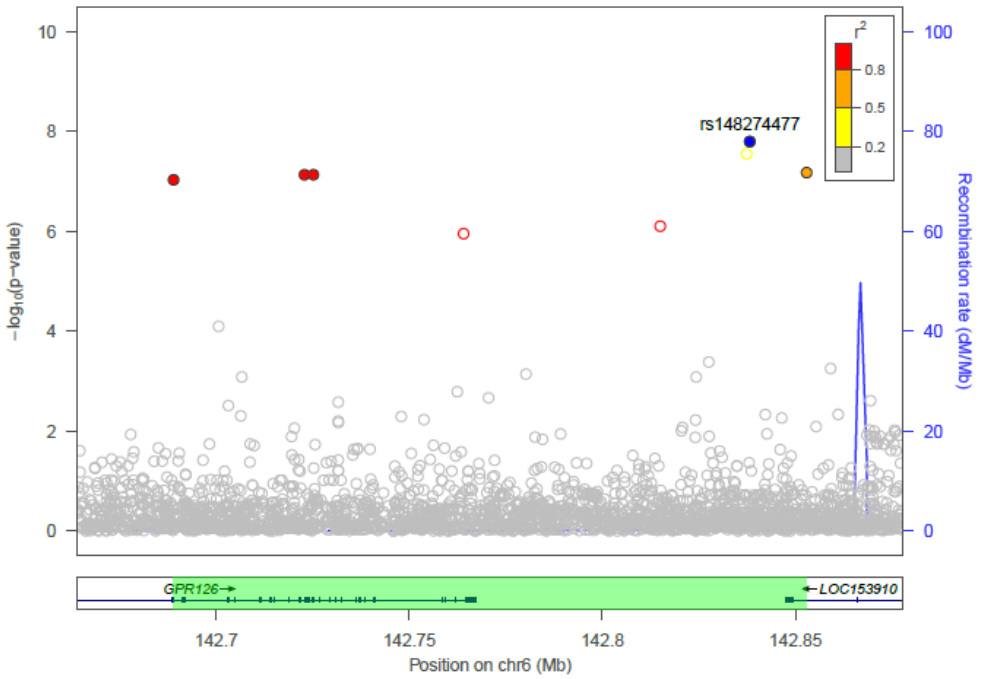


Figure 3-1 Region plot from stage 1 of the GWAS showing association signal with FEV1/FVC at the chromosome 6q24 locus from Wain et al (2017).[55] The x axis shows the chromosomal location annotated with gene names and the y axis shows the -log p value. The sentinel SNP (rs148274477, p=4.48 x 10^{-17} from combined stage 1 and stage 2 analyses) is highlighted in blue and SNPs in LD are colour coded, with SNPs in high LD (r2>0.8) shown in red.

As alluded to in the general introduction, GWAS alone cannot identify the causal SNP(s) driving an association signal and further genetic and bioinformatic analyses are required to unravel the often complex relationships between genotype and phenotype. For example, further analyses may include eQTL analysis, i.e. the effect of a variant on gene and/or protein expression, linkage disequilibrium (LD) analysis and identification of deleterious variants. Improved imputation and a greater understanding of the pattern of inheritance of genetic variation thanks to endeavours like the HapMap project increases the number of SNPs that can be interrogated with GWAS. However, even with these advances identifying causal SNPs is still challenging. Coding region variants are often considered as candidate causal genes due to their effects on protein structure/ function, which can be examined by functional studies. While coding variants present as obvious targets for further analysis, non-

coding variants can also have profound effects. [68] This can be demonstrated by the *HHIP* example discussed in the general introduction in which two SNPs within an *HHIP* enhancer region from the haplotype associated with COPD risk have been shown to modulate *HHIP* expression. These variants can be examined thanks to projects such as ENCODE, which aims to annotate functional elements of the non-coding genome such as promoters and enhancers of gene transcription. These datasets help identify effects of variants that lie within the non-coding genome, which may alter the regulation of gene expression.

In addition to the physical chromosomal location of a variant and possible functional elements it may lie in, LD is another factor which may influence the interpretation of an association signal. LD is the difference between the observed frequency and the expected frequency for random association of a combination of alleles at different loci. The degree of LD is shown using the r^2 value (0-1) with the higher the number representing the higher the likelihood that the alleles are inherited together. LD allows a number of variants at the same locus to be associated with a phenotype when only one might be causative.[140]

A further consideration is expression quantitative trait loci (eQTL) effects. An eQTL is a locus that contributes to altered expression of mRNA or protein for a gene. Non-coding variants located within promoter or enhancer regions of the genome may result in altered expression of a gene, which is why projects such as ENCODE are so important. Taking all these factors into consideration, it is important therefore to analyse the GWAS signal appropriately and to analyse the potential contribution of all variants in an association signal.

This chapter aims to identify the potential effects of variants in the 6q24 association signal and identify candidate causal variants for follow up analyses. To do this, there are a number of *in silico* tools and datasets are openly available. Variants in LD with the sentinel SNP can be identified using openly available tools such as HaploReg [140] and LDlink [141]. In addition, eQTL datasets such as GTEx [142] and the Blood eQTL browser by Westra et al (2013) [143] are widely available.

An important factor to consider when attempting to identify the role of a particular gene in a phenotype, is to determine any known role of the gene including whether the gene and protein are expressed in the tissue or cell types of interest. As detailed in the general introduction we know that *GPR126* is expressed throughout mouse embryonic development including in the lung. [77, 103] Expression is also seen in the inner ear of zebrafish [105] and in mouse and zebrafish Schwann cells throughout development and in adulthood. [101, 102] The gene is known to play vital roles in the nerves [97, 101, 102, 104], cardiovascular system [103] and inner ear [105] and *GPR126* knockout in animal models is fatal due to cardiovascular failure, yet any potential role in the lung is as yet unknown.

Similarly, the role of *GPR126* in the human lung has not been identified. However, variants in the gene are associated with a number of human traits, aside from pulmonary phenotypes, such as scoliosis [115], height [114], periodontitis [144] and certain ear phenotypes [145] as well as mutations in the gene being a known cause of Arthrogryposis Multiplex Congenita (AMC).[117] Interestingly, the missense variant resulting in the Ser123Gly substitution, rs17280293, has recently been shown to be associated with the diffusing capacity of the lung ($p=1.4\times10^{-10}$) measured by carbon monoxide uptake (D_{LCO}) and per alveolar volume (V_A). [146] It was also shown by qPCR that GPR126 expression was lower in lung tissue patients with decreased diffusing capacity and in COPD subjects compared to healthy controls. [146] A list of GPR126 sentinel variants shown to be associated with a number of different human phenotypes accessed through the GWAS catalogue (https://www.ebi.ac.uk/gwas) is shown in Table 3-1. As shown in the table all of these SNPs are in low LD ($r^2<0.1$) with the sentinel SNP for FEV_1/FVC suggesting that they are independent signals.

Variant and risk allele	Rare allele frequency	Reported trait	PubMed ID	First Author	R^2 with rs148274477
rs148274477-T	**0.0263**	**Lung function (FEV_1/FVC)**	**28166213**	**Wain LV**	**1**
rs17280293-A	**-**	**Diffusing capacity of the lung (D_{LCO}/V_A)**	**30049742**	**Terzikhan N**	**0.686**
rs1329705-?	0.8	Airflow obstruction (FEV_1)	22837378	Wilk JB	0.086
rs3817928-A	0.78	Lung function (FEV_1/FVC)	29394082/ 23284291	Burkart KM/ Hancock DB	0.085
rs9399401-T	0.72	Chronic obstructive pulmonary disease	28166215	Hobbs BD	0.03
rs225694-?	NR	Adult height	20881960	Lango Allen H	0.022
rs4637667-A	0.342	Adolescent idiopathic scoliosis	26211971	Ogura Y	0.022
rs7763064-A	0.29	Height	20881960	Lango Allen H	0.022
rs7763064-A	0.288	Birth length	25281659	van der Valk RJ	0.022
rs3748069-A	0.74	Height	18391951	Gudbjartsson DF	0.019
rs6570509-T	0.2874	Hip circumference adjusted for BMI	25673412	Shungin D	0.015
rs7741741-T	0.72	Height	23563607	Berndt SI	0.015
rs6570507-?	0.43	Scoliosis	23666238	Kou I	0.01
rs6570507-?	0.42	Height	19343178	Soranzo N	0.01
rs6570507-A	0.38	Height	25429064	He M	0.01
rs6570507-A	0.3186	Waist circumference	28443625	Justice AE	0.01

rs6570507-A	0.286	Waist circumference adjusted for body mass index	25673412	Shungin D	0.01
rs6570507-G	0.71	Height	20397748	Liu JZ	0.01
rs7753012-G	0.3072	Height	28270201	Nagy R	0.005
rs7753012-G	0.3	Lung function (FEV_1)	28166213	Wain LV	0.005
rs7753012-G	0.3018	Lung function (FEV_1/FVC)	28166213	Wain LV	0.005
rs4896582-A	0.27	Height	18391950	Lettre G	0.004
rs4896582-A	0.299	Height	25282103	Wood AR	0.004
rs4896582-G	0.6606	Waist circumference	28448500	Graff M	0.004
rs6570508-G	0.6695	Waist circumference adjusted for body mass index	28448500	Graff M	0.004
rs9496398-?	NR	Neutrophil count	21507922	Ramsuran V	0.002

Table 3-1 A list of GPR126 sentinel variants associated with human phenotypes as reported in the GWAS catalogue (https://www.ebi.ac.uk/gwas/home, accessed April 2019). The sentinel SNP for FEV1/FVC is shown in red and the degree of LD with all other SNPs are shown.

From the literature it is reported that *GPR126* mRNA is expressed in the foetal lung [100] and in human endothelial cells. [147] In this chapter, I assessed GPR126 mRNA and protein expression in the lung and in airway relevant cells using qPCR, RNA sequencing and immunohistochemistry.

3.1.1 Aims

The overall aim of this chapter is to identify potential causal variants to further define the association signal including identification of the extent of LD in the region and identification of candidate causal variants for functional follow-up analyses. I also sought to identify GPR126 mRNA and protein expression patterns in the lung and in airway cells.

Specifically, the aims were:

- To determine the extent of the LD in the association signal.
- To identify potentially causal SNPs from the 6q24 locus using *in silico* bioinformatic analyses.
- Assess the potential functional effects of variants in this association signal.
- Identify GPR126 mRNA and protein expression in the human lung as a prelude to functional studies.

3.2 METHODS

3.2.1 Bioinformatic analysis of rs148274477 signal

To identify plausible variants underlying the association between *GPR126* and FEV_1/FVC, bioinformatic analysis was carried out to identify SNPs in linkage disequilibrium (LD) with rs148274477. HaploReg v4.1 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) [140] and LDlink (https://ldlink.nci.nih.gov/) [141] were used to annotate SNPs in LD with the sentinel SNP rs148274477. HaploReg v4.1 uses LD information from the 1000 genomes project to identify linked SNPs along with chromatin state and protein binding motifs from the Roadmap Epigenomics and ENCODE (https://encodeproject.org) projects.[140] The program also gives information on sequence conservation, the effect of variants on regulatory motifs and eQTL data.

SNPs identified as potentially damaging, e.g. having a known eQTL effect, lying within a known regulatory element or resulting in a missense mutation (a change in amino acid sequence) are taken forward for further analysis. For example, variants resulting in amino acid sequence changes can be analysed using *in silico* tools to predict functional effects on protein structure/ function.

3.2.1.1 Analysis of Variants in Linkage Disequilibrium with rs148274477

LDlink is a collection of bioninformatic tools which uses data from phase 3 of the 1000 Genomes Project. LDproxy was used to analyse LD information as well as predict consequences of genetic variants in the LD block from RegulomeDB (http://www.regulomedb.org/). [141, 148] RegulomeDB is a database that annotates SNPs with known and predicted regulatory features in the genome. These include regions of DNAase hypersensitivity, transcription factor binding sites and promoter and enhancer regions which have been reported in public datasets such as GEO (https://www.ncbi.nlm.nih.gov/geo/) and ENCODE and in published literature. [148]

3.2.1.2 eQTL Analysis of SNPs in LD with rs148274477

Each SNP within the haplotype block from LDproxy and Haploreg was analysed for potential eQTL effects. eQTL analysis was carried out using available online resources GTEx [142], Blood eQTL browser by Westra et al [143], HaploReg v4.1 [140] and in a lung eQTL data set.[149]

3.2.1.3 Predicting Functional Effects of GPR126 Variants using in silico tools

To predict potentially pathogenic results of the rs17280293 (Ser123Gly) substitution on protein structure and function, *in silico* analysis was carried out using a number of freely available online tools. PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) [150], SIFT (http://sift.bii.a-star.edu.sg/) [151] and Mutation Taster (http://www.mutationtaster.org/) [152] were used to assess the effects of the amino acid substitution. SIFT makes predictions based on the degree of conservation and PolyPhen-2 predicts how damaging a change is by assessing the potential effects of the substitution on protein structure and function. Mutation Taster is designed to predict functional consequences of amino acid substitutions, short indels (insertions/deletions) and synonymous variants taking into account population genetics data.

3.2.2 GPR126 expression in human cells and tissue by quantitative-PCR

RNA was isolated from the cell lysate of BEAS2B cells using the GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) as per the protocol outlined in section 2.2.1. From the resulting RNA, cDNA was synthesised using the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitorgen) using the standard protocol provided by the manufacturer (Section 2.2.3). Additional cDNA samples were provided by Dr Amanda Henry (Division of Respiratory Medicine, University of Nottingham, UK) which were as follows: HASM (donor 0501) and HBEC (donor 4F0507) as well as whole tissue cDNA from Brain, Kidney, Placenta and Lung (RNA from Ambion). Additional HBEC cDNA from cells at passage 3 was provided by Dr Sangita Bhaker (Division of Respiratory Medicine, University of Nottingham, UK).

Using the method outlined in chapter 2, qPCR was performed on the cDNA samples to assess *GPR126* expression levels with *GPR126* TaqMan® PDAR from Thermo Fisher Scientific and an 18s PDAR (Life technologies) housekeeper. Relative gene expression was calculated using the formula: relative mRNA expression = $2^{-\Delta CT}$, where ΔCT is the CT value from the gene of interest PDAR minus the CT value of the housekeeper. (132)

3.2.3 Identification of GPR126 Expression in Human Airway Cells by RNA Sequencing

RNA was collected from five HASM donors (D006, D007, D0023, D0063 and D00529 – donor information can be found in section 2.1.2) at passage 3 from 3 wells of a 6 well plate by Dhruma Thakker and Dr Michael Portelli (Division of Respiratory Medicine, University of Nottingham, UK). RNA was isolated from cell lysate using the RNeasy Mini Kit (Qiagen) as per the manufacturer's protocol. For RNA-seq data for HBECs, RNA was collected from eight donors at passage 2 (D1, D2, D3, D4, D7, D8, D9 and D11 – full donor information is detailed in section 2.1.2) and isolated by Jon O'Loughlin (Division of Respiratory Medicine, University of Nottingham). Extracted RNA was analysed for quantity and quality by nanodrop and sent to Oxford Genomics Centre for RNA sequence analysis as described in the general methods.

In brief, RNA sequencing data received from Oxford Genomics Centre (75bp reads, 25-30 million reads per sample) was processed through quality control steps beginning with concatenation of sample files, trimming of the adaptor sequences using Scythe, quality trimming using Sickle and production of FastQC files to assess the quality of the samples. The samples were then aligned to the human genome using TopHat (build GRCh37) and a transcript assembly was run using CuffLinks which assigns genes to the TopHat files. CuffNorm was then run to obtain expression values in FPKM for both gene and isoform expression. The full pipeline for RNA-seq data processing is covered in detail in section 2.4.

3.2.4 GPR126 expression in GEO datasets

The Gene Expression Omnibus (GEO) is a public functional genomics data repository which supports uploads from array- and sequence-based data for public use. GEO datasets were accessed (https://www.ncbi.nlm.nih.gov/geo/) and used to assess whether *GPR126* mRNA expression varied between cases of COPD and asthma and healthy controls. Using RNA sequencing data obtained from dataset GSE57148 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57148), *GPR126* expression between healthy lung and COPD lung was compared.

The originating study is an analysis of gene expression in whole lung tissue from patients with COPD and subjects with normal spirometry by RNA-seq.[153] In the study, RNA was extracted from lung tissue from 98 COPD subjects and 91 subjects with normal spirometry.[153] The subjects in the study were patients registered in the Asan Biobank who required resection for lung cancer. The inclusion criteria were a $FEV_1/FVC < 0.7$ for the COPD group and normal spirometry for the control group. *GPR126* FPKM values were downloaded from GEO and values from the healthy group and the COPD group were compared using a Mann-Whitney test.

In addition, using microarray data from GEO datasets, *GPR126* expression in healthy control and asthma epithelium was compared between asthma and control samples. Dataset GSE43696 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43696), contains microarray data from epithelial cells obtained from bronchial brushings from healthy controls (n=20), mild-moderate asthma (n=50) and severe asthma (n=38).[154] Asthmatic subjects met the American Thoracic Society (ATS) criteria for the diagnosis of asthma. Mild-moderate asthma subjects had an FEV_1 >60% predicted with or without low-moderate dose inhaled corticosteroid. Severe asthma subjects met the ATS definition for severe refractory asthma including the use of high-dose inhaled corticosteroid and/or use of oral corticosteroid with continuing symptoms and/or chronic airflow inflammation. Healthy control subjects had no history of respiratory disease or

respiratory infection, normal lung function and no evidence of bronchial hyper-responsiveness.

The GSE89809 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89809) [155] dataset also contained microarray expression data from epithelial cells obtained by endobronchial epithelial brushings from healthy control subjects and asthmatic subjects. The study participants were as follows: 19 healthy non-atopic subjects, 15 mild asthmatic patients on β_2-agonists alone, 17 subjects with moderate asthma on inhaled corticosteroid and 14 subjects with severe asthma with persistent symptoms despite high-dose inhaled corticosteroids.

From GSE43696 values from probe 1553025_PM_at were used and from GSE89809 probe, 2013 A_24_P214756 was used as these probes covered all *GPR126* transcripts. Values between cases and controls were compared using an ANOVA test.

3.2.5 GPR126 Protein Expression in Human Lung Tissue

Immunohistochemistry was used to identify GPR126 protein expression in seven lung tissue donors and one small intestine donor. Small intestine tissue was used as a positive control tissue as GPR126 protein expression has been previously identified in the small intestine. [156] Donor information and ethics approval numbers for all donors are documented in Table 3-2.

A standard protocol was used as detailed in section 2.3.1using the Dako REAL EnVision detection system. In brief, slides were cut from paraffin embedded blocks prepared by the pathology department (Queen's Medical Centre). Slides were washed in a series of washes from histoclear to 100% ethanol. Antigen retrieval was achieved by steaming in sodium citrate buffer. Peroxidase blocking solution (Agilent Dako) was applied to the slides before slides were incubated with primary antibody (Anti-GPR126 antibody produced in rabbit (0.05mg/ml), Sigma HPA017346) at a 1:25 dilution in antibody diluent (Agilent Dako) or isotype (Rabbit IgG isotype, Invitrogen 10500C) diluted from 3mg/ul to 0.05mg/ml then 1:25 in diluent. The primary antibodies were left to

incubate for one hour at room temperature followed by a half-hour incubation with the secondary antibody (Agilent Dako). The substrate-chromogen solution (DAB) (Agilent Dako) was applied to the slides and the slides were counterstained with Mayer's Haematoxylin before the slides were run back through the series of alcohols to histoclear. Slides were mounted using Vectamount (Vector Laboratories). Slides were imaged using a Nikon Diaphot 300 fluorescent microscope. A sub-section of slides; lung tissue (TMM140) and small intestine tissue were imaged by nanozooming by the pathology department (Queen's Medical Centre). Images were analysed using the NDP.view2 software.

Tissue type	TMM number	Study name and number	Source	Ethics number
Human small intestine	TMM248 (D1)	Study 42 Obtaining lung tissue samples for laboratory research	Histopathology Dept, Nottingham University Hospitals NHS Trust (HTA licence 12265-8/10)	08/H0712/34
Human lung	TMM140 (D2)	Study 42 Obtaining lung tissue samples for laboratory research	Histopathology Dept, Nottingham University Hospitals NHS Trust (HTA licence 12265-8/10)	08/H0407/1
Human lung	TMM139 (D3)	Study 42 Obtaining lung tissue samples for laboratory research	Histopathology Dept, Nottingham University Hospitals NHS Trust (HTA licence 12265-8/10)	08/H0407/1
Human lung	TMM377 (D4)	Study 54 Nottingham Research Biorepository - STAT	Nottingham Health Science Biobank, City Hospital, Nottingham	15/NW/0685
Human lung	TMM434 (D5)	Study 54 Nottingham Research Biorepository - STAT	Nottingham Health Science Biobank, City Hospital, Nottingham	15/NW/0685
Human lung	TMM458 (D6)	Study 54 Nottingham Research Biorepository - STAT	Nottingham Health Science Biobank, City Hospital, Nottingham	15/NW/0685
Human lung	TMM454 (D7)	Study 54 Nottingham Research Biorepository - STAT	Nottingham Health Science Biobank, City Hospital, Nottingham	15/NW/0685
Human lung	TMM328 (D8)	Study 54 Nottingham Research Biorepository - STAT	Nottingham Health Science Biobank, City Hospital, Nottingham	15/NW/0685

Table 3-2 Details of tissue used for immunohistochemistry.

3.3 RESULTS

3.3.1 Missense Variant Ser123Gly is in Linkage Disequilibrium with Sentinel SNP rs148274477

LD analysis was carried out to identify SNPs inherited with sentinel SNP rs148274477 (MAF=0.03) and to identity potential effects of variants within this haplotype (figure 3:5). The SNP rs148274477 was entered into HaploReg V4.1 and LDproxy with a low stringency r^2 value of 0.2 set as more stringent values (i.e. higher r^2 values) did not yield many results. So to further interrogate the region, a low r^2 value was set to capture more SNPs in this LD block.

Figure 3-2 shows the LD plot from LDproxy which shows the SNPs in LD with the query SNP (blue) and their r^2 values on the y axis. The SNP in highest LD with rs148274477 is a missense variant resulting in amino acid substitution (Ser123Gly) rs17280293 with MAF=0.04. This SNP lies within a domain (CUB domain) thought to be important for ligand binding. [98]

A further 8 SNPs were in LD (r^2>0.2) with the sentinel SNP (Table 3-4). All of these variants were intergenic or intronic, with most not showing any RegulomeDB features indicating that these variants do not lie in known or predicted regulatory elements. Variant rs148274477 was also run on HaploReg v4.1 to assess the potential functional effects of SNPs within this LD block using ENCODE data. Missense variant rs17280293 is predicted to effect a number of motifs and also scores highly on RegulomeDB (Table 3-5). RegulomeDB scores SNPs based on whether they in any identified regulatory regions which is summarised in Table 3-3.

Score	Supporting data
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding / DNase peak
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak
2b	TF binding + any motif + DNase Footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
4	TF binding + DNase peak
5	TF binding or DNase peak
6	other

Table 3-3 RegulomeDB scoring criteria

Table 3-4 shows the output from LD proxy using rs148274477 as the query SNP. Only 9 SNPs are in LD with rs148274477 (r^2>0.2), only one of which results in a functional variation. The SNPs are also annotated with chromosomal position, the alleles and their frequency, distance from the query SNP, r^2 value showing the degree of disequilibrium, any functional effect of the variation and the regulomeDB score which scores variants based on any known or predicted regulatory features. SNP rs55771139 is predicted affect a region important for transcription factor binding or a region of DNase hypersensitive. rs17280293 is the SNP in highest LD with rs148274477 (r^2 = 0.9262) resulting in missense variant in GPR126 Ser123Gly which is also predicted to affect transcription factor binding and DNase hypersensitivity sites.

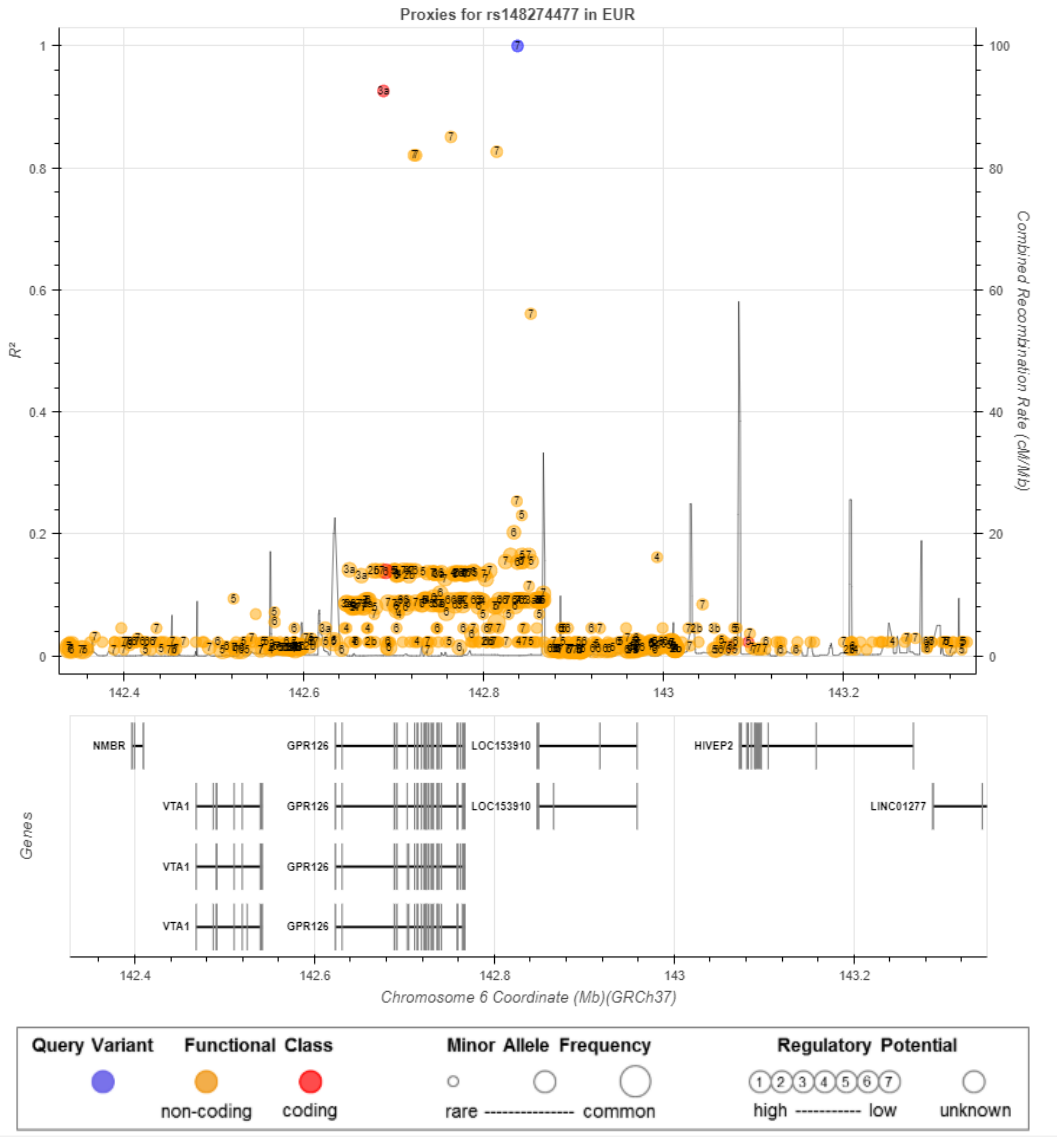


Figure 3-2 *Schematic from LDproxy showing degree of LD of SNPs with rs148274477 in a European population.* [141] *The x axis represents chromosomal location with gene location and exons shown.* R^2 *value is represented on the y axis with the higher the linkage, the higher the SNP. Rs148274477 is shown in blue with non-coding SNPs in LD in yellow and one missense variant, rs17280293, in red. Rs17280293 is the SNP is highest LD with rs148274477.*

RS Number	Chr	Position (GRCh37)	Alleles	MAF	Distance	R^2 (LDLink)	Functional Class	RegulomeDB prediction
rs17280293	6	142688969	(A/G)	0.0408	149204	0.9262	missense	**3a (TF binding + any motif + DNase peak)**
rs9403386	6	142764073	(A/C)	0.0487	74100	0.8509	NA	**7 - No annotation**
rs9403391	6	142814991	(C/T)	0.0477	23182	0.8266	NA	**7 - No annotation**
rs73780219	6	142722866	(G/A)	0.0457	115307	0.8208	NA	**7 - No annotation**
rs73780221	6	142725182	(G/C)	0.0457	112991	0.8208	NA	**7 - No annotation**
rs117865353	6	142852897	(G/A)	0.0239	14724	0.561	NA	**7 - No annotation**
rs117386235	6	142837372	(T/C)	0.0109	801	0.2537	NA	**7 - No annotation**
rs55771139	6	142842841	(G/A)	0.0099	4668	0.2304	NA	**5 - TF binding or DNase peak**
rs13192074	6	142834078	(G/A)	0.1769	4095	0.2027	NA	**6 - Other**

Table 3-4 Results from LDproxy showing SNPs in LD (r2>0.2) with rs148274477. MAF=minor allele frequency. RegulomeDB [148] annotates SNPs with known and predicted regulatory elements genome and scores the SNP accordingly.

HaploReg uses ENCODE data to annotate the regions in which the SNPs lie. Table 3-5 shows the SNPs in LD (r^2>0.2) with rs148274477. The table is annotated with potential promoter, enhancer and DNase sites as well as binding sites for proteins and motifs. This information is also elaborated on with the addition of the tissue/ cell type.

Variant	SiPhy	Promoter	Enhancer	DNAse	Proteins	Motifs	dbSNP
rs117386235						BRCA1,TAL1,TCF11::MafG	
rs117865353			ESDR	BRN			intronic
rs148274477							
rs17280293	✓	HRT, GI, LNG, CRVX	LNG, FAT, SKIN, LIV, GI, BRN, HRT, MUS, BONE	ESDR, IPSC,ADRL, BRN,BRN, HRT,LNG,MUS, MUS, THYM,LNG,SKIN	P300	Brachyury,HNF1,Hmx,Nkx2,ZEB1	missense
rs55771139						Foxj2,Foxp1,Pax-4,RREB-1,Spz1	
rs73780219			GI, LIV			CEBPB,CEBPD,COMP1,NRSF,YY1,Zfp410	intronic
rs73780221						Ascl2,PU.1	intronic
rs9403386						Evi-1,GATA,Hoxa5	intronic
rs9403391						DBP,HNF1,Pax-6	

Table 3-5 SNPs in r2>0.2 LD with rs148274477 from HaploReg v4.1. SiPhy predicts the level of conservation of the SNP. Data from ENCODE shows regions with known or predicted functional motifs (enhancers, DNAase) and in which tissues they may be relevant. Also shown is predicted proteins bound and motifs bound. The final column shows SNP information from dbSNP.

3.3.2 SNPs are not eQTLs for GPR126 or Other Genes

To assess whether these variants had an impact on gene expression, eQTL databases were used. Firstly, GTEx was used using all SNPs in Table 3-4 and Table 3-5 as query SNPs. The analysis showed no eQTL hits for any SNP which suggests that there are no known eQTL effects of these SNPs. To further investigate any possible eQTL effects of the SNPs, the Blood eQTL Browser was used. Again, no SNPs showed any significant eQTL effects on expression of *GPR126* or any other gene in any tissue.

3.3.3 Ser123Gly (rs17280293) is Predicted to be Damaging to Protein Function.

In silico predictors were utilised to assess the effect of amino acid substitution Ser123Gly. SIFT, which makes predictions based on amino acid conservation, predicted the change to be damaging to protein function. PolyPhen-2, which bases predictions on amino acid effects on structure and function of the protein, predicted the substitution to possibly damaging but Mutation Taster predicts that the SNP is a polymorphism due to both the “physico-chemical characteristics of amino acids” and population genetics data. The full results and explanations are shown in Table 3-6.

Tool	Score	Prediction	Prediction Explanation
SIFT	0.01	DAMAGING *Warning! Low confidence	SIFT looks at the composition of amino acids and computes the score. A SIFT score is a normalized probability of observing the new amino acid at that position, and ranges from 0 to 1. A value of between 0 and 0.05 is predicted to affect protein function.
PolyPhen-2	0.879/1	Possibly damaging	This prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. PolyPhen-2 predicts the functional significance of an allele replacement from its individual features by Naïve Bayes classifier trained using supervised machine-learning. For a mutation, PolyPhen-2 calculates Naïve Bayes posterior probability that this mutation is damaging and reports estimates of false positive rate (FPR, the chance that the mutation is classified as damaging when it is in fact non-damaging) and true positive rate (TPR, the chance that the mutation is classified as damaging when it is indeed damaging). A mutation is also appraised qualitatively, as benign, possibly damaging, or probably damaging based on pairs of false positive rate (FPR) thresholds.
Mutation Taster	Prediction confidence score (simple_aae) = 0.5/1 AA substitution score = 56/215	Polymorphism	The probability score is the probability of the prediction, i.e. a value close to 1 indicates a high 'security' of the prediction. The AA score is taken from an amino acid substitution matrix (Grantham Matrix) which takes into account the physico-chemical characteristics of amino acids and scores substitutions according to the degree of difference between the original and the new amino acid. Scores may range from 0.0 to 215. If an alteration is a 'true' SNP (as confirmed by the existence of each of the three genotypes AA, AB, BB in the HapMap data or by presence in TGP in homozygous state in > 4 cases), it is automatically predicted to be a polymorphism. As all three genotypes are reported in population databases, this SNP is automatically scored as a polymorphism.

Table 3-6 Prediction of the effects of Ser123Gly substitution on protein function by SIFT, PolyPhen-2 and Mutation Taster

3.3.4 GPR126 mRNA is expressed in the Lung and Airway Relevant Cells

Quantitative-PCR (Taqman) was performed on cDNA made from RNA from lung tissue and a range of HASM and HBEC donors as described in section 3.2.3 (Figure 3-3). The Taqman GPR126 PDAR purchased spans exons 18-19 or 19-20 depending on the transcript and covers all known protein coding variants of GPR126.

Additional tissue and cell samples were also tested for *GPR126* mRNA expression as references. Data shown is normalised to whole lung tissue expression. *GPR126* mRNA is expressed in the lung and in airway structural cells HASM and HBEC. Highest levels of expression is seen in the placenta with mRNA also detected in the kidney, BEAS2B and at low levels in brain.

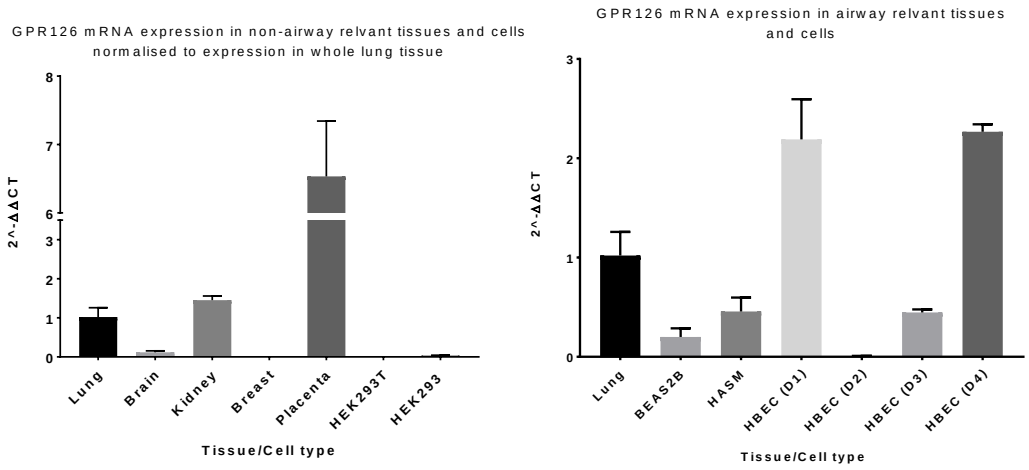


Figure 3-3 Total GPR126 mRNA expression in a range of human cell types and tissues. A) GPR126 expression in non-lung related cells and tissues normalised to GPR126 expression in the lung. B) GPR126 expression again normalised to whole lung. Expression is seen in BEAS2B, HASMs and in 3 of 4 HBEC donors.

3.3.5 GPR126 Expression in GEO RNA-seq Datasets

RNA sequencing data was downloaded from GEO dataset GSE57148. This contains gene expression values as FPKM in lung tissue from 91 controls with normal spirometry and 98 COPD patients. [153] Data for *GPR126* was extracted from the database and gene expression was compared with cases and controls. The results are shown in Figure 3-4. *GPR126* expression is significantly higher in COPD patients versus controls (p=0.0164), although the increase is modest. The mean FPKM in controls is 9.1 (min=2.1, max=18.4) while 10.1 (min=1.7, max=22.5) is the average FPKM for COPD patients.

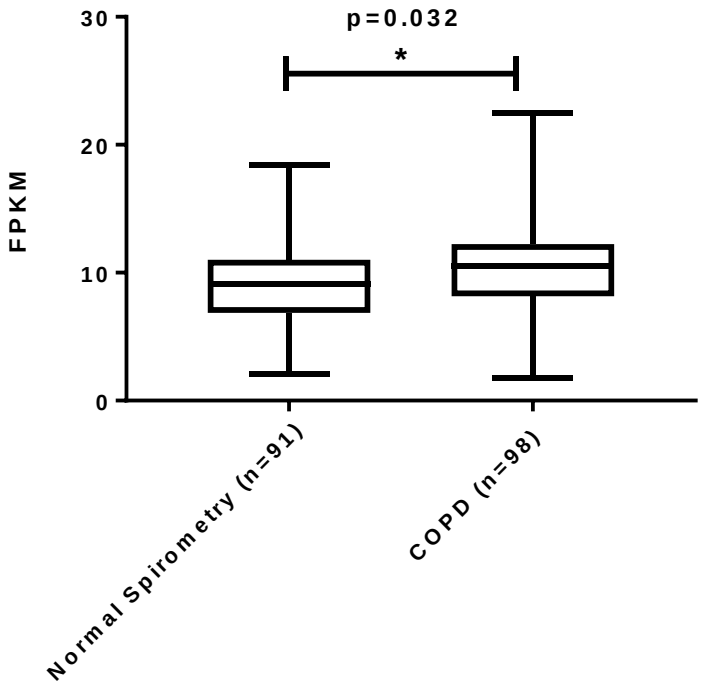


Figure 3-4 FPKM values from control and COPD patients were compared using an unpaired T-test. Data was normally distributed as determined by a Shapiro-Wilk normality test. GPR126 expression is elevated in lung tissue from COPD patients versus lung tissue from individuals with normal spirometry (p=0.0164). Data was downloaded from GEO dataset GSE57148.

Microarray expression data was also downloaded from GEO datasets GSE43696 and GSE89809. These datasets show gene expression in epithelial cells from healthy controls and asthma patients with varying severity.

No difference was identified between cases and controls in either dataset Figure 3-5.

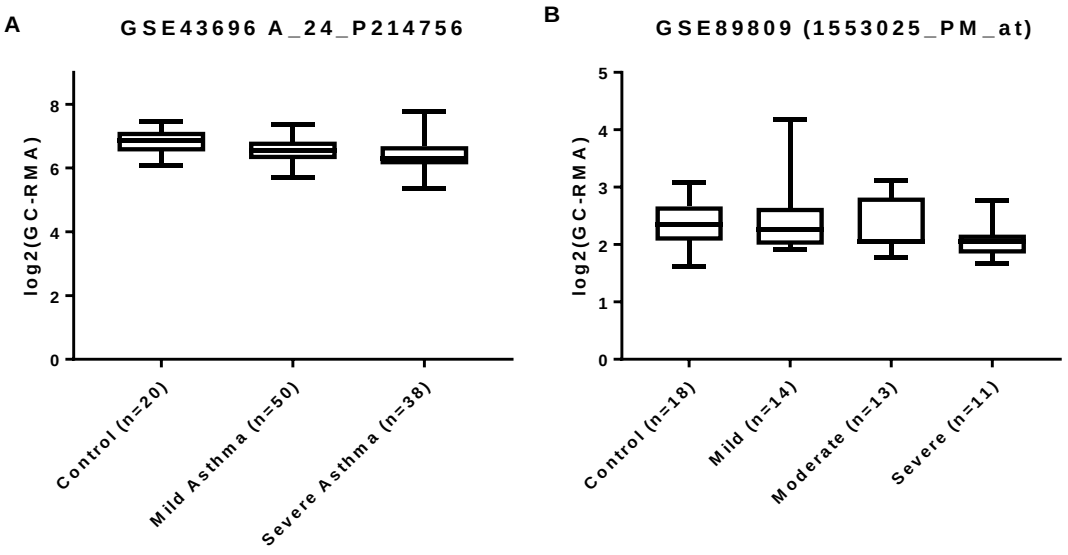


Figure 3-5 GPR126 expression array data from bronchial epithelial cells from GEO datasets GSE43696 and GSE89809. There is no difference in GPR126 expression in either dataset. Values across the groups were compared using an ANOVA with Tukey's multiple comparisons test (left, parametric and a Kruskal-Wallis test with Dunn's multiple comparisons test (right, non-parametric). Data was tested for normality using a Shapiro-Wilk normality test.

3.3.6 RNA Sequencing Quality Control Results

The sequencing data was assessed for quality using FastQC with the trimmed reads used for downstream analysis. The sequencing data was processed as outlined in section 2.4.1. Graphs showing representative data for the FastQC analysis are shown in the appendices of this thesis. The Sequencing data passed all quality control steps outlined in section 2.4.1.3 .

3.3.7 RNA-sequencing Data of GPR126 mRNA Expression in HASM and HBEC

RNA was extracted from five HASM donors at passage 3. RNA was isolated from cell lysates and RNA was sent to Oxford Genomics Centre for

sequencing. Sequencing data was analysed using the Cuffnorm package to determine FPKM expression values in each donor.

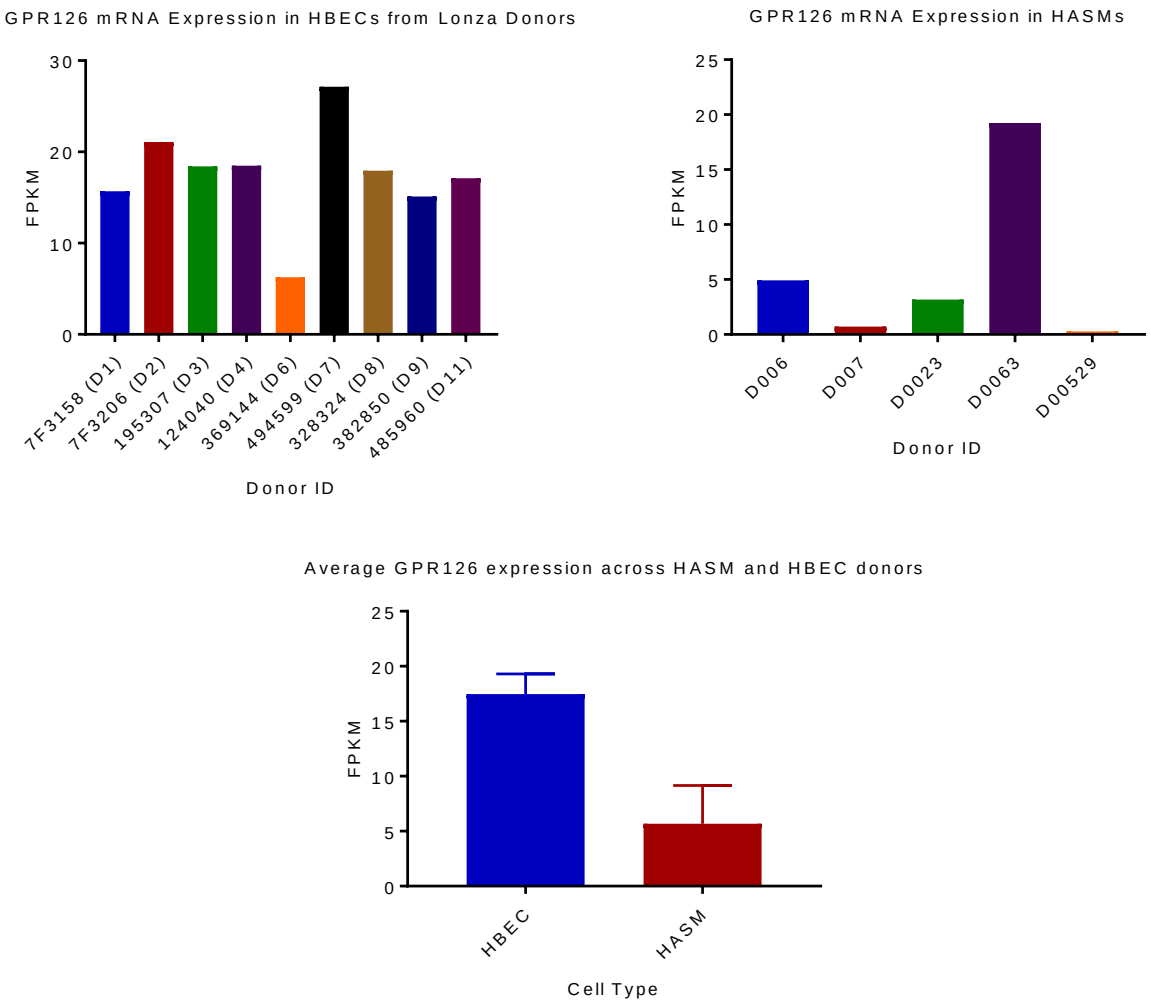


Figure 3-6 RNA sequencing data showing FPKM expression values of GPR126 in individual HBEC (top left) and HASM (top right) donors. Average FPKM values for all donors is shown (bottom) for HBECs (n=9, mean = 17.5, min = 6.3, max = 27.1) and HASMs (n=5, mean = 5.7, min = 0.3, max = 19.2).

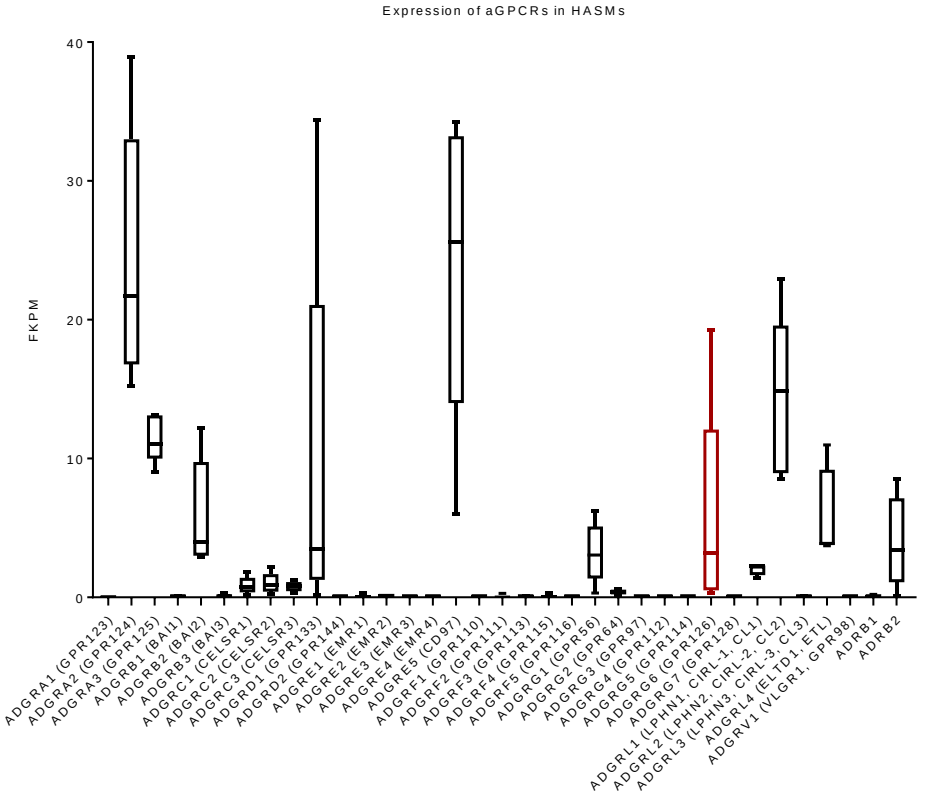
FPKM (fragments per kilobase per million) and TPM (transcripts per million) are two units used in RNA-seq experiments. FPKM is a normalised figure which takes into account sequencing depth and gene length, both of which would result in more reads per gene. TPM does a similar normalisation to account for sequencing depth and gene size but gives a more accurate figure

of the proportion of reads that come from a specific gene/ transcript and thus removes variation between samples. [157]

GPR126 expression is seen in all HBEC donors with FPKM values above 10 in all but one donor. Highest expression was seen in donor 494599. In HASMs, there was more variation in *GPR126* mRNA expression. Donor 0063 showed relatively high levels of expression whereas donor 00529 showed no expression of the gene and the remaining donors all had FPKM values less than 10.

3.3.7.1 Expression of adhesion-GPCRs in Primary Human Airway Cells

In order to put the GPR126 expression values from these RNA-seq experiments into context, we looked at the expression levels of each member of the adhesion-GPCR family of receptors as well as the β1 and β2 adrenergic receptors. The role of β_2 adrenoreceptors and β_2 agonists in airway disease is well documented. β_2 agonists are a major component of the treatment for both asthma and COPD. Activation of the β_2 receptors, located on airway smooth muscle cells, by an agonist results in muscle and airway relaxation. The expression values of these physiologically receptors in our datasets provides a frame of reference for the GPR126 expression values. In HASMs, the average FPKM value for the β_2 adrenergic receptor across the donors is 3.97 (min=0.08, max=8.52) compared to the GPR126 FPKM value of 5.672 (min=0.29, max=19.23). In airway epithelial cells, β_2 receptors also have important roles which include enhanced ciliary beat and mucous clearance. [158] The average expression of this receptor in HBECs is 14.86 (min=8.53, max=32.76) while comparatively the average FPKM for GPR126 is 17.47 (min=6.25, max=27.14). Figure 3-7 shows these results in full with GPR126 highlighted in red.


Expression of aGPCRs in HASMs
FKPM
40
30
20
10
0
ADGRA1 (GPR123)
ADGRA2 (GPR124)
ADGRA3 (GPR125)
ADGRB1 (BAI1)
ADGRB2 (BAI2)
ADGRB3 (BAI3)
ADGRC1 (CELSR1)
ADGRC2 (CELSR2)
ADGRC3 (CELSR3)
ADGRD1 (GPR133)
ADGRD2 (GPR144)
ADGRE1 (EMR1)
ADGRE2 (EMR2)
ADGRE3 (EMR3)
ADGRE4 (EMR4)
ADGRE5 (CD97)
ADGRF1 (GPR110)
ADGRF2 (GPR111)
ADGRF3 (GPR113)
ADGRF4 (GPR115)
ADGRF5 (GPR116)
ADGRG1 (GPR56)
ADGRG2 (GPR64)
ADGRG3 (GPR97)
ADGRG4 (GPR112)
ADGRG5 (GPR114)
ADGRG6 (GPR126)
ADGRG7 (GPR128)
ADGRL1 (LPHN1, CIRL-1, CL1)
ADGRL2 (LPHN2, CIRL-2, CL2)
ADGRL3 (LPHN3, CIRL-3, CL3)
ADGRL4 (ELTD1, ETL)
ADGRV1 (VLGR1, GPR98)
ADRB1
ADRB2

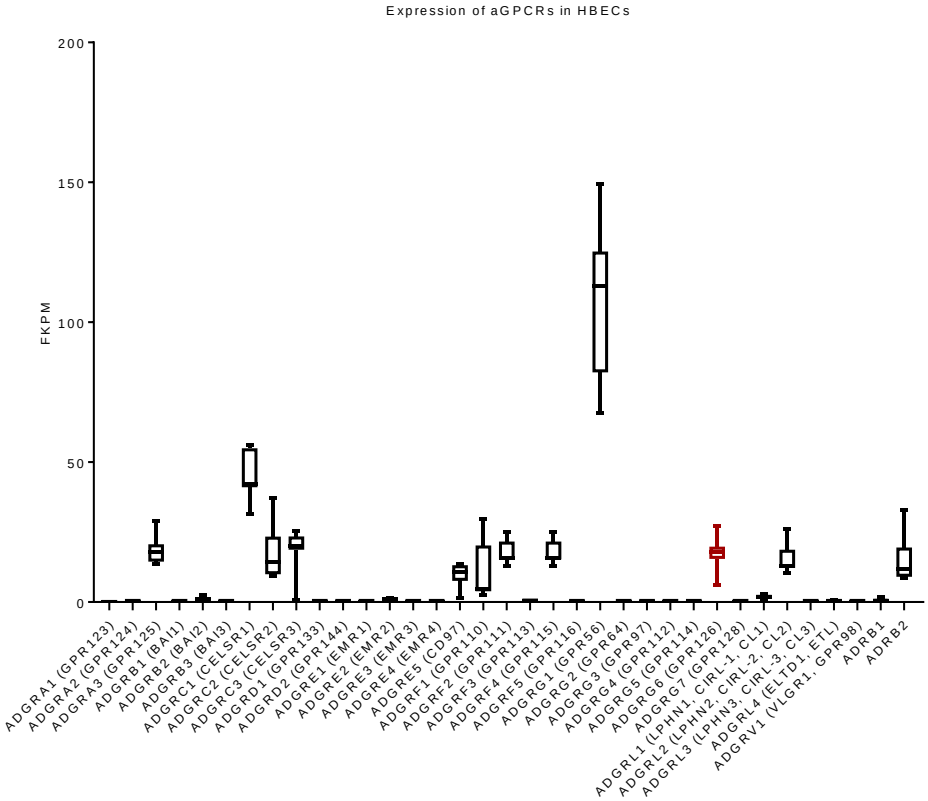

Expression of aGPCRs in HBECs
FKPM
200
150
100
50
0
ADGRA1 (GPR123)
ADGRA2 (GPR124)
ADGRA3 (GPR125)
ADGRB1 (BAI1)
ADGRB2 (BAI2)
ADGRB3 (BAI3)
ADGRC1 (CELSR1)
ADGRC2 (CELSR2)
ADGRC3 (CELSR3)
ADGRD1 (GPR133)
ADGRD2 (GPR144)
ADGRE1 (EMR1)
ADGRE2 (EMR2)
ADGRE3 (EMR3)
ADGRE4 (EMR4)
ADGRE5 (CD97)
ADGRF1 (GPR110)
ADGRF2 (GPR111)
ADGRF3 (GPR113)
ADGRF4 (GPR115)
ADGRF5 (GPR116)
ADGRG1 (GPR56)
ADGRG2 (GPR64)
ADGRG3 (GPR97)
ADGRG4 (GPR112)
ADGRG5 (GPR114)
ADGRG6 (GPR126)
ADGRG7 (GPR128)
ADGRL1 (LPHN1, CIRL-1, CL1)
ADGRL2 (LPHN2, CIRL-2, CL2)
ADGRL3 (LPHN3, CIRL-3, CL3)
ADGRL4 (ELTD1, ETL)
ADGRV1 (VLGR1, GPR98)
ADRB1
ADRB2

Figure 3-7 Expression of the adhesion-GPCRs and the β1 and β2 adrenergic receptors in comparison to GPR126 (red) in both HASM (top) and HBEC (bottom) donors. Data shown is median with 25th and 75th percentile with the minimum and maximum values.

3.3.7.2 GPR126 Isoform Expression

Cuffnorm also gives information on isoform expression. GPR126 has four known full length protein coding transcripts (see section 1.3). In HASMs, each isoform shows very modest expression with all having average FPKM values <3 and most <1 across all donors. The α1 variant showing the highest levels of expression across the five donors (Figure 3-7). The highest expression of most isoforms is in the donor (D063) which shows highest overall GPR126 expression with the sum of GPR126 transcripts totalling 19.23 FPKM.

Gene	Class code*	Nearest ref ID	Length	D006	D007	D0023	D0063	D00529	Average FPKM	Protein product (RefSeq)
GPR126	j	ENST00000230173	6583	2.44	0.38	0.34	8.26	0.19	2.32	a1
GPR126	j	ENST00000230173	6670	0	0.19	0.54	3.55	0	0.86	a1
GPR126	j	ENST00000230173	4973	2.49	0	1.61	0	0.08	0.84	a1
GPR126	=	ENST00000296932	6849	0	0.02	0.68	1.41	0	0.42	a2
GPR126	=	ENST00000541199	844	0	0.01	0	1.99	0	0.4	Protein coding. No RefSeq data
GPR126	=	ENST00000230173	7026	0	0	0	1.83	0	0.37	a1
GPR126	=	ENST00000545477	657	0	0	0	0.57	0	0.11	No Protein
GPR126	=	ENST00000538281	591	0	0	0	0.47	0	0.09	No Protein
GPR126	=	ENST00000540208	652	0	0	0	0.37	0.02	0.08	No Protein
GPR126	=	ENST00000463637	498	0	0.06	0	0.28	0	0.07	No Protein
GPR126	=	ENST00000367609	6887	0	0	0	0.28	0	0.06	b1
GPR126	=	ENST00000508295	583	0	0	0	0.21	0	0.04	Protein coding. No RefSeq data
GPR126	=	ENST00000415128	1610	0	0.05	0	0	0	0.01	Protein coding. No RefSeq data
GPR126	=	ENST00000367608	6879	0	0	0	0	0	0	b2
GPR126	=	ENST00000435011	540	0	0	0	0	0	0	Protein coding. No RefSeq data
GPR126	=	ENST00000472054	1247	0	0	0	0	0	0	No Protein
GPR126	=	ENST00000497898	715	0	0	0	0	0	0	No Protein
TOTAL				4.94	0.73	3.17	19.23	0.29		

Table 3-7 GPR126 transcript expression in five HASM donors. TCONS_00208723, aligning to transcript ENST00000230173 which codes for the a1 variant is the most abundant. Highest expression of most isoforms is seen in donor 0063, the donor with the highest overall GPR126 expression. Class code legend: j - Potentially novel isoform (fragment): at least one splice

junction is shared with a reference transcript. = - Complete match of intron chain.

Transcript	Gene	D1	D2	D3	D4	D7	D8	D9	D11	Average FPKM	Gene product (RefSeq)
ENST00000230173	GPR126	0.166	0.289	0.107	0.218	0.224	0.134	0.087	0.208	0.179	a1
ENST00000230173	GPR126	0.000	0.113	0.034	0.000	0.074	0.000	0.000	0.021	0.030	a1
ENST00000230173	GPR126	0.053	0.028	0.127	0.077	0.096	0.084	0.103	0.028	0.074	a1
ENST00000230173	GPR126	0.158	0.288	0.133	0.288	0.231	0.151	0.181	0.256	0.211	a1
ENST00000230173	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	a1
ENST00000230173	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	a1
ENST00000230173	GPR126	0.167	0.171	0.232	0.190	0.139	0.192	0.183	0.120	0.174	a1
ENST00000230173	GPR126	2.039	3.109	2.236	3.492	2.667	0.615	2.234	0.884	2.159	a1
ENST00000367608	GPR126	0.001	0.035	0.000	0.000	0.000	1.262	0.202	0.030	0.191	No Protein
ENST00000296932	GPR126	0.533	1.287	0.157	0.475	0.001	2.172	0.108	1.117	0.731	a2
ENST00000367609	GPR126	7.722	8.416	9.235	7.581	10.402	5.352	6.571	7.916	7.899	b1
ENST00000415128	GPR126	0.928	0.665	1.406	1.090	1.042	0.665	0.737	1.158	0.961	Protein coding, No RefSeq data
ENST00000541199	GPR126	0.000	0.000	0.198	0.000	1.608	0.048	0.000	0.000	0.232	Protein coding, No RefSeq data
ENST00000435011	GPR126	1.158	0.425	0.649	0.738	5.540	2.677	0.861	0.000	1.506	Protein coding, No RefSeq data
ENST00000545477	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	No Protein
ENST00000508295	GPR126	1.519	2.784	1.231	1.619	1.858	1.400	2.566	2.046	1.878	Protein coding, No RefSeq data
ENST00000540208	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	No Protein
ENST00000463637	GPR126	0.303	0.372	0.368	0.462	0.641	0.502	0.136	0.376	0.395	No Protein
ENST00000538281	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.000	0.007	No Protein
ENST00000472054	GPR126	0.001	0.000	0.000	0.000	0.676	0.094	0.000	0.343	0.139	No Protein
ENST00000497898	GPR126	0.949	3.108	2.302	2.245	1.944	2.581	1.095	2.600	2.103	No Protein
ENST00000497898	GPR126	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	No Protein
TOTAL		**15.695**	**21.089**	**18.413**	**18.476**	**27.143**	**17.929**	**15.116**	**17.104**		

Table 3-8 GPR126 transcripts in 8 HBEC donors. Transcript ENST00000367609 showed the highest transcript expression across all donors which codes for the b1 isoform.

GPR126 isoform expression varies in HBECs from HASMs. In HBECs, the highest expressed transcript is the one which codes for the b1 isoform (FPKM = 7.899), which is the longest isoform.

3.3.7.3 RNA-sequencing Data from GTEx

RNA sequencing data from GTEx shows *GPR126* mRNA is expressed in the lung, which is the second highest value in the human tissues after the bladder (Figure 3-8). [142] *GPR126* has four known full length protein coding transcripts (a1, a2, b1 and b2), Figure 3-9 shows the relative abundance of these transcripts in the GTEx RNA seq dataset with the a2 transcript (ENST00000296932.12, NM_001032394, NP_001027566) showing the highest abundance in lung tissue. [142]

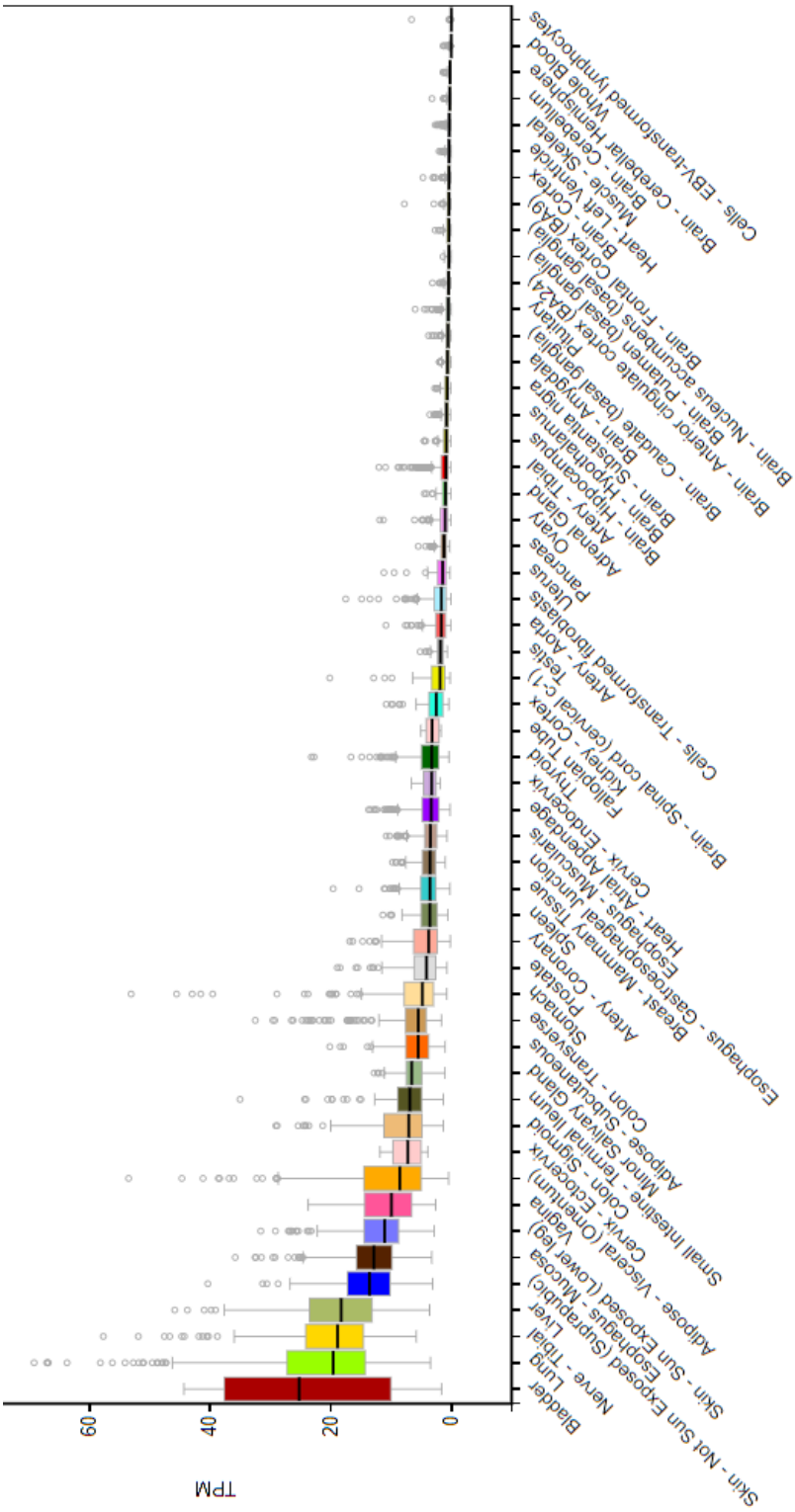


Figure 3-8 GPR126 RNA expression in TPM (transcripts per million) from GTEx. [142] The highest level of gene expression is seen in the bladder, followed by the lung. Box plots show the median with 25th and 75th percentiles.

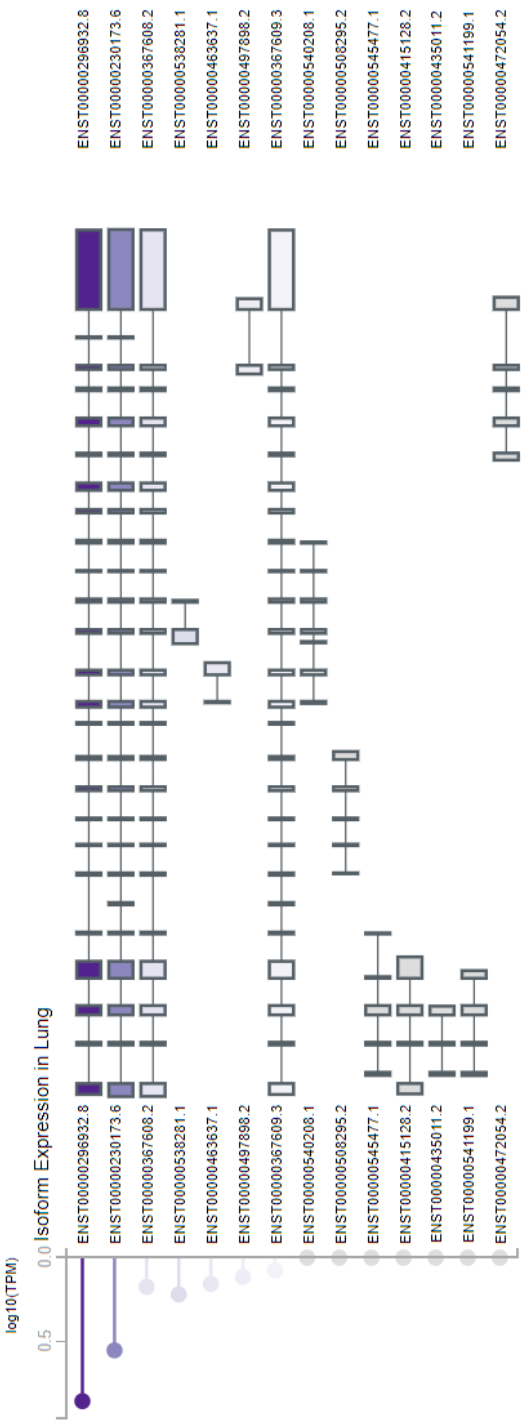


Figure 3-9 A schematic from GTEx showing *GPR126* isoform expression in the lung. [142] The most abundant transcript in the lung codes for the a2 variant of the protein which lacks exon 6.

Similarly, Protein Atlas shows *GPR126* mRNA expression in a range of human tissues including the lung. [156] GPR126 protein expression is not detected in the lung samples but protein expression is seen at high levels in muscle tissues, particularly skeletal and smooth muscle and in the gastrointestinal tract, most abundantly in the small intestine. [156]

3.3.8 Immunohistochemistry of GPR126 protein expression in lung tissue

In order to establish the expression of GPR126 protein in the human lung IHC was performed initially on small intestine and human lung tissue (TMM140). Small intestine tissue was identified as a positive control tissue from the Human Protein Atlas[156] and was used to validate the antibody at the following dilutions; 1:25, 1:50, 1:100, 1:250 and 1:1000. A final antibody concentration of 1:50 was used.

GPR126 expression was seen in small intestine and in the lung (Figure 3-10, Figure 3-12, Figure 3-12 & Figure 3-13). In the lung expression was seen in smooth muscle bundles surrounding the airways. When further lung tissue donors were tested for GPR126 protein expression, it was difficult to detect. Figure 3-12 A) shows expression seen around an airway but the other two lung donors show reduced or absent GPR126 expression.

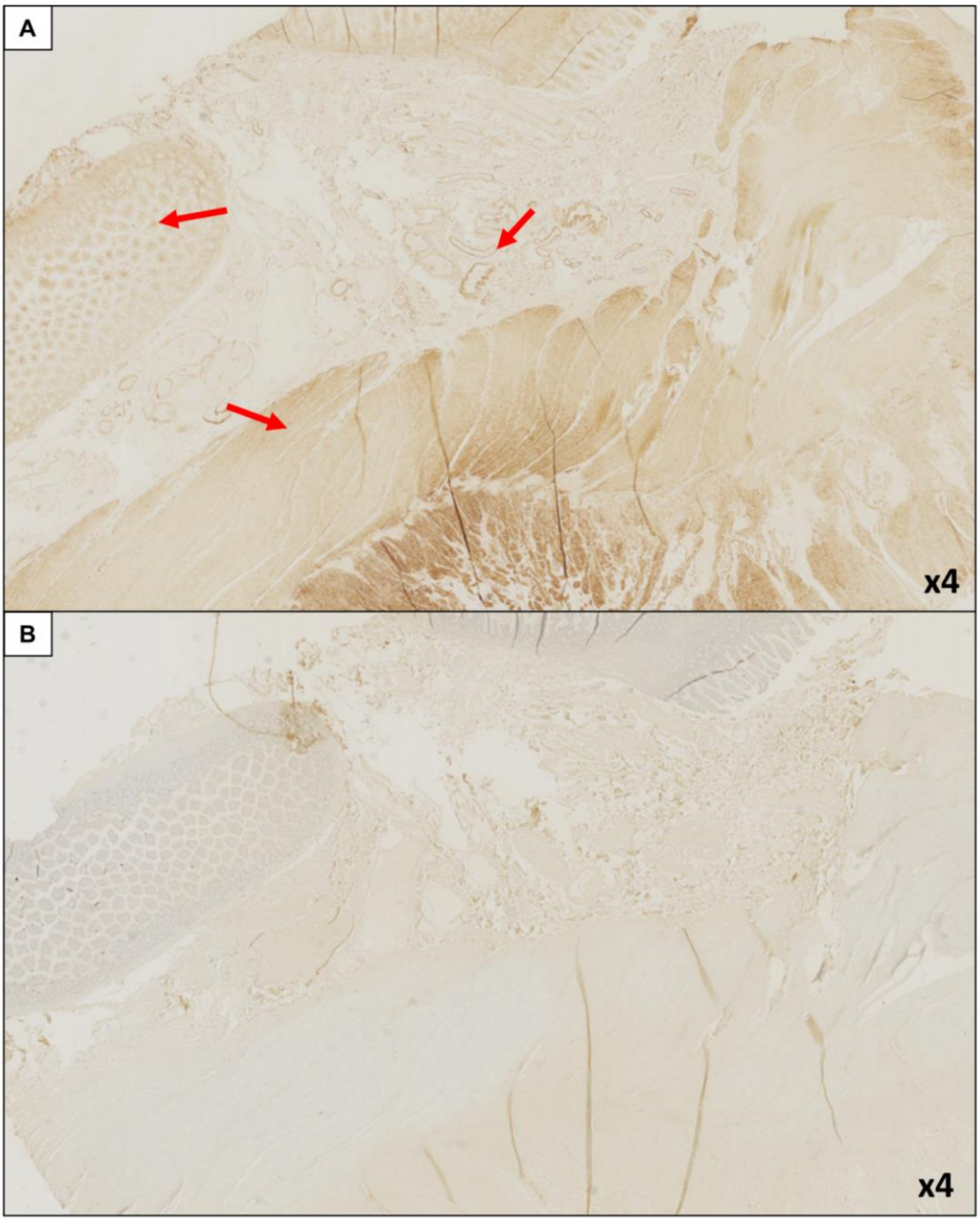


Figure 3-10 Immunohistochemistry showing GPR126 staining in small intestine. A) anti-GPR126 antibody. B) Isotype control. Slides were zanodoomed in the pathology department in Queen's Medical Centre. Lung tissue is D2.

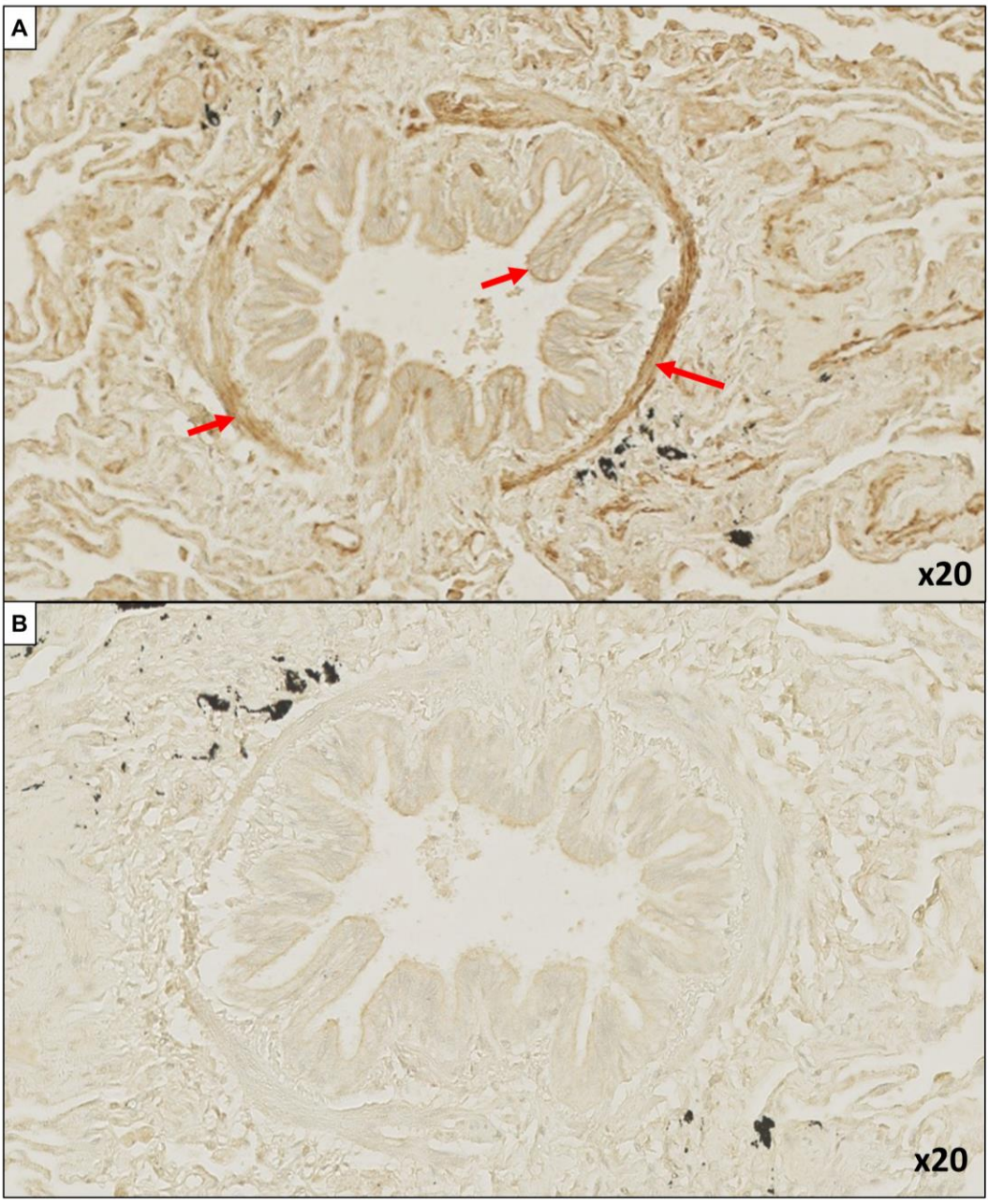


Figure 3-11 Immunohistochemistry showing GPR126 staining in human lung. A) anti-GPR126 antibody. B) Isotype control. Slides were zanodoomed in the pathology department in Queen's Medical Centre. Lung tissue is D2.

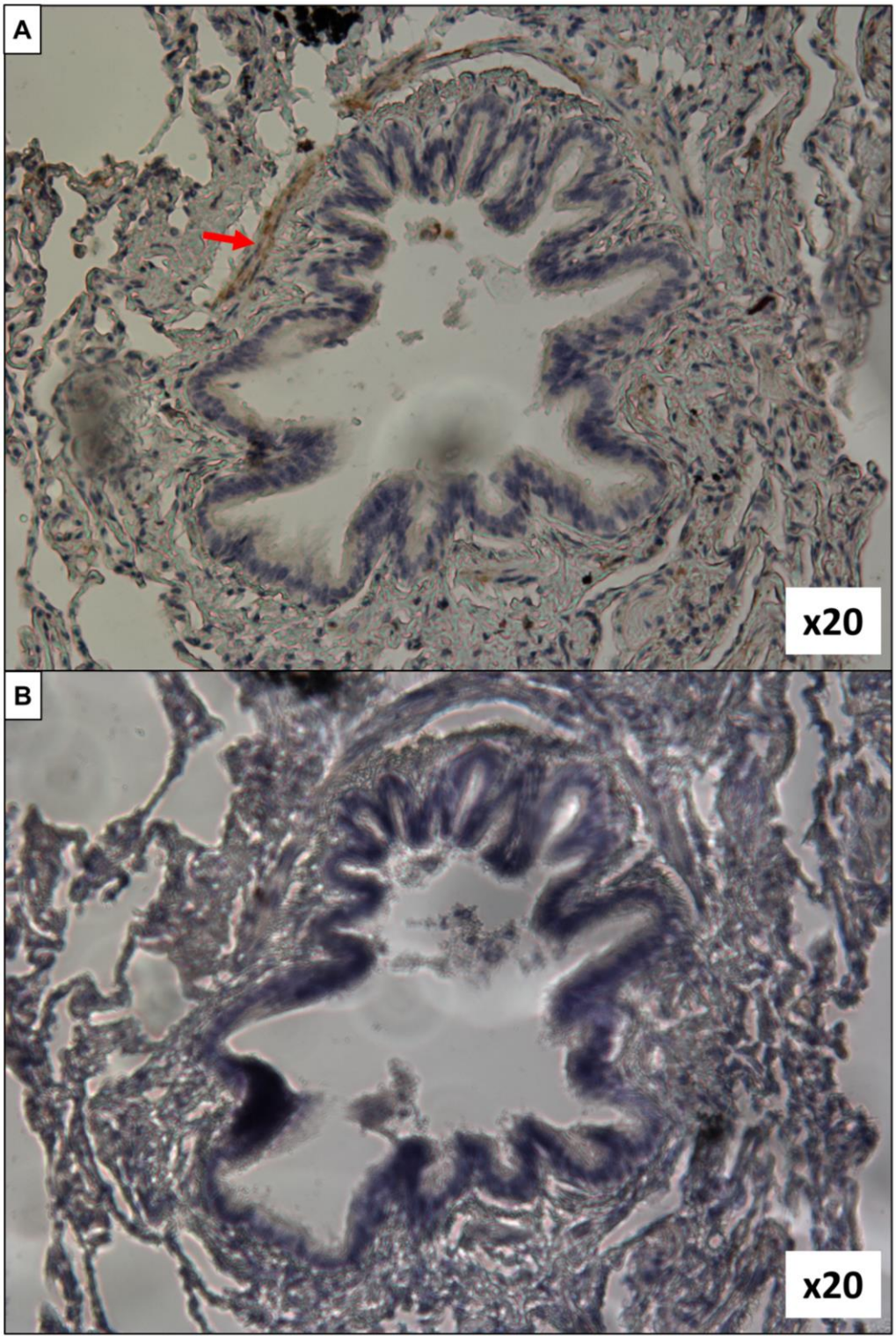


Figure 3-12 Immunohistochemistry on additional lung tissue donor, D2. A) anti-GPR126 antibody. B) Isotype control.

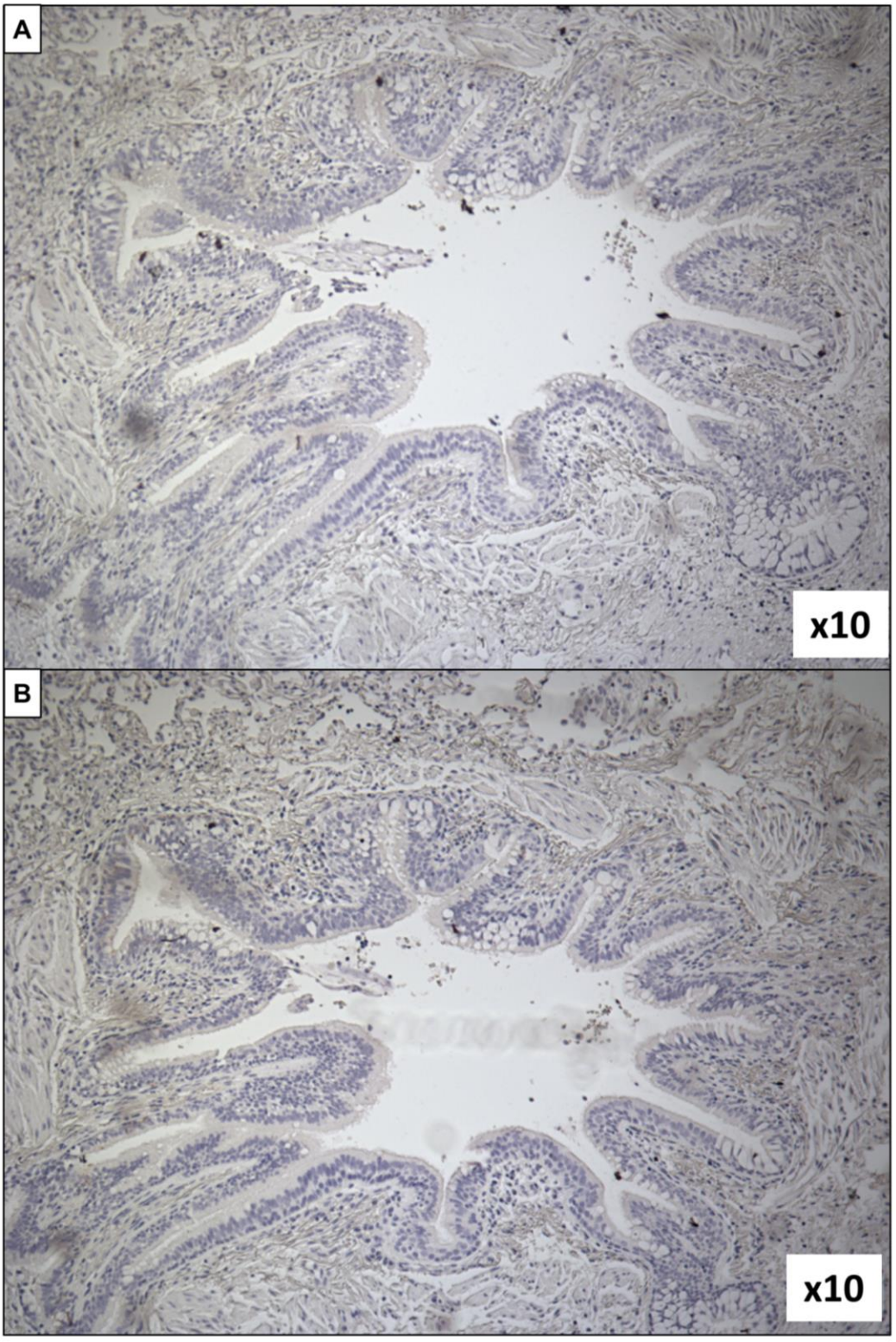


Figure 3-13 Immunohistochemistry on additional lung tissue donor, D5. A) anti-GPR126 antibody. B) Isotype control.

3.4 DISCUSSION

In this chapter, the aim was to identify potentially functional variants in LD with rs148274477 that may underlie the association signal. I also aimed to investigate GPR126 expression in the human lung. Specifically the aims were; to identify potential causal SNPs underlying the association signal by looking at the effect of variants on the coding region, regulatory elements and eQTL effects and to identify GPR126 mRNA expression in the lung and in airway relevant cells using qPCR and RNA sequencing as well as identifying GPR126 protein in lung tissue.

In summary, within the LD block (r^2>0.2 with rs148274477) one variant, rs17280293, lies within the GPR126 coding region and results in the substitution of a serine residue at position 123 with a glycine. I also found that *in silico* tools predict this change to be damaging to protein function. Below the LD threshold of r^2 >0.2 I used in this chapter, there is another missense variant, rs11155242 (r^2 = 0.14) which is also in LD with rs17280293. This SNP was not considered as part of our study due to the r^2 threshold we set but potentially could be a candidate SNP. This SNP results in Lys230Gln and is also reported to be an eQTL for *GPR126* expression in human lung tissue. [146] The variant is predicted to be a benign polymorphism by SIFT, PolyPhen-2 and Mutation Taster.

No SNPs in LD r^2>0.2 were found to be eQTLs for GPR126 or any other genes in any of the eQTL databases searched. In addition, we have provided evidence for GPR126 mRNA and protein expression in the lung, although we saw varied expression seen between different donors in both mRNA and protein, which is to be expected. Analysis of the results from LDproxy and HaploReg v4.1 identifies Ser123Gly (r^2 =0.9262) as a potential credible variant to explain the association signal. This variant is a coding region variant which is predicted to be damaging to protein function. This substitution results in the replacement of a serine residue, a medium sized amino acid with a natural side chain with a glycine, the smallest amino acid.

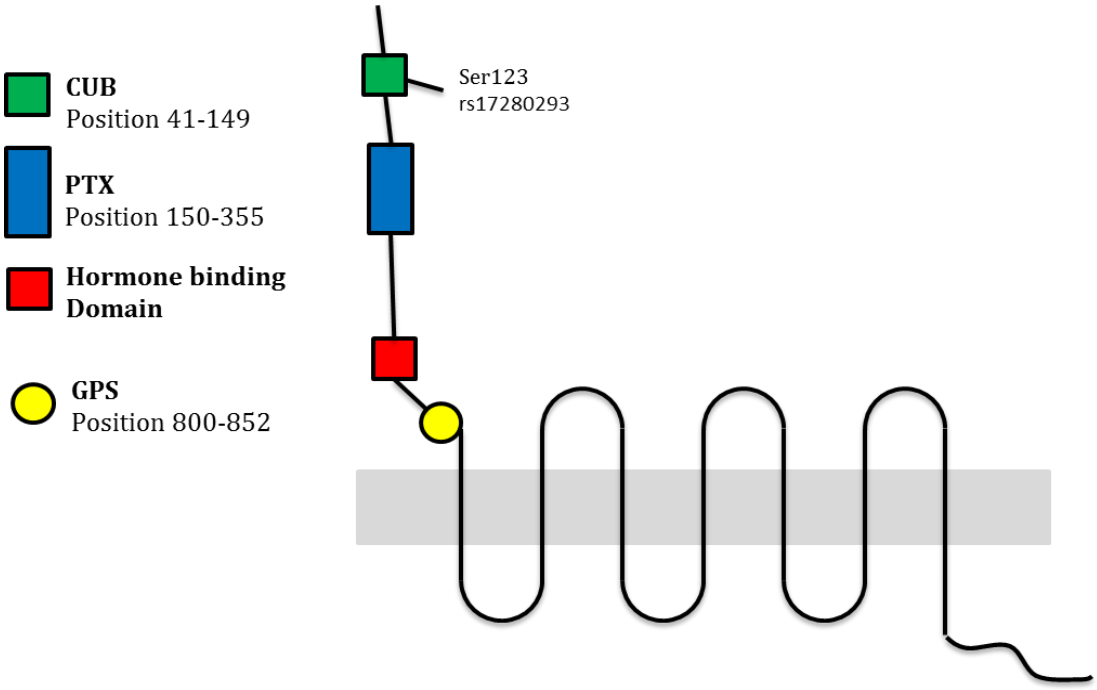


Table 3-9 Representative diagram showing the location of the SNP rs17280293 (Ser123Gly) on the GPR126 receptor.

Interestingly, this substitution lies within the CUB (complement C1r/C1s, Uegf, Bmp1) domain, one of the domains within the GPR126 N-terminal (Figure 3-9). This conserved motif is approximately 110 residues and is involved in a diverse array of functions including but not limited to complement activation [159], axon guidance and angiogenesis [160], cell signalling[161], inflammation[162] and tumour suppression. [163, 164] Included in the list of proteins containing a CUB domain is the Cricetidae sp. (Hamster) serine protease Casp, which is important for degrading collagen types I and IV and fibronectin. Procollagen C-proteinase enhancers are extracellular glycoproteins containing two CUB domains which stimulate processing of fibrillary procollagens.[164] Additionally, the literature shows that collagen type IV is an activating ligand of the zebrafish GPR126 by binding to this CUB domain as well as a PTX domain. [98] I therefore hypothesise that this amino acid substitution may affect the ability of GPR126 to bind collagen type IV and other potential ligands thereby affecting its signalling capabilities.

As covered in the general introduction, airway remodelling is a key feature in COPD which results in structural changes in the lung. It is a complex process that involves the modification of the composition of the ECM including altered expression of genes coding for ECM proteins. [17, 165] It is established that GPR126 interacts with the ECM and is a regulator of ECM gene expression in inner ear development. [105] Collagen type IV and laminin-211, both ECM proteins, are ligands for the receptor. [97, 98, 105] In addition to this, it is known that the CUB domain in GPR126 is important for binding collagen type IV and that variant Ser123Gly lies within this CUB domain. Therefore we speculate that this variant may reduce GPR126 affinity for binding collagen type IV which in turn affects airway remodelling in the lung of patients with COPD. In addition, it is known that angiogenesis is also a contributor of airway remodelling in COPD [166] and interestingly, tumstatin, a fragment of collagen type IV, has antiangiogenic activity.[167] As collagen type IV is a ligand for GPR126 and activates GPR126 signalling perhaps there is another role for GPR126 via tumstatin in regulation of angiogenesis in addition to VEGF signalling. This data collectively implicates rs148274477 as a potential causal SNP in this association signal and there is evidence in a literature to suggest a mechanism by which GPR126 may be involved in disease. The fact that the risk allele is the common one, as outlined in the introduction to this chapter, suggests that GPR126 loss of function by Ser123Gly mutation may be protective. Alternatively, the mutation does not result in loss of function and alters GPR126 signalling/ expression via another mechanism i.e. eQTL effects. Therefore one might predict that GPR126 elevation or over-activity may be damaging and indeed GPR126 mRNA is elevated in lung tissue from COPD patients when compared to healthy controls as shown (Figure 3-4).

GPR126 is expressed in the lung. Gene expression was identified in whole lung tissue and in structural airway cells specifically HASMs and HBECs by both qPCR and RNA seq. There was significant donor variation in both the HASMs and in HBECs in the RNA seq datasets. This data is in agreement with RNA seq data from GTEx, which shows high levels of GPR126 expression in lung tissue. GEO datasets also back up this data, showing GPR126 expression is slightly but significantly elevated in COPD lungs. This is in

contrast to data from a recent study which found GPR126 (key SNP rs17280293) to be associated with diffusing capacity of the lung. This study also found that *GPR126* mRNA levels were lower in patients with decreased diffusing capacity compared to individuals with normal diffusing capacity. Furthermore, *GPR126* expression was also lower in COPD subjects compared to never smoking controls. [146]

While the mRNA expression data is robust, the evidence for protein expression in the lung is less so. Protein Atlas shows that GPR126 in undetectable in the human lung from 3 donors. Our data here shows that GPR126 is expressed in the human lung but only in some donors. In these donors, GPR126 is expressed highly around blood vessels and in smooth muscle around the airways. This is in line with the literature; GPR126 is known to be highly expressed in endothelial cells, specifically human umbilical vein endothelial cells (HUVECs) and it is known to be highly expressed in muscle tissue. [147, 156] Smooth muscle cells and bronchial epithelial cells are important cell types in COPD and airway remodelling. In COPD, smooth muscle surrounding airways is thickened resulting in airway lumen size decreasing in addition to changes in epithelial cell properties (see chapter 1).[168] GPR126 mRNA is expressed in both cell types suggesting a potential role for the gene.

One key limitation in this chapter was the analysis of protein expression in the lung by IHC. The antibody used was validated in small intestine tissue by The Human Protein Atlas. Small intestine tissue was used as a positive control in these experiments and did show positive staining. While I showed some GPR126 staining in the human lung, there was variation between and within the donors used. The degree of variation seen in staining suggests there may be some non-specific staining, suggesting that the antibody needed further validation. Moreover, no GPR126 expression is identified by The Human Protein Atlas. [156] Newly available antibodies should be used to validate these findings as well as identifying a true negative control tissue.

To summarise, in this chapter we have identified a possible causal SNP in the rs14827447 association signal for FEV1/FVC. This variant (Ser123Gly) is predicted to be damaging to protein function as it lies within a conserved

binding domain, which is important for binding collagen type IV. GPR126 is expressed in the lung and in airway structural cells and given the literature on the protein and its known interactions with components of the ECM, the data here supports the hypothesis that GPR126 may be involved in airway remodelling via ECM interactions. I hypothesise Ser123Gly may alter cell/matrix homeostasis leading to aberrant repair mechanisms and airway remodelling resulting in airway obstruction.

4 Investigation of GPR126 Ligand Specificity and Effects of Ser123Gly on cAMP Signalling in a Recombinant Cell System

4.1 INTRODUCTION

As previously discussed, GPR126 has been shown in numerous studies to signal via cAMP pathways in a wide range of biological processes. These signalling pathways are covered in detail in the general introduction, however what follows is a brief summary of GPR126 signalling pathways from the literature (summarised in Figure 4-1). GPR126 belongs to a class of GPCRs known as adhesion GPCRs (aGPCRs) which have long N-termini containing domains important for adhesive functions. Until recently, it was unclear as to whether these receptors functioned purely as adhesion molecules as their name suggests or whether they could also function as GPCRs by coupling to heterotrimeric G proteins and effecting downstream signalling responses. It is now known that this class of receptors can and do act as both adhesion molecules and traditional GPCRs (reviewed in [169]).

In brief, GPCRs are activated upon the binding of an agonist which induces a conformational change in the receptor resulting in activation of the associated heterotrimeric G protein complex. This heterotrimer is comprised of alpha (α), beta (β) and gamma (γ) subunits. Ligand activation of a receptor results in an activated G_{α} subunit which exchanges GTP to GDP triggering the dissociation of G_{α} subunit from the $G_{\beta\gamma}$ dimer and from the receptor. Activation of GPCRs which results in an intracellular increase in cAMP, including GPR126, signal via $G_{\alpha s}$. $G_{\alpha s}$ activates activate adenylyl cyclase which in turn catalyses the conversion of ATP to cAMP, instigating a signalling cascade.

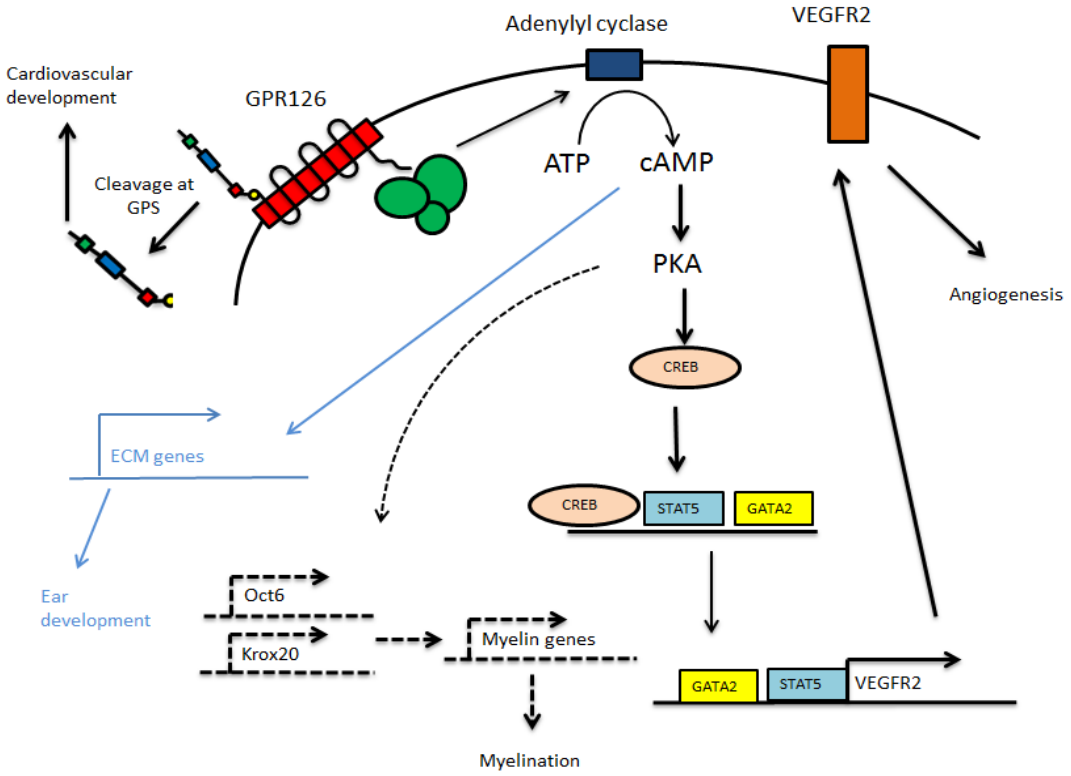


Figure 4-1 Signalling pathways of GPR126 in aspects of development. GPR126 is activated upon ligand binding, which possibly involves detachment of the N-terminal fragment (NTF). A conformational change in the receptor exposes the tethered agonist, the stachel sequence which can activate receptor signalling. GPR126 signals via associated G proteins, G_{as} and G_{ai}, to regulate intracellular cAMP levels. Intracellular cAMP increase leads to the transcription of genes that are important for angiogenesis, myelination and ear development.[81, 82, 102, 104, 105, 170] *In addition, cleavage at the GPS motif results in the NTF becoming dissociated with the C-terminal fragment and acting as a signalling molecule important for cardiovascular development.* [170]

Early studies supported a typical GPCR signalling model for GPR126 with activation of the receptor resulting in an increase in intracellular cAMP presumably via $G_{\alpha s}$ interactions. GPR126 was first shown to signal through cAMP when addition of forskolin, which activates adenylyl cyclase to raise cAMP levels, was able to rescue the GPR126 phenotype in GPR126 mutant zebrafish studies. [101] In addition, cAMP elevation and PKA activation reverses the effects of GPR126 knockout mouse mutants. [104] Studies since have

confirmed this with a number of ligands (Table 4-1) identified for GPR126 resulting in an increase in cAMP levels. [96-98, 104, 171]

Ligand	Measurement	Cell system	Transcript/ Species	Study
Collagen type IV	cAMP: ELISA-based kit (EMD Millipore)	Schwann cells purified from rat sciatic nerve	Endogenous GPR126 expression	Paavola et al (2014) [98]
	cAMP: ELISA-based kit (EMD Millipore)	Exogenous overexpression in HEK293 cells	Full length Zebrafish GPR126	Paavola et al (2014) [98]
	Collagen type IV induced activation not detected by measuring cAMP. CREB-luciferase reporter showed activation of receptor.	Exogenous overexpression in COS-7 cells	Human GPR126	Petersen et al (2015) [97]
Laminin-211	AlphaScreen cAMP assay kit (PerkinElmer Life Sciences)	Exogenous overexpression in COS-7 cells	Human GPR126	Petersen et al (2015) [97]
Col4 Peptide (GPRGKPGVDGYNGSRGDPGYP)	cAMP: Colorimetric competitive immunoassay (Enzo Life Sciences)	Exogenous overexpression in HEK293T cells	Human GPR126	Kuffer et al (2016) [171]
Prion Peptide (KKRPKPGGWNTGGSRYPGQGSPGGNRYP)	cAMP: Colorimetric competitive immunoassay (Enzo Life Sciences)	Exogenous overexpression in HEK293T cells	Human GPR126	Kuffer et al (2016) [171]
Stachel (THFGVLMDLPRSASQL)	cAMP: Alpha Screen cAMP assay kit (PerkinElmer Life Sciences)	Exogenous overexpression in COS-7 cells	Human GPR126. Transcript ID: ENST00000367608 encoding the B2 isoform which is 1222 aa in length.	Liebscher et al (2014) [96]

Table 4-1 Ligands reported to activate GPR126 resulting in an intracellular increase in cAMP.

Paavola et al. (2014) reported a roughly 3 fold increase in cAMP levels in HEK293 cells overexpressing zebrafish GPR126 with collagen type IV

treatment using an ELISA based method.[98] Liebscher et al (2014) reported a 10-15 fold increase in cAMP in COS-7 cells expressing GPR126 when treated with 1mM stachel peptide.[96] In addition to this, Petersen et al (2015) (data in supplementary material) showed a ~30 fold increase in luciferase activity in COS-7 cells overexpressing GPR126. [97] In addition two peptides, one derived from the prion peptide and one a fragment of the collagen type IV peptide have been shown to previously interact with the N-terminal fragment of GPR126 and cause an approximate 2-fold increase in intracellular cAMP. [171]

Petersen et al (2015) showed that purified laminin-211 activated human GPR126 in vitro.[97] Interestingly, in this paper, treatment of transfected cells overexpressing GPR126 with laminin-211 resulted in a dose-dependent decrease in cAMP whereas treatment with laminin-211 and physical shaking of cell cultures resulted in a cAMP increase.[97] The findings from this study suggest that in Schwann cells, laminin-211 binding stabilises GPR126 in the inactive state which suppresses G_s signalling. Following this basal lamina maturation and laminin-211 polymerisation facilitates a conformation change in GPR126 resulting in an active G_s signalling conformation. This process is then thought to result in the removal of the NTF and binding of the stachel sequence. [97]

Interestingly, the N-terminal fragment (NTF) produced by cleavage at the GPS motif appears to have distinct roles from the membrane bound C-terminal fragment (CTF). Most protein domains found in aGPCRs are involved in cell-cell or cell-matrix interactions. [169] However, in the case of GPR126 the NTF has an additional role independent of the CTF as an important regulator of cardiovascular development. [170] The exact mechanisms of this is currently unclear but it is predicted that the NTF acts as a separate signalling molecule once dissociated from the receptor.

GPR126 signalling is activated by a short peptide called the stachel sequence. This amino acid sequence lies within the ectodomain of the receptor and lies C-terminal to the GPS cleavage site and functions as a tethered agonist. Upon structural changes in the receptor, presumably caused by ligand binding, the stachel sequence is exposed to the 7-transmembrane helix domain which

causes G protein activation.[96] This agonistic sequence is a feature present in other aGPCRs and the core sequence is highly similar between aGPCRs.[172] As such, stachel peptides derived from aGPCRs in the same sub-family are able to activate more than one member of the sub-family and this promiscuity also applies to aGPCRs of different sub-families. For example, the stachel peptide of GPR110 (subfamily VI) can activate GPR126 and GPR64, which are both members of sub-family VIII. [172] In addition, GPR126-derived stachel and activate GPR64 and vice-versa, however each peptide exerted a more potent effect on its originating receptor. [172]

The main aim of this chapter was to study GPR126 signalling in a recombinant model and assess functional differences between GPR126 variants Ser123 and Gly123 identified through bioinformatic analysis of the GWAS association signal (chapter 3).

There are a number of methods which have been previously used to measure GPR126 activation including immunoassay methods [97, 98, 171] and reporter gene assays.[97] cAMP accumulation can also be measured, as detailed in the general methods chapter, with radioactively labelled ^{3}H adenine, which is incorporated into cAMP. Tritiated cAMP is then separated from other tritiated compounds by column chromatography. [173] Reporter assays are means to quantify a molecular effect i.e. gene expression or signalling activity into a measurable outcome. Reporter genes offer a chance to study gene regulation and expression rapidly and quantitatively with sensitivity. These assays usually have a wide dynamic range allowing for detection of small or large signals as well as being simple to perform, rapid, which allows for high throughput screening and with outcomes that are simple to measure (e.g. luminescence or fluorescence). To measure gene expression, reporter genes are usually joined to promoter sequences and transfected into cells via an expression vector. Cells can then be assayed for the presence or expression levels of the reporter by directly measuring the protein product or the enzymatic activity of the gene product. For example, in a luciferase assay luminescence is the measurable outcome of the luciferase enzyme's catalysis of its substrate,

luciferin. Reporters are ideally genes that are not endogenously expressed in the cell line of interest.

Reporter genes are commonly used for analysis of *cis* acting transcriptional events. The reporter allows functional identification and characterisation of promoter/ enhancer elements because the expression of the reporter protein is directly correlated to the transcriptional activity of the reporter gene. For these studies, the promoter region of interest is cloned upstream of the reporter gene.

For this project a luciferase reporter was used to measure the cAMP response element. Luciferases are a class of oxidative enzymes that produce bioluminescence as a by-product of the mono-oxygenation of luciferin. CRE is the DNA binding sequence for transcription factor CREB (CRE binding protein). Upon receptor activation and cAMP elevation, CREB locates to the nucleus of the cell and binds to CRE thus promoting its transcription and transcription of the luciferase reporter gene (Figure 4-2). By measuring the activity of the enzyme luciferase, via luminescence, it is possible to indirectly measure the transcriptional activity of CRE and therefore the levels of cAMP signalling.

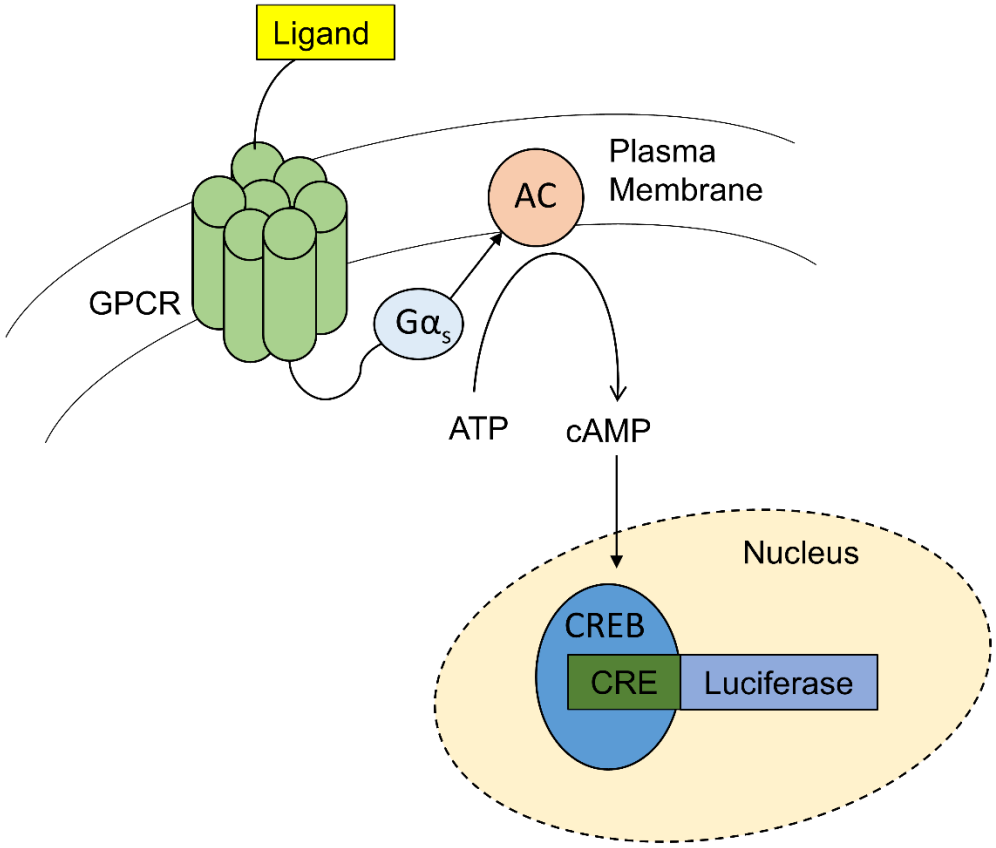


Figure 4-2 Schematic diagram representing the signalling pathways which lead to reporter activation. Upon ligand binding, the GPCR interacts with Gαs which activates adenylyl cyclase (AC), which catalyses the production of cAMP from ATP. This results in the activation of the cAMP response element (CRE) which drives transcription of the luciferase gene.

In this chapter, I generated a recombinant cell model to assess GPR126 signalling in response to ligand interactions. Of particular interest to us is the GPR126 variant rs17280293 (Ser123Gly) which results in a missense variant in the coding region of GPR126. This substitution occurs in the CUB domain, thought to be important for ligand binding. One key objective of this chapter is to produce cell lines overexpressing both GPR126 variants and assess any effect the substitution has on GPR126 signalling. Using ligands identified from the literature I planned to activate receptor signalling and measure the downstream effects via a CRE-luciferase reporter.

4.1.1 Aims

The overarching objective of this chapter was to overexpress GPR126 variants and assess downstream signalling effects of the Ser123Gly variant. To do this required a number of more specific targets.

- To identify ligands which activate GPR126 signalling by evaluating previously reported ligands including components of the ECM, collagen type IV and laminin-211, synthetic peptides derived from collagen type IV and the prion peptide as well as the tethered agonist, the stachel peptide.
- To identify any effects of the Ser123Gly substitution on GPR126 signalling using identified agonists for GPR126 via measurement of cAMP signalling.

4.2 METHODS

4.2.1 Plasmid Production and Preparation

4.2.1.1 Vector Design

Custom designed vectors were designed and purchased from VectorBuilder (https://en.vectorbuilder.com). Two vectors were designed to overexpress the two variants of GPR126 resulting from the rs17280293 variant. These vectors express the Ser123 (vector name: pLV[Exp]-Puro-CMV>hGPR126, Vector Builder ID: VB160108-10075) and Gly123 (vector name: pLV[Exp]-Puro-CMV>hGPR126*, Vector Builder ID: VB160108-10076) variants of GPR126. An additional plasmid was designed without a protein coding open reading frame (ORF) as an empty vector (EV) control which was as otherwise identical to the two protein coding plasmids (vector name: pLV[Exp]-Puro-CMV>ORF_91bp, Vector Builder ID: VB160220-10010). These plasmids are referred to as pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV throughout this thesis. Table 4-3 and Figure 4-3 shows the components of the vector. A vector with a lentivirus backbone was chosen as to facilitate the introduction of *GPR126* to primary human airway cells by lentivirus infection rather than standard transfections, which can be less efficient in primary cells than in cell lines.

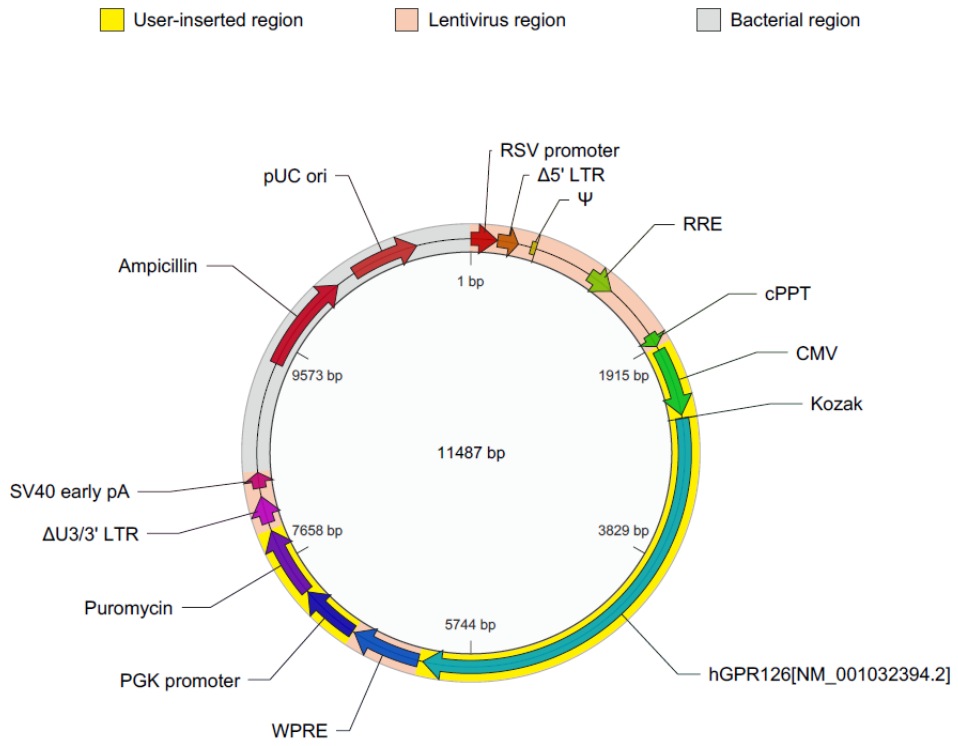


Figure 4-3 The pLV[Exp]-Puro-CMV>hGPR126 (pLV-GPR126^{Ser123}) plasmid from Vector Builder. The lentivirus plasmid contains user entered regions such as the CMV promoter, kozak sequence and the gene of interest to be expressed – in this case GPR126 (NM_001032394.2).

As covered in the general introduction (section 1.3) GPR126 has 4 primary full length transcripts (Table 4-2). A consideration when designing plasmids for recombinant experiments was which transcript and therefore isoform of GPR126 to express. To address this issue I looked up GPR126 isoform expression in GTEx RNA-seq datasets and our own in house RNA-seq datasets from HBECs (the majority of the RNA-seq data, including that from HASMs was unavailable at this point). I chose the transcript coding for the alpha 2 isoform, which is the most abundant transcript in the human lung (see chapter 3 for data).[142] The alpha 2 isoform (ENST00000296932.12, NM_001032394, NP_001027566) lacks exon 6 of the full length protein. [12]

Name	Transcript ID	Base pairs	Isoform	Protein
ADGRG6-201	ENST00000230173.10	7026	Alpha 1	1221aa
ADGRG6-204	ENST00000367609.7	6887	Beta 1	1250aa
ADGRG6-203	ENST00000367608.6	6879	Beta 2	1222aa
ADGRG6-202	ENST00000296932.12	6849	Alpha 2	1193aa

Table 4-2 GPR126 transcript and their corresponding protein isoforms.

In addition to the transcript of interest, puromycin and ampicillin resistance genes were included in the vector to allow selection of stably transfected mammalian cells and selection of bacterial cells respectively. A cytomegalovirus (CMV) promoter was chosen, which is a common promoter used in mammalian expression vectors. The kozak sequence is an important sequence for the initiation of translation and is positioned directly upstream of our gene of interest.

Component name	Nucleotide position	Full name	Description
RSV promoter	1-229	Rous sarcoma virus (RSV) enhancer/promoter	Allows Tat-independent production of viral mRNA.
Δ5' LTR	230-410	HIV-1 truncated 5' LTR	Permits viral packaging and reverse transcription of the viral mRNA.
Ψ	521-565	HIV-1 psi packaging signal	Allows viral packaging.
RRE	1075-1308	HIV-1 Rev response element	Permits Rev-dependent nuclear export of unspliced viral mRNA.

cPPT	1803-1920	Central polypurine tract	Facilitates the nuclear import of HIV-1 cDNA through a central DNA flap.
CMV	1959-2547	Cytomegalovirus promoter	Mammalian expression promoter from the human cytomegalovirus
Kozak	2572-2577	Kozak	The Kozak consensus sequence is a sequence which occurs in eukaryotic mRNA which plays a major role in the initiation translation.
Gene of interest	2578-6159	hGPR126[NM_001032394.2]	GPR126 transcript NM_001032394.2 with A or G at position 367 for each plasmid. In the empty vector this was a 91 base pair long sequence which was non-coding.
WPRE	6198-6795	Woodchuck hepatitis virus posttranscriptional regulatory element	Facilitates effective transcription termination at the 3' LTR.
PGK promoter	6814-7324	Mouse phosphoglycerate kinase promoter	Allows high-level expression of the selection marker in mammalian cell lines.
Puromycin	7337-7936	Puromycin resistance gene	Permits selection of stably transduced mammalian cell lines.
ΔU3/3' LTR	8007-8241	HIV-1 truncated 3' LTR	Allows viral packaging but self-inactivates the 5'LTR for biosafety purposes. The element also contains a polyadenylation signal for transcription termination and polyadenylation of mRNA in transduced cells.
SV40 early pA	8314-8448	SV40 early polyadenation signal	Allows transcription termination and polyadenylation of mRNA.
Ampicillin	9402-10262	Ampicillin resistance gene	Allows selection of the plasmid in E.coli.

pUC ori	10433-11021	pUC origin of replication	Permits high-copy replication and maintenance in E.coli.

Table 4-3. The components comprising the plasmid sequence.

4.2.1.2 Bacterial cell culture

The plasmids were delivered from VectorBuilder in stab culture in cloning host stbl3, a chemically competent strain of *E.coli*. LB broth and LB agar were prepared as explained in the section 2.2.7. Agar plates were pre-warmed in at 37°C in an incubator before a pipette tip was used to streak each of the three stab cultures onto agar plates. Plates were left to incubate overnight at 37°C. One colony was picked from each culture to inoculate 5ml LB broths which were left to incubate overnight at 37°C with shaking at ~200RPM on a Sanyo Orbital Shaker. For maxipreps, LB broth (~100mls) was inoculated with 100µl of the 5ml starter broth and left to incubate overnight at 37°C with shaking at 200RPM.

4.2.1.3 Monarch® Plasmid Miniprep Kit

1.4mls of overnight culture was used for minprep. Plasmid miniprep was carried out as per the manufacturer's instruction (New England Biolabs) as outlined in section 2.2.7.2

4.2.1.4 Qiagen EndoFree Plasmid Maxi Kit

A 100ml overnight broth was inoculated for maxipreps. Plasmid Maxiprep was performed using the Qiagen EndoFree Plasmid Maxi kit following the manufacturer's protocol as described in section 2.2.7.3.

4.2.1.5 Quality control of plasmid DNA

Plasmid DNA obtained from the maxiprep was subjected to a variety of quality control methods outlined below prior to transfection. Plasmid DNA

concentration and quality was assessed using a Nanodrop ND-2000 Full spectrum Spectrophotometer (Thermo Scientific).

4.2.1.5.1 Sanger Sequencing of Plasmid DNA

The primer sequences shown in Table 4-4 were used for Sanger sequencing of the plasmid DNA to assess the sequence of the GPR126 ORF and other plasmid components. The sequencing reaction was carried out by The DNA Sequencing Facility (School of Life Sciences, University of Nottingham) utilising dye terminator chemistry as described in section 2.2.9. The facility use the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). The facility were provided with 5µl DNA plasmid (100ng/µl) and 3µl of primer (10pmole). Sequence analysis was carried out using a 3130xl ABI PRISM Genetic Analyzer (Life Technologies). Analysis of the sequence data was facilitated using GeneScreen (University of Leeds). [136]

Primer name	Sequence (5' -> 3')	Location on plasmid (bp)	Length
Vector Builder F	gcaacagacatacaaactaaagaat	1863	25
VF-2	agctggtttagtgaaccgtca	2523	21
VR-1	gagaggactgggcttccatt	2617	20
VF-3	agctttgaacagctctgcct	3321	20
VR-2	ggagcccaacaacaatttccc	3441	21
VF-4	acctactggggacctgttga	4220	20
VR-3	agttctgcccatcagcagtt	4294	20
VF-5	ccaagaagtgcctcacagtt	5060	20
VR-4	atgtcaggagagttgctgct	5129	20

Vector Builder R	ggagcaacatagttaagaatacc	6277	23

Table 4-4. Primer sequences designed to check plasmid DNA was as expected and to check that the plasmids contained the correct variants.

4.2.1.5.2 Restriction digest of Plasmid DNA

Restriction enzymes are enzymes that recognise short, specific DNA sequences and cleave double stranded DNA at these sites. This allows isolation of specific fragments from larger DNA sequences. Digest sites and the expected fragments produced from a digest reaction were provided by Vector Builder. Restriction digests were carried out using enzymes ApaLI (NEB R0507S) and NdeII (NEB R0111S). For each reaction, 1µg of DNA was required and volumes of DNA required are shown in Table 4-5. The components were mixed in a 1.5ml tube and incubated for 15 minutes at 37ºC. The enzymes used in this section are Time-Saver™ qualified enzymes from New England Biolabs which will digest 1 µg of substrate DNA in 5-15 minutes using 1µl of enzyme.

	1µg DNA (µl)	**Enzyme (10units) (ApaLI/ NdeI/ ScaI-HF®) (µl)**	**Restriction enzyme buffer (µl)**	**Sterile water (µl)**	**Total (µl)**
Ser123	1.5	1	5	42.5	50
Gly123	2.7	1	5	41.3	50
Empty Vector	7.7	1	5	36.3	50

Table 4-5. Restriction digest reaction set up.

The reaction was stopped by addition of 10µl gel loading dye (NEB - B7024S). 20µl of each reaction was loaded into a 0.5% agarose gel and run for approximately an hour at 30V and gels were imaged as previous explained.

4.2.1.5.2.1 Linearization and purification of plasmid DNA

Purified plasmid DNA was first linearized to ensure efficient integration into the genome of the cell. This was achieved using the restriction enzyme ScaI-HF® (New England Biolabs) which has one restriction site in the plasmid in the ampicillin resistance gene which would not be needed in the mammalian cells. The digest was set up in a 1.5ml centrifuge tube and incubated for 15 minutes at 37°C then 65°C for 20 minutes to inactivate the enzyme. The linearized plasmid DNA was purified using the QIAquick PCR Purification Kit following the manufacturer's protocol.

In brief, Buffer PB was added to the DNA at a ratio of 5:1 and the mixture was added to a QIAquick column which was placed into a 2ml collection tube. The column was centrifuged for 60 seconds at 17,000g and the flow through was discarded. The column was washed with 750µl Buffer PE and centrifugation. The flow through was again disposed of and the column was centrifuged once again in the column to remove residue wash buffer. The column was placed into a clean 1.5ml micro centrifuge tube and DNA was eluted with the addition of 50µl Buffer EB (10mM Tris.Cl, pH 8.5) for 1 minute before centrifugation for 1 minute.

4.2.2 Cell Biology Methods

The GloResponse™ CRE-luc2P HEK293 cells (Promega, E8500) is a HEK293 cell line stably expressing the firefly luciferase gene (*luc2P)* under the control of a minimal HSV thymidine kinase promoter with multiple cyclic-AMP response elements (CREs). The *luc2P* gene is a destabilised variant of the luciferase gene with a protein half-life of approximately 1 hour due to the attached PEST sequence. These sequences are peptide sequences which are rich in proline (P), glutamic acid (E), serine (S) and threonine (T) and act as signal peptides for protein degradation. The *luc2P* gene is positioned downstream of the cAMP response elements (CRE) which are activated upon cAMP signalling cascades. The CRE is the DNA response element for the cAMP response element-binding protein (CREB). In response to intracellular cAMP elevation, CREB relocates to the nucleus of the cell and binds to CRE,

promoting gene transcription. The location of the *luc2P* gene downstream of a CRE ensures that luciferase is produced upon activation of the CRE hence making this cell line ideal to study G_s- and $G_{i/o}$-coupled GPCRs, which signal through cAMP and the cAMP response element (CRE).

Thawing, culture and freezing down of the CRE-luc2P HEK293 cells was carried out following the cell culture guidelines from section 2.1. Cells were cultured in DMEM (high glucose) with 10% FCS and hygromycin B (50µg/ml) generally in T75 flasks. Transfected cells were cultured in the presence of puromycin (0.5µg/ml) keeping the cells in selection throughout culture. No antibiotics were used in the assay media.

4.2.2.1 Transfections of CRE-Luc2P HEK293 cells with GPR126 variants and empty vector.

Passage 2 CRE-Luc2P HEK293 cells were transiently transfected with the two GPR126 variants and the empty vector. Transfections were optimised in 12 well plates to ensure there was sufficient protein and RNA for analysis of transfection efficiency. Transfections were carried out using Fugene 6 (Promega) as outlined in the general methods. To optimise transfections a range of Fugene: DNA ratios were used which were 6:1, 4:1. 3:1 and 1.5:1. Transfection master mixes were set up as follows for each well transfected (Table 4-6). Transfected cells were left to incubate for 24 hours at 37°C with 5% CO_2.

	6:1	**4:1**	**3:1**	**1.5:1**
Total volume	50µl	50µl	50µl	50µl
DNA (250ng/µl) amount (µg)	0.5 (2µl)	0.5 (2µl)	0.5 (2µl)	0.5 (2µl)
Fugene volume	3µl	2µl	1.5µl	0.75µl
Media	45µl	46µl	46.5µl	47.25µl

Table 4-6 Mastermixes used for transfections for each well of a 12 well plate.

4.2.2.2 Transfection of CRE-luc2P HEK293 cells with the linearized plasmid

CRE-luc2P HEK293 cells at passage 2 were grown to ~80% confluency in a T75 flask. Cells were then split into individual wells of a 6 well plate at a seeding density of $3x10^5$ cells/ml in 2mls DMEM+10% FCS. The cells were left to incubate at 37°C with 5% CO2 until ~70% confluent, usually 3-4 days post-thawing. Cells were transfected with the linearized plasmids (section 4.2.1.5.2.1) using the transfection master mix shown in Table 4-7. There were three transfection mastermixes for the three plasmids: pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV. One well of a 6 well plate was transfected with 100µl of the appropriate transfection mastermix.

Component	**Volume**
DNA	1ug (4ul)
Fugene	1.5ul
Media	93ul
Total Volume	100ul

Table 4-7. Transfection mastermix for transfections to create a stable cell line.

The concentration of puromycin needed to kill cells not expressing the construct was established by performing a kill curve on the non-transfected CRE-*luc2P* HEK293 cells. The idea being that the lowest concentration that kills all the cells will be used for the selection media. CRE-*luc2P* HEK293 cells were cultured in T75 flasks and treated with 0.25, 0.5, 1, 2, 5 or 10µg/ml puromycin for 10-14 days replacing media as required until the entire cell cultures had died, which was assessed by microscopy.

Transfected cells (denoted P0) were left to incubate for 48 hours at 37°C with 5% CO_2 in DMEM+10% FCS. After two days, cells were trypsinised (P1) and re-plated to 10cm plates with selection media (DMEM + 10% FCS + 0.5µg/ml puromycin) was added to the cells to promote growth of the cells which had successfully integrated the plasmid DNA. Media was replaced everyday as to remove dead cells. Stocks of cell lines overexpressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV were grown from surviving colonies and these cells were continually grown in selection media. Once the population of cells expressing the construct became confluent, they were split 1:5 to T75 flasks (P2) which were used for freezing down as described in the general methods and passaging for experiments.

4.2.2.3 Luciferase assays

Luciferase is a common choice as a reporter for measuring signalling events and gene expression as the functional luciferase enzyme is immediately produced upon translation. This makes luciferase reporters ideal for high throughput assays as they are rapid and sensitive. The luciferase (*luc2P)* gene catalyses the mono-oxygenation of beetle luciferin using ATP as a cofactor - a reaction which produces bioluminescence (Figure 2-1) which can be quantified by measuring luminescence.

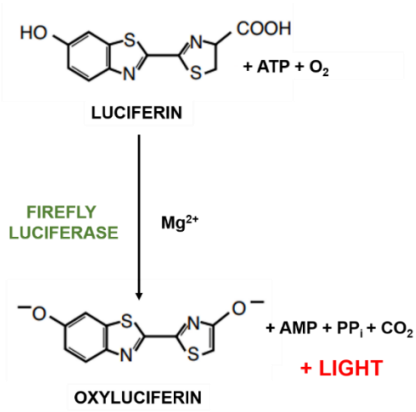


Figure 4-4. Firefly luciferase catalyses the production of oxyluciferin from luciferin, a reaction which produces light.

Cell populations expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV were cultured in complete media with 0.5µg/ml puromycin. Cultures were grown to 80% confluency by which point the cells were trypsinised and resuspended to a density of 2x10^5 cells/ml. The cells were seeded to individual wells of a black-walled 96 well plate (Thermo Fisher Scientific 137101) in 100µl culture media the day before experimentation at passage 3. On the day of experimentation, culture media was replaced with serum-free media.

After the required experimental time period, the Bright-Glo™ Luciferase Assay System (Promega, E2610) was used to lyse cells and produce the luciferase reaction. Figure 2-1 displays the process of the mono-oxygenation of luciferin which is catalysed by luciferase in the presence of Mg2+, ATP and molecular oxygen.

The provided Bright-Glo buffer and Bright-Glo substrate were mixed and 100µl of the solution was added to each well. Plates were incubated at room temperature for at least 2 minutes before luminescence was measured on a Flexstation 3 microplate reader for 0.5 seconds/well.

4.2.2.4 CRE-luc2P HEK293 cell line performance assay

Cells were prepared in black walled 96 well plates as explained in section 4.2.11. Cells were left to adhere until the following day. Cells were stimulated by the adenylyl cyclase activator, forskolin (100µM) and the non-selective β adrenoreceptor agonist isoprenaline (10µM) made up to working concentrations in assay media for 2 hours, 4 hours, 8 hours and 24 hours at 37°C with 5% CO_2. The assay was stopped by adding 100µl of the Bright-Glo™ Luciferase Assay System. The plate was read for luminescence on a Flexstation 3 microplate reader for 0.5 seconds/well.

4.2.2.5 Treatment of transfected cells with collagen type IV

Collagen type IV (Sigma C7521) has been identified as a GPR126 agonist which initiates a cAMP signalling response. [98] Following the protocol outlined in Paavola et al (2014)[98], collagen type IV was reconstituted in PBS to a stock concentration of 1mg/ml. Collagen type IV was dissolved in PBS at room

temperature for 3 hours with rotation. A working concentration of 3µg/ml was achieved by diluting stock aliquots in assay media. Cells transfected with pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV and non-transfected CRE-*luc2P* HEK293 cell lines were seeded to black walled 96 well plates. The following day, culture media was replaced with serum free media (SFM), SFM with collagen type IV (3µg/ml) as performed by Paavola et al (2014) [98] or SFM with forskolin (100µM). Cells were left to incubate before the assay was stopped with the addition of 100µl Bright Glo assay reagent.

4.2.2.6 Treatment of transfected cells with laminin-211

Laminin-211 interacts with GPR126 and this interaction is thought to be important for Schwann cell development. [97] It is hypothesised that laminin-211 binding of GPR126 can prevent or promote GPR126 mediated cAMP accumulation depending on whether it is polymerised or not. [97] Cells were seeded to wells of a 96 well plate as previously described. Laminin-211 was obtained from Biolamina (LN211-02) and used at a working concentration of 200ng/µl in SFM as described by Petersen et al (2016). [97] The day after seeding, cells were treated with laminin-211 and incubated for 4 hours before the assay was stopped and luminescence was read.

4.2.2.7 Extraction and use of human airway smooth muscle extracellular matrix to treat cells overexpressing GPR126

Adhesion-GPCRs are known to interact with components of the ECM and GPR126 specifically has been reported to interact with ECM components collagen type IV and laminin-211. [97, 98, 105] The hypothesis behind this set of experiments is that exogenous addition of ECM may activate GPR126 signalling.

To tackle this, two approaches were used; 1) Removal of the ECM from HASMs and application of the extracted proteins to transfected cells seeded to individual wells of a 96 well plate and 2) De-cellularisation of HASM ECM and subsequent seeding of cells on top of the de-cellularised ECM.

4.2.2.7.1 Preparation of Extracellular Matrix from Human Airway Smooth Muscle Cells (HASMs)

HASMs (donor 1515) at passage 2 were grown in T162 flasks in DMEM + 10% FCS to confluence plus an additional 3 days to ensure maximum matrix deposition. Cultures were washed with Phosphate Buffered Saline (PBS). Cells were removed (de-cellularisation) from the ECM using a method from Philp (2016) [174] which was adapted from a method from Lu et al., (2012). [175] Cultures were then incubated at room temperature with 50mM NH_4OH + 0.05% Triton X-100 (Sigma Aldrich) in PBS for 15 minutes or until cells were no longer visible under a Nikon Diaphot 300 Inverted Microscope. The 50mM NH_4OH + 0.05% Triton X-100 solution was aspirated and the cultures were washed 3 times with PBS. All liquid was then aspirated and ECM was extracted using the following protocol.

4.2.2.7.2 Extraction of the Extracellular Matrix

Proteins of the ECM display variability in solubility and therefore separate processes are needed to extract the different proteins. The method of extracting the extracellular matrix was provided by Dr Christopher Philp (University of Nottingham) as outlined in his thesis which was adapted from a method from Ngoka (2008).[174, 176] Two buffers were used for this process:

> Extraction buffer 1: PBS (pH 7.4), 0.1% Sodium dodecyl sulphate (SDS), 0.25% Na-deoxycholate, complete Mini Protease Inhibitor Cocktail + EDTA (1 tablet per 10ml extraction buffer).
>
> Extraction buffer 2: PBS (pH 7.4), 5.0 M Urea, 2.0M Thiourea, 50 mM DTT, 0.1% SDS.

Buffer 1 was applied to the prepared, decellularised cultures at ~1ml per 9.5cm^2 (~15mls per T162 flask) and flasks were incubated for 5 minutes on a platform rocker at room temperature. The liquid was aspirated from the flask and collected into 10ml universal tubes. Buffer 2 was then directly applied to the flask (~15mls per flask) and incubated at room temperature on a platform rocker. The plate was scraped using a cell scraper and the liquid was collected

in 10ml universal tubes. Each of the ECM extracts was then subject to dialysis using Slide-A-Lyzer™ Dialysis Cassettes (12-30ml, 2K molecular weight cut off) in PBS for 24 hours. Samples were aliquoted and stored at -80°C until further use.

4.2.2.7.3 Quantification of protein using the Pierce™ BCA Protein Assay Kit

The protein extracted from HASM ECM was quantified using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) using the standards provided and following the manufacturer's instructions exactly.

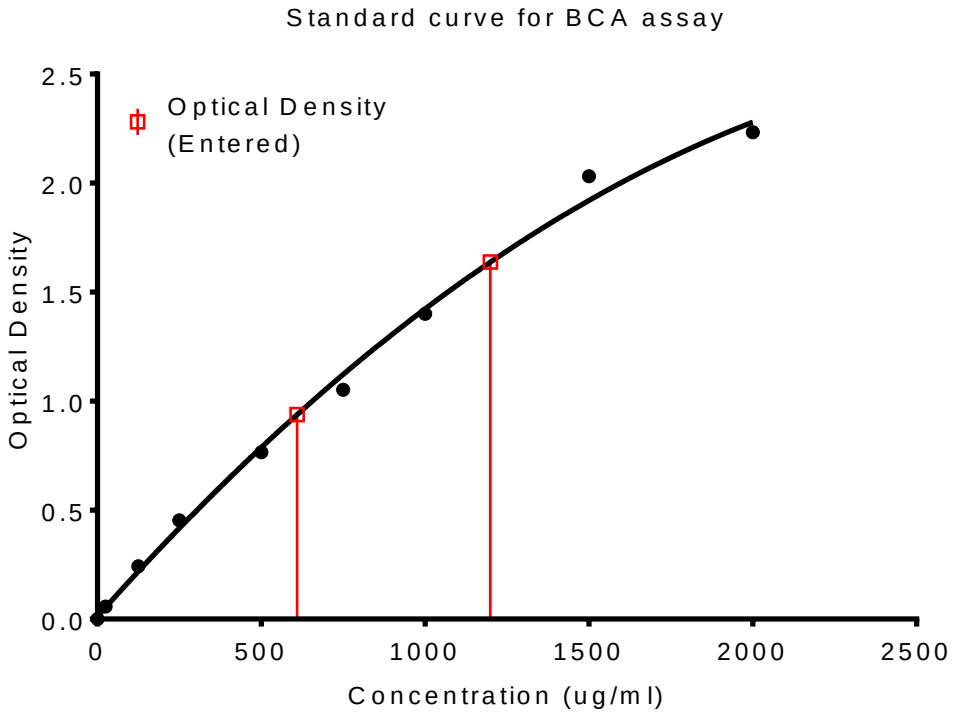
This assay kit is a formulation based on bicinchoninic acid (BCA) and allows the colorimetric detection and quantification of total protein in a sample. This method combines the reduction of Cu^{+2} to Cu^{+1} by protein in an alkaline medium with the colorimetric detection of Cu^{+1} using a unique reagent containing bicinchoninic acid. The reaction produces a purple-coloured product which is formed by the chelation of two molecules of BCA with one Cu^{+1}. This complex exhibits a strong absorbance at 562nm.

Standard solutions are provided in the kit and are made up as follows:

A. 300ul stock
B. 375 stock + 125 diluent
C. 325 stock +325 diluent
D. 175 B +174 diluent
E. 325 C + 325 diluent
F. 325 E + 325 diluent
G. 325 F + 325 diluent
H. 100 G + 400 diluent
I. 400 diluent

25µl of each sample (standards and extracts 1 and 2) was added to wells in duplicate with 200µl of working reagent provided in the kit. The contents of the wells were mixed using a plate shaker for 30 seconds and absorbance was measured at 562nm using a Flexstation 3 microplate reader (Molecular Devices).

ECM was extracted from a T162 flask from confluent HASM cells. The standards provided with the BCA assay kit (Thermo Fisher) were used to produce a standard curve (Figure 4-5) from which the concentration of protein in the extracts were extrapolated. The ECM was extracted from culture plates using two different buffers – resulting in two different extracts. Both samples lay on the standard curve. Extract 1 had a protein concentration of 609µg/ml and extract 2 had a concentration of 1199µg/ml.



	ug/ml (Interpolated)	Optical Density (Entered)
Extract 1	609.265	0.938
Extract 2	1199.267	1.638

Figure 4-5. The standard curve generated from BSA standards allowing extrapolation of the protein concentration of extracellular matrix extracted from HASMs.

4.2.2.7.4 Application of ECM extracts to cells overexpressing GPR126

CRE-LUC2P HEK293 cells expressing GPR126 or EV were seeded to wells of a black-walled 96 well plate. The day after plating cells were treated with:

- ECM extract 1 (10µg/ml)
- Buffer 1 alone
- ECM extract 2 (10µg/ml)
- Buffer 2 alone
- ECM extracts 1 and 2 combined (10µg/ml)
- Only buffer 1 and 2 combined
- Serum free media
- Isoprenaline (10µM) positive control

ECM, buffer and isoprenaline were diluted in SFM to the concentrations shown. Media was aspirated off the cells and replaced with 100ul media containing each treatment. After 4 hours incubation at 37°C the plates were removed and assayed for luciferase as outlined in section 4.2.2.3.

4.2.2.7.5 Plating cells overexpressing GRP126 onto decellularised HASM ECM

HASM ECM was prepared as detailed in section 4.2.2.7.1.CRE-luc2P HEK293 cells expressing pLV-GPR126^{Ser123} or pLV-EV were cultured in T75 where trypsinised as previously described. These cultures were resuspended to 2x10^5 cells/ml and 100µl of cell suspension was added to each well of the 96 well plate containing the HASM ECM or untreated wells. Cells were left to incubate for 4, 8 or 24 hours at 37°C at 5% CO_2. After incubation, 100µl Glo response assay system solution was added to the wells and left to stand for 2 minutes. Whole contents of the well was then transferred to a black walled 96 well plate and plates were measured for luminescence.

4.2.2.8 Two peptides derived from collagen type IV and the prion peptide used as potential GPR126 agonists.

In addition to collagen type IV and laminin-211, two peptides, one derived from collagen type IV (col4) and one derived from the prion peptide (P28) have been identified as GPR126 ligands.[171] The collagen type IV peptide and prion peptide share a moderate homologous region which is essential for the activity of both proteins. The KKRPKPG motif in P28 is similar to the GPRGKPG motif

of the GPR126 agonist, collagen type IV. The peptides were synthesised by Thermo Fisher Scientific and the amino acid sequences are shown in Table 4-8. A peptide containing KKRPKPG derived from the prion peptide has been shown to be sufficient to induce a GPR126-dependent cAMP response in vitro and in vivo. (171)

Peptide name	Sequence
P28	KKRPKPGGWNTGGSRYPGQGSPGGNRYP
Col4	GPRGKPGVDGYNGSRGDPGYP
Stachel	THFGVLMDLPRSASQL

Table 4-8 Amino acid sequences of peptides used in luciferase assays

4.2.2.9 Activation of GPR126 expressing cells with the stachel peptide – a tethered agonist for GPR126

Cultures of cells expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV were prepared in black-walled 96 well plates as previously described. The stachel peptide (Table 4-8) was synthesised by Thermo Fisher Scientific. The peptide was reconstituted in sterile H_2O to a stock concentration of 10mM as recommended by the supplier and further diluted in SFM to working concentrations of 1mM, 100µM, 10µM, 1µM, 0.1µM for luciferase assays. On the day of assaying, culture media was removed and replaced with SFM or SFM containing stachel (1mM), isoprenaline (10µM) or forskolin (100µM). Cells were left to incubate for 5 hours at all doses or for 2, 5, 8 and 24 hours with the maximal dose (1mM).

4.2.2.10 Pre-treatment of cells with PKA inhibitor H89

H89 is a protein kinase inhibitor which exerts the most potent effect on protein kinase A (PKA) as a competitive antagonist of ATP at its binding site on PKA. pLV-GPR126^{Ser123} and pLV-EV cells were brought up from -80 storage into

T75 flasks. When confluent, cells were split to 96 well plates at $1x10^5$ cells/ml in 100ul of DMEM + 10% FCS. The day after seeding, media was replaced with 100µl serum free media containing H89 (Sigma B1427) (10uM final solution) or plain assay media. After 2 hours, the cells were washed three times with assay media. Agonists made up to final concentration in assay media were then added to the cells for a 4 hour incubation. The agonists used were isoprenaline (10uM), forskolin (100uM) and Stachel (1mM). The assay was stopped with the addition of 100µM of Bright Glo Luciferase assay reagent.

4.2.3 Characterisation of transfected cells by qPCR

Following transfections, cells were lysed and RNA was extracted using the GenElute™ Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) and cDNA was synthesised using the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitorgen). These methods are described in detail in section 2.2.3.

Quantitative PCR was used to assess *GPR126* mRNA expression in transiently and stably transfected cells using the standard method used throughout this thesis and detailed in section 2.2.6. The *GPR126* TaqMan® PDAR (Thermo Fisher), 18s (Life Technologies) and GAPDH (Thermo Fisher) housekeeping PDARs were used. Relative gene expression was calculated using the formula: relative mRNA expression = $2^{-\Delta CT}$, where ΔCT is the CT value from the gene of interest PDAR minus the CT value of the housekeeper. (132)

4.2.4 cAMP Accumulation Assay on pLV-GPR126 and pLV-EV Transfected Cells Treated with Stachel Peptide

HEK293 cells overexpressing GPR126 were treated with the stachel peptide and assayed for cAMP for 10, 60, 120 and 240 minutes using column chromatography as explained in section 2.1.8.

A solution containing 3-isobutyl-1-methylxanthine (IBMX) (Sigma I5879) was made up in serum free media to a final IBMX concentration of 100µM. To each well 270µl of IBMX containing medium was added to the cells which were left

to incubate for 15 minutes in the cell incubator. Isoprenaline stock solution was made up to 100µM in serum free media and 30µl of this stock solution was applied to the relevant wells to give a final working concertation of 10µM. Stachel peptide was prepared as outlined in section 4.2.2.9 and 30µl of stachel solution was added to relevant wells to give a final working concentration of 1mM. Plates were left to incubate for 10minutes, 1 hour, 2 hours or 4 hours and assays were stopped by adding 50µl of concentrated hydrochloric acid. Plates were taped together and frozen at -20°C until columns were run.

The columns were set up and run as described in the general methods chapter and vials were counted on a Packard Scintillation counter using a $^{3}H/^{14}C$ dual count program for 5 minutes per vial. Three standards were made up containing 100 100µl stock ^{14}C-cAMP + 5ml 0.1M imidazole + 10ml scintillation fluid.

4.2.5 Statistical Analyses

4.2.6 Column statistics

Graphpad Prism 7.04 software was used to calculate column statistics including mean values, standard deviation (SD) and standard error of the mean (SEM) and the median, minimum and maximum values and 25% and 75% percentiles.

4.2.6.1.1 ANOVA

Column analyses was carried out using an ANOVA with multiple comparisons tests to compare effect of treatment on cell responses to untreated cells and compare *GPR126* expression levels to non-transfected control cells.

4.2.6.2 Grouped analyses - 2 way ANOVA

A 2-way ANOVA was used to investigate the effect of more than one variable e.g. the effect of treatment at different time points for example the effect of isoprenaline or forskolin treatment on CRE-luciferase response in cells transfected with different constructs.

A detailed description of these statistical analyses can be found in the general methods chapter.

4.3 RESULTS

4.3.1 Optimisation of the CRE-Luc2P Assay with Isoprenaline and Forskolin

To establish the kinetics of the CRE-luciferase system and to establish the most appropriate time points for the luciferase assays, passage 2 CRE-luc2P HEK293 cells were treated with agonists to increase intracellular cAMP for 2, 4, 8 and 24 hours. CRE-Luc2P HEK293 cells were treated with the adenylyl cyclase activator forskolin (100µM) and isoprenaline (10µM), an adrenergic receptor agonist, to produce a cAMP response from the cells. The data shown in Figure 4-6 is n=5 raw data in relative light units (RLU). Unstimulated cells show baseline levels of luminescence with mean values ranging from 103-120 RLU with background luminescence ~10 RLU. Statistically significant increases in luminescence is seen for forskolin at 4 ($p<0.0001$), 8 ($p<0.0001$) and 24 hours ($p=0.0053$) and for isoprenaline treatment at 2 ($p=0.0291$), 4 ($0<0.0001$) and 8 hours ($p=0.0314$). The greatest increase in luminescence for both stimuli was seen at 4 hours (forskolin mean = 28997±3060, isoprenaline mean = 12583±745).

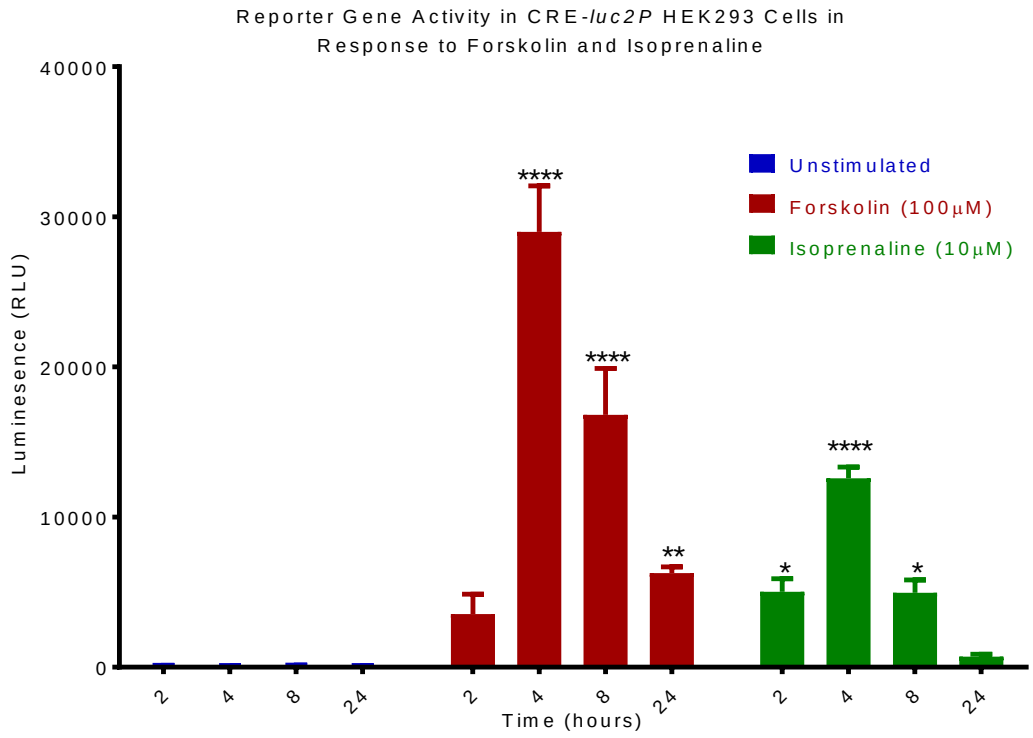


Figure 4-6. Response of luciferase reporter in response to forskolin and isoprenaline over a time course. A CRE driven luciferase response is seen at all time-points assayed with both forskolin and isoprenaline stimulation. The biggest response for both forskolin and isoprenaline was seen after 4 hours incubation. Isoprenaline and forskolin responses were compared to unstimulated cells at the same time point with a 2-way ANOVA with Dunnett's multiple comparisons test.

4.3.2 Plasmid DNA Quality Control

4.3.2.1 Plasmid DNA Quantification and Quality Control

Sample ID	Nucleic Acid Conc. (ng/µl)	260/280	260/230
Blank	0	0.02	-0.02
Ser123	698.3	1.88	2.27
Gly123	368.4	1.89	2.26

Empty Vector	139	1.88	2.19

Table 4-9. Concentrations and 260/280 and 260/230 absorbance ratios for the plasmid DNA.

Quantification of plasmid DNA concentration after the maxiprep was achieved using a Nanodrop ND-2000 Full spectrum Spectrophotometer. The concentrations (ng/µl) and the 260/280 and 260/230 ratios are shown in Table 4-9. The 260/280 and 260/230 ratios are used to determine the 'purity' of the DNA samples with a 260/280 value of ~1.8 and 260/230 value of 2-2.2 being considered pure for DNA. The concentration for the Ser123 and Gly123 plasmids were adequate with the empty vector concentration low for a Maxiprep.

4.3.2.2 Restriction Digest

The plasmids were then digested with ApaLI and NdeI and run on a 1% agarose gel. Digest of the plasmids produced bands of expected sizes (Figure 4-7). Lanes 1 to 3 show digestion of all three plasmids (two variants plus empty vector) with the ApaLI enzyme. Lanes 5 and 6 show digestion of the two variant plasmids with the NdeI enzyme. The emtpy vector plasmid does not have a restriction digest site for NdeI.

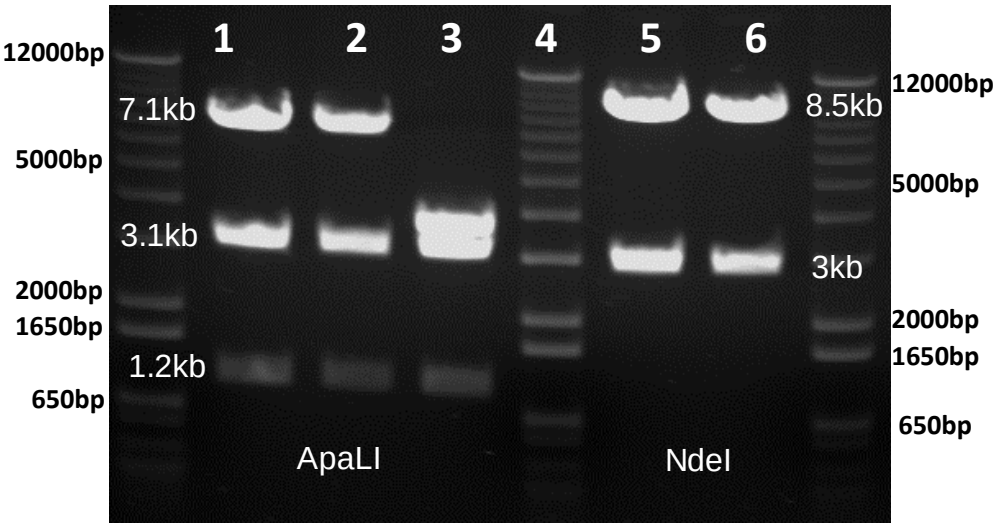


Figure 4-7 Restriction digest of plasmids with restriction enzymes ApaLI and NdeI. Digest with ApaLI of the Ser123 and Gly123 plasmids produced expected bands of 7.1kb, 3.1kb, and 1.2kb (lanes 1 and 2) and three bands of 3.6kb, 3.1kb and 1.2kb were produced as expected from digest of the control plasmid. Digest of the Ser123 and Gly123 plasmids with NdeI produced bands of 8.5kb and 3kb.

4.3.2.3 Sequencing of Plasmid DNA

The GPR126 ORF in the plasmid was fully sequenced by Sanger Sequencing by the Sequencing facility at the University of Nottingham. Sequencing of the plasmid confirmed the base pair change c.367A>G in the Gly123 plasmid and confirmed the wild-type sequence in the Ser123 plasmid (Figure 4-8).

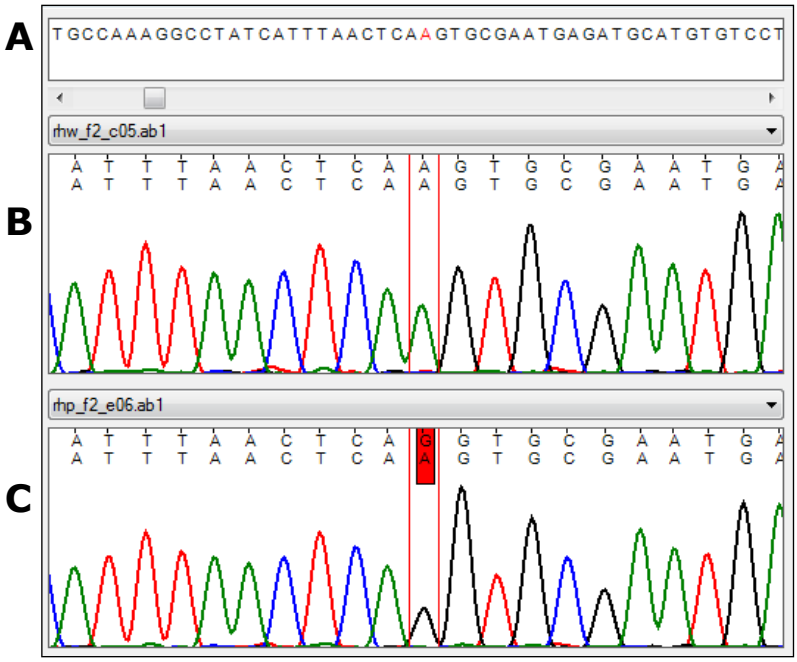


Figure 4-8. Sequencing results showing base pair change c.367A>G in the Gly123 plasmid. The GPR126 cDNA sequence (ENST00000296932) was obtained from Ensembl (A) and used as a reference sequence for the Ser123 plasmid (B) and the Gly123 plasmid (C).

4.3.3 GPR126 Signalling in CRE-luc2P HEK293 Cells Transiently expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV.

4.3.3.1 GPR126 mRNA Expression in Cells Transfected with GPR126 Constructs

CRE-*luc2P* HEK293 cells were used as a host for GPR126 overexpression. RNA was extracted from cells 24 hours post transfection. Four different Fugene: DNA ratios were used in order to find the optimum transfection conditions.

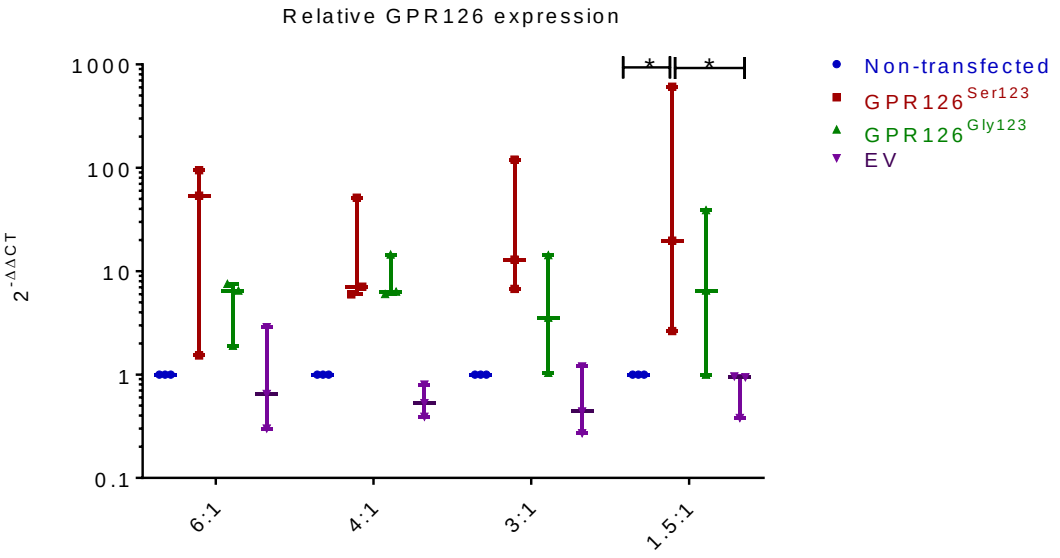


Figure 4-9 shows the results of n=3 transient transfections in CRE-*luc2P* HEK293 cells. There is increased *GPR126* expression in both the cell lines transfected with the *GPR126* construct and no increased *GPR126* expression seen in the cells transfected with the EV construct. This increase does not reach statistical significance in the majority of the conditions possibly due to the high variability seen in the data. In addition, there is no statistically significant difference in *GPR126* expression seen between the two *GPR126* vectors.

The only statistically significant increases in GPR126 expression was seen in the 1:5.1 ratio conditions. GPR126^{Ser123} mean expression significantly increased over non-transfected cells (mean = 209.94 ± 198.813, p=0.0335) and was significantly increased over empty vector expression (p=0.0332). The Fugene: DNA ratio which on average produced the highest levels of GPR126 expression was 1.5:1 however there is a wide spread of data particularly in the pLV-GPR126^{Ser123} transfected cells.

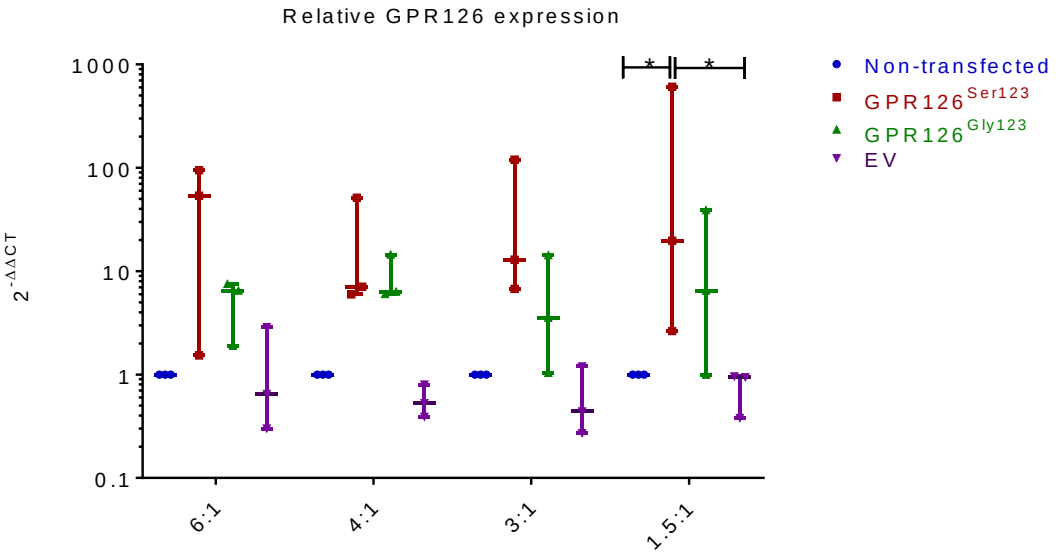


Figure 4-9 GPR126 expression in cells transfected with GPR126^{Ser123}, GPR126^{Gly123} and EV 24 hours after transfection (n=3). Data is normalised to non-transfected cells. Expression levels were statistically compared using a 2-way ANOVA with Tukey's multiple comparisons test.

CRE-*luc2P* HEK293 cells were transfected with pLV-GPR126 constructs and the pLV-EV constructs as previously described in 96 well plates. The cells were treated with collagen type IV and forskolin was used as a positive control (Figure 4-10). Forskolin produced an increase in CRE-luciferase in all conditions although this failed to reach statistical significance versus unstimulated controls in any of the transfection conditions at 4 (EV mean = 137±84; Ser123 mean = 87±56; Gly123 mean = 130±87) or 8 hours (EV mean = 158±110; Ser123 mean = 39±25; Gly123 mean = 49±31). There was no response seen in transfected cells to collagen type IV treatment versus unstimulated cells at 4 (EV = 1.191±0.2, Ser123 = 1.043±0.243, Gly123 = 0.937±0.263) or 8 hours incubation (EV = 1.427±0.028, Ser123 = 1.380±0.332, Gly123 = 0.786±0.108). Data shown in fold change in luminescence.

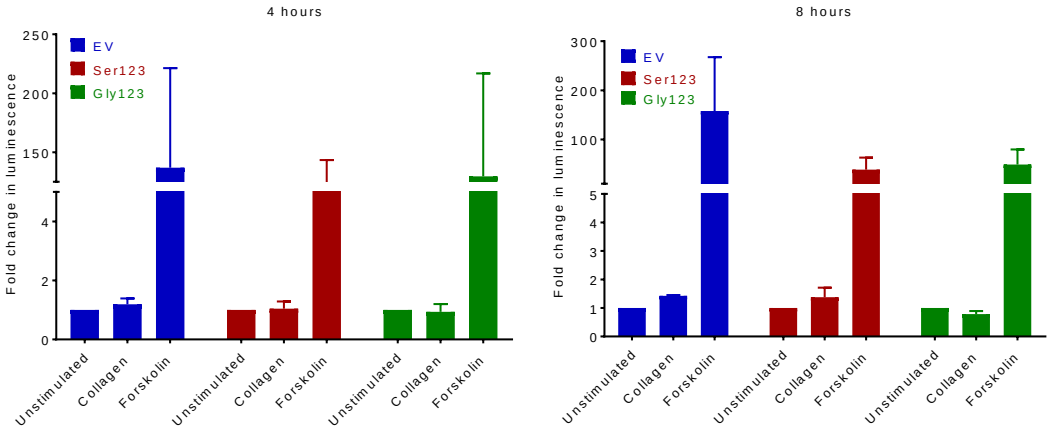


Figure 4-10. Luciferase reporter activity in CRE-luc2P HEK293 cells transiently overexpressing GPR126 variants and EV. Cells were treated with collagen type IV (3µg/ml) or forskolin (10µM) for 4 or 8 hours before the assay was stopped. Data shown is fold change over unstimulated cells. Raw data was analysed using a 2-way ANOVA with Dunnet's multiple comparisons test.

4.3.4 Production of a Mixed Population of Cells Stably Expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV.

To address the issue of variable expression of GPR126 in the transient cells, cell lines were created which would stably transfect pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and the pLV-EV constructs. Three cell lines were produced by linearization of the plasmid to facilitate integration of the construct into the genome prior to transfection.

4.3.4.1 Linearization Restriction Digest

Linearization of the plasmid was achieved by using a restriction enzyme, ScaI-HF®, which cuts in the middle of the ampicillin gene. Figure 4-11 shows the results of the linearization with only one band on the gel per plasmid. The linearized plasmids were used to stably transfect CRE-*luc2P* HEK293 cell cultures.

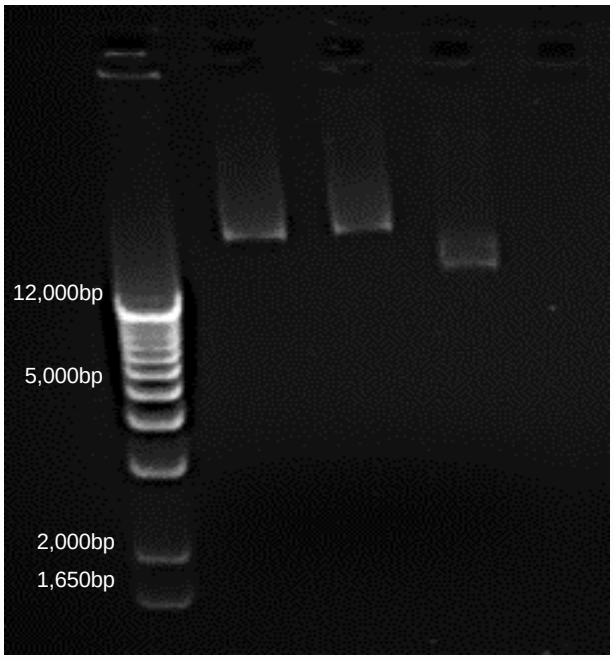


Figure 4-11. Gel showing the three plasmids after linearization with NEB Time-Saver™ Qualified Restriction Enzyme ScaI-HF® showing only one band per plasmid as expected showing that only one digest look place. From left to right: Ladder, Ser123, Gly123, EV.

4.3.4.2 Characterisation of Cell Lines Expressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123}, pLV-EV constructs by qPCR

Cells transfected by linearized plasmid (section 4.2.1.5.2.1) were cultured in selection media (0.5µg/ml puromycin) until all non-transfected cells had been killed (approximately 7 days) with media changes every day. The remaining cells were cultured in selection and grown to produce liquid nitrogen stocks in T75 flasks. RNA was extracted and isolated from cell lysate of passage 3 (post-transfection) CRE-*luc2P* HEK293 cells which had been transfected with pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV and non-transfected cells. cDNA was synthesised from the RNA and Taqman qPCR was performed to assess the levels of expression of GPR126 in the transfected cell lines. GPR126 expression is increased in both cell lines which were transfected with the GPR126 construct.

The pLV-GPR126Ser123 transfected cells show an average 18 fold increase (SEM±3) in *GPR126* mRNA expression over non-transfected cells ($p<0.0001$) (Figure 4-12). The Gly123 cell line shows GPR126 expression approximately 12 fold (SEM+0.5) over non-transfected ($p<0.0001$). The cells expressing the Ser123 construct express higher levels of GPR126 than the Gly123 construct expressing cells ($p= 0.0082$). There is no difference between *GPR126* mRNA expression in the empty vector versus the non-transfected cells.

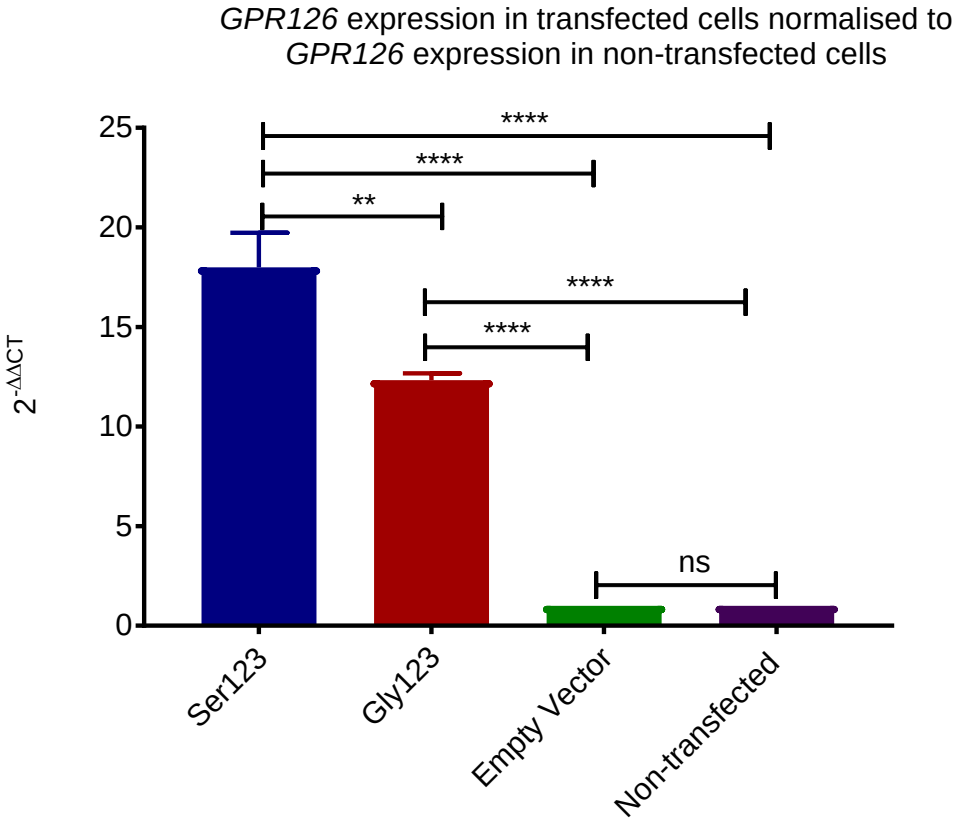


Figure 4-12 GPR126 Gene expression normalised to non-transfected cells using the $2^{-\Delta\Delta CT}$ formula (n=3). Expression values were compared using an ANOVA with Tukey's multiple comparisons test.

4.3.5 Establishing GPR126 Agonists by Luciferase Assays in a Mixed Population of Cells Expressing GPR126

4.3.5.1 Treatment of Recombinant Cell Lines with Collagen type IV

Based on evidence found in the literature for GPR126, collagen type IV was selected as a potential activating ligand for the receptor.[98] Transfected cells were treated with 3µg/ml collagen type IV for 4 and 8 hours. Cells were lysed with the addition of the Bright-Glo™ Reagent and the plate was for 2 minutes at room temperature. Luminescence was measured using the Flexstation 3 for 1 second per well. Figure 4-13 shows the response of the Ser123, Gly123 and EV cells to collagen type IV (3µg/ml) and forskolin (100µM) as a positive control. Data is n=5 and is normalised to unstimulated cell luminescence reading.

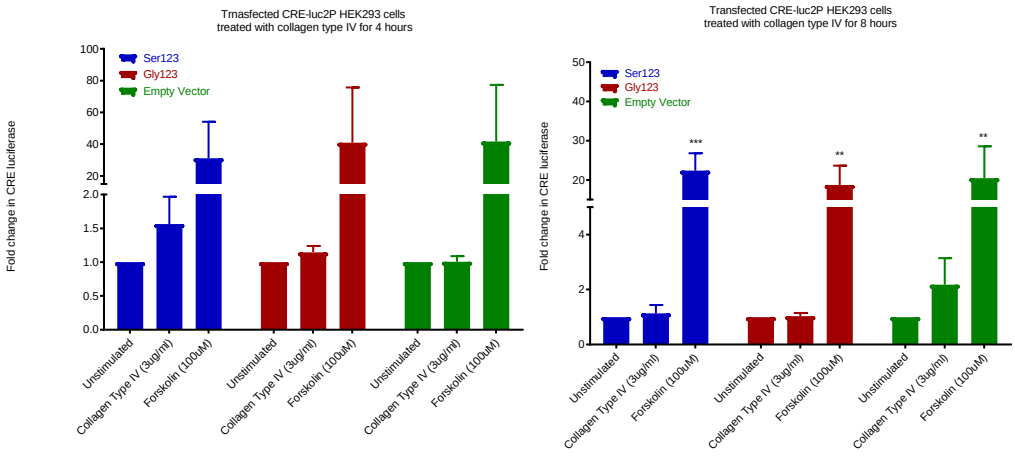


Figure 4-13. CRE-luciferase response of transfected cells to incubation with collagen type IV for 4 and 8 hours (n=5). Raw data was compared using a 2-way ANOVA with Tukey's multiple comparisons test. The transfected cells show no response to collagen type IV activation.

At 4 hours incubation mean fold increase in luciferase activity in response to forskolin was 41.7±35.6 for EV cells, 41±35 for Gly123 and 31.2±22 for Ser123 cells although this data did not reach statistical significance. At 8 hours, the forskolin response versus unstimulated reaches statistical significance for all

conditions (p=0.0004 for Ser123, p=0.0029 for Gly123 and p=0.0012 for EV). Treatment with collagen type IV did not initiate a statistically significant change in reporter activity in any cell line at either 4 or 8 hours. There is a hint of an increase in reporter activity to collagen type IV in the pLV-GPR126^{Ser123} cell line at 4 hours (mean = 1.565±0.404) but this is not statistically significant.

4.3.5.2 Treatment of Recombinant Cell Lines with Laminin-211

Transfected cells were subjected to treatment from laminin-211, another potential GRP126 ligand identified in the literature.[97] Cells were treated with 200ng/ml laminin-211 or forskolin for 4 hours. As with the collagen, there is no significant response to laminin-211 while the forskolin response is 50-100 fold in all cell lines (Figure 4-14). These values however do not reach statistical significance due to large variation within the data.

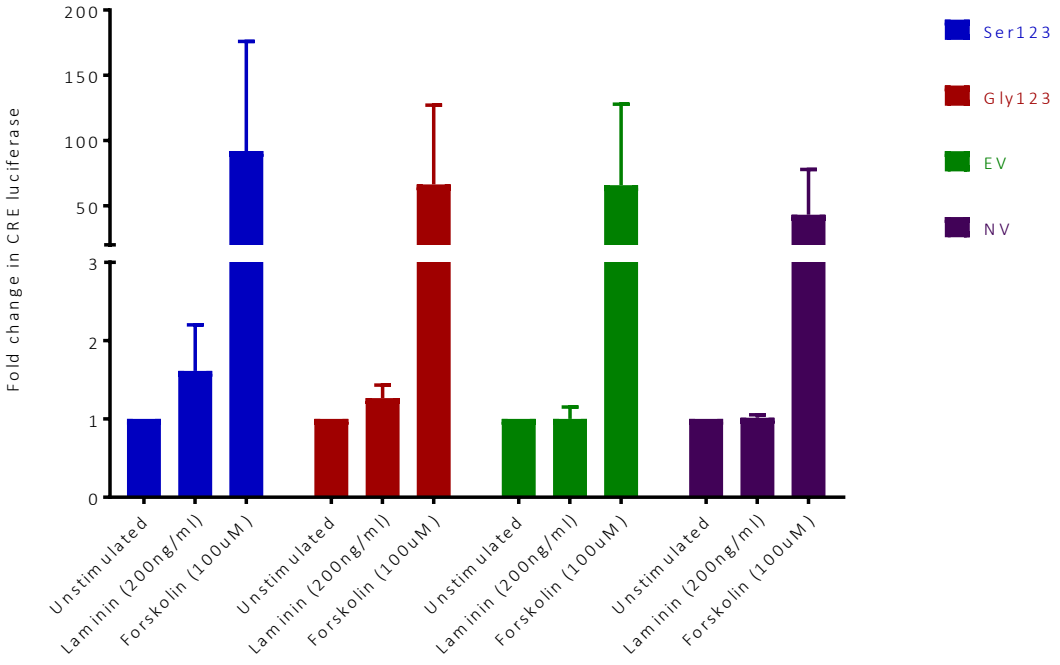


Figure 4-14. Fold change in reporter gene activity in transfected cells in response to laminin 211 (200ng/ml) and forskolin (100µM) (n=4). Raw data was compared using a 2-way ANOVA with Tukey's multiple comparisons test.

4.3.5.3 Treatment of recombinant cell lines with peptides derived from collagen type IV and the prion peptide

Two short peptides derived from the collagen type IV peptide and the prion peptide (P28) have been previously identified to initiate a cAMP response inCRE-luc2P HEK293 cells overexpressing human GPR126. These two sequences share a region of moderate homology which is thought to be essential for the activity of both peptides. GPR126^{Ser123}, GPR126^{Gly123} and EV were treated with the peptides for 4 hours before being measured for luminescence.

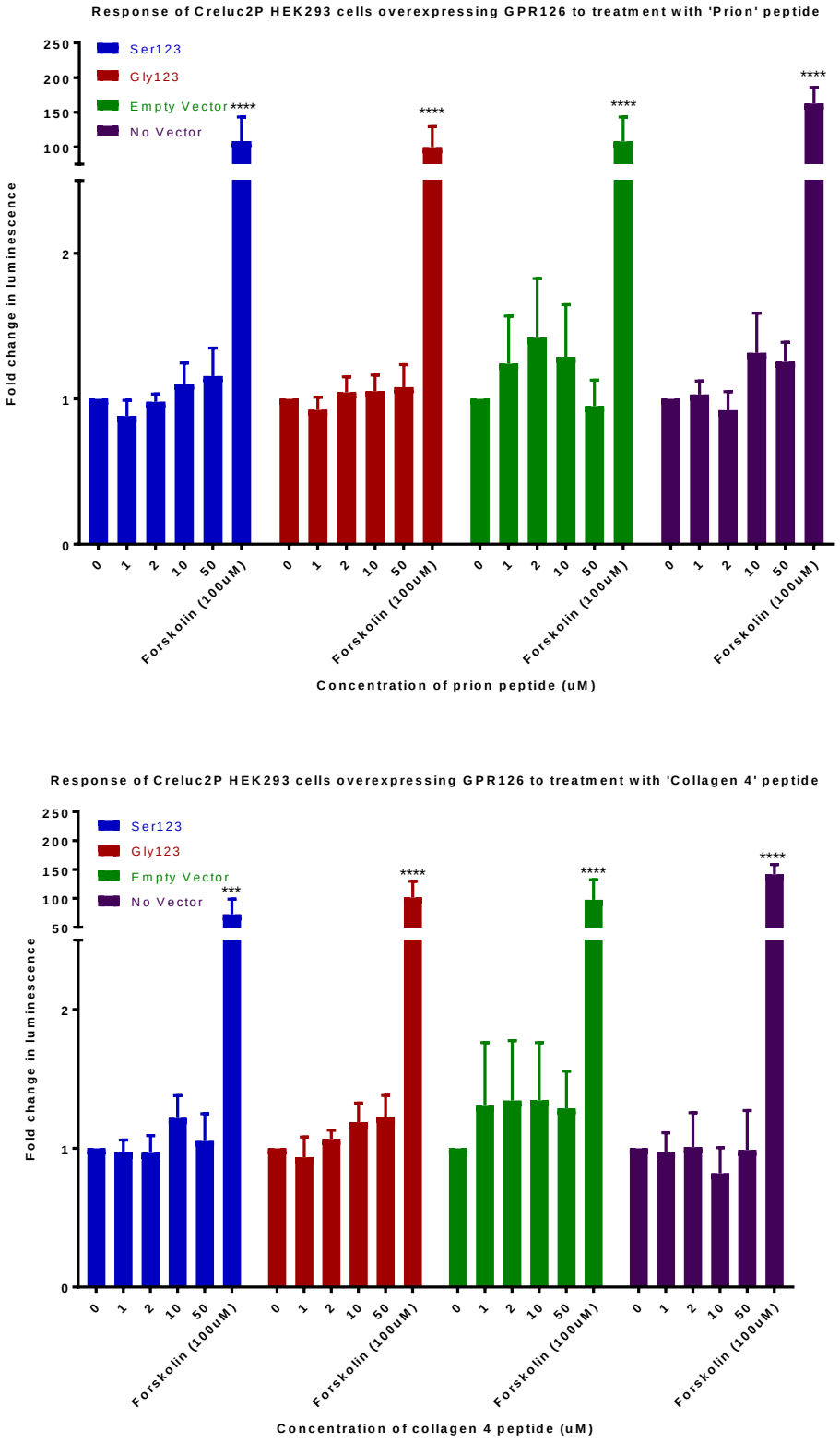


Figure 4-15 CRE luciferase response of transfected cells to the amino acid sequence derived from the prion peptide (P28) and collagen type IV. Data is

normalised to untreated cells and data shown is n=5. Raw data was analysed using a 2-way ANOVA with Dunnett's multiple comparisons test.

Treatment of the cells with neither peptide at 1, 2, 10 or 50µM instigated a statistically significant response in reporter activity after 4 hours of treatment (Figure 4-15).

4.3.5.4 Using the ECM extracted from HASMs as a potential GPR126 activator

Collagen type IV and laminin-211 are extracellular proteins which are both reported in the literature as ligands for GPR126. CRE-luc2P HEK293 cells overexpressing pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123} and pLV-EV were treated with ECM extracts produced by HASM cells.

4.3.5.4.1 Transfected CRE-luc2P HEK293 cells treated with ECM extracts from HASM culture

ECM extracts were diluted to a working concentration of 10µg/ml before being applied to cultures. Serum free media containing the ECM extracts and dialysed buffers were applied to the cells which were left to incubate for 4 hours. Figure 4-16 shows the response of the cells to ECM addition. There is no response seen in the transfected or empty vector cells while the addition of the buffers seem to be causing a reduction in luciferase activity. The graph shows log fold change in order to incorporate the isoprenaline response which is up to 100-fold over baseline versus non-treated cells (*GPR126* transfected cells $p=0.0002$, EV cells $p<0.0001$).

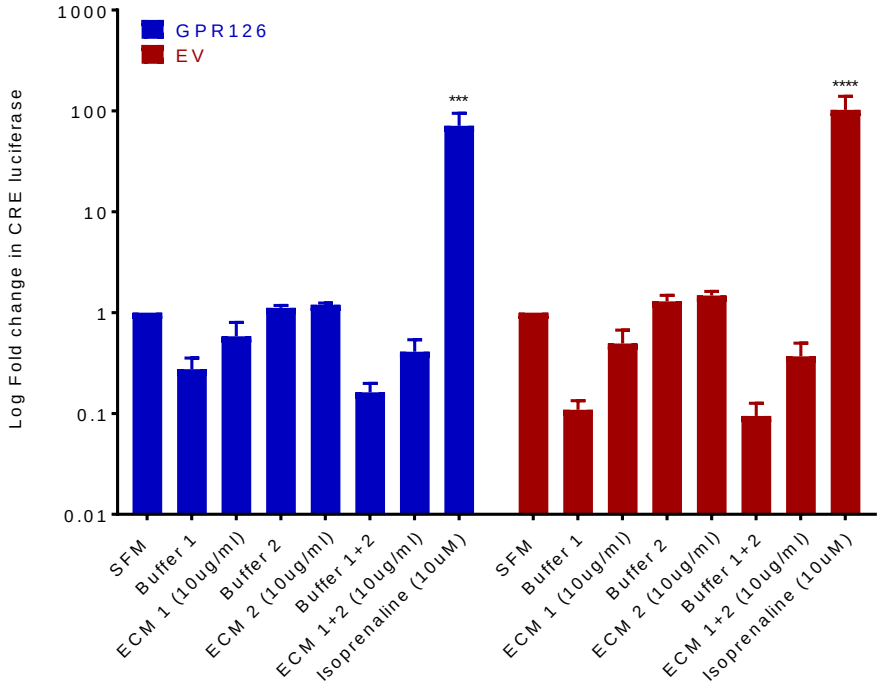


Figure 4-16 Luciferase activity in GPR126^{Ser123} and EV cells in response to extracts of HASM ECM (n=5). Cells were treated with dialysed ECM extracts or the dialysed buffer in which the ECM was diluted in or isoprenaline for 4 hours. Raw data was analysed using a 2-way ANOVA with Dunnett's multple comparisons test.

4.3.5.4.2 Transfected CRE-luc2P Cells Seeded to HASM ECM

HASMs were grown to confluence plus 3 additional days of culture to allow maximum deposition of ECM on the wells of a 96 well plate. The ECM was de-cellularised as explained and CRE-luc2P HEK293 cells expressing GPR126^{Ser123} or EV were seeded at a seeding density of 2 x 10^5 cells/ml in 100µl culture medium to ECM coated wells of the 96 well plate. Cells were left to incubate for 4, 8 and 24 hours before the assay was stopped and the cells were lysed by adding 100µl of the Bright Glo luciferase assay reagent.

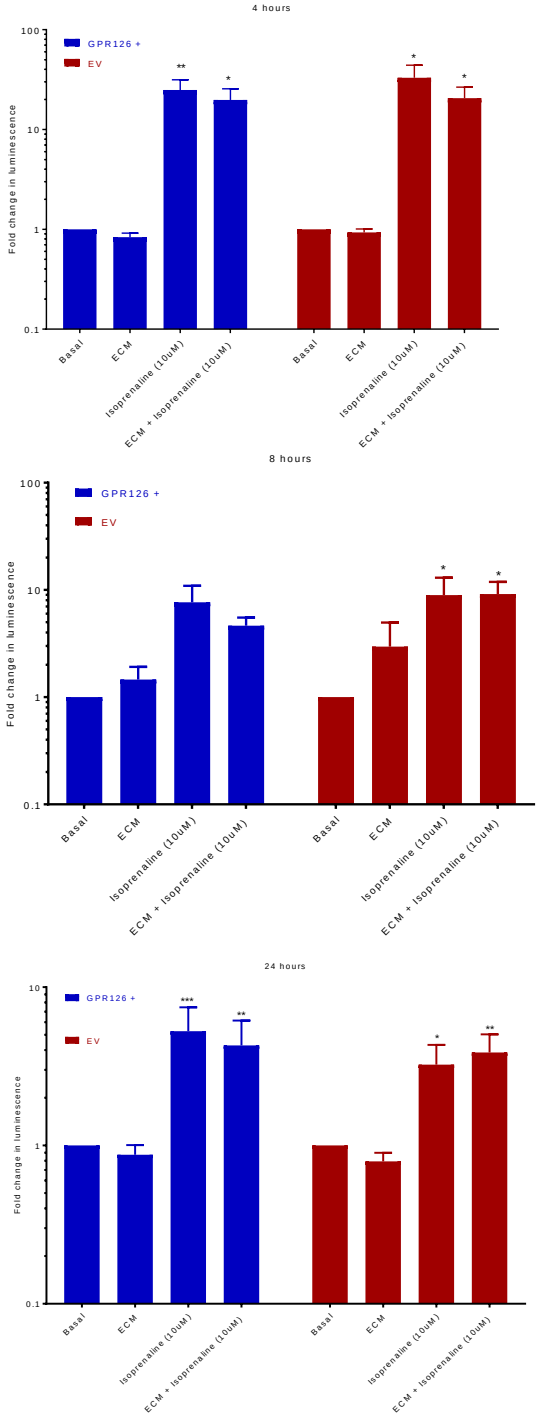


Figure 4-17. Luciferase activity from transfected cells plated onto de-cellularised HASM ECM. Cells were plated on top of ECM and then left to incubate for 4, 8 and 24 hours. Cells plated on to ECM showed no increase in reporter activity over untreated cells. Statistical analysis was performed on raw data using a 2-way ANOVA with Tukey's multiple comparisons test.

Cells transfected cells with *GPR126* or EV showed no statistically significant response to being seeded onto the ECM at 4, 8 or 24 hours while there was an isoprenaline response at every time point, this was most prominent after 4 hours in keeping with previous results.

4.3.5.5 The Tethered Agonist, Stachel, Activates GPR126 Signalling

The stachel peptide is a short amino acid sequence which is a part of the receptor itself, acting as a tethered agonist. pLV-GPR126^{Ser123}, pLV-GPR126^{Gly123}, pLV-EV and non-transfected cells were treated with the stachel peptide at a range of concentrations (100µM-1mM) for 5 hours (Figure 4-18).

Figure 4-18 shows luciferase reporter activity as fold change over untreated cells. There is a dose-dependent increase in reporter activity in the two cell lines expressing GPR126 (max response = 9.42±1.253 fold, $p < 0.0001$ for 750µM dose in Gly123 cells) while there is no response in the empty vector cells or the non-transfected cells. There is no response to stachel at 100µM and reporter activity is seen at 250µM (not significant), 500µM, 750µM and 1mM in both GPR126 cell lines. There is no significant difference in reporter activity between the two GRP126 variant cell lines at any dosage. At 750µM stachel concentration the mean response for the Ser123 variant was 7.240±1.165 compared to 9.42±1.253 in the Gly123 variant ($p=0.0546$). At the 1mM concentration mean values were 8.92±1.580 for Ser123 and 6.88±1.435 for Gly123 ($p=0.0812$).

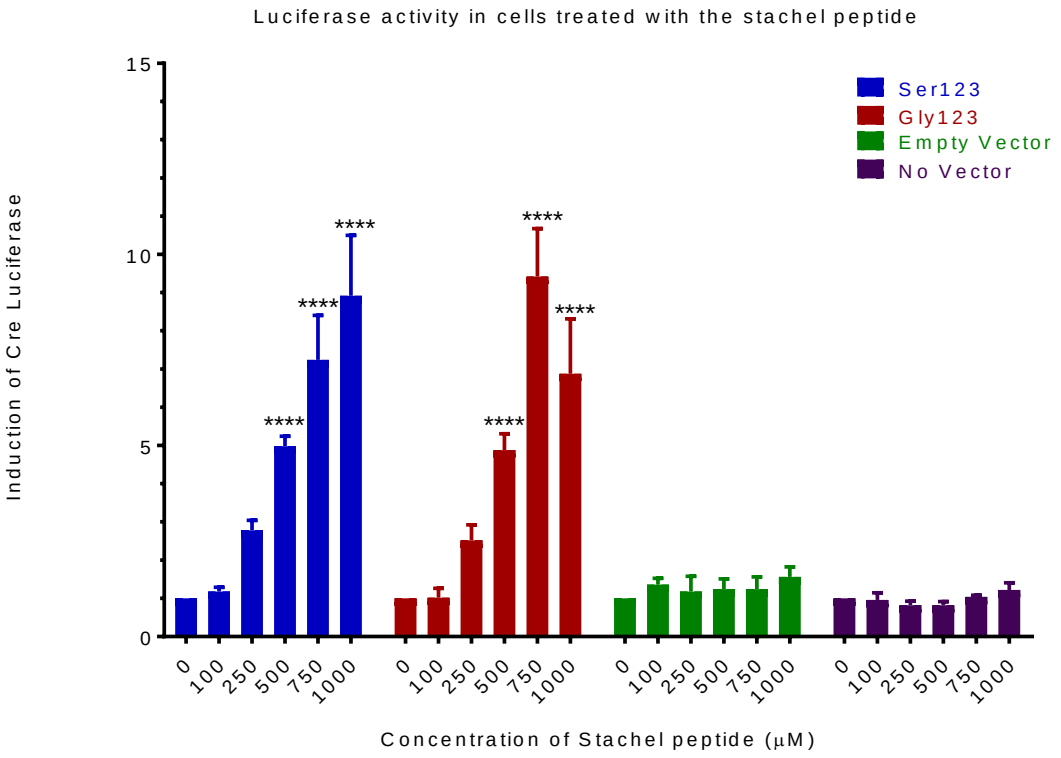


Figure 4-18 Response of transfected cells to increasing concentrations of the stachel peptide at 5 hours of incubation (n=5). Cells were treated with increasing doses of the stachel peptide before the assay was stopped. Significant responses in luciferase activity is seen for doses 500μM and higher for both GPR126 cell lines. There is no response to stachel peptide in the empty vector cells or the non-transfected cells. A 2-way ANOVA with Dunnett's multiple comparisons test was used to test luminescence changes on raw data by stachel activation against untreated cells.

4.3.5.6 GPR126 response to stachel in the presence of H89 – a PKA inhibitor

To test if it would be possible to inhibit stachel mediated GPR126 signalling by inhibiting PKA, cells were pre-treated with 10μM H89. Figure 4-19 shows the response of pLV-GPR126^{Ser123} which have been pre-treated with H89 and then forskolin, isoprenaline or stachel for 4 hours.

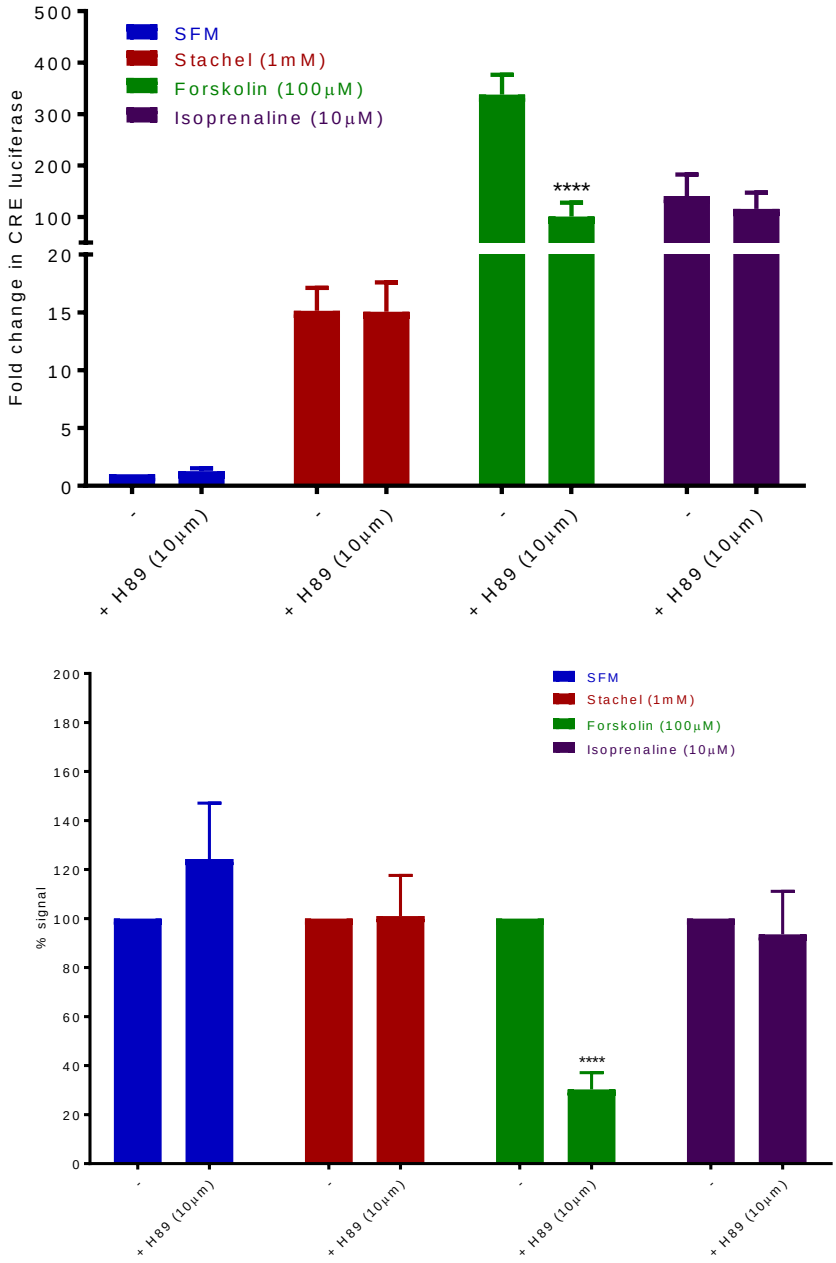


Figure 4-19 Response of pLV-GPR126^{Ser123} cells to H89 (10µM) pre-incubation prior to treatment with various agonists. Cells were pre-incubated for 2 hours before being treated with stachel (1mM), isoprenaline (10µM) or forskolin (100µM) Left showing fold change in response to serum free media only. Right shows % of the signal compared to cells not pre-treated with H89. For each condition, raw data from cells pre-treated with H89 were compared to cells not pre-treated with a 2-way ANOVA with Sidak's multiple comparisons test.

There is a clear inhibition of the forskolin luciferase signal, indicating inhibition of PKA, with the pre-incubation with H89 reducing the signal to ~30% (mean without H89 = 338±38.5, mean with H89 pre-treatment = 100.7±27.1) of the response ($p<0.0001$). However, there is no inhibition of the signal from stachel ($p>0.9999$) or isoprenaline ($p=0.7731$).

4.3.5.7 cAMP accumulation in CRE-luc2P HEK293 cells overexpressing GPR126

Due to the differences between direct cAMP measurement and a luciferase reporter, I expected to see differences in response time and signal magnitude i.e. I expected the luciferase assays to show a slower response but be more sensitive and therefore show a greater magnitude of change over a longer period of time. To address this and to attempt to assess how comparable direct cAMP measurement and luciferase assays are I performed a cAMP assay using column chromatography on theGPR126 recombinant HEK293 cells treated with stachel peptide.

To complement the signalling data obtained from the CRE-luciferase assays, I also measured cAMP directly using column chromatography (Figure 4-20). GPR126^{Ser123} cells were treated with stachel peptide (1mM) for 10, 60, 120 and 240 minutes. There was an increase in cAMP seen at all time-points with statistical significance seen at 1 hour (mean fold increase 3.1±0.872, $p=0.0204$) and 4 hours incubation (mean fold increase = 2.514±0.927, $p=0.0055$). The degree of cAMP increase is lesser than seen using the CRE-luciferase reporter, in which I saw a 10-15 fold increase in response to stachel activation.

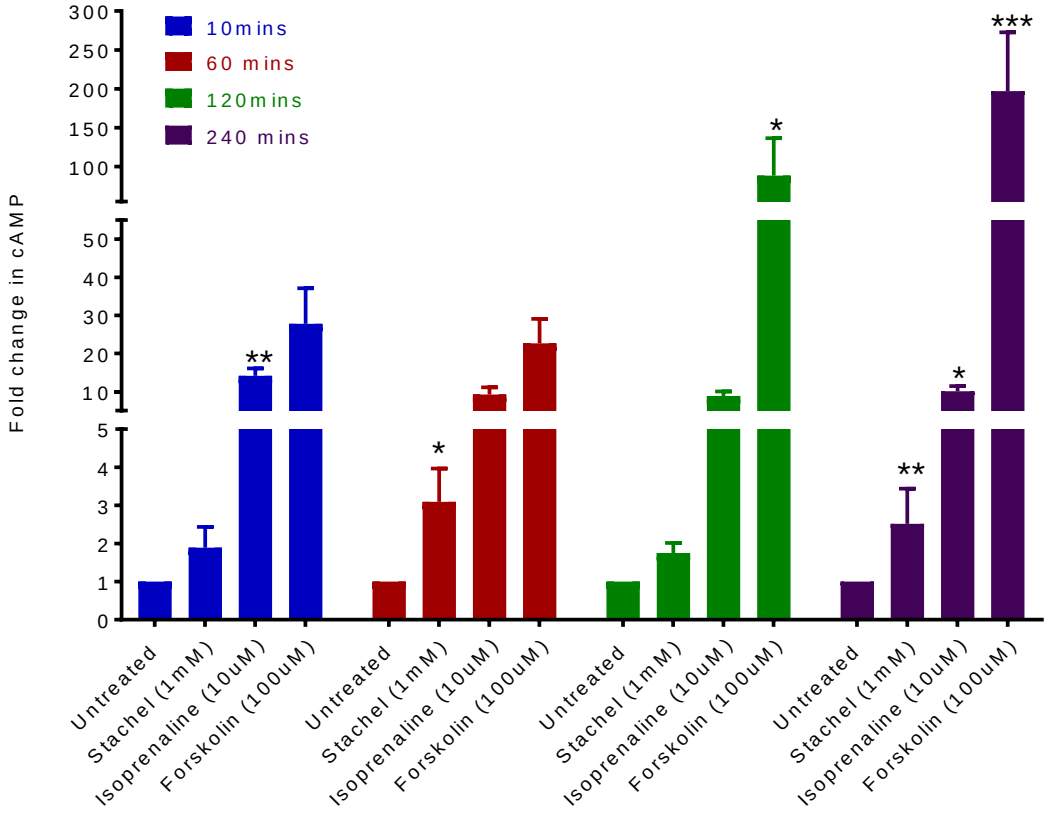


Figure 4-20. cAMP measurements in GPR126^{Ser123} cells in response to stachel peptide treatment for 10, 60, 120 and 240 minutes with pre-treatment with IBMX. Treated cells were compared to untreated cells at the same time point using an ANOVA with Dunnett's multiple comparisons test.

4.4 DISCUSSION

In this chapter I sought to study GPR126 signalling activity using a range of potential ligands in a recombinant cell system expressing a CRE-luciferase reporter and GPR126 variants. My main aims were to confirm the activity of previously reported GPR126 ligands and investigate the function effects of the Ser123Gly substitution using two methods of studying cAMP signalling: via a CRE-luciferase reporter and by direct measurement of cAMP using column chromatography. To achieve these aims I produced a mixed population of cells stably overexpressing the two GPR126 variants Ser123 and Gly123 and an empty vector construct and treated the cells with collagen type IV, laminin-211, two peptides – one derived from collagen type IV and one from the prion peptide – and the stachel peptide. In summary, these cells responded to activation with the stachel peptide, producing a ~15 fold increase in luciferase but previously reported ligands did not initiate a significant response. This finding was replicated, albeit with a smaller response, by direct measurement of cAMP.

In the luciferase assays there was a 10-15 fold increase in luciferase activity upon treatment of the cells with the stachel peptide (1mM). This is in keeping with data from Petersen et al (2015) who showed a ~30 fold increase in luciferase under similar experimental conditions.[97] In addition, direct measurement of cAMP showed statistically significant increases after 1 hour (mean fold change = 3.1±0.8) and 4 hours (mean fold change = 2.5±0.9) of treatment with the stachel peptide. In comparison, Liebscher et al (2014) reported a 10-15 fold increase in cAMP in COS-7 cells expressing GPR126 when treated with 1mM stachel peptide. [96] The direct cAMP measurement is ~1/4 the signal of the CRE-luciferase suggesting that the luciferase assays are more sensitive than direct cAMP measurement as expected.

Together, these data from the stachel activation of GPR126 show that the recombinant CRE-LUC2P HEK293 cells have a functional GPR126 receptor capable of signalling via cAMP mediated pathways. This data provides

confidence in the system and demonstrates that modest increases in intracellular cAMP can be detected via the CRE-luciferase reporter.

As the stachel peptide was added in a high dose (1mM) compared to the standard agonists such as isoprenaline, it would have been prudent to test for potential toxic effects of such a high dosage. Cells were examined by microscopy and no signs of cell death were observed at the point of cell lysis using the Bright-Glo™ assay reagent. However, because there was no response seen in the empty vector cells in response to stachel activation, we can be confident that the cAMP driven gene response is GPR126 specific and not as a result of cell toxicity.

Aside from the stachel peptide, the other reported ligands tested did not result in an increase in CRE-luciferase in the recombinant CRE-LUC2P HEK293 cells. Treatment of cells with Collagen type IV did not result in a CRE-luciferase response. Reportedly, treatment of CRE-LUC2P HEK293 cells overexpressing zebrafish GPR126 with collagen type IV results in a ~3 fold increase in cAMP.[98] One obvious explanation for a lack of response in my recombinant cells could be the differences between the GPR126 sequence in zebrafish and human. The CUB and PTX domains in the N-terminal fragment of GPR126 is thought to be the important region for ligand binding. Figure 4-21 shows the Clustal Omega alignment of the human CUB/PTX domain and the Zebrafish amino acid sequence.[177] The human and zebrafish variants share only 47% sequence identity, a discrepancy which may explain the ineffectiveness of collagen type IV in our assays.

```
CUB/PTX         ------------------------------------CRVVLSNPSGTFTSPCYPNDY     21
Zebrafish       ISFISGRWWRWKFQNTLAVFLLLICLSTSVAQSCQSSTSCNVVLTDSQGSFTSPCYPNDY     60
                                                    *.***:: .*:**********

CUB/PTX         PNSQACMWTLRAPTGYIIQITFNDFDIEEAPNCIYDSLSLDNGESQTKFCGATAKGLSFN     81
Zebrafish       PPSQSCNWTIQAPAGFIVQITFLDFELEEAQGCIYDRVVVKTGTSDAKFCGLTANGLTLN     120
                * **:* **::**:*:*:**** **::*** .**** : :..* *::**** **:**::*

CUB/PTX         SSANEMHVSFSSDFSIQKKGFNASYIRVAVSLRNQKVILPQTSDAYQVSVAKSISIPELS     141
Zebrafish       STGNVMEVFFNSDFSVQKKGFHISYKQVAVTLRNQKVTMPKSSKT-ILRVSNSISIPVLT     179
                *:.* *.* *.****:*****: ** :***:****** :*::*.:  : *::***** *:

CUB/PTX         AFTLCFEATKVGHEDSDWTAFSYSNASFTQLLSFGKAKSGYFLSISDSKCLLNNALPVKE     201
Zebrafish       AFTVCFEIARTAQKAT-ETIFTLSDAAGTSILAFEKTSNGMELFIGASYCSVDNLLTSSD     238
                ***:*** ::..:: :  * *: *:*: *.:*:* *:..*  * *. * * ::* *  .:

CUB/PTX         KEDIFAESFEQLCLVWNNSLGSIGVNFKRNYETVPCDSTISKVIPGNGKLLLGSNQ----     257
Zebrafish       ----ITATMKPLCLTWTKSSGLIGVYFGGHYFSSICSASQIYTLQSGGLLQIAGKGSSSV     294
                    :: ::: ***.*.:* * *** *  :* :  *.::   .: ..* * :..:

CUB/PTX         -NEIVSLKGDIYNFRLWNFTMNAKILSNLSCNVKGNVVDWQNDFWNIPNLALKAESNLSC     316
Zebrafish       SVDDQNLDGFIYNFRLWDHAMLSSELSALTCDTVGNVVDWDHSYWTIPGSSTQTDSTLSC     354
                  :  .*.* *******:.:* :. ** *:*:. ******::.:*.**. : :::*.***
```

Figure 4-21 Sequence similarity between the human CUB/ PTX domains and the Zebrafish sequence. Sequences were aligned with Clustal Omega. [177]

Furthermore, Petersen et al (2015) (data presented in the supplementary material) treated COS-7 cells overexpressing human GPR126 with collagen type IV and failed to measure a response by measuring cAMP directly.[97] They did however see a response in luciferase reporter activity when the cells were transfected with a CREB-luciferase reporter.[97] This response was only approximately 2 fold over baseline and this is not reported to have reached statistical significance. This is very much in line with the data presented in this chapter. While collagen type IV may be an important ligand for GPR126, these data indicate that it does not initiate a measureable or robust response from the human receptor. This could be due to transcript differences effecting ligand binding and therefore signalling or there are additional factors required for GPR126 activation as well as collagen type IV binding.

Similarly, laminin-211 treatment resulted in no difference in luciferase activity. Petersen et al (2015) suggest that GPR126 activity and signalling may involve other elements, e.g. mechanical forces that I did not consider in our assays.[97] Perhaps this extra factor of mechanical stimulation may also be important when collagen type IV is the ligand. In addition to these two ligands, I tested two peptides, 'P28' derived from the prion peptide and col4 – a fragment of the

collagen type IV peptide. These peptides have been shown to previously interact with the N-terminal fragment of GPR126 and cause an approximate 2-fold increase in intracellular cAMP. [171] I tested these peptides in our recombinant cells overexpressing GPR126 and failed to see an increase in reporter activity.

One key difference in our experimental design for most of these experiments was that the previous studies mentioned mostly measured cAMP accumulation whereas I measured luciferase activity for the majority of ligands. While both methods are appropriate to study G protein signalling activity, they are methods which measure different outcomes. By measuring cAMP, you are directly measuring the accumulation of the second messenger itself within the cell and therefore you would expect a faster response time from receptor activation. You would also expect this signal to be transient unless a PDE inhibitor such as IBMX was added to prevent cAMP degradation. On the other hand, the luciferase reporter relies on cAMP accumulation after receptor activation followed by a signalling cascade which results in translation of the luciferase gene and expression of the protein. Therefore, by measuring luciferase I am measuring the downstream effects of receptor activation rather than the direct response to receptor activation. The luciferase reporter I used (*luc2P)* has a rapid response time and I saw a peak signal at ~4 hours but clearly the kinetics of the two methods is different. On the basis of this and the data comparing direct cAMP measurement and a CRE-luciferase reporter, I would expect the reporter to be the more sensitive method and therefore it seems like the different measurable outcome is not the reason I didn't see a response from these ligands.

It is difficult to explain the reason I did not see GPR126 signalling activity in response to many of the ligands reported in the literature. I showed that GPR126 mRNA is overexpressed in those cells transfected with the GPR126 constructs. However, a limitation of this chapter is that I failed to convincingly measure GPR126 protein due to technical difficulties with protein detection methods. Therefore perhaps there is not an adequate concentration of receptor on the cell surface to be activated by some of these potential ligands.

Although, activation of GPR126 by the stachel peptide shows that the GPR126 protein is being produced, localising the membrane and is physiologically active and our data is not too dissimilar from that in the literature. In addition, I know that this response is GPR126 specific as there was no response seen in empty vector or in non-transfected controls.

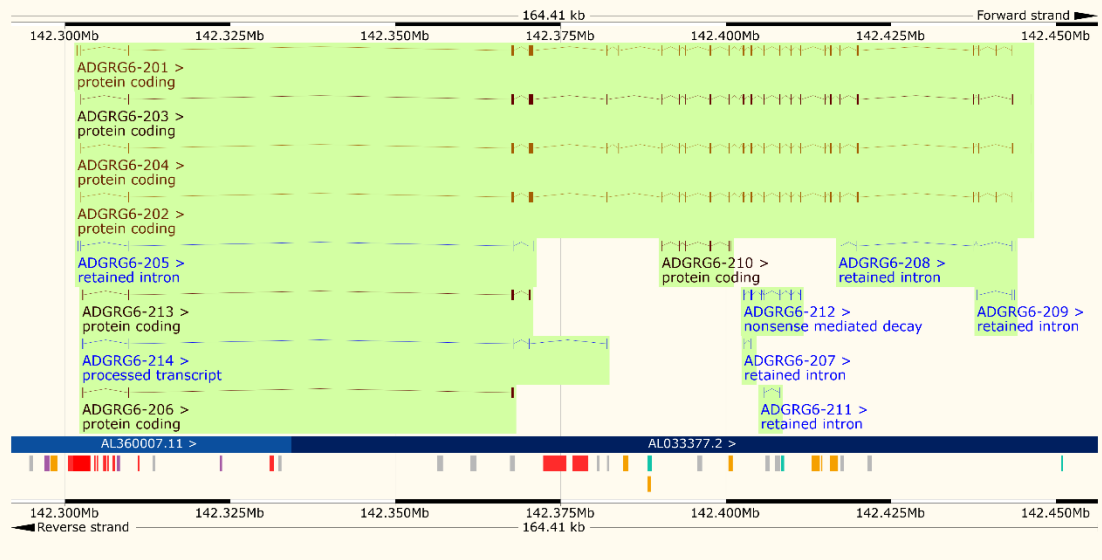


Figure 4-22 GPR126 transcripts from Ensembl.[76] The transcript coding for the alpha 2 transcript used in this chapter is denoted by ADGRG6-202 which lacks exon 6.

One difference may have been the GPR126 transcript used in our constructs and the constructs used by other research groups. As mentioned, I decided upon using the transcript which codes for the alpha 2 isoform of the receptor as this was the transcript most abundant in the lung in GTEx RNA-seq datasets. This seems like a sensible choice as our primary aim is to identify the role of GPR126 in the lung and to identify any mechanism by which GPR126 may affect lung function. This transcript lacks exon 6 of the 'full-length' transcript (Figure 4-23 and Figure 4-22). This exon is 84 nucleotides in length and codes for a region of the receptor in the N—terminal domain. The region coded for by exon 6 begins at amino acid 380, which is downstream of the important binding domains, the CUB and PTX domains (amino acids 41-356) and is upstream of the GPS cleavage site and the stachel sequence. It is currently unclear whether this region contains sequences important for ligand

interaction – no regulatory or binding domains are listed on UniProt – but it is an important consideration as possible explanation for our data.

The function of exon 6 in GPR126 is unclear. BLAST results for this sequence only returns hits for GPR126 isoforms. It is plausible that this exon may be important for ligand binding as it lies close to the CUB/PTX domains that are thought to be important for collagen type IV and laminin-211 binding or for signal transduction.

```
b1      mmfrsdrmwschwkwkpspllflfalyimcvphsvwgcancrvvlsnpsgtftspcypnd      60
b2      mmfrsdrmwschwkwkpspllflfalyimcvphsvwgcancrvvlsnpsgtftspcypnd      60
a1      mmfrsdrmwschwkwkpspllflfalyimcvphsvwgcancrvvlsnpsgtftspcypnd      60
a2      mmfrsdrmwschwkwkpspllflfalyimcvphsvwgcancrvvlsnpsgtftspcypnd      60
        ************************************************************

b1      ypnsqacmwtlraptgyiiqitfndfdieeapnciydslsldngesqtkfcgatakglsf      120
b2      ypnsqacmwtlraptgyiiqitfndfdieeapnciydslsldngesqtkfcgatakglsf      120
a1      ypnsqacmwtlraptgyiiqitfndfdieeapnciydslsldngesqtkfcgatakglsf      120
a2      ypnsqacmwtlraptgyiiqitfndfdieeapnciydslsldngesqtkfcgatakglsf      120
        ************************************************************

b1      nssanemhvsfssdfsiqkkgfnasyirvavslrnqkvilpqtsdayqvsvaksisipel      180
b2      nssanemhvsfssdfsiqkkgfnasyirvavslrnqkvilpqtsdayqvsvaksisipel      180
a1      nssanemhvsfssdfsiqkkgfnasyirvavslrnqkvilpqtsdayqvsvaksisipel      180
a2      nssanemhvsfssdfsiqkkgfnasyirvavslrnqkvilpqtsdayqvsvaksisipel      180
        ************************************************************

b1      saftlcfeatkvghedsdwtafsysnasftqllsfgkaksgyflsisdskcllnnalpvk      240
b2      saftlcfeatkvghedsdwtafsysnasftqllsfgkaksgyflsisdskcllnnalpvk      240
a1      saftlcfeatkvghedsdwtafsysnasftqllsfgkaksgyflsisdskcllnnalpvk      240
a2      saftlcfeatkvghedsdwtafsysnasftqllsfgkaksgyflsisdskcllnnalpvk      240
        ************************************************************

b1      ekedifaesfeqlclvwnnslgsigvnfkrnyetvpcdstiskvipgngkllllgsnqnei      300
b2      ekedifaesfeqlclvwnnslgsigvnfkrnyetvpcdstiskvipgngkllllgsnqnei      300
a1      ekedifaesfeqlclvwnnslgsigvnfkrnyetvpcdstiskvipgngkllllgsnqnei      300
a2      ekedifaesfeqlclvwnnslgsigvnfkrnyetvpcdstiskvipgngkllllgsnqnei      300
        ************************************************************

b1      vslkgdiynfrlwnftmnakilsnlscnvkgnvvdwqndfwnipnlalkaesnlscgsyl      360
b2      vslkgdiynfrlwnftmnakilsnlscnvkgnvvdwqndfwnipnlalkaesnlscgsyl      360
a1      vslkgdiynfrlwnftmnakilsnlscnvkgnvvdwqndfwnipnlalkaesnlscgsyl      360
a2      vslkgdiynfrlwnftmnakilsnlscnvkgnvvdwqndfwnipnlalkaesnlscgsyl      360
        ************************************************************

b1      iplpaaelascadlgtlcqatvnspsttpptvttnmpvtnridkqrndgiiyrisvviqn      420
b2      iplpaaelascadlgtlcq-------------------------dgiiyrisvviqn      392
a1      iplpaaelascadlgtlcqatvnspsttpptvttnmpvtnridkqrndgiiyrisvviqn      420
a2      iplpaaelascadlgtlcq-------------------------dgiiyrisvviqn      392
        ******************                          *************

b1      ilrhpevkvqskvaewlnstfqnwnytvyvvnisfhlsagedkikvkrsledeprlvlwa      480
b2      ilrhpevkvqskvaewlnstfqnwnytvyvvnisfhlsagedkikvkrsledeprlvlwa      452
a1      ilrhpevkvqskvaewlnstfqnwnytvyvvnisfhlsagedkikvkrsledeprlvlwa      480
a2      ilrhpevkvqskvaewlnstfqnwnytvyvvnisfhlsagedkikvkrsledeprlvlwa      452
        ************************************************************

b1      llvynatnntnlegkiiqqkllknneslde glrlhtvnvrqlghclameepkgyywpsiq      540
b2      llvynatnntnlegkiiqqkllknneslde glrlhtvnvrqlghclameepkgyywpsiq      512
a1      llvynatnntnlegkiiqqkllknneslde glrlhtvnvrqlghclameepkgyywpsiq      540
a2      llvynatnntnlegkiiqqkllknneslde glrlhtvnvrqlghclameepkgyywpsiq      512
        ************************************************************

b1      pseyvlpcpdkpgfsasricfynatnplvtywgpvdisnclkeanevanqilnltadgqn      600
b2      pseyvlpcpdkpgfsasricfynatnplvtywgpvdisnclkeanevanqilnltadgqn      572
a1      pseyvlpcpdkpgfsasricfynatnplvtywgpvdisnclkeanevanqilnltadgqn      600
a2      pseyvlpcpdkpgfsasricfynatnplvtywgpvdisnclkeanevanqilnltadgqn      572
        ************************************************************
```

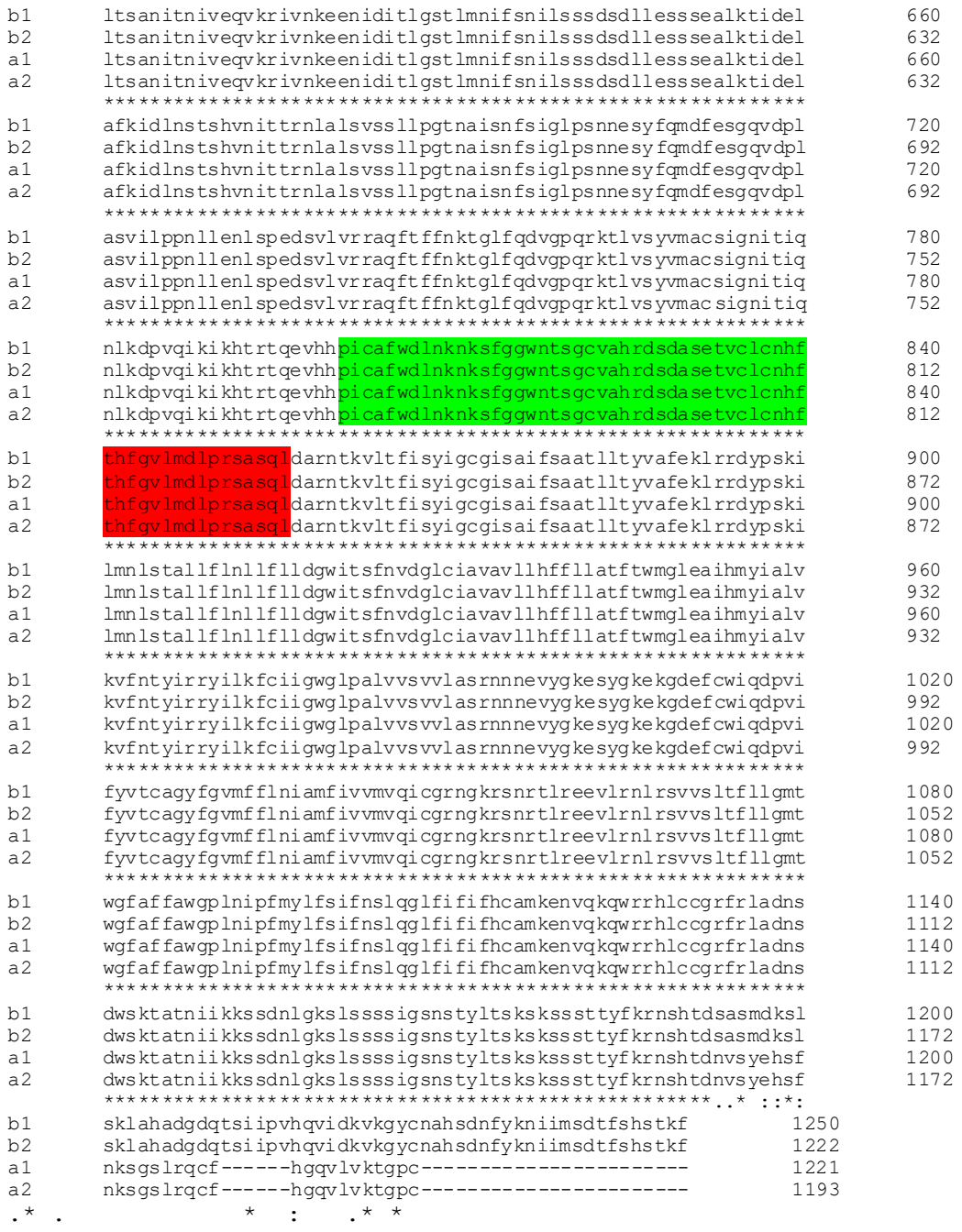


Figure 4-23 Clustal Omega[177] alignments of the four full length protein coding transcripts of GPR126. The Ser123Gly variation is highlighted in yellow. CUB domain is purple and PTX is cyan and the stachel peptide is in red. Asterisks highlight were amino acids are the same between transcripts and dashes represent missing amino acids. The alpha 2 and beta 2 transcripts lack exon 6 compared to the other two transcripts. The GPS motif is highlighted in green.

While I did identify a ligand with the potential to activate GRP126 specific signalling, I did not see any significant signalling differences between the two GPR126 variants of interest. One explanation for this is that the Ser123Gly substitution lies in the CUB domain and the stachel binds much further downstream on the receptor where the endogenous stachel sequence would bind to activate receptor signalling. Therefore it is plausible that the Ser123Gly substitution wouldn't affect signalling activity when the receptor is activated with the stachel sequence. Whether the substitution in the important binding domain has impact on a traditional ligand binding method of activation is something I failed to assess due to the difficulty in finding a ligand to activate GPR126.

One aspect that could have been accounted for more effectively was that of transfection efficiency. The Ser123 variant was expressed at higher levels than the Gly123 variant (Figure 4-12) which could be explained in two ways. Firstly, this could be due to transfection efficiency i.e. the Ser123 transfection was more efficient hence more expression was seen relative to the empty vector control. Alternatively, this could be due to SNP effects on mRNA expression, i.e. eQTL effects.

While there was no significant difference in signalling observed between the two variants, the signalling activity could be effected by the level of gene expression. Therefore it is difficult to compare the two signalling responses particularly given that the Ser123 plasmid was more highly expressed. To account for this a dual reporter (e.g. luciferase) system would have been ideal. A dual reporter refers to the simultaneous measurement of two individual reporter enzymes. In the case of a dual luciferase assay, the system relies on two variants of the luciferase gene, *Renilla* and firefly luciferase.

In these experiments, there is the experimental reading which alters in response to specific experimental conditions (e.g. CRE-luciferase) and the co-transfected control reporter (*Renilla* luciferase) acts as an internal control that serves as the baseline response. This normalisation minimises variability caused by differences in transfection efficiency.

From chapter 3, I know that GPR126 is expressed in the human lung and in airway structural cells, HASMs and HBECs. I used RNA-seq data (Chapter 3) which shows which donors of HASMs and HBECs express high levels of GPR126 and I can select these donors for further analysis. The next step was to use the stachel peptide as a tool to activate GPR126 signalling in primary human airway cells and assess the response of exogenous GPR126 expression by RNA-seq.

5 GPR126 signalling in primary human airway cells

5.1 INTRODUCTION

In the previous chapters I described *GPR126* mRNA and protein expression in the human lung and in Human Airway Smooth Muscle Cells (HASMs) and Human Bronchial Epithelial Cells (HBECs). In addition, immunohistochemistry showed GPR126 protein expression in human lung with particularly strong staining seen in smooth muscle surrounding the airways. However, there was considerable donor variation with some donors showing no staining. Moreover, Protein Atlas shows no GPR126 staining in the lung (www.proteinatlas.org).[156]

In chapter 4, I examined GPR126 function in a recombinant cell line and identified that the stachel peptide, a tethered agonist, activated GPR126 activity with no significant difference in signalling between the Ser123 and Gly123 variants. In this chapter, I moved away from a recombinant model of GPR126 expression in a cell line and focused on endogenous GPR126 expression and signalling in HASMs and HBECs in order to try and gain insight into the role of GPR126 in these cells and provide greater insight into the role of GPR126 in the lung.

HASM cells are, for the purposes of this thesis, grown from tissue explants from tissue donors undergoing lung transplantation or resection. Airway smooth muscle (ASM) surrounds the lumen of the bronchi accounting for 25% of cellular mass in the upper airways and 4-5% in the lower airways. Smooth muscle plays vital roles in the regulation of bronchomotor tone and in the control of the airway calibre.[178, 179] ASM cells are crucial for the contractile function of the airways. Alteration of the function and/or the mass of ASM can lead to airway inflammation, bronchial hyper-responsiveness and airway remodelling which are characteristics in both COPD and asthma. [180] ASM plays a key role in airway remodelling, which is as a result of processes such as hypertrophy, hyperplasia and cell migration. It is also thought that ASM plays a role in the regulation of airway inflammation as these cells are capable of producing inflammatory cytokines and chemokines. [181] ASM controls bronchoconstriction with contraction of the smooth muscle resulting in a

reduction of the airway diameter. This function of ASM causes airflow obstruction in COPD and asthma and is the reason why the most widely used medications for asthma and COPD are bronchodilators. Bronchodilators, such as salbutamol, a β_2 agonist, act on airway smooth muscle causing it to relax and alleviate bronchoconstriction.

The airway epithelium is the frontline defence barrier between the host and the environment and therefore is important as the first structural defence against environmental insults such as inhaled cigarette smoke, the primary risk factor for development of COPD. [182, 183] This barrier is comprised of apical tight junctions and basolateral adherens junctions which regulate transport of ions and molecules and restrict permeability to inhaled pathogens and environmental insults.[184] Epithelial cells have key roles in pathogen recognition and mucociliary clearance. The airway epithelium expresses a broad range of pattern recognition receptors (PRRs) which recognise pathogen-associated molecular patterns (PAMPs) from microbial threats and danger-associated molecular patterns (DAMPs) released from cells upon tissue damage, cell death or cell stress. [185] PRR activation leads to an immune response from the host including the release of cytokines and chemokines that attract and activate innate and adaptive immune cells.[185] The activation of the epithelium is an important step for the activation of dendritic cells, cells which help coordinate the immune response. [186]

Epithelial dysfunction may be relevant in COPD where repeated injury is caused by cigarette smoke exposure and subsequent inflammation may compromise barrier function. Damage or destruction to the epithelium exposes sub-epithelial components to potential pathogens and hazards and alters the normal function of the epithelial cells themselves, potentially contributing to COPD development.[183] Epithelium disruption may lead to alteration of signalling pathways involved in repair and inflammatory responses. Cultured bronchial epithelial cells from ex-smoking COPD patients display abnormalities with reduced capacity to from epithelial junctions during air liquid interface (ALI) differentiation *in vitro*. [187] In addition, cigarette smoking induces down-

regulation of expression of junction proteins in the airway epithelium and increased mucosal permeability *in vivo.* [187]

Cell lines can be used to overexpress a gene or protein of interest in a recombinant system. While this can provide an idea of gene/protein function, it is in a largely artificial system and using primary human cells offers an opportunity to study a gene or protein function in a more physiologically relevant environment *in vitro.*

As discussed in the previous chapters, cell lines are cells derived from a multicellular organism that have been engineered to avoid normal cellular senescence and proliferate endlessly allowing prolonged use of the cells *in vitro*. For example, HEK293 cells used in the previous chapter are derived from a human embryonic kidney from a healthy aborted foetus. These cells were transformed with sheared adenovirus 5 DNA which resulted in the incorporation of ~4.5kb of viral genome into chromosome 19 of the human genome in the HEK cells. [118] Cell lines are important and useful for functional genetics due to their proliferative properties which make them easy to culture and easy to transfect with plasmids. However, with these benefits come limitations. The properties of cell lines which make them ideal for functional genetics and cell biology mean that they are not representative of the *in vivo* environment from which they originate from. Cell lines are often genetically and phenotypically very different from their tissue of origin due to genetic engineering to immortalise them and due to years of culturing. The over-expression of a gene in such an artificial system – i.e. a cell which has an unusual karyotype and properties and expression of the gene at artificially high levels – limits the degree of biological insight that can be gained.

Primary cells offer an opportunity to study a gene of interest within the desired context as they are usually obtained from patients directly. For example, as described above HBECs can be obtained from bronchial brushings and HASMs can be grown from tissue explants. Because of this they are much more representative of the cell in the organ of interest compared to cell lines. However there are more challenges associated with culturing primary cell cultures. Donor variability can result in inconsistent data due to natural

phenotypic and genotypic difference between individuals. These cells also undergo phenotypic and genotypic drift in vitro the longer they are kept in culture. The major challenge is that these cells have a finite lifespan and can be difficult to culture or manipulate making them unsuitable for large-scale experiments such as drug screening.

While the role of GPR126 in these airway structural cells is unknown, there is literature to suggest importance of other adhesion-GPCRs in airway biology. GPR116 (ADGRF5) is an orphan adhesion-GPCR expressed in alveolar type I and II epithelial cells and pulmonary endothelial cells.[188] GPR116 is an essential regulator of alveolar surfactant homeostasis by regulating surfactant secretion and reuptake in alveolar type II (AT2) cells by $G_{q/11}$ signalling pathways.[188] Alveolar surfactant, secreted by alveolar type II epithelial cells, is a lipid-protein complex that reduces surface tension at the air-liquid interface which prevents atelectasis. Alterations in the surfactant concentration or its function can impair ventilation and cause injury. Additionally, deletion of GPR116 results in the development of airway inflammation and this is possibly via its role in the regulation of genes encoding CCL2 in lung endothelial cells.[189]

5.1.1 Aims

The aim of this chapter was to determine and characterise the capacity of primary human airway smooth muscle and bronchial epithelial cells to signal via GPR126. An additional aim was to determine the donor dependency of this response given the variable levels of GPR126 mRNA and protein expression observed among different donors.

5.2 METHODS

5.2.1 Culture of Human Airway Smooth Muscle (HASM) Cells for ^{3}H-cAMP Accumulation Assays and CRE-luciferase Transfections

A full and detailed protocol on the cell culture methods used in this thesis can be found in the general methods chapter. HASMs were cultured following the basic protocols outlined in the general methods. All HASMs used in this chapter were at passage 3 or 4. Donors D6, D63, D0304 and D1515 were used for optimisation experiments. For initial experiments with the stachel peptide D6 and D1515 were used and for the final stachel experiments D63 was used. For CRE-luciferase transfections and experiments, P3 D63 cells were used. Donor information is given in the general methods section.

5.2.2 Culture of Human Bronchial Epithelial Cells (HBEC) for ^{3}H-cAMP Accumulation assays.

HBECs were cultured as described in the general methods. All experiments were performed on donor 7 passage 3 cells. Full details of HBEC culture conditions and donor information can be found in section 2.1.2.

5.2.3 cAMP assays using column chromatography

Passage 3 HBECs and HASMs in 48 well plates were treated with stachel peptide (1mM) and isoprenaline (10µM) for 1 hour with a 15 minute pre-incubation with IBMX. Intracellular cAMP accumulation was measured using column chromatography as explained in section 2.1.8.[124] In brief, cells were grown in 48 well plates until ~80% confluency. Cells were loaded with ^{3}H adenine as previously described. IBMX solution was made up in serum free media to a final IBMX concentration of 100µM. Media containing IBMX was added to the cells which were left to incubate for 15 minutes in the cell incubator. Isoprenaline was added to the wells to a final concentration of 10µM and stachel peptide solution was added to a final concentration of 1mM. Plates were left to incubate for 1 hour. Assays were stopped by adding 50µl of concentrated hydrochloric acid. The columns were set up and run as described

in the general methods chapter and vials were counted on a Packard Scintillation counter.

5.2.4 Preparation and transfection of the CRE-luciferase plasmid (p6CRE/luc)

The p6CRE/luc construct was produced by Scott et al (1999) and glycerol stocks were stored at -80°C.[133] This construct contains six copies of the CRE consensus sequence **(**5'-TGACGTCA-3') which were cloned into vector pTK/luc. This vector contains the firefly luciferase gene downstream of the minimal Herpes Simplex Virus Thymidine Kinase promoter allowing measurement of cAMP driven gene expression.[133] Bacterial cell culture, plasmid preparation and plasmid quality-control were carried out as described in section 2.2.7.

Passage 2 HASMs (D63) were grown in T75 flasks until ~80% confluent before being split to individual wells of a 96 well plate at a density of $2x10^5$ cells/ml in 100µl culture media. HBECs (passage 2, D7) were seeded to the same density also in 96 well plates. Cells which were ~80% confluent were transfected with the p6CRE/luc plasmid using Fugene 6 (Promega) as covered in detail in section 2.1.6.

5.2.5 CRE-luciferase assays in HBECs and HASMs

I also examined cAMP signalling using a luciferase reporter based method, similar to the method used in chapter 4. This was primarily to complement the data from direct cAMP measurement and allow a more direct comparison between the recombinant cell line experiments and the primary cell experiments. Reporter based methods are measurements of gene transcription which is a different measurement than direct measurement of the second messenger itself.

Luciferase assays were carried out using the Bright-Glo™ Luciferase Assay System (Promega) as previously described. In the previous chapter I used a Promega cell line (GloResponse™ CRE-luc2P HEK293) which contains a destabilised luciferase gene (luc2P) which has a rapid response time and is

degraded within ~2 hours. The gene in the p6CRE/luc is a firefly luciferase gene which is not destabilised and therefore the signal accumulates over time requiring a longer assay time window.

Media was removed from cells in culture and replaced with SFM containing isoprenaline (10µM), forskolin (100µM) or stachel peptide (1mM). Plates were left for 24 hours in the incubator at 37°C with 5% CO_2. At the end of the assay, cells were lysed with the Bright-Glo™ Assay Reagent and read for luminescence on a Flexstation 2 for 1 second per well.

5.2.6 Statistical analyses

All data was analysed using GraphPad Prism 7.04. When comparing two groups of data, a student's t-test was used (parametric) or a Mann-Whitney test (non-parametric) was used based on the normality of the data. To analyse variance across 3 or more groups, a paired Friedman test was used, which is a non-parametric test used to determine differences in treatments across multiple groups. Dunn's multiple comparisons test is a post hoc comparisons test to test the differences between the individual means of each group. For a detailed description of this statistical analyses can be found in the general methods chapter.

5.3 RESULTS

5.3.1 Cyclic-AMP Accumulation in Primary Human Airway Cells Treated with the Stachel Peptide

Cyclic-AMP accumulation was measured in primary human airway cells treated with the GPR126 agonist, the stachel peptide, using column chromatography. Specific donors of HASMs (donor D63) and HBECs (donor D7) were chosen based on their levels of GPR126 expression as identified by RNA-sequencing (Chapter 3). These donors showed the highest levels of GPR126 expression with FPKM values of 27 and 19 respectively.

5.3.1.1 Optimisation of cAMP accumulation assays identifies that experiments can be scaled down without a loss in a measurable response

To reduce the quantity of stachel peptide used, which required high concentrations and specialist synthesis, I optimised the experiments, scaling-down the assays from a 24 well protocol to a 48 well protocol. This optimisation also included varying the volume of media in which the assays were performed. The β-adrenergic receptor agonist isoprenaline (10µM) was used to increase intracellular cAMP and was used as a positive control for the assay. Experiments were also performed with and without IBMX (100µM) pre-treatment for 15 minutes prior to isoprenaline treatment for 1 hour.

Figure 5-1 shows cAMP accumulation in 24 well plates with 1mL media and 0.5mL media as well as cAMP accumulation in 48 well plates with 0.5mL media and 0.3mL media. Without IBMX, the results between all experimental conditions was comparable with a 10-20 fold increase in cAMP seen across all conditions, although this was not statistically significant likely due to variation between experiments. With IBMX pre-treatment, all conditions showed a statistically significant increase in cAMP after isoprenaline treatment. However, the cells treated in 24 well plates and 1mL media showed a considerably higher fold change in cAMP compared to the other three conditions (182.772 ± 73.371 fold increase, n=5, p=0.003), which were all comparable showing a mean fold increase in cAMP ranging from 25.8 to 42.8.

This data suggests that using a 48 well plate and 0.3mL media volume for the assay will be appropriate and will allow me to minimise the quantity of peptide used in the following experiments. It is also clear the IBMX should be included to maximise the measureable response from the cells.

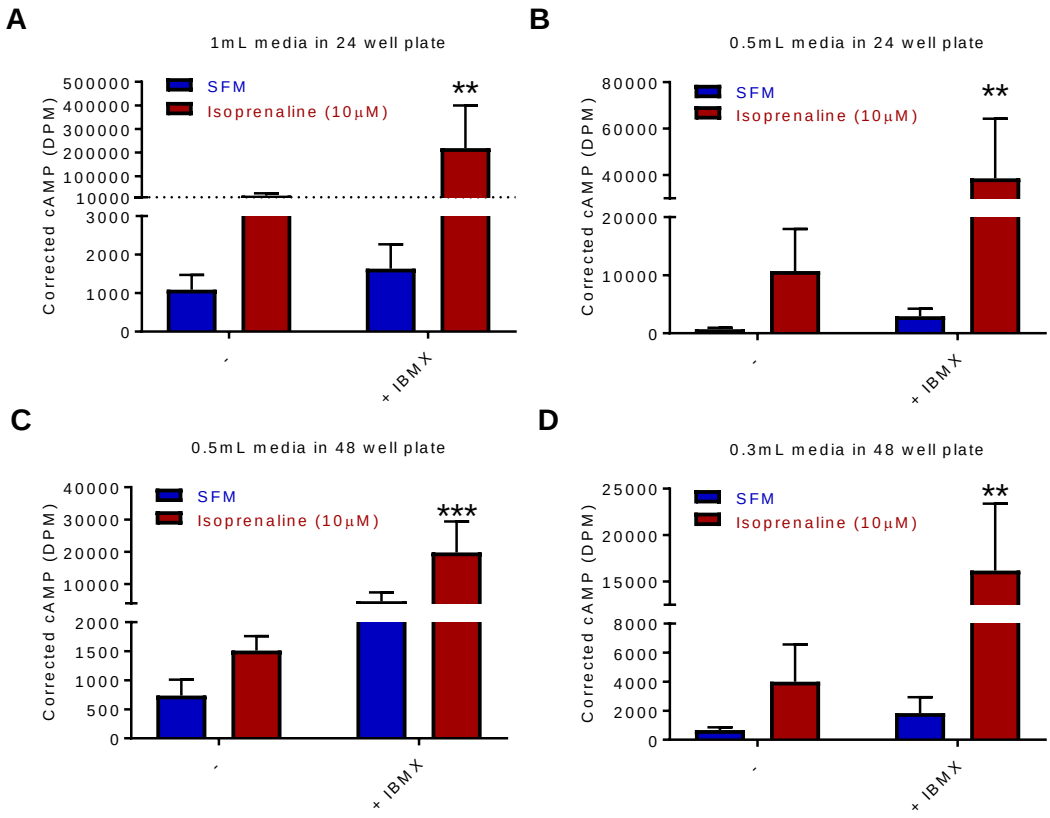


*Figure 5-1 cAMP accumulation in HASM cells treated with isoprenaline (10µM) in different volumes of media and different well sizes with and without IBMX (n=5). In all well sizes and media volumes, cells pre-treated with IBMX before isoprenaline treatment resulted in a statistically significant increase in cAMP. Data was analysed using a Friedman's test with Dunn's multiple comparisons test. All data was compared to SFM controls without IBMX pre-treatment. *p=0.0332, **p=0.0021, ***p=0.0002*

5.3.1.2 Stachel activation of GPR126 induces a ~20 fold increase in intracellular cAMP in HASMs

Passage 3 HASMs (D63) were cultured in 48 well plates in 0.3mL culture medium until 80-90% confluent. Cells were pre-treated with IBMX for 15 minutes before they were treated with agonists (stachel or isoprenaline) for 1 hour (n=5). In response to stachel treatment, passage HASMs produced a mean fold increase in cAMP of 21.5 ± 3.4 (p=0.0079). Positive control isoprenaline resulted in a 112.6 ± 23.6 fold increase in cAMP (p=0.0079) (

Figure 5-2). These data indicate that the GPR126 receptor is expressed in HASMs and has the capacity to signal via cAMP mediated pathways.

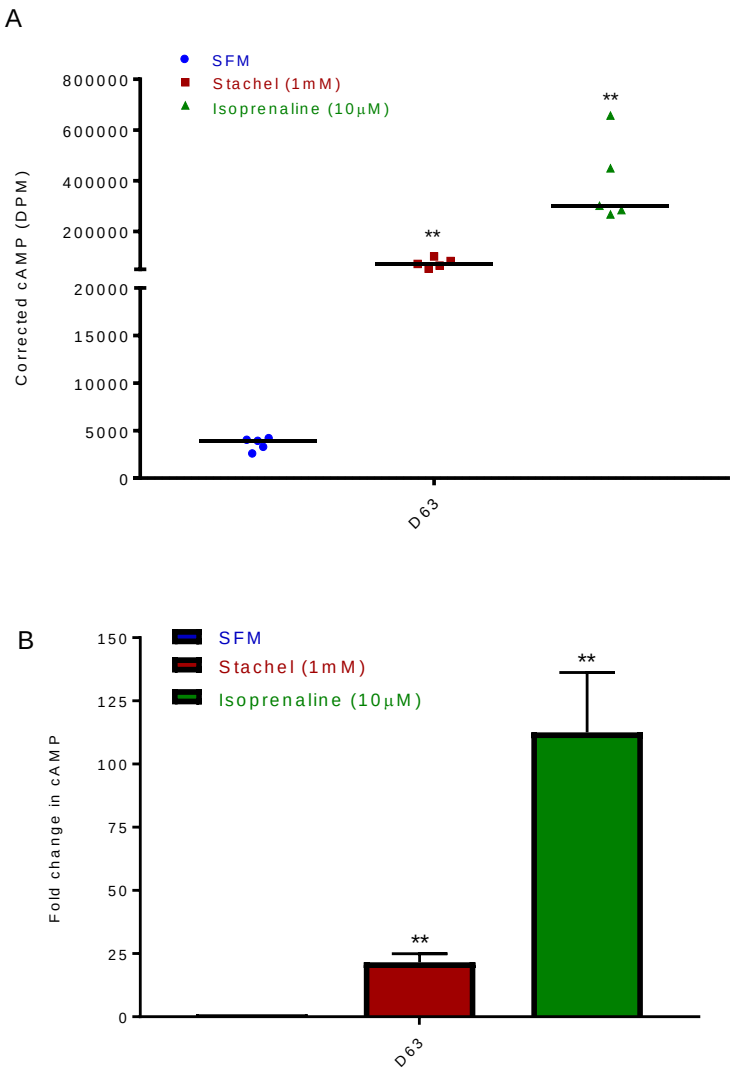


*Figure 5-2 cAMP accumulation in HASM cells pre-treated with IBMX and then treated with Stachel (1mM) or isoprenaline (10µM) for 1 hour (n=5). A) Scatter plot shows corrected cAMP accumulation (DPM) with the median and B) shows normalised fold change data (± SEM). Comparisons were performed to SFM and carried out using a Mann-Whitney test. *p=0.0332, **p=0.0021, ***p=0.0002*

5.3.1.3 HBECs respond to stachel but to a lesser extent than HASMs

In order to identify if GPR126 could also signal via cAMP in HBECS, D7 cells were treated with the stachel peptide (1mM) for 1 hour with IBMX pre-incubation. Donor 7 was the HBEC donor with the highest levels of GPR126 expression in our RNA-seq datasets with an FPKM expression value of 27.143.

Passage 3 HBECs were cultured in 48 well plates in 0.3mL culture medium until ~80% confluent before being treated with agonists (n=5). Isoprenaline and forskolin were used a positive controls for the assay.

HBECs did not respond to stimulation to the same magnitude that HASMs did. The positive controls, forskolin and isoprenaline resulted in mean fold increases in cAMP of 3.3 ± 0.6 (p=0.004) and 2.0 ± 0.5 (p=0.0079) respectively. Stachel treatment resulted in a 2.1 ± 0.5 mean fold increase in cAMP (p=0.0278). Figure 5-3 panel A shows corrected cAMP (DPM) to display the spread of the data and mean fold change + SEM is shown in panel B.

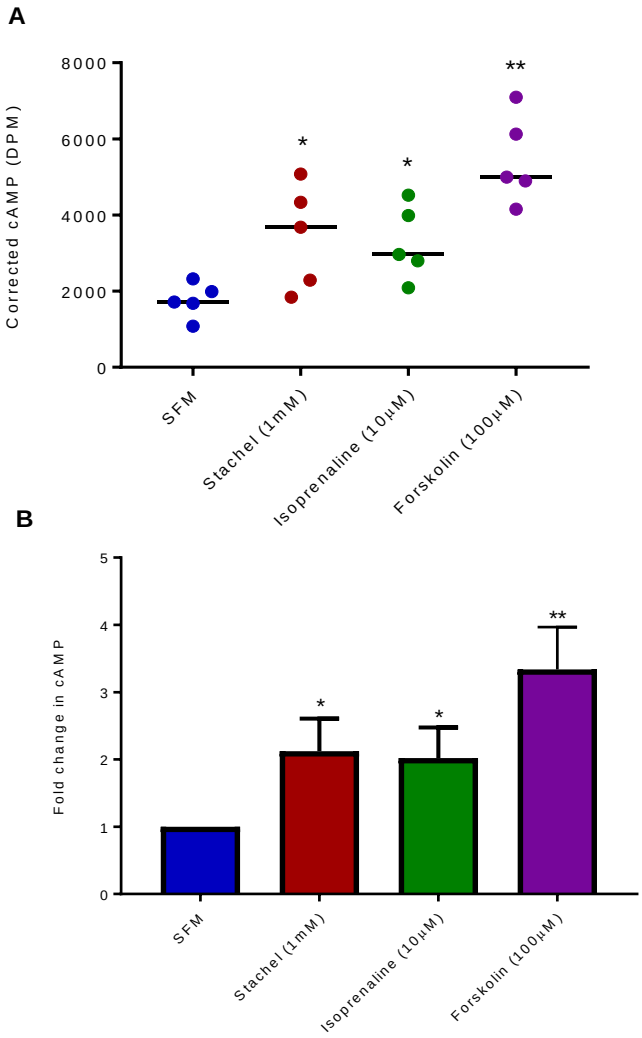


*Figure 5-3 cAMP accumulation in HBECs pre-treated with IBMX and then treated with Stachel (1mM), forskolin (100µM) or isoprenaline (10µM) for 1 hour. A) Corrected cAMP means from 5 experiments and median. B) Mean fold change in cAMP ± SEM. Corrected cAMP values were compared to SFM with a Mann-Whitney test. There is no difference in cAMP levels between individual groups. *p=0.0332, **p=0.0021, ***p=0.0002*

5.3.2 CRE-luciferase transfections in primary human airway cells

In chapter 4 recombinant HEK293s expressing GPR126 produced a ~10 fold increase in luminescence but a 3-4 fold difference measuring cAMP directly, highlighting I may see differences between the two methods.

Passage 3 HASMs (D63) were also transfected with the p6CRE/luc plasmid. Following 48 hours incubation, cells were treated with stachel peptide, isoprenaline or forskolin for 24 hours. The transfected basal cells show an increase in luminescence over non-transfected cells as they have the reporter where the non-transfected cells are not expressing luciferase. There is no increase in luciferase upon stachel treatment (mean fold change = 1.048). Isoprenaline (mean fold change 4.3 ± 0.7) and forskolin (mean fold change 5.4 ± 1.3) show significant increases in CRE-luciferase (Figure 5-4).

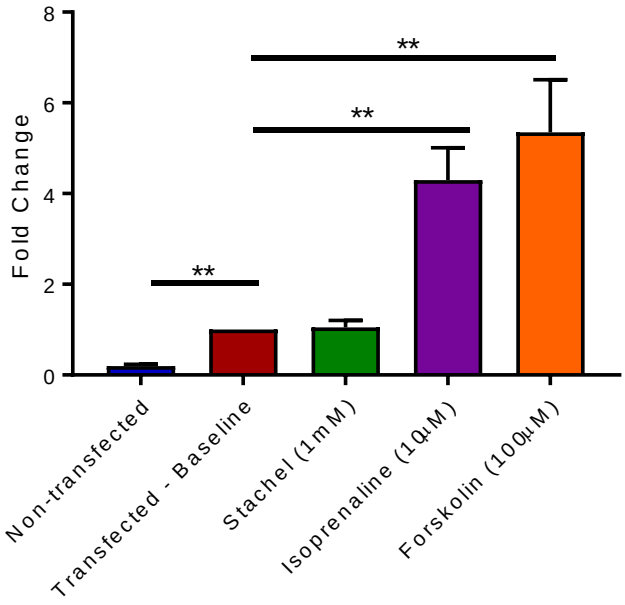


Figure 5-4 Luminescence readings from HASM cells transfected with a CRE-luciferase plasmid (n=6). Mean fold change in luminescence ± SEM. Raw data was analysed using a Mann-Whitney test. Transfected cells showed a significant increase in luminescence and isoprenaline and forskolin produced significant increases in CRE driven luminescence versus transfected but untreated cells. The stachel peptide did not cause an increase in luminescence.

Passage 3 donor 7 HBECs were transfected with the p6CRE/luc construct and treated with forskolin, isoprenaline or stachel peptide for 24 hours. Transfection of the HBECs resulted in a trend towards an increase in

luminescence and following addition of agonists, no additional increase in luminescence was observed (Figure 5-5).

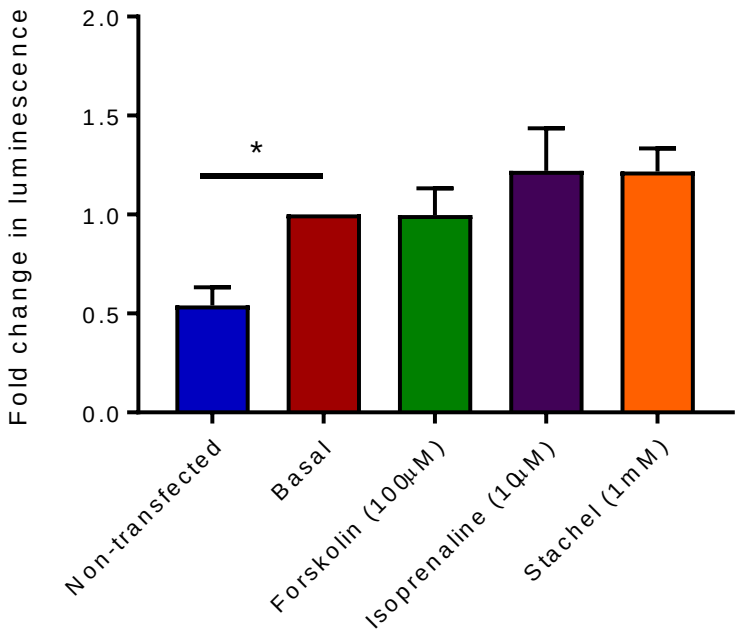


Figure 5-5 Luminescence readings from HBECs transfected with a CRE-luciferase plasmid. Data shown is mean fold change compared to transfected – baseline condition ±SEM. Mean luminescence readings were compared to transfected – baseline condition using a Friedman test with Dunn's multiple comparisons test. CRE-luciferase assays in HASMs. Data shown is from 6 experiments. No conditions were statistically different from the basal, non-treated condition.

5.3.3 Adhesion GPCR expression in primary human airway cells

The following data was presented in chapter 3, but is presented here as some of the data is pertinent to the discussion.

When analysing this data it is important to note that the stachel peptide of some aGPCRs can activate additional aGPCRs to the one they originate from. For example, the GPR126 stachel sequence has been show to activate GPR64 but with a smaller magnitude of effect. [172] Therefore, in order to provide translation of these findings I profiled the expression levels of the entire family

of aGPCRs in the HASM and HBECs used in this chapter. Table 5-1 shows the expression levels of adhesion GPCRs in FPKM from HASMs and HBECs.

aGPCR	Average FPKM in HBEC (n=7)	Average FPKM in HASM (n=5)
ADGRA1 (GPR123)	0.0	0.0
ADGRA2 (GPR124)	0.1	24.3
ADGRA3 (GPR125)	18.6	11.4
ADGRB1 (BAI1)	0.1	0.0
ADGRB2 (BAI2)	1.2	5.9
ADGRB3 (BAI3)	0.0	0.1
ADGRC1 (CELSR1)	45.5	0.9
ADGRC2 (CELSR2)	17.6	1.0
ADGRC3 (CELSR3)	19.0	0.8
ADGRD1 (GPR133)	0.0	9.6
ADGRD2 (GPR144)	0.0	0.0
ADGRE1 (EMR1)	0.0	0.1
ADGRE2 (EMR2)	1.0	0.1
ADGRE3 (EMR3)	0.0	0.0
ADGRE4 (EMR4)	0.0	0.0
ADGRE5 (CD97)	10.0	24.0
ADGRF1 (GPR110)	10.6	0.0
ADGRF2 (GPR111)	17.9	0.1
ADGRF3 (GPR113)	0.2	0.0
ADGRF4 (GPR115)	17.9	0.1
ADGRF5 (GPR116)	0.0	0.0
ADGRG1 (GPR56)	107.3	3.2
ADGRG2 (GPR64)	0.1	0.4
ADGRG3 (GPR97)	0.0	0.0

ADGRG4 (GPR112)	0.0	0.0
ADGRG5 (GPR114)	0.0	0.0
ADGRG6 (GPR126)	17.5	5.7
ADGRG7 (GPR128)	0.0	0.0
ADGRL1 (LPHN1, CIRL-1, CL1)	1.9	2.0
ADGRL2 (LPHN2, CIRL-2, CL2)	15.3	14.4
ADGRL3 (LPHN3, CIRL-3, CL3)	0.1	0.0
ADGRL4 (ELTD1, ETL)	0.2	6.0
ADGRV1 (VLGR1, GPR98)	0.1	0.0
ADRB1	0.7	0.1
ADRB2	14.9	4.0

Table 5-1 Expression (in FPKM) of the adhesion GPCR family of receptors in HASMs (n=5) and HBEC (n=9). GPR126 and GPR64 are highlighted in red.

5.4 DISCUSSION

The aims of the work described in this chapter were to characterise the capacity of primary human airway smooth muscle and bronchial epithelial cells to signal via cAMP through GPR126 activation. A stachel mediated increase of cAMP was observed in both HBECs and HASMs although the magnitude of cAMP accumulation was greater in HASMs. Furthermore, a CRE-luciferase reporter was used to measure cAMP driven gene expression in both cell types. Neither showed a significant increase in CRE-luciferase activity upon activation with the stachel peptide. However, HASMs did produce a CRE response after isoprenaline and forskolin activation.

Stachel activation of GPR126 in HASMs resulted in a mean fold increase in cAMP of 21.51 (min = 12.5, max = 31.8). While the possible functions of GPR126 signalling in airway smooth muscles is unclear, the role of other aGPCRs can give insights into its potential function. CD97 (*ADGRE5*) is an adhesion GPCR expressed in smooth muscle which contains epidermal growth factor (EGF) like domains in its NTF. This protein plays a role in cell migration, cell adhesion as well as leukocyte recruitment, activation and migration. The extracellular EGF-like repeats mediate binding to chondroitin sulphate and the cell surface complement regulatory protein CD55.[89] CD97 is expressed in our HASM donors (mean FPKM = 24) and the other most highly expressed aGPCRs include *ADGRA2, ADGRL2, ADGRA3, ADGRD1, ADGRL4* and *ADGRB2* as well as *GPR126* (all >5 FPKM), none of which have known roles in airway smooth muscle or the lung.

While HBECs were activated by GPR126, this activation resulted in a response with reduced magnitude than activated HASMs showing a response around 1/10th that from HASMs. However, the cAMP response to stachel peptide was equal to that produced by isoprenaline suggesting that these cells may have a lesser capacity for cAMP signalling. In contrast, stachel activation produced a response ~1/5th that produced by isoprenaline treatment. Therefore while HASMs produced a far greater response over baseline, it could be argued that the response in HBECs may be as physiologically important.

The modest response produced by HBECs in response to stachel activation compared to HASMs is in contrast to the level of *GPR126* mRNA expression in each cell type. The mean FPKM from 9 HBEC donors is 17.5 (min = 6.3, max = 27.1) and from 5 HASM donors the mean FPKM is 5.3 (min= 0.3, max = 19.2). Generally, cellular concentrations of proteins correlate with the abundance of their corresponding mRNA. However, the Pearson correlation coefficient of mRNA to protein is ~0.4, implying that only 40% of the variation in protein concentration is due to mRNA abundance so while *GPR126* HBEC mRNA is high, the protein expression may not correlate.[190] In addition, the protein may be expressed but not localise to the plasma member unless under certain conditions. Other contributors to variance in protein abundance are post-transcriptional and translational regulation and protein degradation. The lack of evidence of GPR126 protein expression and localisation is a limitation of this chapter. Immunofluorescence was attempted to detect GPR126 receptor expression and localisation however the protocol could not be optimised to give results of sufficient quality.

Adhesion GPCRs have been shown to play important roles in epithelial cells adhesion GPCRs are involved in polarity and have been shown to be key regulators of planar cell polarity of epithelial cells.[191] As mentioned in the introduction for this chapter, GPR116 is expressed in alveolar type I and II epithelial cells and is a regulator of pulmonary surfactant secretion and uptake, although its ligand is unknown.[188] However, this gene is not expressed in the bronchial epithelium [189] and indeed is not expressed in our HBEC donors.

Other aGPCRs that are expressed at high levels (FPKM >10) in our HBEC donors include ADGRG1, which mediates cell-matrix adhesion and has collagen type III as a ligand and ADGRC1 (CELSR1). ADGRC1 is a gene required for normal lung branching morphogenesis and ADGRC1 mutants show disruption of epithelial integrity, aberrant cytoskeletal remodelling and the mutant endoderm does not branch normally in response to FGF10.[192] ADGRC1 is also expressed in the epithelium during lung development.[192] Other highly expressed receptors from HBEC RNA-seq datasets include

ADGRC3, ADGRA3, ADGRF4 and ADGRF2 which have less defined roles and no known function in the lung.

In order to assess GPR126 signalling activity in human primary cells in a method more comparable to the recombinant cell data from chapter 4, I transfected HBECs (D7) and HASMs (D0063) with the p6CRE/luc plasmid. I also hypothesised this method to be more sensitive given data in the previous chapter where the magnitude of change observed measuring CRE-luciferase was greater (~10 fold) compared to measuring cAMP directly (~2 fold).

Neither HASMs nor HBECs produced a significant response to stachel treatment. Both cell types showed an increase in basal levels of CRE-luciferase after transfection with the luciferase plasmid. However, there was a statistically significant response in HASMs to isoprenaline and forskolin stimulation. There may be a number of reasons as to why there was no response. Firstly, the transfection efficiency may be poor which means that the CRE-driven gene response may not result in sufficient luciferase production. This would limit the magnitude of the response and therefore the effects of increased cAMP may not be seen due to a lack of sensitivity. Secondly, the response from these cells may not elicit a large transcriptional response and the method used is not sensitive enough to detect it.

To address this issue further experiments optimising these transfections would have been ideal, however this was not possible due to time constraints. For example, to maximise the efficiency of plasmid uptake, viral infection with a host such as lentivirus would likely prove more efficient in transferring the plasmid DNA into the cell. In addition, I used time points based on previous methods[133] where optimising the transfection incubation period and treatment period may have yielded different results.

With this data, there is a caveat that the stachel peptide can activate other aGPCRs due to the highly conserved sequence of these aGPCR tethered agonist regions.[172] However, the only known aGPCR that the GPR126 stachel can cross-react with is GPR64, a member of the same family. While

RNA-seq data shows that GPR64 is not expressed in these samples it is unknown whether the GPR126 stachel can activate other adhesion GPCRs.

In future studies it would be worth assessing other outputs aside from cAMP accumulation to test this hypothesis. From the literature, it is known that GPR126 interacts with the extracellular matrix (ECM) and is involved in ECM gene regulation and has adhesive properties.[97, 98, 105] The ECM provides a scaffold for cells and provides biochemical and biomechanical signals that are required for a number of cellular processes such as growth, migration, differentiation, survival, morphogenesis sand homeostasis.[193-195] Therefore it would be plausible that GPR126 activation and/or knock-down may affect these processes. It would be interesting to assess the effect of GPR126 activation and knock-down on cell properties such as proliferation or cell migration.

In summary, these data show that HASMs and HBECs express an active form of GPR126 which is capable of signalling when activated with the stachel peptide. Based on the roles of adhesion GPCRs in developmental biology and known roles of certain receptors in the lung (GPR116, ADGRC1 and CD97), GPR126 and other aGPCRs are likely to be via interactions with components of the ECM to mediate cell adhesion, orientation, and positioning during development, immune responses, and tumour growth.[169] Together, these functions suggest a role for this receptor class and GPR126 in control of cellular organisation of the lung.

This work provided evidence that GPR126 is functional in HASMs and HBECs. Therefore further work aimed to assess the downstream targets of GPR126 activation in HASMs by RNA-seq (Chapter 6). HASMs were chosen for this as they produced a larger and more robust response to stachel activation. This allowed me to look at genes and pathways that are effected by GPR126 activation and give insight into potential biological pathways GPR126 is involved in, in airway smooth muscle. In addition, this experiment was designed to assess the cAMP-PKA dependency of this response by including a PKA inhibitor, H89 to block this pathway.

6 Analysis of GPR126 Downstream Signalling and Potential Functions using Transcriptomics

6.1 INTRODUCTION

Many adhesion-GPCRs are still considered to be orphan receptors although there is good evidence they play important roles in the immune, pulmonary, cardiac, nervous and musculoskeletal systems. [90] GPR126 is known to play roles in the latter three but the functional role of GPR126 in the pulmonary system is unknown. In this chapter, I set out to understand the potential functions of GPR126, and the cAMP-PKA dependency of this response, using transcriptomics in airway smooth muscle cells.

Cyclic-AMP has three main intracellular targets, protein kinase A (PKA), the exchange protein activated by cAMP (Epac), and cyclic nucleotide–gated ion channels (CNGCs). These signalling pathways are covered in detail in the general introduction.

GPR126 is an important receptor in a number of developmental pathways including nerve development, angiogenesis and cardiovascular development. These pathways have already been discussed in detail in previous chapters of this thesis. While it is known that activation of GPR126 can result in an increase in intracellular cAMP in at least some cell types, the downstream effectors of GPR126 activation are largely unknown, particularly in the field of lung biology. GPR126 can couple to G proteins, Gαs or Gαi, to signal via cAMP mediated pathways but it can also signal via its cleaved N-terminal fragment (NTF), independent of the membrane bound fragment.[170] The NTF is essential and sufficient for cardiovascular development, although the mechanism by which this occurs is unclear.[170] GPR126 signals via cAMP mediated pathways to elicit a number of downstream responses, e.g. GPR126 in Schwann cells is responsible for transcriptional regulation of key genes essential for myelination, *Oct6* and *Krox20*. [101]

In order to evaluate the contribution of cAMP to the transcriptional response of HASMs to GPR126 activation, a PKA inhibitor was used to assess effects of PKA-independent signalling. N-[2-(p-bromocinnamylamino)ethyl]-5-isoquinoline sulphonamide (H89) is a pharmacological inhibitor of PKA and therefore has been widely used to study PKA signalling activity.[87] H89 is a

derivative of isoquinoline which was developed from the nonspecific protein kinase G inhibitor, H8.[196] It functions as a competitive inhibitor of ATP at its binding site on the PKA catalytic subunit which prevents the cAMP-dependent phosphorylation of PKA substrates.[197]

Transcriptomics is used to study differential gene expression between different conditions, e.g. during stages of development or with and without a particular stimulus. It is an important tool to help interpret functional elements of the genome and determine functional components of cells and tissues. Transcriptomics can be used to determine all types of transcripts including mRNAs (protein coding RNA), non-coding RNAs and small RNAs. This is useful as it can help determine the transcriptional structure of genes including start sites, 3' and 5' ends and splicing patterns.[198] Gene expression varies depending on, amongst other factors, cell type, time and stimuli so where some genes are expressed ubiquitously, others may only be expressed in highly specific conditions.

Next generation sequencing (NGS) technologies have revolutionised sequencing research due to their high-throughput capabilities, high sensitivity and high speed. Since its inception, NGS has been used for the characterisation of genomes, epigenomes and transcriptomes.[199] NGS has the advantage over previous technologies of avoiding bias introduced by probe selection in array assays, cross-hybridisation and signal saturation-induced detection dynamic range limitation. In addition, NGS provides large datasets with resolution down to the single nucleotide level providing a more precise and comprehensive method of sequencing. [199]

RNA sequencing is a widely used method for the exploration of the transcriptome. RNA sequencing uses cDNA generated from whole RNA molecules, followed by library construction and then massively parallel deep sequencing.[198] High-throughput RNA sequencing has numerous advantages over other comparable technologies such as microarray and traditional Sanger sequencing. Firstly, RNA-seq offers greater resolution than microarray allowing resolution at the single base pair level and increased dynamic range. Moreover, RNA-seq is hypothesis generating and the advantage of this is that

it doesn't require prior knowledge of the genome as microarrays do. This allows the identification of novel transcripts and assessment of tens of thousands of genes rather than selected targets.

In previous chapters, I have shown that GPR126 is expressed in airway smooth muscle cells (HASMS). I have also shown that HASMs respond to stachel activation determined by an increase in cAMP, however the functional implications of this activation is unknown.

The overall aim of this chapter was to determine pathways and genes differentially expressed following GPR126 activation in HASMs to gain biological insight into the potential role of GPR126 in this lung-relevant cell type.

Specifically the aims were:

- To use transcriptomics and global gene expression to further define the potential functions of GPR126 in airway cells by identifying downstream genes and pathways that are altered following receptor activation.
- To determine the cAMP-PKA dependency of this response by pre-incubating cultures with the PKA inhibitor H89.

6.2 METHODS

6.2.1 Cell Biology Methods

6.2.1.1 Cell Culture of Human Airway Smooth Muscle Cells

Passage 2 donor D0063 HASMs were removed from liquid nitrogen storage and cultured in T75 flasks until ~80% confluent and cultured as detailed in section 2.1.2.1. Full details on HASM isolation and donor details are described in the general methods chapter.

6.2.1.2 Treatment of HASMs and Isolation of Cell Lysates for RNA-seq Analysis

Passage 3 D0063 HASMs were cultured in 6 well plates until ~80% confluent. At this point, cells in appropriate wells were pre-treated with the PKA inhibitor H89 (10µM) dissolved in H_2O for 2 hours at 37°C with 5% CO_2. Cell cultures were then treated with stachel peptide (1mM) or a scrambled peptide control (1mM) for 4 or 24 hours at 37°C and 5% CO_2. The concentration of 1mM was chosen as this was the dose used that showed the largest CRE-luciferase response in the HEK293 recombinant cell line. This was also the dose that produced a measurable increase in HASMs and HBECs in chapter 5. After the appropriate incubation period, media was removed from cells and cultures were washed with PBS. Cell lysis buffer was applied to the cells and cells were removed from the culture wells with a cell scraper. Cell lysates were stored at -80°C until further use.

6.2.2 Molecular Biology Methods

6.2.2.1 RNA extraction

RNA was extracted from thawed cell lysates as described in the general methods chapter using the Sigma GenElute™ Total RNA Purification Kit. Isolated RNA samples were stored at -80°C until further use.

6.2.2.2 RNA QC

RNA was assessed for quality and quantity with a Nanodrop ND-2000 Full spectrum Spectrophotometer and the RIN scoring of samples was calculated using Agilent 2100 Bioanalyser with the Agilent RNA 6000 Nano Kit as explained in section 2.2.8.2

6.2.2.3 Library Preparation and RNA-sequencing

Library preparation and sequencing was performed by Oxford Genomics Centre. Library preparation begins with enrichment of samples for mature polyA transcripts which enriches the prepared libraries for protein coding transcripts. Paired end sequencing was performed using an Illumina HiSeq4000 to generate ~25 million reads per sample. Full details of this method can be found in section 2.2.10

6.2.3 Analysis of RNA-seq data

RNA sequencing data in the form of fastq files was received from Oxford Genomics Centre (75bp reads, 25-30 million reads per sample) and was processed through the pipeline detailed in section 2.4.1.[138] In brief, the fastq files were trimmed for adapter sequences and to remove the ends of sequences where there was poor quality sequence data. The trimmed files were then assessed for the quality of sequencing with FASTQC as previously explained. Trimmed fastq files were then aligned to the genome using TopHat2[139], which produces .bam files for further analysis. The trimmed fastq files were processed through cufflinks and cuffmerge for transcript assembly.

6.2.3.1 Baseline Gene and Isoform Expression with Cuffnorm

BAM files from TopHat, along with the merged file from cuffmerge and cufflinks, was used to generate gene and isoform expression values for each sample.[139] The program that does this is Cuffnorm which gives normalised FPKM gene and isoform expression values for each individual sample.

6.2.3.2 Differential Expression with Cuffdiff

TopHat generated BAM files were then used for differential expression analysis with Cuffdiff.[139] Cuffdiff is an algorithm used to find differentially expressed genes and transcripts by testing the observed log fold change of a gene's expression against the null hypothesis of no change. Cuffdiff compares the mean gene expression values of replicates in one group (e.g. treated) to the mean value of replicates in another (e.g. untreated) to determine whether there is a statistical difference between them. Significantly differentially expressed genes are considered to be those that met 5% FDR.

6.2.3.3 Gene Set Enrichment Analysis

There are limitations with focusing on individual differential gene expression as generally this means focusing on genes that show the most significant or genes which show the greatest fold change. The limitations of this are that because this is essentially an arbitrary significance level, biologically relevant changes in gene expression may not meet this threshold. Moreover, looking at

genes in isolation does not necessarily allow the interpretation of biological functions and/or pathways.

Gene set enrichment analysis (GSEA) is a method that is used to examine gene expression data by looking at gene sets based on published data or co-expression studies.[200] The aim of GSEA is to determine whether members of a gene set tend to occur toward the top (or bottom) of a ranked list of genes, in which case the gene set is correlated with the phenotypic class.

GSEA has a number of advantages compared to looking for insight in single gene changes. Firstly, it identifies pathways and processes changed which makes interpretation of large-scale experiments far easier and more translational. Analysing gene sets is often easier to interpret than analysing high scoring genes, which may be poorly annotated or not reproducible.[200] In addition, GSEA can detect modest changes in individual gene expression by boosting the signal to noise ratio if gene sets (i.e. related genes) exhibit strong cross-correlation and leading edge analysis can help define gene subsets to elucidate the results. The method derives its power by focusing on gene sets, that is, groups of genes that share common biological function, chromosomal location, or regulation.[200]

The GSEA software was downloaded from the Broad Institute website (http://software.broadinstitute.org/gsea/index.jsp).[200] The software requires a phenotype file, which assigns each sample to a specific phenotype and an expression dataset. The expression dataset is generated by Cuffnorm, as previously described and altered to a format which was acceptable to the software. One issue with the Cuffnorm output file is that some genes are allocated the same expression value and appear on the same row - due to overlapping transcripts the program cannot differentiate the two genes. R studio was used to generate scripts that remove this problem by separating rows which have two or more genes separated by a comma and placing them on new rows with identical expression values. In addition, some genes have multiple entries and therefore an additional script was written to merge these rows and average the expression values from each row. These lines of code are presented in the appendices.

The GSEA software gives the user a number of outputs and the primary result is the enrichment score. This reflects the degree to which a gene set is overexpressed at the top or bottom of a ranked list of genes. The enrichment score is calculated by walking down the ranked list of genes, increasing running-sum statistic when a gene is in the gene set and decreasing it when it is not. The magnitude of the increment depends on the correlation of the gene with the phenotype. The enrichment score is the maximum deviation from zero and a positive score indicates gene set enrichment where a negative score indicates gene set enrichment at the bottom of the ranked list. The enrichment plot is provided to show a graphical view of the enrichment score for a gene set (Figure 6-1).

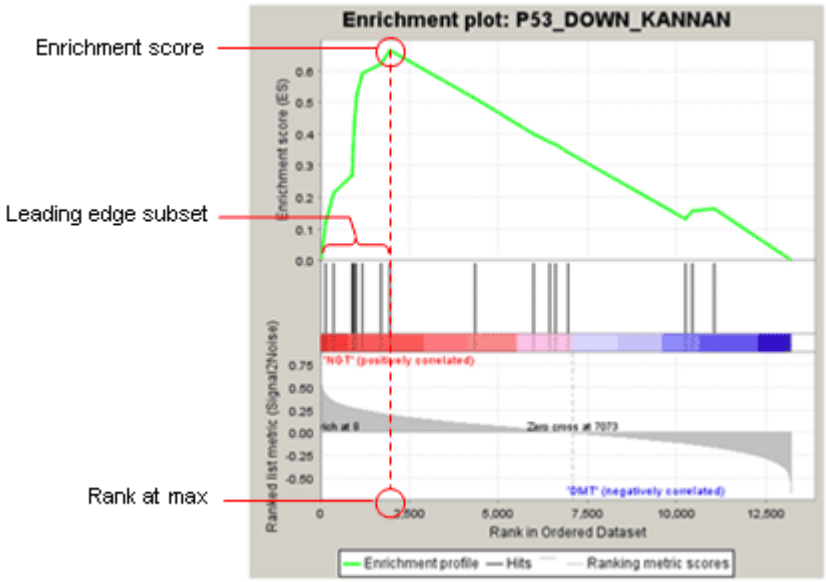


Figure 6-1 A typical plot showing the enrichment score for a gene set. The top portion of the plot shows the enrichment score for the gene set as the analysis walks down the ranked list of genes. The score at the peak of the plot (furthest score from 0) is the enrichment score for the given gene set. The middle section of the plot shows where the members of the gene set appear in the ranked list where the leading edge subset is the list of genes that contribute most to the enrichment score. The bottom section of the plot shows the value of the ranking metric as one moves down the list of ranked genes. The ranking metric measures a gene's correlation with a phenotype. A positive value

indicates correlation with the first phenotype (i.e. treated) and a negative value indicates correlation with the second phenotype (i.e. control).

The statistical significance of the enrichment score is then estimated and adjusted for multiple hypothesis testing – this accounts for differences in gene set size and in correlations between gene sets and the expression dataset. The leading-edge subset of genes are the genes that account for the enrichment signal and these genes can be used to highlight biologically relevant pathways. GSEA calculates this enrichment score, p value and false discovery rate (FDR) for each gene set. The software generates a report which details the number of gene sets enriched in each phenotype (<25% FDR) and the number of gene sets enriched with a p value <1% and <5%. A 5% FDR cut-off point was used for the analyses in this chapter.

Gene set enrichment analysis (http://software.broadinstitute.org/gsea/index.jsp) was used to identify pathways which were significantly altered between cells treated with the stachel peptide compared to the scrambled control. GSEA uses MSigDB Collections which contains 17810 gene sets. I used two gene set databases to identify potential pathways involved in GPR126 activation.

Firstly, GSEA provides access to curated datasets which are curated from various sources such as pathway databases, the literature and known domains. The 'C2' curated gene set databases is inclusive and includes all curated datasets divided into two sub-collections, CGP (chemical and genetic perturbations) and CP (canonical pathways).

The other gene set database used was the hallmark gene sets. Hallmark gene sets summarize specific well-defined biological states or processes. These gene sets were generated by a computational methodology based on identifying gene set overlaps and retaining genes that display coordinate expression. The hallmark gene sets reduce noise and redundancy. This database contains 50 gene sets, with information from over 4,000 original overlapping gene sets from v4.0 MSigDB collections.[201]

6.2.3.4 Graphical Visualisation of RNA-seq Results

Volcano plots and heatmaps were used to visually represent differentially expressed genes, which were generated with R studio. For the code used, ggplot2 and ggrepel were needed and installed using the commands presented in the appendices. Venny 2.1 was used to produce venn diagrams to show overlapping genes between different conditions (http://bioinfogp.cnb.csic.es/tools/venny/index.html).[202]

6.3 RESULTS

6.3.1 Quality Control of Samples for RNA-sequencing

HASM cells (D063) were treated with stachel peptide, a scrambled control peptide or serum free media for 4 or 24 hours, with and without H89 pre-treatment (n=4 independent experiments). RNA was extracted and used for RNA-sequencing analysis.

Table 6-1 shows the concentration and wavelength ratios (260/280 and 260/230) for RNA samples used as determined by nanodrop and RIN scores were determined as outlined in the general methods chapter. RIN scoring was carried out by the Post Genomics Facility (University of Nottingham). All samples were deemed to be of sufficient quality for sequencing by Oxford Genomics Centre (RIN >8).

	Treatment	+/- H89	Time	Conc.	260/280	260/230	RIN
N=1	SFM	-	4 hours	146.6ng/µl	2.03	1.78	>8
	SFM	+	4 hours	238.2ng/µl	2.08	1.15	>8
	Stachel	-	4 hours	230ng/µl	2.07	1.32	>8
	Stachel	+	4 hours	228.6ng/µl	2.06	1.49	>8
	Scrambled	-	4 hours	307.4ng/µl	2.13	0.72	>8
	Scrambled	+	4 hours	244.5ng/µl	2.06	1.77	>8
	SFM	-	24 hours	214.2ng/µl	2.07	0.63	>8
	SFM	+	24 hours	104.5ng/µl	1.99	0.72	>8
	Stachel	-	24 hours	169.3ng/µl	2.04	1.75	>8
	Stachel	+	24 hours	202ng/µl	2.04	1.93	>8
	Scrambled	-	24 hours	197.1ng/µl	2.03	1.83	>8
	Scrambled	+	24 hours	230.8ng/µl	2.06	1.62	>8
N=2	SFM	-	4 hours	287.9ng/µl	2.06	1.86	9.4
	SFM	+	4 hours	260.1ng/µl	2.06	1.93	9.6
	Stachel	-	4 hours	267.4ng/µl	2.06	1.55	9.4
	Stachel	+	4 hours	279.2ng/µl	2.04	2.06	9.5
	Scrambled	-	4 hours	296.9ng/µl	2.04	1.97	9.3
	Scrambled	+	4 hours	210.3ng/µl	2.04	1.91	9.6
	SFM	-	24 hours	262ng/µl	2.05	1.75	9.4
	SFM	+	24 hours	185.9ng/µl	2.06	1.99	6.9
	Stachel	-	24 hours	250.2ng/µl	2.06	2.05	6.8
	Stachel	+	24 hours	224.6ng/µl	2.05	1.33	8.3
	Scrambled	-	24 hours	237.7ng/µl	2.05	1.65	9.3
	Scrambled	+	24 hours	203.1ng/µl	2.03	1.94	9.7

N=3	SFM	-	4 hours	275.9ng/µl	2.07	2.07	2.3
	SFM	+	4 hours	241.6ng/µl	2.06	2.15	>8
	Stachel	-	4 hours	233ng/µl	2.07	2.03	>8
	Stachel	+	4 hours	252.6ng/µl	2.06	1.78	>8
	Scrambled	-	4 hours	292.5ng/µl	2.07	1.39	>8
	Scrambled	+	4 hours	244.6ng/µl	2.06	1.64	9.3
	SFM	-	24 hours	244.6ng/µl	2.06	1.64	8.6
	SFM	+	24 hours	194.7ng/µl	2.07	2.07	>8
	Stachel	-	24 hours	211.8ng/µl	2.06	1.75	8.9
	Stachel	+	24 hours	194.4ng/µl	2.07	0.89	9.6
	Scrambled	-	24 hours	244.1ng/µl	2.08	0.83	9.9
	Scrambled	+	24 hours	243.8ng/µl	2.05	2.06	9.3
N=4	SFM	-	4 hours	199.8ng/µl	2.06	1.41	8.6
	SFM	+	4 hours	131.1ng/µl	2.05	1.05	8.9
	Stachel	-	4 hours	138.9ng/µl	2.03	0.96	8.7
	Stachel	+	4 hours	162.4ng/µl	2.06	0.86	8.3
	Scrambled	-	4 hours	184.4ng/µl	2.04	1.02	9.1
	Scrambled	+	4 hours	139.3ng/µl	2.05	0.74	8.5
	SFM	-	24 hours	78.1ng/µl	2.07	0.73	9
	SFM	+	24 hours	104.3ng/µl	2.11	0.73	9.5
	Stachel	-	24 hours	114.2ng/µl	2.01	1.01	9.1
	Stachel	+	24 hours	119.1ng/µl	2.03	1.03	7.4
	Scrambled	-	24 hours	117.2ng/µl	2.08	1.19	8.2
	Scrambled	+	24 hours	90.8ng/µl	2.04	0.93	8.5

Table 6-1 RNA sample quality control information. SFM = Serum Free Media. Samples labelled >8 were scored by Oxford Genomics Centre.

6.3.2 FastQC Quality Control of Sequencing Data

The FastQC tool was used to assess quality of sequencing data. This tool provides a number of outputs such as per base sequence quality, per sequence quality scores and sequence length distributions. These outputs are covered in detail in the general methods chapter. An example of high quality sequence quality, representative of all the samples used in this chapter, is displayed in the appendices.

6.3.3 GPR126 Expression Determined by Cuffnorm

Cuffnorm was used to generate normalised expression values (FPKM) for individual genes and isoforms for each sample. *GPR126* mRNA was expressed in all conditions, and is expressed at higher levels 24 hours after activation with the satchel peptide in all conditions including at basal (media only control) except the scrambled + H89 incubation (Figure 6-2).

There is no statistically significant difference in *GPR126* expression between the different conditions at 4 hours. At 24 hours however, stachel treatment results in increased GPR126 levels ($p=0.0071$) compared to basal and this is attenuated in cells which have been pre-treated with H89 prior to stachel treatment. Additionally, scrambled treatment with pre-incubation with H89 resulted in a significant decrease in GPR126 expression compared to media only control ($p<0.0001$).

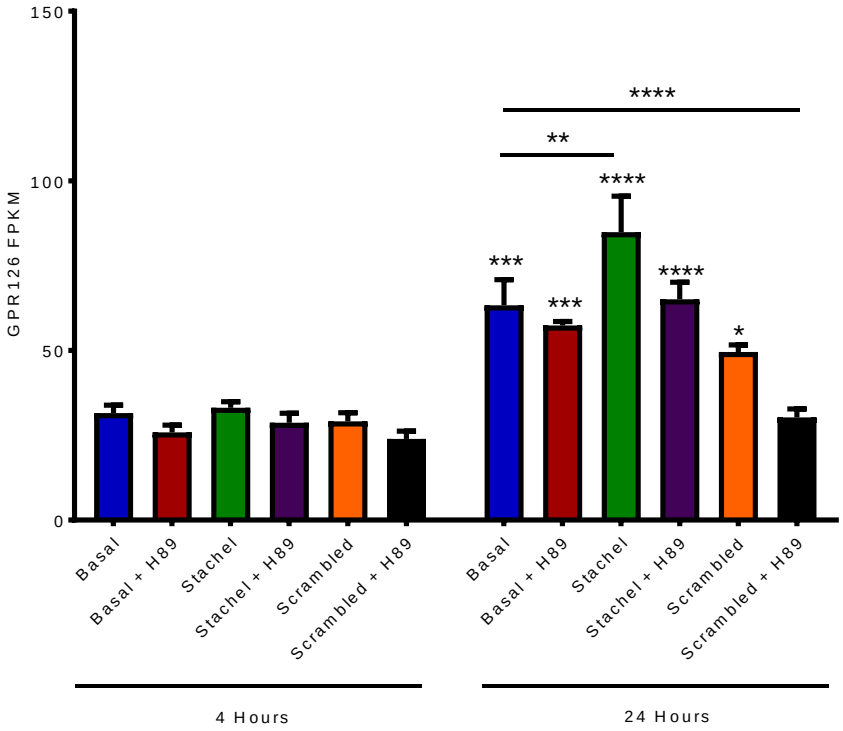


Figure 6-2 GPR126 mRNA expression increases over time in cultured HASM which is augmented by satchel peptide and is cAMP dependent. GPR126 is expressed in all samples. GPR126 is increased in all conditions excluding the scrambled + H89 condition. FPKM values for each condition were compared between the two time points using a 2 way ANOVA with Sidak's multiple comparisons test. In addition, FPKM values were compared to basal within each time point for each condition using a 2 way ANOVA with Dunnett's multiple comparisons test. Data shown is mean ± SEM from 4 experiments.

These data show that GPR126 expression increases with time and as a response to stachel activation. Moreover, GPR126 expression, baseline and stimulated, appear to be cAMP-PKA sensitive given that the stachel activation with H89 treatment showed no increase in GPR126 expression over baseline.

6.3.4 Differential gene expression in HASMs in response to stachel

To identify genes and pathways that may be downstream of GPR126 activation, I activated HASMs (n=4) with the stachel peptide for 4 and 24 hours and used Cuffdiff to identify differentially expressed genes at 5% FDR. To

determine the PKA dependency of this response, a PKA inhibitor (H89) was included and this response was compared to the response of cells treated with the stachel peptide alone.

6.3.4.1 Stachel peptide induces a significant transcriptional response in HASM cells at 4 hours compared to control peptide.

Treatment with the stachel peptide resulted in 514 differentially expressed genes, out of a total of 13,809 comparisons run by Cuffdiff, compared to the scrambled control (5% False Discovery Rate (FDR)). Some comparisons could not be run due to genes with expression levels that were too low.

Of these, 206 were upregulated and 308 were downregulated. The output from the Cuffdiff is shown in the volcano plot in Figure 6-3. The y axis is capped at $p = 5x10^{-5}$, this is a function of the Cuffdiff software which does not provide a lower p value than this.

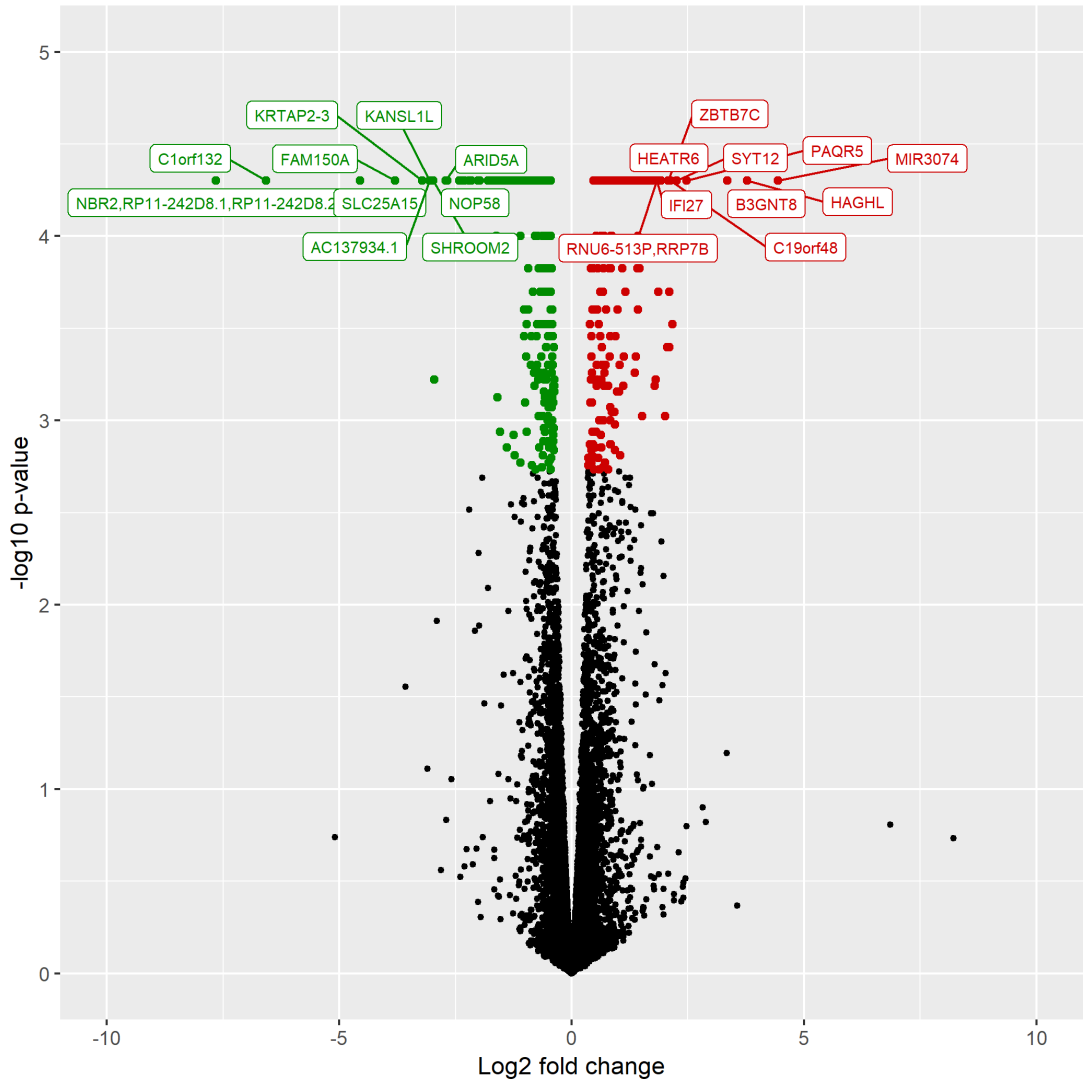


Figure 6-3 Volcano plot showing differentially expressed genes between scrambled control and stachel treated HASM cells after 4 hours. Labelled genes are top 10 up- and down-regulated ranked firstly by p value then by fold change. Coloured dots are genes which reached 5% FDR. Red represents upregulated genes and green represents downregulated genes.

The top 20 upregulated genes, which are ordered firstly on p-value (low-high) and then log2 fold change (high-low), are displayed in Table 6-2 and downregulated genes are listed in Table 6-3. Among the top upregulated genes are those involved in DNA binding and regulation including *ZBTB7C (4.2 fold increase), ZNF813 (3.5 fold increase)* and *PPARGC1A (2.8 fold increase)*; a number of other plasma membrane proteins such as *TMEM100 (3.5 fold increase), SLC16A9 (3.5 fold increase), PAQR5 (5.5 fold increase);*

and genes involved in cytokine signalling pathways including *IL33 (3 fold increase)* and *IFI27 (3.6 fold increase)*.[203] *ADAM8* (3.3 fold increase) is a gene that is implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including muscle development, and neurogenesis.[203] *FGF18 (*3 fold increase*)* is a member of the fibroblast growth factor family, members of which possess mitogenic and cell survival activities, and are involved in a variety of biological processes, including embryonic development, cell growth, morphogenesis, tissue repair, tumour growth, and invasion.[203] There are also a number of genes which are involved in or downstream of the cAMP signalling pathway. *PPARGC1A* is a transcriptional co-activator which can regulate the activity of the cAMP response element binding protein (CREB) and is thought to be important in muscle fibre determination. PDE3A (3. 2 fold increase) is a phosphodiesterase, a class of proteins important for cAMP degradation and therefore important for cAMP signalling regulation.[203] *PDE3A* is also thought to be important for regulating vascular smooth muscle contraction and relaxation.[203]

Gene	Fold change	Gene summary
MIR3074	21.7	**Micro RNA** (post-transcriptional regulation). Involved in DNA damage response.
HAGHL	13.8	**Hydroxyacylglutathione Hydrolase Like** - hydrolase activity and hydroxyacylglutathione hydrolase activity
B3GNT8	10.2	**UDP-GlcNAc:BetaGal Beta-1,3-N-**Acetylglucosaminyltransferase 8 – Protein metabolism and O-linked glycosylation
PAQR5	5.5	**Progestin And AdipoQ Receptor Family Member 5** – Steroid binding. Plasma membrane progesterone. Acts through Gi mediated pathway. Involved in oocyte maturation.
SYT12	4.8	**Synaptotagmin 12** – Belongs to a family of proteins that mediate calcium-dependent regulation of membrane trafficking in synaptic transmission.
C19orf48	4.4	**Chromosome 19 Open Reading Frame 48**
ZBTB7C	4.2	**Zinc Finger And BTB Domain Containing 7C** – DNA binding
HEATR6	3.8	**HEAT Repeat Containing 6**
IFI27	3.6	**Interferon Alpha Inducible Protein 27** - Related pathways are Cytokine Signalling in immune system and Innate Immune System. RNA polymerase II activating transcription factor binding and laminin binding.
RNU6-513P, RRP7B	3.6	**Pseudogene**

SLC16A9	3.5	**Solute Carrier Family 16 Member 9** - Symporter activity and monocarboxylic acid transmembrane transporter activity
TMEM100	3.5	**Transmembrane Protein 100** - Plays a role during embryonic arterial endothelium differentiation and vascular morphogenesis through the ACVRL1 receptor-dependent signalling pathway.
ZNF813	3.5	**Zinc Finger Protein 813**
ADAM8	3.3	**ADAM Metallopeptidase Domain 8** - Members of this family are membrane-anchored proteins structurally related to snake venom disintegrins. Implicated in a variety of biological processes involving cell-cell and cell-matrix interactions, including fertilization, muscle development, and neurogenesis. May be involved in cell adhesion during neurodegeneration, and it is thought to be a target for allergic respiratory diseases, including asthma.
PDE3A	3.2	**Phosphodiesterase 3A** - GMP-inhibited cyclic nucleotide phosphodiesterase (cGI-PDE) family. cGI-PDE enzymes hydrolyse cAMP and cGMP, and play critical roles in many cellular processes by regulating the amplitude and duration of intracellular cyclic nucleotide signals. The encoded protein mediates platelet aggregation and also plays important roles in cardiovascular function by regulating vascular smooth muscle contraction and relaxation.
RASD2	3.0	**RASD Family Member 2** - Ras superfamily of small GTPases
FGF18	3.0	**Fibroblast Growth Factor 18** - FGF family members possess broad mitogenic and cell survival activities, and are involved in a variety of biological processes, including embryonic development, cell growth, morphogenesis, tissue repair, tumour growth, and invasion.
IL33	3.0	**Interleukin 33** - Cytokine that binds to and signals through the IL1RL1/ST2 receptor which in turn activates NF-kappa-B and MAPK signalling pathways in target cell. Has known roles in lung inflammation in asthma and COPD.
SNHG17	2.9	**Small Nucleolar RNA Host Gene 17**
PPARGC1A	2.8	**PPARG Coactivator 1 Alpha** - Transcriptional coactivator that regulates the genes involved in energy metabolism. This protein can interact with, and regulate the activities of, cAMP response element binding protein (CREB) and nuclear respiratory factors (NRFs). It provides a direct link between external physiological stimuli and the regulation of mitochondrial biogenesis, and is a major factor that regulates muscle fibre type determination.

Table 6-2 Top 20 upregulated genes at 4 hours (out of 206) by stachel treatment versus scrambled control treatment ranked firstly by q value and

then by fold change. All genes have p=5.00E-05, q=0.00244. Gene functions and descriptions were obtained from GeneCards [(204)] and Entrez.[(203)]

Of the downregulated genes, there are genes which are involved in development and cell growth (*ARID5A* – 6.5 fold decrease*)*, cell migration (*SHROOM2* – 1.8 fold decrease*),* cell adhesion, proliferation and differentiation (*CTGF* – 4 fold decrease*)* and two genes which code for keratin fibres (*KRTAP2-3* – 9.3 fold decrease*, KRTAP1-5* – 4.5 fold decrease*). DMD* (3.4 fold decrease) is another gene of interest as this gene codes for a protein which is an important for a component of the cytoskeleton and extracellular matrix. [(203)]

Gene	Fold Change	Gene function
NBR2	202.1	Neighbour Of BRCA1 LncRNA 2
C1orf132	95.2	RNA gene
SLC25A15	23.4	Solute Carrier Family 25 Member 15 - mitochondrial carrier family. essential component of the urea cycle, and functions in ammonium detoxification and biosynthesis of the amino acid arginine
FAM150A	13.9	ALK And LTK Ligand 1 - Ligand for receptor tyrosine kinase LTK
KRTAP2-3	9.3	Keratin Associated Protein 2-3 – KRTAPs make up hair keratin intermediate filaments.
AC137934.1	8.3	Unknown
KANSL1L	8.3	KAT8 Regulatory NSL Complex Subunit 1 Like – Transcriptional regulation.
SHROOM2	7.9	Shroom Family Member 2 - Implicated in amiloride-sensitive sodium channel activity. Depletion of this gene results in an increase in endothelial sprouting, migration, and angiogenesis.
ARID5A	6.5	AT-Rich Interaction Domain 5A - roles in development, tissue-specific gene expression, and regulation of cell growth.
NOP58	6.4	NOP58 Ribonucleoprotein
ATP8B1	5.3	ATPase Phospholipid Transporting 8B1 - member of the P-type cation transport ATPase family.

ANKRD1	5.0	Ankyrin Repeat Domain 1 - localized to the nucleus of endothelial cells and is induced by IL-1 and TNF-alpha stimulation. May be a transcription factor.
RNA5SP283	4.9	RNA, 5S Ribosomal Pseudogene 283
HDAC9	4.6	Histone Deacetylase 9 - Histones play a critical role in transcriptional regulation, cell cycle progression, and developmental events. Histone acetylation/deacetylation alters chromosome structure and affects transcription factor access to DNA.
KRTAP1-5	4.5	Keratin Associated Protein 1-5 - The KAP proteins form a matrix of keratin intermediate filaments which contribute to the structure of hair fibres.
PTGS2	4.0	Prostaglandin-Endoperoxide Synthase 2 – The key enzyme in prostaglandin biosynthesis
CTGF	3.9	Cellular Communication Network Factor 2 - mitogen that is secreted by vascular endothelial cells. Plays a role in chondrocyte proliferation and differentiation, cell adhesion in many cell types, and is related to platelet-derived growth factor
NLRP10	3.9	NLR Family Pyrin Domain Containing 10 - This protein plays a regulatory role in the innate immune system.
KLHL38	3.5	Kelch Like Family Member 38
DMD	3.4	Dystrophin - The encoded protein forms a component of the dystrophin-glycoprotein complex (DGC), which bridges the inner cytoskeleton and the extracellular matrix. Deletions, duplications, and point mutations at this gene locus may cause Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), or cardiomyopathy.

Table 6-3 Top 20 downregulated genes at 4 hours (out of 308) by stachel treatment versus scrambled control treatment ranked firstly by q value and then by fold change. All genes have p=5.00E-05, q=0.00244. Gene functions and descriptions were obtained from GeneCards [(204)] and Entrez.[(203)]

Heatmaps were produced in order to assess the variability across independent experiments and identify genes which were reproducibly differentially expressed across all experiments.

Figure 6-4 shows the top 50 genes which showed the highest fold change in expression (up or down) after being ranked on q value after 4 hours stachel treatment compared to scrambled control. The heatmap shows the relative

gene expression in each experimental replicate. There is heterogeneity observed between biological replicates however there are a set of genes which show reproducible changes in expression across experimental replicates. For example, ADAM8, PAQR5, TMEM100 and SLC16A9 are reproducibly upregulated by stachel treatment and CYR61, KRTAP2-3, ARID5A and SHROOM2 are reproducibly downregulated across experimental replicates.

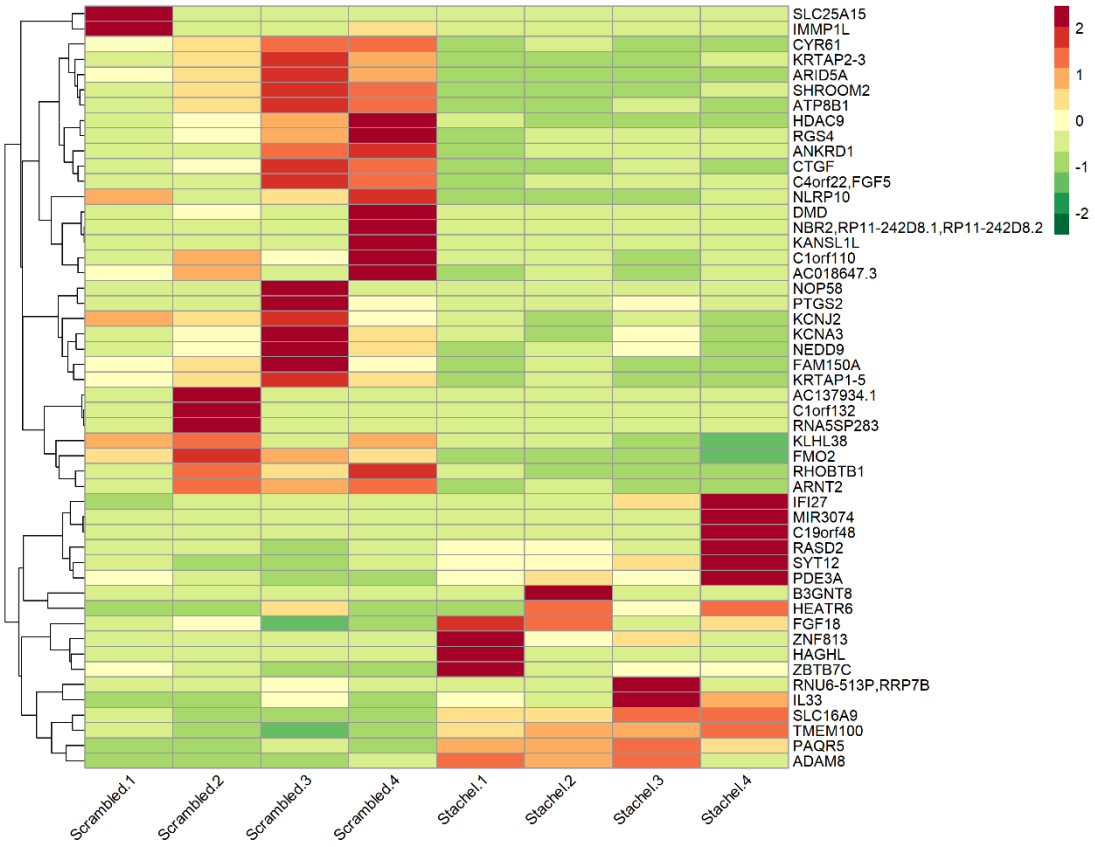


Figure 6-4 Heatmap of differentially expressed genes between scrambled control group and stachel treated cells after 4 hours illustrating reproducibility of differential expression across experimental replicates. Genes shown are the top 50 genes ranked firstly by p value then by fold change. Values are normalised across rows as to show changes in expression of individual genes across all the samples.

6.3.4.2 *Scrambled peptide treatment leads to non-specific alterations in gene expression*

In order to assess the impact of the exogenous addition of a peptide to the cell system, a scrambled control peptide was used to treat the HASMs. After 4 hours incubation, treatment with the scrambled control peptide resulted in 184 differentially regulated genes, with 107 being upregulated and 77 downregulated compared to a media only control – fewer genes than were differentially expressed after stachel treatment. In order to assess the specific effect of the stachel peptide and remove the genes which were differentially regulated by peptide treatment irrespective of the specificity of the peptide, Venny was used to assess which genes were present in both groups.[202] Of the 206 genes upregulated by the stachel peptide, only 6 were also upregulated in the scrambled control group. These genes were: *RNU6-513P/ RRP7B, SLC2A13, IL1RN, CRISPLD2, GFPT2 and ODC1*. Of 308 genes downregulated by stachel activation only 1 gene was also downregulated in the cells treated with a scrambled control peptide (*GDF5*).

These data suggest that exogenous peptide addition to the cell cultures results in non-specific changes in gene regulation. While there was little overlap in differential gene expression between the scrambled and stachel activated cells, it is possible that a number of genes differentially expressed by the stachel peptide will be non-specific.

6.3.4.3 *There is a distinct but overlapping transcriptional response to stachel activation at 24 hours compared to 4 hours*

After 24 hours of stachel treatment, there were 530 differentially expressed genes (5% FDR) out of 13,883 loci tested, with 170 of these being upregulated and 360 downregulated (Figure 6-5).

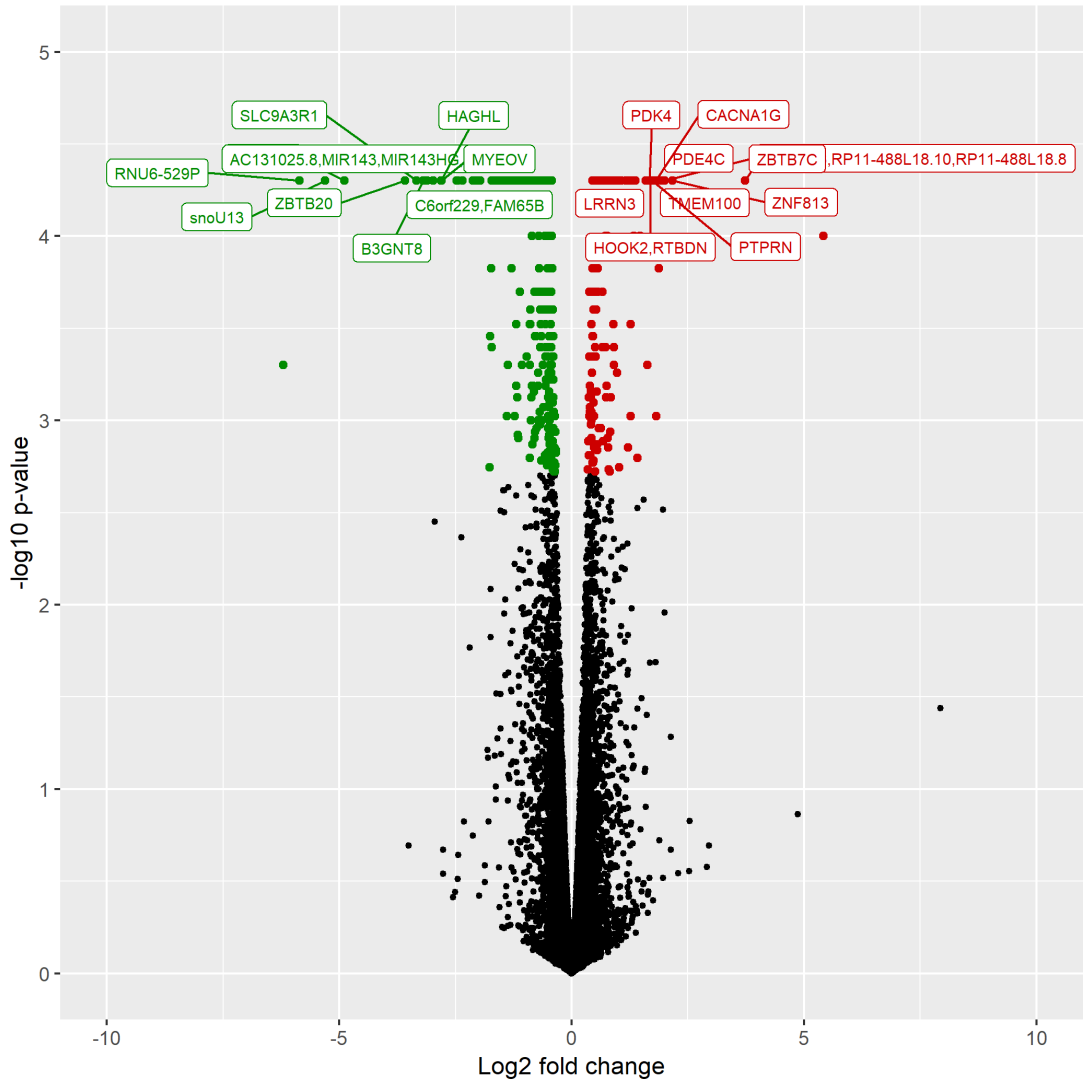


Figure 6-5 Volcano plot showing differentially expressed genes after stachel treatment versus scrambled control for 24 hours. Labelled genes are top 10 up- and down-regulated ranked firstly by p value then by fold change. Coloured dots are genes which reached 5% FDR. Red represents upregulated genes and green represents downregulated genes.

Of the top differentially regulated genes, there are a number of genes which may warrant further investigation due to their known roles in airway biology, aGPCR biology or in cAMP signalling. There is a phosphodiesterase (*PDE4C* – 3.7 fold change) which hydrolyses cAMP. There is also a calcium voltage-gated ion channel (*CACNA1G* – 3.4 fold change) which may be involved in smooth muscle contraction. *HTRA3* (2.6 fold change) is a serine protease that cleaves components of the ECM. [203] GPR64 is also upregulated.

Interestingly, this receptor can be activated by the stachel peptide of GPR126 and therefore may be upregulated as a response to stachel activation. [172]

Gene	Fold Change	Gene Function
MIR3916,RP11-488L18.10,RP11-488L18.8	13.29	**Micro-RNA**. Post-transcriptional regulation
ZNF813	4.50	**Zinc Finger Protein 813**
ZBTB7C	4.00	**Zinc Finger And BTB Domain Containing 7C** – DNA binding
TMEM100	3.95	**Transmembrane Protein 100** - Plays a role during embryonic arterial endothelium differentiation and vascular morphogenesis through the ACVRL1 receptor-dependent signalling pathway.
PDE4C	3.73	**Phosphodiesterase 4C.** This PDE hydrolyzes the second messenger, cAMP
CACNA1G	3.50	**Calcium Voltage-Gated Channel Subunit Alpha1 G.** Mediates the entry of calcium ions into excitable cells. Involved in a variety of calcium-dependent processes, including muscle contraction, hormone or neurotransmitter release, gene expression, cell motility, cell division, and cell death.
PTPRN	3.31	**Protein Tyrosine Phosphatase, Receptor Type N.** A member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signalling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation.
HOOK2,RTBDN	3.25	**Hook Microtubule Tethering Protein 2.** Hook proteins are cytosolic coiled-coil proteins that contain conserved N-terminal domains, which attach to microtubules, and more divergent C-terminal domains, which mediate binding to organelles. **Retbindin.** The protein encoded by this gene is preferentially expressed in the retina and may play a role in binding retinoids and other carotenoids as it shares homology with riboflavin binding proteins
PDK4	3.20	**Pyruvate Dehydrogenase Kinase 4.** Encodes a mitochondrial protein with a histidine kinase domain. This protein is located in the matrix of the mitrochondria and inhibits the pyruvate dehydrogenase complex by phosphorylating one of its subunits, thereby contributing to the regulation of glucose metabolism
LRRN3	3.02	**Leucine Rich Repeat Neuronal 3.**

PLP1	2.58	**Proteolipid Protein 1.** Transmembrane proteolipid protein that is the predominant component of myelin.
HTRA3	2.58	**HtrA Serine Peptidase 3.** Serine protease that cleaves beta-casein/CSN2 as well as several extracellular matrix (ECM) proteoglycans such as decorin/DCN, biglycan/BGN and fibronectin/FN1. Inhibits signalling mediated by TGF-beta family proteins possibly indirectly by degradation of these ECM proteoglycans
YPEL4	2.52	**Yippee Like 4.**
RNF112	2.52	**Ring Finger Protein 112.** A member of the RING finger protein family of transcription factors. The protein is primarily expressed in brain.
SLITRK6	2.45	**SLIT And NTRK Like Family Member 6.** This protein functions as a regulator of neurite outgrowth required for normal hearing and vision
IL33	2.43	**Interleukin 33** - Cytokine that binds to and signals through the IL1RL1/ST2 receptor which in turn activates NF-kappa-B and MAPK signalling pathways in target cells
FSIP1,GPR176	2.41	**Fibrous Sheath Interacting Protein 1.** **G Protein-Coupled Receptor 176.** Orphan receptor involved in normal circadian rhythm behaviour. Acts through the G-protein subclass G(z)-alpha and has an agonist-independent basal activity to repress cAMP production.
RBMY2TP	2.39	**RNA Binding Motif Protein.**
RASL12	2.37	**RAS Like Family 12.** GTP binding.
GPR64	2.30	**Adhesion G Protein-Coupled Receptor G2.** Orphan receptor. Could be involved in a signal transduction pathway controlling epididymal function and male fertility.

Table 6-4 Top 20 upregulated genes (out of 170) by stachel treatment versus scrambled control after 24 hours treatment ranked firstly by p-value and then by log2 fold change. p=5.00E-05, q=0.002314. Gene functions and descriptions were obtained from GeneCards [(204)] and Entrez.[(203)]

There are a number of genes which have potential roles in airway biology. Of note from the downregulated genes, *SLC9A3R1* (10.1 fold change) is a sodium/hydrogen exchanger regulatory cofactor which interacts with and regulates proteins including the cystic fibrosis transmembrane conductance

regulator and GPCRs such as the β2-adrenergic receptor, both of which are important receptors in the lung.[203] In addition, *SLC9A3R1* is an important linker protein between membrane and cytoskeletal proteins. *SERPINE1* is an inhibitor of fibrinolysis and has known to be involved in lung fibrosis and emphysema.[205] *TM4SF20* (5.4 fold change) is involved in a number of cell processes via interactions with integrins.[203] ACAN (5.1 fold change) is a component of the ECM, *LAD1* (3.9 fold change) is a component of basement membranes and *TNNT2* encodes a protein that regulates muscle contraction in response to intracellular calcium ion concentration.[203]

Gene	Fold Change	Gene Function
RNU6-529P	58.0	**RNA, U6 Small Nuclear 529, Pseudogene**
snoU13	39.5	**Small Nucleolar RNA, C/D Box 13 Pseudogene 1**
HEATR6	29.7	**HEAT Repeat Containing 6.**
ZBTB20	12.0	**Zinc Finger And BTB Domain Containing 20.** Transcription factor which acts as a transcriptional repressor and plays a role in many processes including neurogenesis, glucose homeostasis, and postnatal growth.
SLC9A3R1	10.1	**SLC9A3 Regulator 1.** Sodium/hydrogen exchanger regulatory cofactor. The protein interacts with and regulates various proteins including the cystic fibrosis transmembrane conductance regulator and G-protein coupled receptors such as the beta2-adrenergic receptor and the parathyroid hormone 1 receptor. The protein also interacts with proteins that function as linkers between integral membrane and cytoskeletal proteins. Mutations in this gene result in hypophosphatemic nephrolithiasis/osteoporosis type 2
B3GNT8	9.1	**UDP-GlcNAc:BetaGal Beta-1,3-N**-Acetylglucosaminyltransferase 8 – Protein metabolism and O-linked glycosylation
AC131025.8, MIR143, MIR143HG, MIR145	8.7	**Micro-RNAs**
C6orf229,FAM65B	7.9	**RHO Family Interacting Cell Polarization Regulator 2**. Inhibitor of the small G protein RhoA
HAGHL	7.0	**Hydroxyacylglutathione Hydrolase Like** - hydrolase activity and hydroxyacylglutathione hydrolase activity

MYEOV	6.9	**Myeloma Overexpressed.**
SERPINE1	5.6	**Serpin Family E Member 1.** Principal inhibitor of tissue plasminogen activator (tPA) and urokinase (uPA), and hence is an inhibitor of fibrinolysis.
TM4SF20	5.4	**Transmembrane 4 L Six Family Member 20.** Members of this family function in various cellular processes including cell proliferation, motility, and adhesion via their interactions with integrins.
ACAN	5.1	**Aggrecan.** The encoded protein is an integral part of the extracellular matrix in cartilagenous tissue and it withstands compression in cartilage. Mutations in this gene may be involved in skeletal dysplasia and spinal degeneration.
OXTR	4.4	**Oxytocin Receptor.** G-protein coupled receptor family and acts as a receptor for oxytocin.
DCXR	4.3	**Dicarbonyl And L-Xylulose Reductase.** The protein encoded by this gene acts as a homotetramer to catalyze diacetyl reductase and L-xylulose reductase reactions.
TUFM	4.2	**Tu Translation Elongation Factor, Mitochondrial.** Participates in protein translation in mitochondria.
LAD1,TNNT2	3.9	**Ladinin 1.** The protein encoded by this gene may be an anchoring filament that is a component of basement membranes. **Troponin T2, Cardiac Type.** The protein encoded by this gene is the tropomyosin-binding subunit of the troponin complex, which is located on the thin filament of striated muscles and regulates muscle contraction in response to alterations in intracellular calcium ion concentration.
ANO1	3.9	**Anoctamin 1.** Calcium-activated chloride channel (CaCC) which plays a role in trans-epithelial anion transport and smooth muscle contraction. Acts as a major contributor to basal and stimulated chloride conductance in airway epithelial cells and plays an important role in tracheal cartilage development.
MAP3K7CL	3.3	**MAP3K7 C-Terminal Like.** Among its related pathways are Integrated Breast Cancer Pathway.
MYBL2	3.2	**MYB Proto-Oncogene Like 2.** A nuclear protein involved in cell cycle progression.

Table 6-5 Top 20 downregulated genes (out of 360) by stachel treatment versus scrambled control treatment after 24 hours ranked firstly by p-value and then by log2 fold change. p=5.00E-05, q=0.002314. Gene functions and descriptions were obtained from GeneCards [(204)] and Entrez.[(203)]

Figure 6-6 shows the top 50 genes which showed the highest fold change in expression (up or down) after being ranked on q value after 24 hours stachel treatment compared to scrambled control. The heatmap shows the relative gene expression in each individual sample. As with the 4 hour treatment groups, there is some heterogeneity observed between samples of the same treatment group with a difference between the two groups being driven by one sample showing very high or low expression. However, there are also genes which show reproducible changes in expression across 4 individual experiments. A number of the upregulated genes such as PLP1, ZBTB7C, TMEM100, PTPRN, LRRN3 and YPEL4 show reproducible upregulation after 24 hours and RTKN2, CTGF, HAPLN3, HSPB3, KRTAP1-5, FAM83D, ACTA2, and SERPINE1 are reproducibly downregulated after 24 hours.

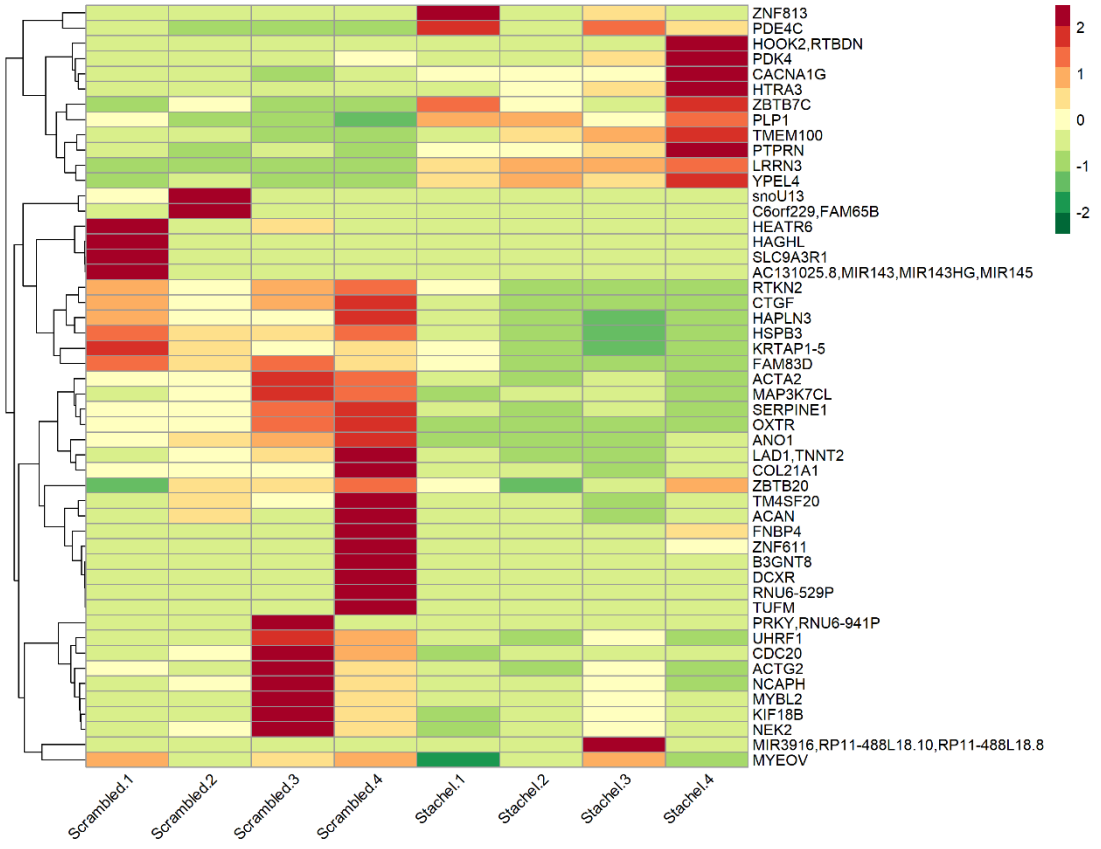


Figure 6-6 Heatmap of differentially expressed genes between scrambled control group and stachel treated HASMs after 24 hours. Genes shown are the top 50 genes ranked firstly by p value then by fold change.

6.3.4.4 *Scrambled peptide treatment leads to non-specific alterations in gene expression at 24 hours.*

Again, to assess the impact of the scrambled peptide, this treatment group was compared to a media only control. After 24 hours of treatment, scrambled peptide resulted in 101 differentially regulated genes vs. media only control (5% FDR) out of 13,716 comparisons. 66 of these were upregulated and 35 were downregulated. Venny was used to compare the stachel response to the scrambled response.[202] Of 170 genes upregulated by stachel, 0 were upregulated in the scrambled treatment. Of the 360 genes downregulated by stachel treatment, 1 gene was also downregulated by scrambled treatment (*HEATR6*).

While there is almost no overlap between the scrambled peptide treatment and stachel treatment, this data suggests that the addition of a peptide to HASMs in culture results in non-specific differential gene expression.

Therefore, this must be taken into account when analysing this data and focus must be on genes which are reproducible across individual experiments and genes which show reproducible changes in expression at both 4 and 24 hours. While it is expected for there to be different transcriptome profiles at the two times due to early and late responders, replication at two time points provides some confidence that these are real effects.

6.3.4.5 *Approximately 20% of genes shows sustained differential expression across 4 and 24 hour time points post stimulation*

HASMs that were treated for 4 and 24 hours have distinct but overlapping transcription profiles. Differentially expressed genes from cells treated with stachel peptide at 4 and 24 hours were compared with Venny.[202] Of the upregulated genes, there are 61 genes in common between the two time points (19.3%) as shown in Figure 6-7 providing confidence that these genes are a result of GPR126 activation.

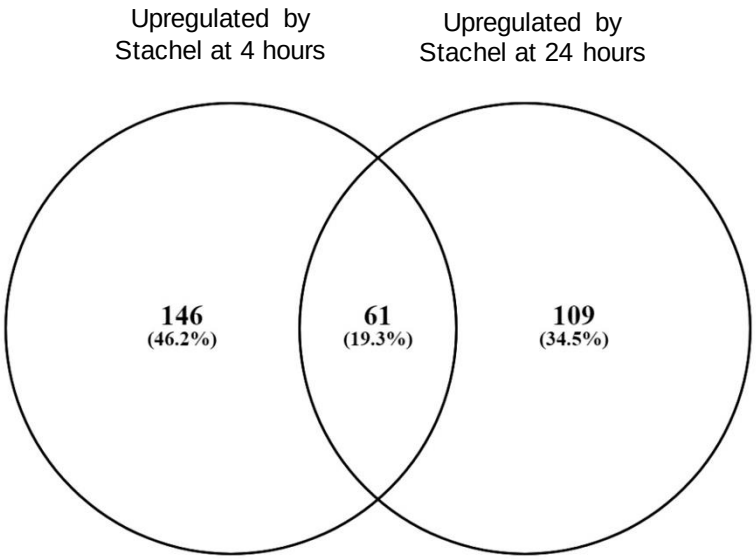


Figure 6-7 Venn diagram showing the upregulated genes in common between cells treated with stachel after 4 hours incubation and 24 hours. There are 61 genes in common.

Table 6-6 shows the 13 genes that showed a >2 fold change in gene expression after both 4 and 24 hours out of the total of 61 genes upregulated at both time points. A number of these genes were not in the top 20 ranked genes described in the previous sections of this chapter. ADCY4 codes for a member of the adenylyl cyclase family of proteins, which are the enzymes required for cAMP production.[203] S1PR1 is a membrane receptor that is involved in endothelial cell differentiation and activation of the receptor induces cell-cell adhesion. [203, 206] PLXNA4 codes for a protein involved in roles in the nervous system.[207] BDKRB1 codes for bradykinin receptor. Bradykinin is produced in response to inflammation and plays a role in bronchoconstriction.[208] AQP3 codes for a small transmembrane water/glycerol channel that facilitates the uptake of hydrogen peroxide which may have roles in inflammatory pathways. [209]

Gene	q_value after 4 hours	Fold change after 4 hours	q_value after 24 hours	Fold change after 24 hours
PAQR5	0.00244	5.53	0.00627	3.66

ZBTB7C	0.00244	4.21	0.00231	4.00
ZNF813	0.00244	3.48	0.00231	4.50
TMEM100	0.00244	3.52	0.00231	3.95
CACNA1G	0.00244	2.22	0.00231	3.50
PDK4	0.00244	2.27	0.00231	3.20
IL33	0.00244	2.97	0.00231	2.43
ADCY4,RIPK3,RP 11-934B9.3	0.00639	2.68	0.00441	2.53
LRRN3	0.00244	2.07	0.00231	3.02
SLITRK6	0.00244	2.33	0.00231	2.45
S1PR1	0.00244	2.63	0.00231	2.00
RASL12	0.00244	2.12	0.00231	2.37
PLXNA4	0.00244	2.37	0.00231	2.05

Table 6-6 13 genes were upregulated by ≥2 fold after both 4 and 24 hours, out of the 61 genes overexpressed at both time points.

In the downregulated gene sets, 121 genes are downregulated after both 4 and 24 hours (22.1%) as shown in Figure 6-8.

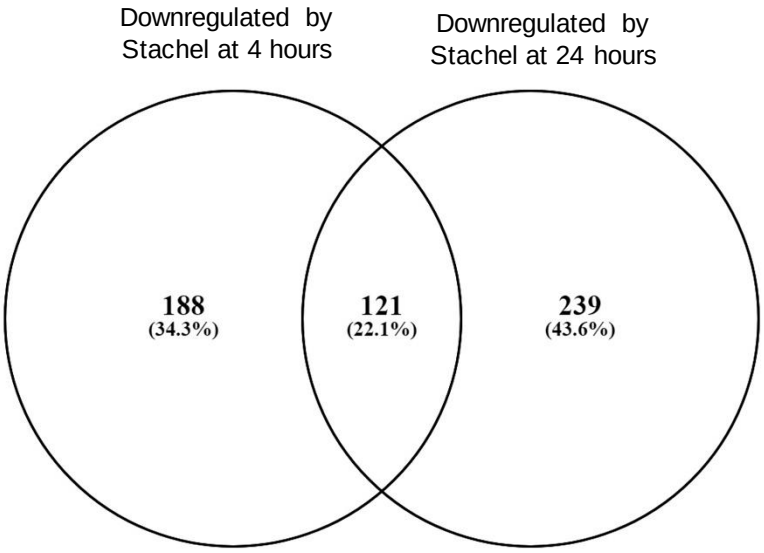


Figure 6-8 Venn diagram showing the overlap of downregulated genes in cells treated with stachel after 4 hours incubation vs. 24 hours incubation. There are 121 genes downregulated at both time points.

13 genes were downregulated by >2 fold at both 4 and 24 hours (Table 6-7) out of 121 genes that were significantly downregulated a both time points. There are a number of genes which were not in the top 20 gene lists but are significantly downregulated at both time points >2 fold.

AURKA, a mitotic serine/threonine kinase which is involved in mitosis and meiosis and is essential for cell proliferation. [210] EDN1, which codes for endothelin-1 a vasoconstrictor produced by vascular endothelial cells. NEDD9 is a protein involved in a number of signalling pathways involved in cell proliferation, survival and migration. [211, 212] HSPB3 codes for a heat shock protein; SEMA7A is involved in axon growth and plays roles in inflammation, [213, 214] and TNFRSF12A promotes angiogenesis and endothelial cell proliferation. It may also have roles in cell adhesion.[215]

Gene	q value after 4 hours	Fold change after 4 hours	q value after 24 hours	Fold change after 24 hours
AC018647.3	0.00244	3.27	0.002314	2.48
ATP8B1	0.00244	5.35	0.002314	2.49
AURKA	0.00244	2.01	0.002314	2.46
CTGF	0.00244	3.94	0.002314	2.58
EDN1	0.00244	2.60	0.014276	3.29
HSPB3	0.00244	2.10	0.002314	2.63
KRTAP1-5	0.00244	4.45	0.002314	3.03
NEDD9	0.00244	3.01	0.023881	2.25
OXTR	0.00244	2.34	0.002314	4.37
RTKN2	0.00244	2.58	0.002314	2.72
SEMA7A	0.00244	2.23	0.002314	2.26
SERPINE1	0.00244	2.42	0.002314	5.56

TNFRSF12A	0.00244	2.25	0.002314	2.13

Table 6-7 13 genes were downregulated by ≥2 fold after both 4 and 24 hours, out of the 121 genes downregulated at both time points.

6.3.5 Pathway based analyses identifies a potential role for GPR126 in inflammation and control of cellular proliferation

Gene set enrichment analysis (http://software.broadinstitute.org/gsea/index.jsp) was used to identify pathways which were significantly altered between cells treated with the stachel peptide compared to the scrambled control. Two genes sets were used for this analysis, the hallmark gene sets, which summarise specific well-defined biological processes and the curated gene sets which contain gene sets curated from a variety of sources such as pathway databases and the literature.

6.3.5.1 Hallmark dataset analyses shows GPR126 downregulates pathways involved in cell proliferation

There were 9 gene sets that were downregulated (5% FDR) by GPR126 activation in the hallmark gene sets with no gene sets upregulated at both time points. No gene sets were enriched at 4 hours. The gene sets downregulated at 24 hours (5% FDR) are involved in cell cycle regulation and pathways involved in cell proliferation (Table 6-8).

Gene set	Pathway	Size	ES	NES	q value (FDR)	Leading Edge	Time point
Hallmark	E2F targets	199	-0.71	-2.04	0	tags=66%, list=14%, signal=76%	24 hours
Hallmark	MYC targets	199	-0.7	-1.97	0	tags=67%, list=19%, signal=83%	24 hours
Hallmark	G2M checkpoint	199	-0.68	-1.93	0	tags=60%, list=12%, signal=68%	24 hours
Hallmark	MTORC1 signalling	199	-0.62	-1.74	0.001	tags=63%, list=18%, signal=76%	24 hours
Hallmark	MYC targets v2	58	-0.68	-1.65	0.0033	tags=62%, list=18%, signal=75%	24 hours

Hallmark	Androgen response	100	-0.57	-1.51	0.018	tags=52%, list=15%, signal=61%	24 hours
Hallmark	Unfolded protein response	112	-0.56	-1.49	0.0183	tags=58%, list=18%, signal=70%	24 hours
Hallmark	Cholesterol Homeostasis	73	-0.59	-1.49	0.0207	tags=51%, list=10%, signal=57%	24 hours
Hallmark	Mitotic Spindle	199	-0.52	-1.46	0.0254	tags=46%, list=14%, signal=53%	24 hours

Table 6-8 Nine gene sets are downregulated by GPR126 activation in the hallmark gene sets. Size = number of genes in the gene set after filtering out those not in the expression dataset. ES = enrichment score for the gene set. This is the degree to which the gene set is overrepresented at the top or bottom of the ranked list of genes in the expression dataset. NES = normalised enrichment score. q-value = false discovery rate – the estimated probability that the NES represents a false positive. Leading edge subset = subset of members that contribute the most to the ES. Tags = the percentage of genes contributing to the ES. List = percentage of genes in the ranked gene list before the peak in the running ES. This gives an indication of where in the list the ES is attained. Signal = the ES strength that combines the two previous statistics.

6.3.5.2 Curated dataset analyses identifies a number pathways regulated by GPR126 are involved in the immune response and cell proliferation

There was one pathway upregulated at 5% FDR in the cells treated with the stachel peptide for 4 hours (Table 6-9). There were no pathways enriched after 24 hours incubation. The one pathway upregulated was beta defensins, which are antimicrobial peptides implicated in the resistance of epithelial surfaces to microbial colonization.

Gene set	Pathway	Size	ES	NES	q value (FDR)	Leading Edge	Time point
Curated	Beta Defensins	40	0.88	1.9	0	tags=5%, list=10%, signal=6%	4 hours

Table 6-9 Pathway upregulated by stachel activation of HASMs at both 4 and 24 hours (5% FDR) in the curated gene data sets. Size = number of genes in the gene set after filtering out those not in the expression dataset. ES = enrichment score for the gene set. This is the degree to which the gene set is overrepresented at the top or bottom of the ranked list of genes in the expression dataset. NES = normalised enrichment score. q-value = false discovery rate – the estimated probability that the NES represents a false positive. Leading edge subset = subset of members that contribute the most to the ES. Tags = the percentage of genes contributing to the ES. List = percentage of genes in the ranked gene list before the peak in the running ES. This gives an indication of where in the list the ES is attained. Signal = the ES strength that combines the two previous statistics.

Using the curated datasets, 192 pathways were significantly downregulated in the stachel samples at 5% FDR after 24 hours. There were no statistically significant (<5% FDR) downregulated pathways after 4 hour incubation.

Many of the downregulated pathways are involved in cancer phenotypes, cell proliferation (EGFR, Myc) and cell cycle control (MTOR signalling, G2M checkpoint). Other pathways include TNC, which is a gene which is expressed in the extracellular matrix of various tissues during development, disease or injury and mitotic spindle formation pathways.

Gene set database	Pathway	Size	ES	NES	q value (FDR)	Leading Edge	Time point
Curated	Breast cancer grade 1 vs 3	146	-0.8	-2.2	0	tags=80%, list=13%, signal=92%	24 hours
Curated	Cervical cancer proliferation cluster	137	-0.8	-2.2	0	tags=76%, list=11%, signal=85%	24 hours

Curated	TNC targets	138	-0.79	-2.16	0	tags=71%, list=11%, signal=80%	24 hours
Curated	EGFR Signalling	244	-0.74	-2.13	0	tags=68%, list=12%, signal=77%	24 hours
Curated	Cell proliferation	138	-0.77	-2.11	0	tags=72%, list=15%, signal=85%	24 hours
Curated	WTAP targets	299	-0.72	-2.1	0	tags=58%, list=12%, signal=66%	24 hours
Curated	Acinar development	270	-0.72	-2.1	0	tags=73%, list=16%, signal=86%	24 hours
Curated	Transformed by RHOA	526	-0.68	-2.1	0	tags=60%, list=14%, signal=69%	24 hours
Curated	Hypoxia	282	-0.72	-2.09	0	tags=67%, list=14%, signal=77%	24 hours
Curated	Core serum response	206	-0.73	-2.08	0	tags=68%, list=14%, signal=80%	24 hours
Curated	Lung cancer poor survival	438	-0.69	-2.07	0	tags=64%, list=16%, signal=75%	24 hours
Curated	Cell cycling genes	143	-0.75	-2.05	0	tags=72%, list=12%, signal=82%	24 hours
Curated	Doxorubicin resistance	53	-0.85	-2.05	0	tags=87%, list=11%, signal=98%	24 hours
Curated	Epithelial mesenchymal transition	173	-0.73	-2.04	0	tags=67%, list=13%, signal=77%	24 hours
Curated	CML dividing vs normal quiescent	178	-0.72	-2.03	0	tags=54%, list=9%, signal=59%	24 hours
Curated	Melanoma metastasis	157	-0.74	-2.03	0	tags=69%, list=14%, signal=80%	24 hours
Curated	MYCN targets with E-Box	759	-0.65	-2.03	0	tags=60%, list=17%, signal=71%	24 hours
Curated	YBX1 targets	276	-0.69	-2.03	0	tags=62%, list=14%, signal=72%	24 hours

Curated	Response to GSK3 inhibitor (SB216763)	353	-0.67	-2.02	0	tags=58%, list=15%, signal=67%	24 hours
Curated	Neocortex basal radial GLIA	185	-0.72	-2.02	0	tags=69%, list=11%, signal=77%	24 hours
Curated	Intestine probiotics	548	-0.65	-2	0	tags=63%, list=18%, signal=76%	24 hours
Curated	Glutamine deprivation	333	-0.68	-2	0	tags=63%, list=16%, signal=75%	24 hours
Curated	IL6 deprivation	98	-0.77	-2	0	tags=76%, list=15%, signal=89%	24 hours

Table 6-10 Pathways downregulated by stachel activation of HASMs at both 4 and 24 hours (5% FDR). At 24 hours, 192 downregulated pathways using the curated pathways, pathways shown are those with NES >2 after being ranked on q value. Size = number of genes in the gene set after filtering out those not in the expression dataset. ES = enrichment score for the gene set. This is the degree to which the gene set is overrepresented at the top or bottom of the ranked list of genes in the expression dataset. NES = normalised enrichment score. q-value = false discovery rate – the estimated probability that the NES represents a false positive. Leading edge subset = subset of members that contribute the most to the ES. Tags = the percentage of genes contributing to the ES. List = percentage of genes in the ranked gene list before the peak in the running ES. This gives an indication of where in the list the ES is attained. Signal = the ES strength that combines the two previous statistics.

Pathway analyses suggests that GPR126 activation leads to downregulation of a number of biological pathways. These pathways seem to be involved in cell cycle checkpoints and cellular proliferation. The one pathway which was significantly upregulated by GPR126 activation was the beta defensins pathway, implicating a role for GPR126 in the immune response.

6.3.6 The stachel mediated response is partly cAMP-PKA dependent

In order to assess the proportion of the stachel-driven transcriptional response that was cAMP-PKA driven, samples pre-treated with the PKA inhibitor H89 before stachel treatment were compared to samples treated just with the scrambled peptide. A total of 677 genes were differentially regulated when H89 was added to cells treated with stachel compared to cells treated only with a scrambled control after 4 hours and 616 after 24 hours. This is a greater number of genes than were differentially regulated by stachel activation alone, which is not surprising given the number of genes and pathways cAMP is known to regulate.

After 4 hours, a total of 206 genes were upregulated by stachel activation compared to scrambled control and of these 80 (39%) where no longer upregulated, indicating cAMP-PKA dependency with 126 (61%) still upregulated after H89 pre-incubation, suggesting that the expression of these genes is potentially regulated independent of the cAMP-PKA pathway (Figure 6-9). Of the 308 genes downregulated by stachel activation at 4 hours, 187 (61%) were also downregulated after H89 pre-treatment with 121 (39%) no longer downregulated (Figure 6-9), suggesting PKA dependency.

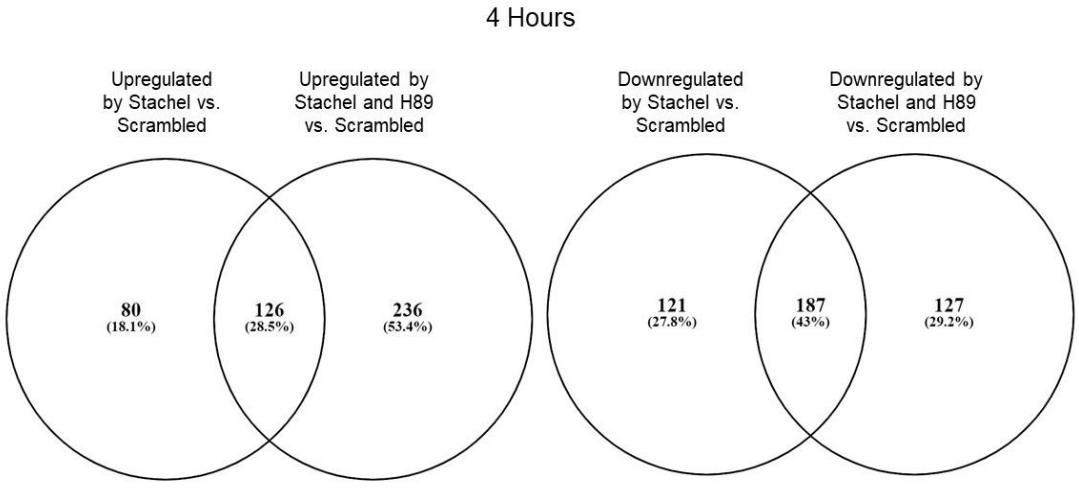


Figure 6-9 Overlap of genes which were differentially expressed after stachel peptide and those which were differentially expressed after stachel treatment with H89 pre-incubation, both compared to scrambled control after 4 hours. There are 126 upregulated and 187 genes downregulated in common between

the two conditions. 39% of the response for both up and down regulated genes was PKA dependent.

There were 236 and 127 additional genes up and down regulated respectively by treatment with stachel (4 hours) after H89 pre-treatment. These genes were not differentially expressed with stachel treatment. Therefore, it is probable that differential expression of these genes are due to other effects of PKA blocking or due to off target effects.

At 24 hours, out of the 170 genes upregulated by stachel peptide, 76 (45%) were no longer upregulated with H89 pre-incubation and of the 360 downregulated genes at 24 hours, 127 (35%) were no longer downregulated with H89 addition (Figure 6-10). There were 203 and 86 genes up and down regulated respectively by H89 pre-treatment prior to stachel treatment independent of those genes upregulated by stachel peptide alone.

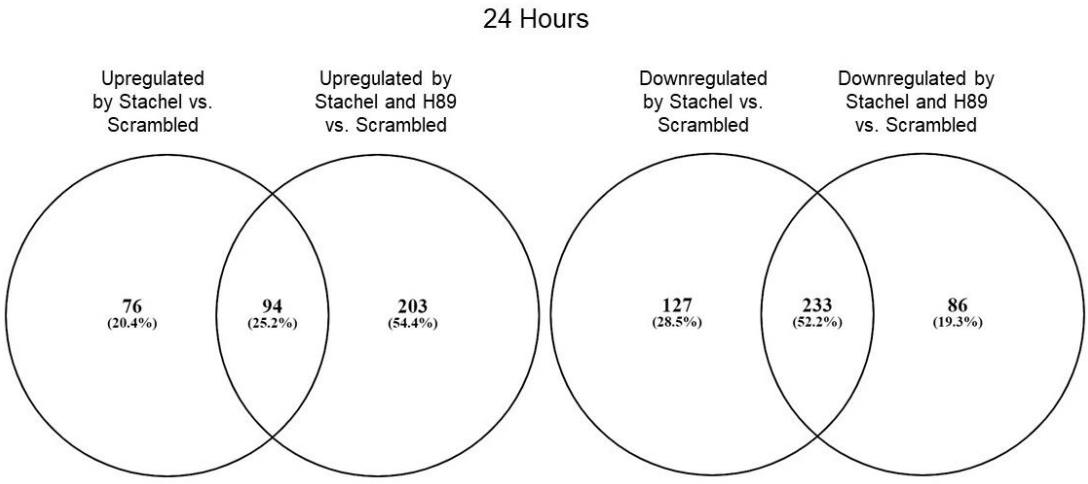


Figure 6-10 Overlap of genes which were upregulated and downregulated by stachel peptide and those which were upregulated after stachel treatment with H89 pre-incubation, both compared to scrambled control after 24 hours. There are 94 upregulated and 233 genes downregulated in common between the two conditions. 45% and 35% of upregulated and downregulated genes respectively were no longer upregulated with H89 pre-treatment.

6.3.6.1 *Top differentially expressed genes at both time points are mostly PKA independent*

The top differentially expressed genes, as identified by their expression being ≥2 fold difference after stachel treatment at both time points (Table 6-6 & Table 6-7), were then examined for whether their expression level changed after pre-incubation with H89.

Of the 13 genes upregulated ≥2 fold 4 (ZNF813, ADCY4, SLITRK6 and PLXNA4) were no longer significantly upregulated after 4 hours stachel treatment after H89 pre-treatment. At 24 hours 3 genes were no longer upregulated (PAQR5, ZNF813 and PDK4). The fact that 3 genes' upregulation was stopped at 4 hours but then upregulated again at 24 hours perhaps suggests a transient effect of the H89 pre-treatment. However, additional genes PAQR5 and PDK4 were no longer upregulated at 24 hours but not at 4 hours.

Gene	4 Hours		24 hours	
	Fold change after stachel	Fold change with H89 pre-incubation	Fold change after stachel	Fold change with H89 pre-incubation
PAQR5	5.5	3.2	3.7	**1.6**
ZBTB7C	4.2	2.7	4.0	2.9
ZNF813	3.5	**1**	4.5	**1.1**
TMEM100	3.5	3.7	4.0	3.3
CACNA1G	2.2	1.9	3.5	3.1
PDK4	2.3	1.9	3.2	**1.1**
IL33	3.0	4.9	2.4	2.8
ADCY4,RIPK3,RP11-934B9.3	2.7	**1.9**	2.5	4.3
LRRN3	2.1	1.8	3.0	3.1
SLITRK6	2.3	**1.4**	2.5	2.0
S1PR1	2.6	2.1	2.0	2.4
RASL12	2.1	3.4	2.4	3.2
PLXNA4	2.4	**1.1**	2.1	1.8

Table 6-11 Genes upregulated by 2 fold at both time points are mostly PKA independent. Genes in red are those which were no longer significantly differentially expressed after H89 pre-treatment.

Of the 13 genes downregulated at both time points ≥2 fold by stachel peptide versus scrambled control, 2 were PKA dependent (CTGF and SERPINE1) at 4 hours and at 24 hours only one gene (NEDD9) was no longer downregulated by H89 pre-treatment.

Gene	4 Hours		24 hours	
	Fold change after stachel	Fold change with H89 pre-incubation	Fold change after stachel	Fold change with H89 pre-incubation
AC018647.3	3.3	3.9	2.5	4.1
ATP8B1	5.4	3.8	2.5	2.6
AURKA	2.0	1.6	2.5	3.1
CTGF	3.9	**1.6**	2.6	3.0
EDN1	2.6	2.7	3.3	2.9
HSPB3	2.1	1.9	2.6	2.0
KRTAP1-5	4.5	2.3	3.0	2.9
NEDD9	3.0	3.1	2.3	**1.5**
OXTR	2.3	1.7	4.4	2.6
RTKN2	2.6	3.1	2.7	2.7
SEMA7A	2.2	2.3	2.3	2.6
SERPINE1	2.4	**1.3**	5.6	3.4
TNFRSF12A	2.3	1.5	2.1	2.4

Table 6-12 Genes downregulated by 2 fold at both time points are mostly PKA independent. Values in red represent the genes which were no longer significantly differentially expressed after H89 pre-treatment.

These data suggest that the response of airway smooth muscle cells to satchel is in part cAMP-PKA dependent and that the upregulated response is more dependent on this pathway than those genes which are downregulated.

6.3.6.2 H89 alone has an effect on gene expression

To assess the impact of the effect of the PKA inhibitor on gene expression, cells treated with H89 alone were compared to cells cultured in serum free media. A total of 263 genes were differentially expressed after pre-incubation with H89 after 4 hours. Of these, 174 were upregulated and 89 were downregulated. After 24 hours, 205 genes were differentially expressed with 62 of these downregulated and 143 upregulated.

Pathway analysis revealed that these genes were involved in a number of different cellular processes (Table 6-13). Upregulated pathways include NFKB and MTOR signalling pathways as well as pathways involved in inflammation. Downregulated pathways include c-myc pathways and pathways involved in cell cycle regulation.

Direction	Time point	Gene set	ES	NES	FDR q-val	FWER p-val
Up	4	HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.648	1.820	0.000	0.000
Up	4	HALLMARK_INFLAMMATORY_RESPONSE	0.638	1.781	0.001	0.001
Up	4	HALLMARK_KRAS_SIGNALING_UP	0.584	1.644	0.003	0.007
Up	4	HALLMARK_IL6_JAK_STAT3_SIGNALING	0.654	1.649	0.003	0.005
Up	4	HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.656	1.621	0.003	0.011
Up	24	HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.851	2.046	0.000	0.000
Up	24	HALLMARK_MTORC1_SIGNALING	0.598	1.619	0.004	0.007
Up	24	HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.543	1.455	0.059	0.160
Up	24	HALLMARK_ANDROGEN_RESPONSE	0.568	1.418	0.070	0.244
Up	24	HALLMARK_COMPLEMENT	0.504	1.358	0.123	0.464
Down	4	HALLMARK_PROTEIN_SECRETION	-0.505	-1.317	0.369	0.501
Down	4	HALLMARK_UNFOLDED_PROTEIN_RESPONSE	-0.495	-1.300	0.220	0.557
Down	4	HALLMARK_E2F_TARGETS	-0.431	-1.221	0.316	0.826
Down	4	HALLMARK_MYC_TARGETS_V1	-0.407	-1.151	0.448	0.964
Down	4	HALLMARK_MYC_TARGETS_V2	-0.467	-1.123	0.454	0.980
Down	24	HALLMARK_MYC_TARGETS_V1	-0.576	-1.691	0.007	0.007
Down	24	HALLMARK_G2M_CHECKPOINT	-0.535	-1.570	0.015	0.028
Down	24	HALLMARK_E2F_TARGETS	-0.521	-1.523	0.021	0.059
Down	24	HALLMARK_OXIDATIVE_PHOSPHORYLATION	-0.503	-1.466	0.026	0.094
Down	24	HALLMARK_NOTCH_SIGNALING	-0.564	-1.253	0.170	0.561

Table 6-13 GSEA results showing the top 5 up and down regulated gene sets at 4 and 24 hours. Gene sets were ranked firstly on q-value then on enrichment score.

These pathways show that PKA inhibition has an effect on the cells which may affect cell homeostasis and this information should come into consideration when analysing the effect of H89 on the GPR126 signal.

6.4 DISCUSSION

The main aims of this chapter were to determine differentially regulated genes and pathways in HASMs upon stachel activated GPR126 in order to provide new insight into GPR126 biology and to determine the cAMP-PKA dependency of this response.

GPR126 activation by stachel peptide resulted in 514 and 530 differentially expressed genes at 4 and 24 hours treatment respectively. There was a significant overlap of the transcriptome profiles at the two time points (~20%) providing confidence in the genes that were differentially expressed at both time points while highlighting there may be early and late response genes. In addition, the GPR126 signal is partly cAMP dependent with approximately 40% of the genes differentially expressed with stachel treatment no longer changed when the PKA inhibitor, H89, was used to pre-treat the cells.

GPR126 mRNA is expressed in HASMs and the level of expression increases over time in culture in all but one condition. Cells were seeded 3-4 days before experimentation and experiments were begun when cells were 80-90% confluent. The increase in expression may be due to the fact that the cells were more confluent and therefore there are more cell-cell and cell-matrix interactions. As the cells proliferate, more matrix is deposited and an increase in matrix may include known ligands for GPR126 such as laminin-211 and collagen type IV. aGPCRs are known to modulate a number of cellular functions including cell migration, adhesion, polarity and guidance via these adhesion interactions.[216] Moreover, it is also known that GPR126 is vital for normal tissue development via interactions with the ECM.[97, 98] Therefore, GPR126 may be upregulated to facilitate these functions when there is increased cell-cell and cell-matrix contact.

In addition, *GPR126* mRNA was elevated further with cells treated with the stachel peptide and that this elevation was reversed when cells were pre-incubated with a PKA inhibitor. This effect was only seen after 24 hours of treatment with the stachel peptide. This could potentially be explained by a positive feedback loop whereby signalling via the receptor results in increased

receptor expression. In addition, the fact that this effect was then reversed when cells treated with stachel were also pre-treated with H89 suggests this might be the case. In addition, this effect was seen with the aGPCR, GPR64 which can also be activated by the stachel sequence from GPR126. This may also represent a feedback loop for this receptor and indicate there may be off-target effects of stachel activation.

There were distinct transcriptome profiles at the two time points with a ~20% overlap of genes. Therefore while the two time points may identify early and late response genes, there is increased confidence in the genes differentially regulated at both time points are real effects of GPR126 activation. These genes are the focus of this discussion as well as those with known roles in airway biology.

There were 61 genes upregulated at both time points with 13 genes upregulated by ≥2 fold (*PAQR5, ZBTB7C, ZNF813, TMEM100, CACNA1G, PDK4, IL33, ADCY4/RIPK3, LRRN3, SLITRK6, S1PR1, RASL12,* and *PLXNA4*). There were 13 genes downregulated by stachel activation ≥2 fold (*AC018647.3, ATP8B1, AURKA, CTGF, EDN1, HSPB3, KRTAP1-5, NEDD9, OXTR, RTKN2, SEMA7A, SERPINE1* and *TNFRSF12A*) compared to a scrambled control.

Of these genes, *ZNF813, ADCY4/RIPK3, SLITRK6* and *PLXNA4* showed PKA dependency at 4 hours and only ZNF813 maintained this effect at 24 hours. Additional genes PAQR5 and PDK4 showed PKA dependency at 24 hours but not at 4. There were fewer downregulated genes which showed PKA dependency. At 4 hours only *CTGF* and *SERPINE1* were no longer downregulated after H89 pre-treatment however at 24 hours both of these genes were significantly downregulated even with H89 pre-treatment. At 24 hours only *NEDD9* showed PKA dependency.

In chapter 4, H89 pre-treatment did not attenuate the CRE response from HEK293 cells treated with stachel peptide. However, response to isoprenaline was attenuated indicating that the H89 is functioning to block the cAMP-PKA

signal. Therefore, potentially GPR126 may signal via cAMP but via pathways independent of PKA, such as via Epacs.[217]

Chronic inflammation of the airways is a key feature of COPD and the expression of chemokines and cytokines is important in regulating airway inflammation.[181] Importantly, COPD is associated with an abnormal inflammatory response of the lung to toxic particles or gases. Airway smooth muscle has been found to secrete a variety of chemokines and cytokines to perpetuate airway inflammation. [181] The data presented in this chapter suggest that activation of GPR126 could play a role in airway inflammation.

A single pathway was significantly upregulated by GPR126 activation, which was the beta defensins pathway. Beta defensins are antimicrobial peptides that are involved in the resistance of epithelial surfaces to microbes. Airway smooth muscle cells may be involved in regulating inflammatory responses via production of inflammatory mediators such as cytokines and growth factors and human beta defensins have been shown to induce IL-8 release and induce apoptosis in airway smooth muscle.[218]

IL33 (upregulated at both time points) is a well characterised cytokine that plays an important role in several chronic inflammatory diseases by promoting inflammatory cytokine production.[219] Serum levels of IL-33 and its receptor ST2 are elevated in patients with COPD compared to control subjects and peripheral blood lymphocytes from COPD patients release more IL33 in response to stimulus than healthy controls. [220]

An additional gene only upregulated at 4 hours was *ADAM8.* This gene is involved in cell-cell and cell-matrix adhesion interactions. This gene has been shown to be associated with asthma and airway inflammation.[221, 222] *ADAM8* deficiency increases cigarette smoke induced lung inflammation, emphysema and mucus cell metaplasia indicating an anti-inflammatory effect of this gene. [223]

Inflammation and airway remodelling are two key processes in the pathogenesis of COPD. Airway remodelling results from a wide range of mechanisms involving smooth muscle cell hyperplasia, increased activation of

fibroblasts and myofibroblasts, increased deposition of the ECM and aberrant inflammatory responses. Airway smooth muscle undergoes alterations in health and disease characterised by cell proliferation, hypertrophy, and migration, and the expression of chemokines and cytokines is important in regulating airway inflammation.[181] Furthermore, the smooth muscle mass is altered in COPD due to hypertrophy and hyperplasia. [181] Two genes in particular suggest a role for GPR126 in airway remodelling.

Firstly, CTGF is a gene downregulated by GPR126 activation which has roles in inflammation and tissue remodelling. This ECM protein recruits fibroblasts to wound sites, along with other inflammatory mediators such as TGF-β and IL-8/CXCL8. CTGF is not constitutively expressed in resting lung fibroblasts but is overexpressed during wound repair by fibrotic mediators such as TGF-β, thrombin and ET-1.[224] *CTGF* has a known role in pulmonary fibrosis and airway remodelling and has been shown to modulate a number of signalling pathways involved in cell adhesion and migration, angiogenesis, myofibroblast activation, and extracellular matrix deposition, which together can result in tissue remodelling and fibrosis. [225] Elevated CTGF expression contributes to expression of ECM proteins, cell migration, and the myofibroblastic phenotype in tissue repair.[224] In mouse lung tissue, CTGF expression and promoter activity is elevated after bleomycin challenge. [226] CTGF exerts profibrotic effects and contributes to lung fibrosis via transcriptional regulation of COL1A2. [226]

Secondly, *SERPINE1* (PAI-1) (downregulated by GPR126 activation) is upregulated in emphysema lung and increased expression of PAI-1 suppresses fibrinolysis leading to pathological fibrin deposition and tissue damage.[227, 228] Moreover, PAI-1 knockout mice display an emphysema-like phenotype, possibly due to an imbalance in protease-antiproteases activity in the lung – a known factor in emphysema development.[205] PAI-1 is upregulated in the lung during hypoxia [229], after exposure to LPS or cigarette smoke extract [230] and after oxidative stress via NF-κB. [231] Elevated levels of PAI-1 are also observed in the sputum of COPD patients and these levels were more elevated in more severe COPD cases. [231, 232] PAI-1 plays a critical role

in fibrosis development in a number of tissues including in the lung [233] and is produced by smooth muscle cells as well as endothelial cells and fibroblasts.[227] PAI-1 expression is regulated by inflammatory cytokines, growth factors and hormones.[227]

PAQR5 (upregulated at both time points) is a membrane bound progesterone receptor which is expressed in mouse lung. [234, 235] There has been a role suggested for its ligand, progesterone in the lung. Interestingly, progesterone has been shown to inhibit TGF-β1 and block the subsequent upregulation of CTGF and PAI-1 (both genes are also downregulated by GPR126 activation) in lung epithelial cells. [236] The TGF-β-Smad signalling pathway is involved in lung development as well as being a key pathway in airway inflammation and airway remodelling.[237] TGF-β is involved in normal tissue repair following lung injury but its expression is increased in a number of pulmonary pathologies including acute and chronic lung diseases. [237] Upregulation of PAQR5 by GPR126 activation may therefore result in downregulation of these TGF-β pathways and may explain the downregulation of CTGF and PAI-1 (*SERPINE1*) by GPR126 activation. These potentially effected pathways may present a role for GPR126 in airway inflammation and remodelling.

Collectively, these differentially expressed genes and pathways highlight a potential novel role for GPR126 in regulating the connected processes of airway remodelling and airway inflammation – both key processes in the pathophysiology of COPD. While it is unclear whether activation of GPR126 results in enhanced inflammation or remodelling, aberrant GPR126 signalling or expression may contribute to an imbalance of these pathways leading to disease processes.

As well as potential roles in inflammation and remodelling, GPR126 activation results in the differential expression of genes and pathways involved in cell proliferation and cell cycle control. GPR126 activation predominantly resulted in the downregulation of pathways involved in proliferation and cell cycle regulation. One downregulated pathway was “TNC targets”. TNC codes for tenascin-C which is an adhesion-modulating ECM molecule that is highly expressed in tumour stroma and stimulates tumour cell proliferation. [238]

Similarly, EGFR signalling is an important pathway for cell growth, survival, proliferation and differentiation. [239] Moreover, dysregulation of EGFR signalling plays roles in airway disease and plays an important role in epithelial changes in the bronchial epithelium in COPD possibly via regulation of mucus production. [240, 241] Other pathways downregulated include the mTOR pathway, involved in cell cycle control [242], E2F signalling, which is a transcription factor which regulates the G1/S transition in the cell cycle and Myc signalling which is an oncogenic transcription factor.[243]

A number of differentially expressed genes are involved in cell proliferation and the cell cycle. *ZBTB7C* (upregulated) codes for a transcription factor that regulates cell proliferation via increased glutamine metabolism. [244] PDK4 (upregulated) codes for a mitochondrial protein which contributes to the regulation of glucose metabolism which is important for cell growth. [245] *PPARGC1A* is similarly upregulated at both time points and is also involved in metabolism and mitochondrial function. NEDD9 (downregulated) is a protein involved in a number of signalling pathways involved in cell proliferation, survival and migration. [211, 212] It is a mediator of integrin signalling, which is involved in cell movement and adhesion to the ECM. [246] Interestingly, NEDD9 binds to Aurora A kinase (another gene downregulated by GPR126) at the centrosome, promoting its activity and allowing cells to enter mitosis. [247] Aurora A kinase is coded by AURKA, a mitotic serine/threonine kinase which is involved in mitosis and meiosis and is essential for cell proliferation. [210]

Furthermore, a number of the genes upregulated by GPR126 activation include a number of other membrane receptors. *CACNA1G* codes for the calcium gated channel unit subunit alpha 1 G. These calcium channels are involved a variety of calcium-dependent processes including muscle contraction, cell motility, cell division and cell death. S1PR1 is a receptor which is a key regulator of angiogenesis and has a role in cell adhesion.[206] It is also essential for limb development and for embryonic cardiac development – two developmental processes GPR126 has previously been shown to be involved in. [81, 103, 248]

These genes highlight potential roles for GPR126 in airway smooth muscle function as well as roles in cell proliferation and migration. Proliferation and migration of airway smooth muscle are key features of airway remodelling and downregulation of these pathways by GPR126 activation may result in increased airway smooth muscle mass, a predominant feature of airway remodelling in COPD.

The fact that the response to stachel peptide was only partly (~40%) cAMP-PKA dependent suggest that GPR126 can signal via other signalling pathways. While these pathways are currently unclear, it is possible that the genes which were not PKA dependent could be regulated by the N-terminal fragment of GPR126, which has known signalling capabilities independent of the cAMP pathway or via alternate cAMP pathways independent of PKA. [170, 217]

The limitations of this study are firstly, H89 is not specific to PKA inhibiting a number of other kinases such as S6K1, MSK1, ROCKII, PKBα and MAPKAP-K1b.[87] Furthermore, GPR126 was activated by the stachel peptide – which is an exogenous peptide and not a classic ligand for the receptor. Whether this represents the physiological conditions that GPR126 would be activated by, and whether this would affect the transcriptome profile is something to consider. There is also a caveat that aGPCR tethered agonist sequences, like the stachel, can cross-activate other aGPCRs due to the highly conserved sequence.[172] The GPR126 stachel sequence can activate another aGPCR of the same family, GPR64, however RNA-seq data shows that GPR64 is not expressed in these samples.

My original hypothesis was that Ser123Gly, is the causal variant underlying the association signal. The functional implications of the Ser123Gly (c.367A>G) substitution are unknown but genetic association studies shows that the Ser variant (common allele (A), frequency = 97.3%) is associated with decreased lung function (beta = -0.1803) and is the risk allele for COPD.[56] Therefore, the effects of this variant may be deleterious to GPR126 function (as predicted by *in silico* analysis) but protective for lung function. Moreover, GPR126 is elevated in lung tissue from patients with COPD. It is therefore

plausible that loss of function of GPR126 may be beneficial and increased expression of GPR126 or over-activity results in differential expression of pathways involved in COPD development.

Previous work in this thesis did not identify a detrimental effect on the cAMP signalling capacity of GPR126 due to the Ser123Gly variant. As previously discussed, the mode of activation by the stachel peptide does not depend upon the binding region where the Ser123Gly variant resides which may be masking the effect of the variant on cAMP signalling. Without knowing the effect of this variant on GPR126 signalling, it is difficult to speculate on the process by which it may be involved in lung function decline.

For example, alterations in GPR126 signalling/ expression could result in genes such as *CTGF* and *SERPINE1* no longer being downregulated or downregulated to a lesser extent. Elevated expression of these genes is associated with pulmonary diseases such as emphysema and COPD and particularly airway remodelling. Therefore deregulation of these pathways by GPR126 loss of function may be a processes by which this variant results in reduced lung function. Similarly, pathway analysis revealed that GPR126 predominantly downregulates pathways involved in cell proliferation and cell cycle control. If cell cycle proliferation is usually downregulated by GPR126, then a loss of function may result in increased smooth muscle cell proliferation – a feature in airway remodelling.

GPR126 activation also lead to induction of genes and pathways involved in the inflammatory response. IL33 was upregulated by GPR126 and is expressed at higher levels in asthma and COPD, having a key role in airway inflammation. Loss of function of GPR126 therefore might result in reduced IL33 expression, which may suggest a protective effect.

Together, these data suggest roles for GPR126 in mediating the inflammatory response, ECM regulation and cell proliferation potentially leading to airway remodelling. Deregulation of these pathways via GPR126 loss of function could potentially lead to airway disease. Clearly, much of this data is

hypothesis generating and further studies would be needed to follow up these findings and validate the data.

7 General Discussion

7.1 HYPOTHESIS AND AIMS

At the commencement of this project in late 2015 there were well elucidated roles for GPR126 in a number of developmental processes, particularly in peripheral nerve development.[82, 97, 102, 104] Moreover, roles for the receptor in inner ear and cardiovascular development via interactions with components of the extracellular matrix (ECM) had been identified as well as roles in modulating endothelial cell proliferation and migration through VEGF signalling.[81, 97] By the beginning of this project, collagen type IV, laminin-211 and a number of peptides had been identified as potential ligands for the receptor.[97, 98, 171]

It was also known that adhesion GPCRs function by interacting with adjacent cells and components of the ECM to play important roles in signalling pathways involved in a diverse array of biological processes.[249] The N-Terminal Fragment (NTF) of these receptors enables them to facilitate many cellular functions such as cell adhesion, orientation, and positioning during development, immune responses, and tumour growth.

By late 2015, a number of GWAS had identified associations between genetic variants at the GPR126 locus with lung function.[49, 51, 54] A novel signal at this locus (rs148274477, MAF=3%) was found to be associated with FEV_1/FVC and in high LD (r^2=0.7) with missense variant rs17280293 (Ser123Gly) in GPR126.[54] However, the mechanism driving this association was unknown.[54]

Genetic variant rs17280293 results in amino acid substitution Ser123Gly and is predicted to be damaging to protein function by *in silico* analysis. Therefore, the primary hypothesis of this project was that Ser123Gly may be causative due to altered function of GPR126 and this may be protective with regards to COPD.[54]

Furthermore, I hypothesised that altered GPR126 function or expression may contribute to airway remodelling based on the receptor's known interactions with important components of the basement membrane, collagen type IV and

laminin-211 as well as having known roles in development and repair pathways.[97, 98]

The overarching aim of this project was to identify the driving mechanisms underlying the association signal at GPR126 for lung function and determine the functional relevance of GPR126 and genetic variants in GPR126 to lung biology and COPD.

7.2 MAIN FINDINGS

The work in this thesis identifies GPR126 as the likely causative gene underpinning the 6q24 association signal with lung function. Ser123Gly (r^2 =0.9262 with sentinel SNP) was identified as a credible variant driving this association signal. This was the only coding region variant out of 9 total variants in this LD block and was predicted to be damaging to protein function. Ser123Gly is also an eQTL for GPR126, resulting in decreased expression and the Gly variant is associated with reduced FVC in PheWAS.[250, 251] All other SNPs within this LD block were in non-coding regions of the genome and none were found to be eQTLs for GPR126 or any other genes in any of the eQTL databases searched. This lead to the hypothesis that GPR126 loss of function (or reduced expression), via this mutation, may contribute to reduced lung function and COPD risk.

GPR126 mRNA expression was identified in human lung and in cultured structural airway cells HBECs and HASMs, in concordance with data from GTEx and Protein Atlas which both show high GPR126 mRNA expression in the lung. [142, 156] GPR126 protein expression was identified in the lung using immunohistochemistry, although there was donor variability (n=3).[156] GPR126 is expressed highly around blood vessels and in smooth muscle bundles around the airways. These findings are in line with the literature; GPR126 is known to be highly expressed in endothelial cells and in muscle tissues.[147, 156] In endothelial cells, GPR126 controls cell proliferation, migration, and tube formation [81] where there is no reported role identified in airway smooth muscle.

GPR126 expressing recombinant cells responded to the stachel peptide, a tethered agonist for the receptor, resulting in a ~10 fold increase in cAMP driven luminescence indicating an increase in cAMP driven gene transcription.[96] However, a number of potential GPR126 ligands, including collagen type IV and laminin-211 failed to activate the receptor in this recombinant cell system.[97, 98, 171] This is in contrast to what is reported in the literature and could be down to a number of reasons, i) collagen type IV was

identified as a ligand using the zebrafish clone of GPR126, not human ii) the splice variant I expressed, which was the most abundant in the lung, lacks exon 6 of the full length transcript and iii) reported experiments performed with a human clone only showed very modest responses to ligands, some of which did not reach statistical significance. [97] This suggests that these ligands may not be activating ligands for GPR126 and there may be other as yet unknown conditions/ factors required for GPR126 signalling in humans.

Furthermore, GPR126 activation by the stachel peptide was not genotype dependent. This was true in regards to GPR126 baseline expression and GPR126 signalling activity once activated with stachel peptide. While it may be that the variant is not damaging to protein function, I did not necessarily anticipate to see a difference in signalling activity between the two genotypes due to the expectation that stachel activation bypasses the N-terminal binding domains. The stachel sequence binds downstream (C-terminal) of the Ser123Gly variant which lies in the N-terminal of the receptor. Therefore, if the Ser123Gly variant, which lies in the binding domain thought important for collagen type IV interactions, does affect GPR126 function, it might be expected that this would be via impaired ligand interaction and therefore this effect may be bypassed by stachel activation.

I identified that GPR126 is functional on airway smooth muscle and influences multiple pathways of relevance to lung homeostasis. To define GPR126 activity and potential roles in airway structural cells, which would help provide insight into potential roles of GPR126 in human lung biology, the stachel peptide was used to activate endogenous GPR126 in HASMs and HBECs. Stachel activation resulted in an increase in cAMP after 1 hour treatment in HASMs (~20 fold increase) with a smaller response observed in HBECs (~2 fold).

Novel potential functions for GPR126 in the lung were identified by transcriptome analysis of HASMs activated by the stachel peptide. Genes upregulated by GPR126 activation at both time points are involved in inflammatory pathways (IL33) and downregulated genes include CTGF and SERPINE1, which both have known roles in fibrosis and tissue remodelling.

Gene set enrichment analysis revealed that beta defensin pathways were upregulated by GPR126 activation where genes involved in pathways involved in cancer, cell cycle checkpoints and cell proliferation were downregulated. Altered GPR126 may therefore lead to increased smooth muscle mass due to proliferation (a feature of airway remodelling) and increased expression of genes such as *SERPINE1* which is found to be elevated in COPD patients and CTGF, which contributes to inflammation and tissue remodelling.

Furthermore, GPR126 signalling was found to be only partly cAMP-PKA dependent, with other signalling pathways potentially involved in downstream signalling. The literature on GPR126 is focused primarily on cAMP signalling via $G_{\alpha s}$ and $G_{\alpha q}$, however it is known that the N-terminal fragment, once cleaved, can signal independently of the C-terminal membrane bound fragment.[96-98, 102, 104, 170] This may explain the large number of genes that were still differentially expressed when PKA was blocked. Alternatively, data from chapter 4 (Figure 4-19) showed no decrease in CRE-luciferase in recombinant HEK293 cell lines incubated with H89 while activation of cells with isoprenaline decreased the CRE-luciferase response by ~70%. This may suggest that GPR126 primarily drives transcription independently of PKA, which is an avenue for further research.

To summarise, the work in this thesis has identified a potential casual missense variant associated with reduced lung function. This SNP, Ser123Gly, is in high LD with the sentinel SNP from lung function GWAS' and acts as an eQTL for GPR126 resulting in reduced expression of the gene. I anticipate the Ser123Gly results in altered function of the gene due to its disruptive presence in a ligand binding domain and that loss of function/ reduced expression of GPR126 results in differential regulation of genes/ pathways known to play roles in COPD. Moreover, I identified genes/ pathways regulated by GPR126 activation in primary human airway cells and identified potential pathways by which GPR126 altered function may result in COPD.

7.3 STUDY LIMITATIONS

While *GPR126* mRNA expression was identified in lung tissue, HBECs and HASMs, I was hindered in confidently detecting protein expression due to technical difficulties and a lack of antibodies for human GPR126. However, I am be confident that GPR126 was present on the cell systems used based on the assays showing a response to GPR126 agonist.

One primary limitation of this thesis is that amongst the genetic variants in LD ($r^2 > 0.2$) with the sentinel SNP only the Ser123Gly variant was assessed for potential functional effects on GPR126. Coding region variants present obvious targets for follow-up studies to GWAS as they make up a disproportionate amount of heritability given that the exome makes up only 1.5% of the genome.[68] This means that coding region variants have an enhanced probability of being causal which leads to the assumption that coding region variants identified through GWAS are causal – which may be inaccurate.[68] While a risk variant may be linked with a causative coding variant, it is also likely that variants in non-coding regions have direct functional consequences on nearby genes.[68, 252] In fact, it has recently been suggested that coding region variants previously thought to be casual may represent false leads.[68] Moreover, non-coding region variants can and do play an important role in human traits and disease, with most of the GWAS signals mapping to non-coding regions. [74, 75, 252] Therefore by only looking at protein coding variant, rs17280293 (Ser123Gly) it is possible that I have overlooked potentially causal variants which may be regulating other mechanisms such as gene expression/ alternative splicing.

Additionally, when looking for potential causal variants, I used an r^2 cut off value of 0.2 which excludes missense variant rs11155242 (Lys230Gln) (MAF=2.1%). This variant is in low r^2 with the sentinel SNP (0.1442) but with a D' = 1. The variant lies within the PTX domain, thought important for ligand interaction, but it is predicted to be benign by PolyPhen-2 and Mutation Taster. However, this SNP is an eQTL for GPR126, suggesting that it may contribute to this association signal via altered expression of GPR126.[146] Again, this

highlights that there are other variants that may have contributed to this association signal that were not fully considered in this study.

A major limitation of this thesis is that reported ligands did not activate the receptor in the recombinant cell systems I established. As previously discussed, this then presented difficulties in identifying functional effects of the Ser123Gly variant in regard to ligand binding and signalling activity. The stachel-mediated GPR126 response did not identify a genotype-dependent response in the recombinant cell lines. As explained above, this may be due to the mechanism of stachel activation, which binds downstream of the Ser123Gly substitution. Therefore, it is still unclear as to whether this SNP alters GPR126 function or expression and therefore whether this alteration has an impact on lung function.

One explanation as to why there was no response from these ligands may be due to the isoform overexpressed. The HEK293 cells were engineered to overexpress the alpha 2 isoform, which is the most abundant in the lung based on GTEx RNA-seq data but lacks exon 6 of the full length protein.[142] This spliced region may be important in binding ligands such as collagen type IV or laminin-211. Therefore to assess the full impact of the Ser123Gly substitution, it may have been beneficial to overexpress all four full length protein coding transcripts or at least those predominantly expressed in HASMs (a1) and HBECs (b1), however this data was not available at the beginning of the project. This would also have the added benefit of making the responses seen from the recombinant model and the primary cells more comparable.

Furthermore, the stachel agonistic sequences have been shown to be promiscuous between closely related adhesion GPCRs, which is a potential limitation. [172] While I attempted to take this into account by confirming that the one identified aGPCR (GPR64) that GPR126's stachel sequence could activate was not expressed in HASMs, it may still be possible that the stachel peptide has activating effects on other closely related aGPCRs.

The limitations of the RNA-seq chapter are that firstly, H89 is not specific to PKA. H89 can inhibit a number of other kinases such as S6K1, MSK1, ROCKII,

PKBα and MAPKAP-K1b.[87] PKA is an important signalling molecule in a number of cellular processes and inhibition of it is clearly going to have other effects on the cell which I would expect to be reflected in the transcriptomic response and I have consequently tried to take into account. Moreover, GPR126 was activated by the stachel peptide – which is an exogenous peptide and not a classic ligand for the receptor. Whether this represents the physiological conditions that GPR126 would be activated by, and whether this would affect the transcriptome profile is something to consider.

7.4 FUTURE WORK

Future work would prioritise optimisation of protein assays for western blotting, immunohistochemistry and immunofluorescence which was a technical limitation of this work. This would allow me to expand my analysis to additional cell/ tissue donors from a range of different disease states or lung function measurements. This would allow me to observe whether GPR126 protein levels correlate with disease/ lung function phenotype. As identified in chapter 3 (Figure 3-4) *GPR126* mRNA levels in lung tissue are modestly increased in COPD patients.

As there was no functional effect observed from the Ser123Gly mutation, further work may also include more in-depth analysis of the other variants within the LD block driving the association signal with FEV_1/FVC. This could include similar experiments to those described in this thesis such as overexpression of variant rs11155242 (Lys230Gln). Therefore, using a more structured approach to identify potentially causal SNPs, e.g. identifying a 95% credible set of variants would provide more insight into the potential causality of SNPs in this association signal. These methods have been used previously in GWAS and there are a number of approaches to take. A common approach is to use Bayesian fine-mapping.[253] In this framework, the evidence for association is measured using a Bayes Factor which is used to assign posterior probabilities of causality to each SNP. [253] This would assist in prioritising variants in the association signal for follow-up analyses.

To address whether reported ligands did not activate signalling due to the isoform expressed in the recombinant cell system, all 4 'full-length' transcripts should be expressed to identify if different isoforms have different signalling capacities or if the Ser123Gly variant has a different effect on different isoforms. As the predominant isoforms present in HASMs and HBECs (a1 and b1 respectively) is different from the one transfected into the recombinant cells, it is possible that different isoforms may respond differently to activation from different ligands. This may help to answer the question of why there was no response to 'known' GPR126 ligands and also may help identify functional consequences of the Ser123Gly mutation. A useful addition to this may be to test the ligands such as collagen type IV and laminin-211 which did not activate GPR126 in HEK293 cells on primary human cultures. This would provide useful context for the recombinant work also.

The RNA-seq data provides an avenue for future directions for this project. Firstly, validation of the RNA-seq results would be required. This should include qPCR analysis of the top upregulated and downregulated genes to replicate and provide confidence in the findings.

Additional experiments to compliment the RNA-seq data should include further functional studies on the receptor. From the RNA-seq data, modulation of GPR126 activation changes levels in expression of genes involved in cell proliferation and cell cycle control as identified through GSEA. Therefore a number of assays which look at cell proliferation, metabolic activity[254] and migration[255] may be useful to perform on cells activated with the stachel peptide.

To complement these additional experiments, it would be informative to knock-out GPR126 in primary human cells using a method such as CRISPR/cas9.[256] It would also be useful to assess the impact of this knockdown on cell properties as described above. In addition, assessment of cell adhesion, which can be measured using ECIS (Electric cell-substrate impedance sensing) would be informative of GPR126's function as an adhesion receptor.[257] It will also be beneficial to assess genes and pathways that are dysregulated after GPR126 knock out to compare with the activated receptor data and this could

be achieved with additional RNA-seq analysis or targeted analysis of the genes of interest by qPCR. In addition, it would be useful to use CRISPR to introduce SNPs from the association signal into primary human cells, or alternatively genotype cells, to look at whether these SNPs effect GPR126 function/ expression in primary human airway cells.

Moreover, to assess the role of GPR126 in inflammatory pathways, activating the receptor with the stachel peptide and measuring inflammatory cytokines such as IL33, which was upregulated by GPR126 activation, via ELISA may give insights into potential inflammatory pathways GPR126 is involved in and would validate some RNA-seq findings.

While I identified that part of the GPR126 driven response was cAMP-PKA independent, I did not examine other signalling pathways in detail. Some preliminary work suggested that GPR126 activation may result in a modest increase in intracellular Ca^{2+} (Bo Liu, unpublished data) however further study of additional signalling pathways such as EPACs may prove important to determine the true PKA dependency of GPR126 signalling.

7.5 FINAL CONCLUSIONS

The data presented in this thesis has suggested a number of potential roles for GPR126 in airway biology and disease processes, which will require validation and follow-up analyses. Bioinformatic analysis identified a potential causal variant which I hypothesised was driving the association signal with FEV_1/FVC. Moreover, due to known roles of GPR126 and other aGPCRs, I hypothesised that altered GPR126 function/ expression may alter processes involved in airway remodelling and contribute to disease.

One key aim of the thesis was to identify potential functional effects of the Ser123Gly missense variant which was not achieved due to difficulties finding ligands for the receptor. However, GPR126 is expressed and active in human structural airway cells and is capable of signalling via cAMP mediated pathways. RNA-seq analysis of HASMs activated with the stachel peptide identified a number of genes and pathways that were differentially regulated by GPR126. Approximately 60% of these genes were still upregulated with the inclusion of a PKA inhibitor suggesting that GPR126 may signal via other pathways. This may include its N-terminal fragment which is capable of signalling independently of its role as a GPCR. [170] Differentially expressed genes were involved in pathways such as inflammation, cell proliferation, cell cycle control and interactions with the ECM, including two genes, SERPINE1 and CTGF, with known roles in airway remodelling. [225, 233]

aGPCRs have known roles in a diverse array of biological processes including roles in the nervous system, in which GPR126 plays an important role in myelination and axon sorting, cancer, the immune system, and cardiac and immunoskeletal systems.[90] However, roles for aGPCRs that are expressed in the lung, of which there are several, are largely unknown.[90] Many aGPCRs are not ubiquitously expressed but are distributed in selected tissues playing specific biological functions in these tissues. Indeed, most aGPCRs are expressed at low levels in HASMs and HBECs in the lung (Table 5-1) with GPR126 being the one of the most highly expressed in both cell types. Other highly expressed aGPCRs in HASMs include GPR124, GPR125, BAI1,

GPR133, CD97 and ADGRL2. Many of these genes have roles in angiogenesis (GPR124, GPR125, BAI1) as does GPR126[81] and CD97 is involved in cell adhesion and leukocyte recruitment and migration but none have known roles in airway smooth muscle.

Of the aGPCRs that are expressed in the lung only GPR116 has been studied in detail. GPR116 is expressed in alveolar type I and II epithelial cells and pulmonary endothelial cells but not in the bronchial epithelium.[188] GPR116 is an essential regulator of alveolar surfactant homeostasis by regulating surfactant secretion and reuptake in alveolar type II (AT2) cells by $G_{q/11}$ signalling pathways.[188] Additionally, deletion of GPR116 results in the development of airway inflammation and this is possibly via its role in the regulation of genes encoding CCL2 in lung endothelial cells.[189] ADGRC1 is a gene required for normal lung branching morphogenesis and mutants show disruption of epithelial integrity and aberrant cytoskeletal remodelling.[192] ADGRC1 is also expressed in the epithelium during lung development.[192] Based on the roles of adhesion GPCRs in developmental biology and known roles of certain receptors in the lung GPR126 and other aGPCRs are likely to be via interactions with components of the ECM to mediate cell adhesion, orientation, and positioning during development, immune responses, and tumour growth.[169] Much of this data from the literature correlates with the findings from the RNA-seq data from activated GPR126 which signals roles for GPR126 in the immune system, regulation of the ECM and cell proliferation.

While further studies are required to fully define the role of GPR126 in the human lung in health and disease, this thesis has identified potential novel areas for GPR126 biology. Genetic analysis suggest that the Ser123Gly variant may be driving the association signal with lung function and COPD via altered GPR126 function. While functional studies could not back up this hypothesis, potential roles identified for GPR126 in airway smooth muscle suggest GPR126 may be involved in cell proliferation, inflammation as well as regulatory roles of genes involved in fibrosis. This thesis has led to novel understanding of the pathways GPR126 is involved in and highlight potential pathways where GPR126 loss of function may contribute to airway disease.

8 Appendices

8.1 LIST OF REAGENTS

8.1.1 General Reagents

Material/Reagent	Supplier	Catalogue Code
Sterile Water(bottled)	Baxter	UKF7114
Ethanol	BDH	K3954
Human recombinant laminin-211	Biolamina	LN211-02
dNTPs	Bioline	39026
Dowex	Bio-Rad	142-1351
Dako REAL EnVision Detection System Peroxidase/DAB, Rabbit/Mouse	Dako	K406511-2
Peroxidase-Blocking Solution	Dako	S202386-2
Ethidium Bromide 10mg/ml.	GIBCO	15585-011
Adenosine 3',5'-cyclic phosphate, ammonium salt, [8-14C]-	Hartmann Analytic	MC157
Rb IgG Isotype control antibody	Invitrogen	10500C
Superscript First Strand Synthesis System for RT-PCT kit	Invitrogen	11904018
Monarch® Plasmid Miniprep Kit	New England Biolabs	T1010
Phosphate buffered saline (Dulbecco A) PBS tablets	Oxoid	BR0014
Adenine, [2,8-3H]-,>97%, 5mCi (185MBq)	Perkin Elmer	NET063005MC
Scintilation Fluid Ultima Gold XR	Perkin Elmer	6013119
Fugene 6	Promega	E2691
Qiagen Plasmid Endofree Maxi Kit	Qiagen	12362
Histoclear	Raymond A Lamb	C78-G
6x Orange G loading buffer	Sigma Aldrich	O1625
Alumina	Sigma Aldrich	A9003
Ampicilin	Sigma Aldrich	A-9518

Anti-GPR126 antibody produced in rabbit	Sigma Aldrich	HPA017346
ANTI-GPR126, C-TERMINAL	Sigma Aldrich	SAB4502469
Anti-Rabbit IgG (whole molecule)–FITC antibody produced in goat	Sigma Aldrich	F0382
Collagen from human placenta Bornstein and Traub Type IV, powder	Sigma Aldrich	C7521
Forskolin	Sigma Aldrich	F6886
GenElute™ Mammalian Total RNA Miniprep Kit	Sigma Aldrich	RTN70
IBMX	Sigma Aldrich	I5879
Imidazole	Sigma Aldrich	I0125
Isoprenaline	Sigma Aldrich	I5627
Mayers haematoxylin	Sigma Aldrich	51275
On-Column DNase 1 digestion set	Sigma Aldrich	DNASE70-1SET
Water, for molecular biology, DEPC-treated and sterile filtered	Sigma Aldrich	95284-100ML
1 Kb Plus DNA Ladder	Thermo Fisher Scientific	10787018
18s PDAR	Thermo Fisher Scientific	4310893E
BigDye™ Terminator v3.1 Cycle Sequencing Kit	Thermo Fisher Scientific	4337455
DNase I, Amplification Grade	Thermo Fisher Scientific	18068015
GAPDH PDAR	Thermo Fisher Scientific	4310884E
GPR126 PDAR (Hs0189210-m1)	Thermo Fisher Scientific	4331182
LB agar	Thermo Fisher Scientific	BPE1426-500
LB broth	Thermo Fisher Scientific	BPE1425-500
Taq DNA Polymerase	Thermo Fisher Scientific	18038020

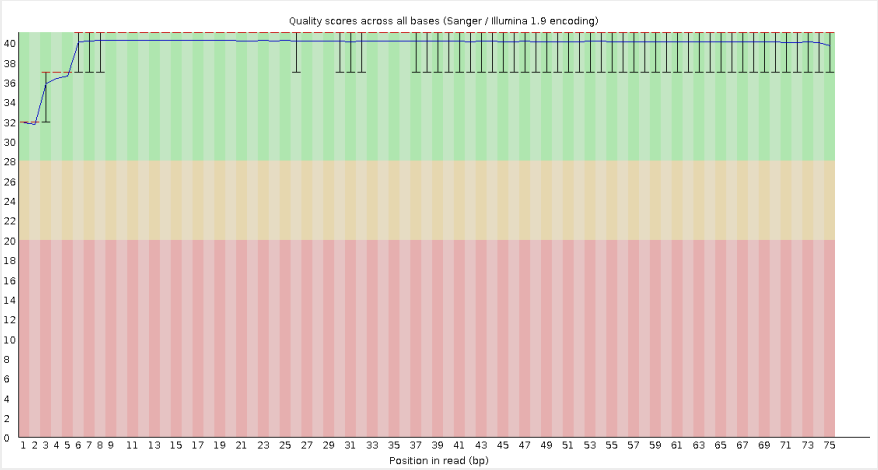
TaqMan® Gene Expression Master Mix	Thermo Fisher Scientific	4369016
UltraPure™ Agarose	Thermo Fisher Scientific	16500500
Vectamount	Vector Laboratories	H5000

8.1.2 Cell Culture Reagents

Material/Reagent	**Supplier**	**Catelogue Code**
Cell Culture vessels	Corning	
Trypsin Inhibitor	Invitrogen	17075-029
Human Bronchial Epithelial Cells	Lonza	CC-2540
Human Airway Smooth Muscle Cells	N/A	
GloResponse™ CRE-luc2P HEK293 Cell Line	Promega	E8500
Antibiotic, antimycotic solution	Sigma Aldrich	A-5955
Dimethyl sulfoxide(DMSO)	Sigma Aldrich	154938
Dulbecco's Modified Eagle's Medium (DMEM) - high glucose	Sigma Aldrich	D5796
Foetal Calf Serum (FCS)	Sigma Aldrich	F-7524
Phosphate buffered Saline (PBS) with Calcium and Magnesium	Sigma Aldrich	D8662
Puromycin dihydrochloride	Sigma Aldrich	P9620
Trypsin-EDTA solution (1x)	Sigma Aldrich	T3924
Trypan Blue	Sigma Aldrich	T-8154
Hydrocortisone stock solution	Stem Cell Technologies	7925
Pneuma cult-ex media	Stem Cell Technologies	5009
Hygromycin B	Thermo Fisher Scientific	10687010
Thermo Scientific™ Nunc™ F96 MicroWell™ Black and White Polystyrene Plate	Thermo Fisher Scientific	136101
Bright-Glo™ Luciferase Assay System	Promega	E2610

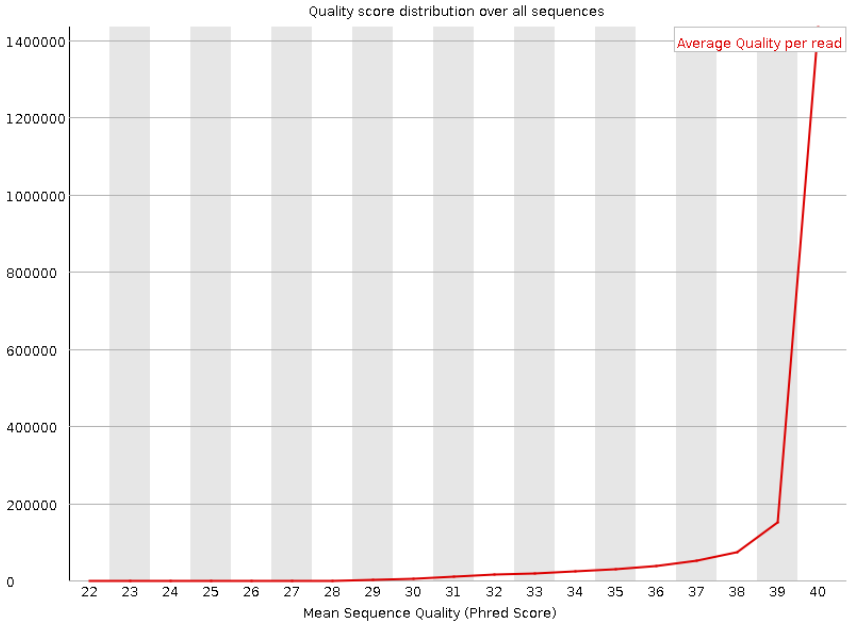
8.2 RNA-SEQUENCING QUALITY CONTROL

Per Base Sequence Quality



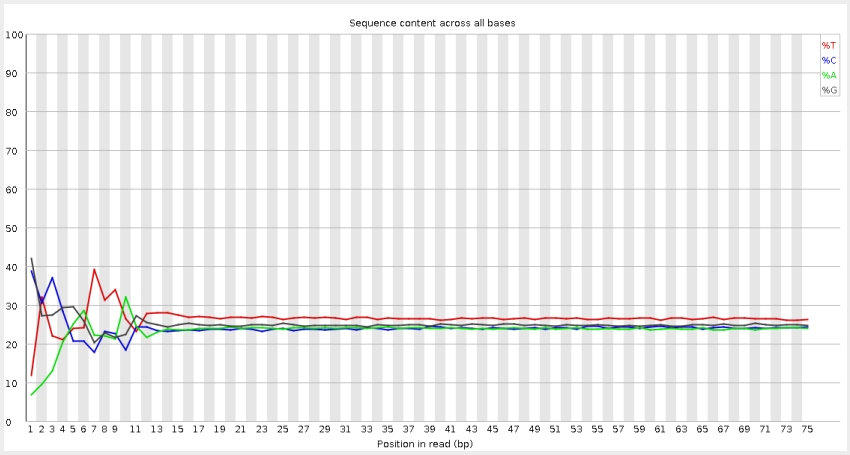
An overview of the range of quality values across all bases at each position in the FastQ file. Very good quality calls are represented in the green section, reasonable calls are orange and calls of poor quality are red. A warning will be issues if the lower quartile for any base is <10 or if the median for any base is <25. A failure will be issued if the lower quartile for any base is <5 or if the medium for any base is <20.

Per Sequence Quality Scores



This allows the visualisation of a subset of sequences which have universally low quality scores. A warning is issued if the most frequently observed mean quality is <27 (0.2% error rate) and a failure if the mean quality is <20 (1% error rate).

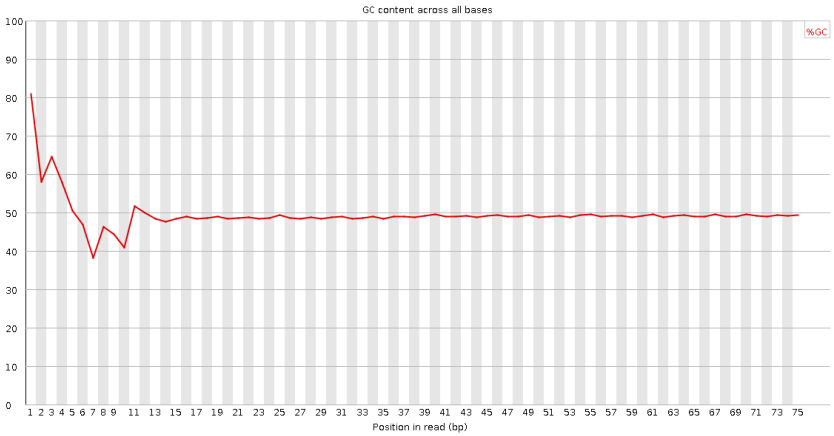
Per Base Sequence Content



This plot shows the proportion of each base position in a file for which each of the four normal DNA bases has been called. In a random library, there should be little to no difference between the different bases of a sequence run so the lines should not deviate from each other. A warning is issued if the

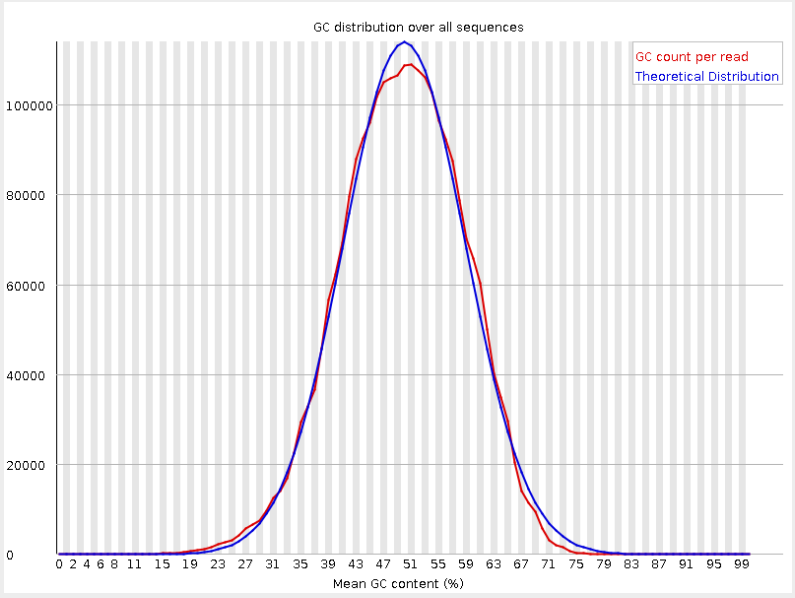
difference between A and T or G and C is >10% and a failure is issued if this is >20%.

Per Base GC Content



This plots out the GC content of each base position. There should be little to no difference between the different bases of a sequence run. The overall GC content should reflect the GC content of the genome. A warning is issued if the GC content of any base deviates >5% from the mean GC content. A failure is issued if this deviates >10%.

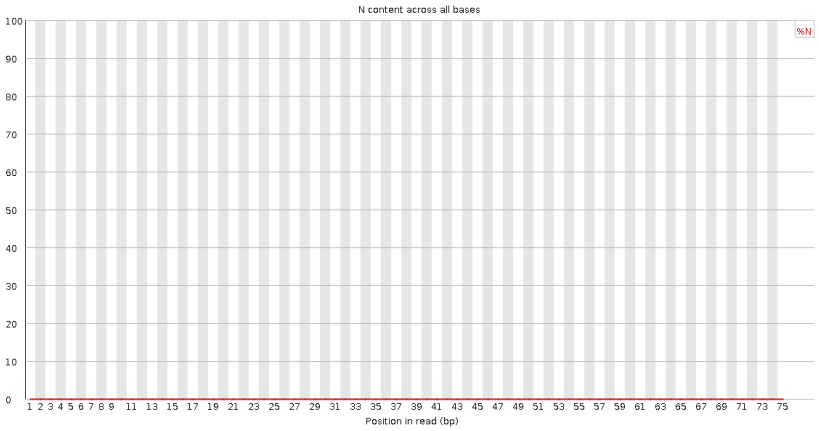
Per Sequence GC Content



This measured GC content across the whole length of each sequence and compares it to a normal distribution of GC

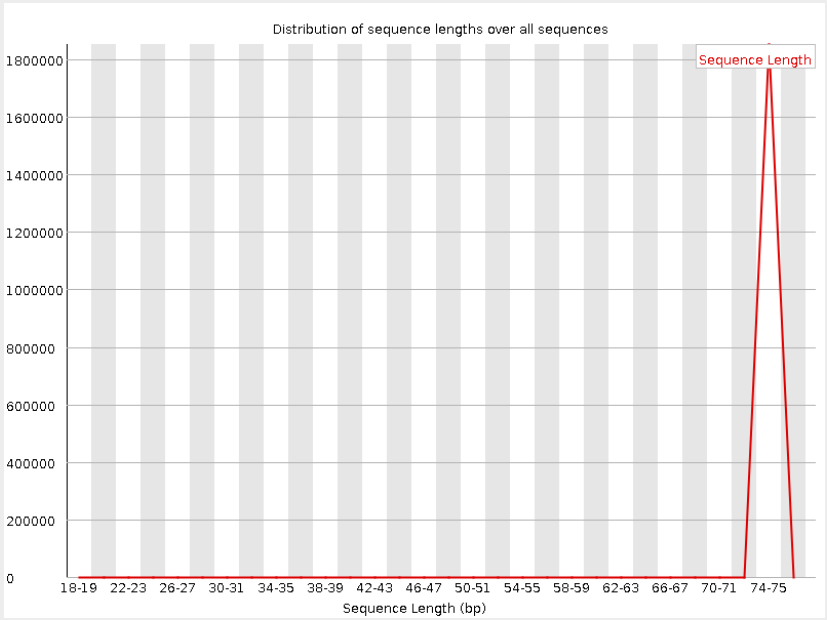
content. In a normal library you would expect to see a normal distribution of the GC content. A warning equates to a scenario in which the sum of the deviations from the normal distribution represents >15% of the reads and a failure is >30%.

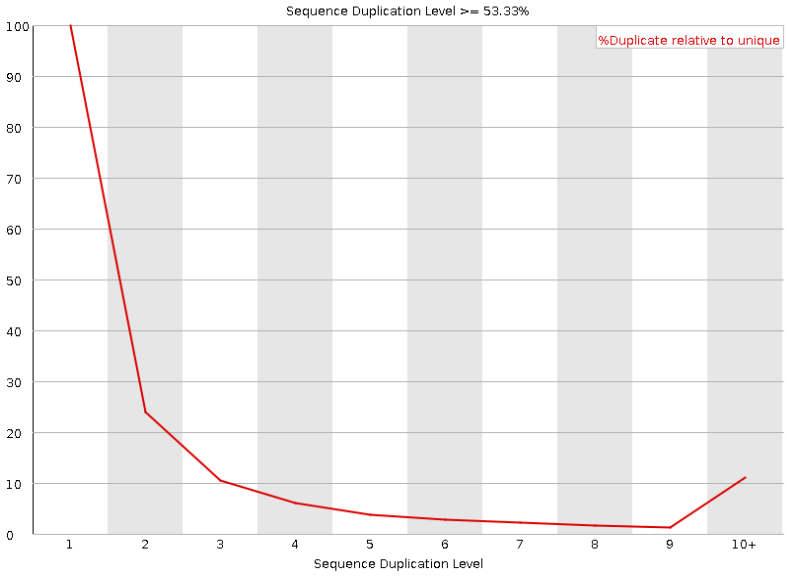
Per Base N Content



If the sequencer is unable to make a base call with sufficient confidence it places an N rather than a conventional base. N content >5% is a warning and >20% is a failure.

Sequence Length Distribution



	This shows the distribution of fragment sizes. A warning is raised if all the sequences are not the same length and a fail is if any of the sequences have zero length.
Duplicate Sequences	 Most sequences in a diverse library will occur only once. Low level duplication may indicate very high level of coverage of the target sequence but high level duplication is more likely due to some kind of enrichment bias. The plot shows the relative number of sequences with different degrees of duplication. A warning is issued if non-unique sequences make up >20% of the total and >50% constitutes a failure.
Over-represented Sequences	A normal library will contain a diverse set of sequences, with no individual sequence making up a tiny fraction of the whole. If a sequence is very over-represented means either that it is highly biologically significant or that the library is contaminated or not as diverse as expected.

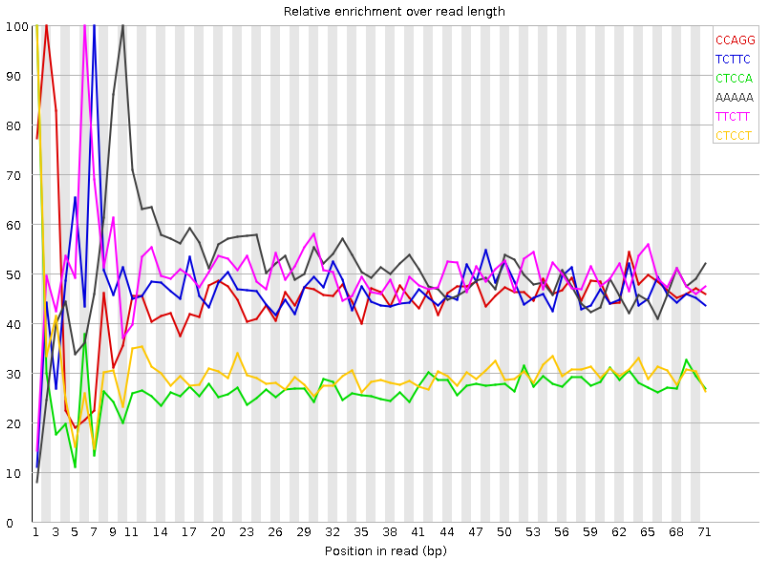
Over-represented Kmers	 This counts the enrichment of every 5-mer within the library. It then calculates an expected level at which this k-mer should have been based on the base content of the library as a whole and then uses the actual count to calculate and observed/expected ratio for that k-mer. A k-mer showing > 3 fold overall enrichment or >5 fold enrichment at any base will be reported. An error is issued if any k-mer is enriched >10 fold at any base position.

Table 8-1 Visual outputs from FastQC analysis of sequence data. Graphs are representative of good quality sequencing data.

8.3 LINUX SCRIPTS FOR RNA-SEQUENCING ANALYSIS

8.3.1 Processing Sequencing data

8.3.1.1 Concatenation of FastQ files

The first step of processing these samples is to concatenate the FASTQ files from each of the sequencing lanes. This combines the sequencing data from each of the multiple lanes of sequencing for each sample. A default file location was set up as the default index for all commands using the cd/ command followed by the required pathway. The following command concatenates the files. R1 and R2 are to denote the forward and reverse reads respectively.

```
zcat file1_R1.fastq.gz file2_R1.fastq.gz file3_R1.fastq.gz | gzip > R1.fastq.gz
```

```
zcat file1_R2.fastq.gz file2_R2.fastq.gz file3_R2.fastq.gz | gzip > R2.fastq.gz
```

8.3.1.2 Adapter and Quality-Trimming

The adaptor sequence is trimmed from the concatenated files using the Scythe programme (https://github.com/vsbuffalo/scythe). This was performed using the following command where pathway is replaced with the desired pathway.

```
/pathway/scythe -a pathway/adapter_file.fasta -q sanger -p 0.1 -o trimfile_R1.fastq file_R1.fastq.gz
```

The first part of the command is the pathway to the scythe programme, the second is the path to the adapter sequence, which is provided by the company. For example, Oxford Genomics uses their in house multiplex adaptors which are based on the Illumina adapters:

```
5' GATCGGAAGAGCACACGTCT "Multiplexing Adapter 1"
```

```
5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT "Multiplexing Adapter 2"
```

Sanger or Solexa qualities can be specified with –q sanger or –q solexa. The –p command sets the contamination rate which is set to 0.05 by default. The output file is named by –o command.

Further quality trimming is achieved using the Sickle programme (https://github.com/najoshi/sickle) which trims reads with a Q score <20.

```
/pathway/sickle pe –f *R1.fastq –r *R2.fastq –t sanger –o qtrimfile_R1.fastq –p qtrimfile_R2.fastq –s qtrim.unpaired.fastq –q 20 –l 20
```

The pe command identifies trimmed reads to be used with –f identifying the forward read and –r identifying the reverse read. The –t command determines the quality scheme to be used (Sanger). The -o command determines the name of the output file where -s determines the name for the singles file that is generated. This file contains reads that passed filter in either the forward of reverse direction but not the other. The –q command defines the PHRED score which is the quality cut off for trimming.

8.3.1.3 FASTQC Analysis of Data

The final step in quality control is the FASTQC analysis which generates graphical outputs to show the quality of sequence data. The command for this is as follows.

```
fastqc *R1.fastq
```

```
fastqc *R2.fastq
```

An example of these visual outputs is shown below in the appendices.

8.3.1.4 Alignment of reads using TopHat2

The TopHat alignment is performed with the following command.

```
tophat –p 4 –G /pathway/GRCh37/Annotation/Genes/genes.gtf –o filename /pathway/Bowtie2Index/genome qtrimfile_R1.fastq qtrimfile_R2.fastq
```

The –G command directs TopHat to the set of gene model annotations in the form of a .gtf file. TopHat will then extract the transcript sequences and use Bowtie to align reads to this transcriptome. Reads that do not fully map to the transcriptome will be mapped on the genome. Reads that do map to the transcriptome are converted to genomic mappings and merged with novel mappings in the final output file. The –p command determines the number of threads used to align reads with 1 being the default and higher numbers speed up the analysis. The trimmed fastq files enter the analysis and the output is .bam and .bed files, the latter of which contains identified insertions, deletions and splice junctions.

8.3.1.5 Transcript Assembly using Cufflinks and Cuffmerge

```
cufflinks –p 2 –o newfilename accepted_hits.bam
```

The commands –p and –o are previously explained and the pathway points the programme to the accepted_hits.bam file generated by TopHat.

Cuffmerge is then used to merge all the transcript files into a single transcriptome assembly. This creates a merged.gtf for the total population of samples and is run with the following commands:

```
cuffmerge –g /pathway/GRCh37/Annotation/Genes/genes.gtf –s /pathway/GRCh37/Sequence/WholeGenomeFasta/genome.fa –p 10 assemblies.txt
```

The –g command directs the path to a reference annotation .gtf file which is obtained from Ensemble and of the same genome build as the sequencing. The input files are merged together with the reference .gtf for the final output. The –s command provides the path to the genomic DNA sequences for the reference. The assemblies.txt file is user generated and lists the assembly file for each sample as follows:

```
./filename1/transcripts.gtf

./filename2/transcripts.gtf
```

./filename3/transcripts.gtf

8.3.1.6 Gene and Transcript Expression (Cuffnorm)

In order to determine the individual expression levels of both gene and isoform expression we run Cuffnorm. This requires the merged.gtf file created by Cuffmerge and a user generated sample sheet directing the programme to the .bam files generated by TopHat. An example sample sheet is as follows:

sample_id group_label

/pathway/filename1/accepted_hits.bam file1

/pathway/filename2/accepted_hits.bam file2

/pathway/filename3/accepted_hits.bam file3

Cuffnorm is run with the following command:

cuffnorm –use-sample-sheet /pathway/merged_asm/merged.gtf –p 10 /pathway/sample_sheet.txt

8.3.1.7 Differential gene expression analysis using CuffDiff

Cuffdiff is used to identify differentially expressed genes and isoforms between different samples or different conditions. This required the merged.gtf file and the file paths to each .bam file to be analysed. Replicate samples are separated by (,) whereas different samples or conditions are separated by a space. In the example below, filename1 and filename 2 are being compared to filename3 and filename4.

cuffdiff –o diff_out /pathway/merged.gtf –p 10 /pathway/filename1/accepted_hits.bam,/pathway/filename1/accepted_hits.bam,/pathway/filename3/accepted_hits.bam,/pathway/filename2/accepted_hits.bam

8.4 R SCRIPTS USED FOR VISUALISATION OF RNA-SEQ DATA

8.4.1 Parsing files for GSEA

Install and load required packages in R studio:

```
library(dplyr)

library(tidyr)
```

Remove rows where there are no gene names:

```
Parse1 <-gene_exp[!gene_exp$gene=="-",]
```

Separate the rows in which there are more than one gene separated by a comma and give the same expression value to each new row

```
Parse2 <-parse1 %>% separate_rows(gene_short_name, sep=",")
```

Remove duplicates (average the FPKMs) and combine to one row:

```
Parse3 <-Parse2 %>%

group_by(gene_short_name) %>%

summarize_each(funs(mean))
```

To save newly created file:

```
Write.table(parse3, “savefile.txt”, sep=”\t”)
```

8.4.2 Volcano Plots in R studio

```
install.packages("ggplot2")

install.packages("ggrepel")

library(ggrepel)
```

```
library(ggplot2)

library(calibrate)
```

The file output from cuffdiff is the gene_exp.diff file. This is the file used to generate the volcano plots used to display differentially expressed genes. It contains the list of genes with mean FPKM values for each condition with log2 fold change, p-value and q-value. Before these files are used, certain rows are removed if they have missing labels (i.e. a "-" instead of a gene name) or if the log2 fold change is an infinite change (labelled "inf" or "-inf") as these points cannot be displayed on a volcano plot. Genes for which the program could not perform a statistical test (labelled "NOTEST") were also removed. This was done using the following code where each step generated a new data frame.

```
diff <-gene_exp[!gene_exp$gene=="-",]

diff2 <-diff[!diff$log2.fold_change.=="Inf",]

diff3 <-diff2[!diff2$status=="NOTEST",]

diff4 <-diff3[!diff3$log2.fold_change.=="-Inf",]
```

The previous step prepares the file for the volcano plot. The next step is to select the data which is to be labelled with red points and gene labels. In this case, genes were labelled which had a q-value < 0.05 and a fold change ≥ 2. A subset of the original list of genes is created using these parameters using the following code.

```
label <- subset(diff4, q_value < 0.05 & abs(log2.fold_change.) > 2)
```

These files are then used to produce the volcano plot. Axes are labelled using "x=" and "y=", geom_point sets the type of graph and sets the ascetics of the points. The functions "xlim" and "ylim" set the limits of the axes and "xlab" and "ylab" label the axes. The geom_text_repel function moves the labels so they don't overlap. The following code assigns these values to the vc function.

```
vc = ggplot(data=diff2, aes(x=log2.fold_change., y=-log10(p_value))) +
```

```
geom_point(alpha=0.4, size=1.75) +

geom_point(data = label, aes(col="red")) +

theme(legend.position="none") +

xlim(c(-10,10)) + ylim(c(0,5)) +

geom_label_repel(data = label, aes(label=gene, col="red"), size=3) +

xlab("Log2 fold change") + ylab("-log10 p-value")
```

The volcano plot is produced by typing the code “vc”.

8.4.3 Heatmaps in R studio

The package pheatmap was used to generate heatmaps for RNA-seq data. The two following packages were installed.

```
library(pheatmap)

library(RColorBrewer)
```

Files were prepared in excel to contain the data required, e.g. expression values for genes and samples of interest.

A data frame is produced from this spreadsheet and converted into a matrix file.

```
df_heatmap <- data.frame(Inducible_heatmap, row.names=1)

matrix_heatmap <- as.matrix(df_heatmap)
```

The heatmap is generated and saved using the following command.

```
pheatmap(matrix_heatmap, scale = "row", legend =TRUE,
cluster_cols=F, fontsize_row = 10, cellheight=10, height=10, width=10,
angle_col = 45, color = rev(brewer.pal(n=11, name="RdYlGn")),
filename = "Inducible_heatmap.jpeg")
```

References

1. Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet (London, England). 2004;364(9435):709-21.

2. Abramson MJ, Perret JL, Dharmage SC, McDonald VM, McDonald CF. Distinguishing adult-onset asthma from COPD: a review and a new approach. International Journal of Chronic Obstructive Pulmonary Disease. 2014;9:945-62.

3. Soriano JB AA, Abate KH, Abera SF, Agrawal A, Ahmed MB, Aichour AN, Aichour I, Aichour MTE, Alam K, Alam N, Alkaabi JM, Al-Maskari F, Alvis-Guzman N, Amberbir A, Amoako YA, Ansha MG, Antó JM, Asayesh H, Atey TM, Avokpaho EFGA, Barac A, Basu S, Bedi N, Bensenor IM, Berhane A, Beyene AS, Bhutta ZA, Biryukov S, Boneya DJ, Brauer M, Carpenter DO, Casey D, Christopher DJ, Dandona L, Dandona R, Dharmaratne SD, Do HP, Fischer F, Gebrehiwot TT, Geleto A, Ghoshal AG, Gillum RF, Ginawi IAM, Gupta V, Hay SI, Hedayati MT, Horita N, Hosgood HD, Jakovljevic MMB, James SL, Jonas JB, Kasaeian A, Khader YS, Khalil IA, Khan EA, Khang YH, Khubchandani J, Knibbs LD, Kosen S, Koul PA, Kumar GA, Leshargie CT, Liang X, El Razek HMA, Majeed A, Malta DC, Manhertz T, Marquez N, Mehari A, Mensah GA, Miller TR, Mohammad KA, Mohammed KE, Mohammed S, Mokdad AH, Naghavi M, Nguyen CT, Nguyen G, Le Nguyen Q, Nguyen TH, Ningrum DNA, Nong VM, Obi JI, Odeyemi YE, Ogbo FA, Oren E, Pa M, Park EK, Patton GC, Paulson K, Qorbani M, Quansah R, Rafay A, Rahman MHU, Rai RK, Rawaf S, Reinig N, Safiri S, Sarmiento-Suarez R, Sartorius B, Savic M, Sawhney M, Shigematsu M, Smith M, Tadese F, Thurston GD, Topor-Madry R, Tran BX, Ukwaja KN, van Boven JFM, Vlassov VV, Vollset SE, Wan X, Werdecker A, Hanson SW, Yano Y, Yimam HH, Yonemoto N, Yu C, Zaidi Z, El Sayed Zaki M, Lopez AD, Murray CJL, Vos T. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Respiratory medicine. 2017;5(9):691-706.

4. Montuschi P. Pharmacological treatment of chronic obstructive pulmonary disease. International Journal of Chronic Obstructive Pulmonary Disease. 2006;1(4):409-23.

5. Alsaeedi A, Sin DD, McAlister FA. The effects of inhaled corticosteroids in chronic obstructive pulmonary disease: a systematic review of randomized placebo-controlled trials. The American journal of medicine. 2002;113(1):59-65.

6. Sutherland ER, Allmers H, Ayas NT, Venn AJ, Martin RJ. Inhaled corticosteroids reduce the progression of airflow limitation in chronic obstructive pulmonary disease: a meta-analysis. Thorax. 2003;58(11):937-41.

7. Salpeter SR. Bronchodilators in COPD: Impact of β-agonists and anticholinergics on severe exacerbations and mortality. International Journal of Chronic Obstructive Pulmonary Disease. 2007;2(1):11-8.

8. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. European Respiratory Journal. 2017;49(3).

9. Viniol C, Vogelmeier CF. Exacerbations of COPD. European Respiratory Review. 2018;27(147).

10. Tønnesen P. Smoking cessation and COPD. European Respiratory Review. 2013;22(127):37.

11. Alam DS, Chowdhury MA, Siddiquee AT, Ahmed S, Clemens JD. Prevalence and Determinants of Chronic Obstructive Pulmonary Disease (COPD) in Bangladesh. Copd. 2015;12(6):658-67.

12. Walia GK, Vellakkal R, Gupta V. Chronic Obstructive Pulmonary Disease and its Non-Smoking Risk Factors in India. Copd. 2016;13(2):251-61.

13. MacNee W. Pathology, pathogenesis, and pathophysiology. BMJ : British Medical Journal. 2006;332(7551):1202-4.

14. MacNee W. Pathogenesis of chronic obstructive pulmonary disease. Proceedings of the American Thoracic Society. 2005;2(4):258-66; discussion 90-1.

15. Snider GLK, J. Thurlbeck, W.M. Bengali, Z.H. The definition of emphysema. Report of a National Heart, Lung, and Blood Institute, Division of Lung Diseases workshop. The American review of respiratory disease. 1985;132(1):182-5.

16. Grainge CL, Lau LC, Ward JA, Dulay V, Lahiff G, Wilson S, et al. Effect of bronchoconstriction on airway remodeling in asthma. The New England journal of medicine. 2011;364(21):2006-15.

17. Roth M. Airway and lung remodelling in chronic pulmonary obstructive disease: a role for muscarinic receptor antagonists? Drugs. 2015;75(1):1-8.

18. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease 2019. 2019.

19. Redline S, Tishler PV, Lewitter FI, Tager IB, Munoz A, Speizer FE. Assessment of genetic and nongenetic influences on pulmonary function. A twin study. The American review of respiratory disease. 1987;135(1):217-22.

20. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. The American review of respiratory disease. 1982;125(4):409-15.

21. McClearn GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. Journal of gerontology. 1994;49(6):264-8.

22. Palmer LJ, Burton PR, James AL, Musk AW, Cookson WO. Familial aggregation and heritability of asthma-associated quantitative traits in a population-based sample of nuclear families. European journal of human genetics : EJHG. 2000;8(11):853-60.

23. Palmer LJ, Knuiman MW, Divitini ML, Burton PR, James AL, Bartholomew HC, et al. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. The European respiratory journal. 2001;17(4):696-702.

24. Hallberg J, Iliadou A, Anderson M, de Verdier MG, Nihlen U, Dahlback M, et al. Genetic and environmental influence on lung function impairment in Swedish twins. Respiratory research. 2010;11:92.

25. Hukkinen M, Kaprio J, Broms U, Viljanen A, Kotz D, Rantanen T, et al. Heritability of lung function: a twin study among never-smoking elderly women. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2011;14(5):401-7.

26. Blanco I, de Serres FJ, Fernandez-Bustillo E, Lara B, Miravitlles M. Estimated numbers and prevalence of PI*S and PI*Z alleles of alpha1-antitrypsin deficiency in European countries. The European respiratory journal. 2006;27(1):77-84.

27. Janciauskiene SM, Bals R, Koczulla R, Vogelmeier C, Kohnlein T, Welte T. The discovery of alpha1-antitrypsin and its role in health and disease. Respiratory medicine. 2011;105(8):1129-39.

28. DeMeo DL, Celedon JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, et al. Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2004;170(12):1294-301.

29. Wilk JB, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, et al. A genome-wide scan of pulmonary function measures in the National Heart, Lung, and Blood Institute Family Heart Study. American journal of respiratory and critical care medicine. 2003;167(11):1528-33.

30. Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, Brown A, et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Human molecular genetics. 2002;11(6):623-32.

31. Celedon JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, et al. The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). Human molecular genetics. 2004;13(15):1649-56.

32. Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. Human molecular genetics. 2003;12(10):1199-210.

33. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, Lasky-Su J, et al. MMP12, lung function, and COPD in high-risk populations. The New England journal of medicine. 2009;361(27):2599-608.

34. Hardin M, Silverman EK. Chronic Obstructive Pulmonary Disease Genetics: A Review of the Past and a Look Into the Future. Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation. 2014;1(1):33-46.

35. Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, Zhao C, et al. HTRA1 promoter polymorphism in wet age-related macular degeneration. Science (New York, NY). 2006;314(5801):989-92.

36. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science (New York, NY). 2001;291(5507):1304-51.

37. Consortium TIH. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.

38. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.

39. Syvanen AC. Toward genome-wide SNP genotyping. Nature genetics. 2005;37 Suppl:S5-10.

40. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Computational Biology. 2012;8(12):e1002822.

41. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. The Lancet Respiratory medicine. 2015;3(10):769-81.

42. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology. 2008;32(4):381-5.

43. Altshuler D, Daly MJ, Lander ES. Genetic Mapping in Human Disease. Science (New York, NY). 2008;322(5903):881-8.

44. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American journal of human genetics. 2017;101(1):5-22.

45. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS genetics. 2009;5(3):e1000429.

46. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS genetics. 2009;5(3):e1000421.

47. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nature genetics. 2010;42(3):200-2.

48. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nature reviews Genetics. 2010;11(7):499-511.

49. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nature genetics. 2010;42(1):45-52.

50. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, et al. Genome-wide association study identifies five loci associated with lung function. Nature genetics. 2010;42(1):36-44.

51. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nature genetics. 2011;43(11):1082-90.

52. Loth DW, Soler Artigas M, Gharib SA, Wain LV, Franceschini N, Koch B, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nature genetics. 2014;46(7):669-77.

53. Tang W, Kowgier M, Loth DW, Soler Artigas M, Joubert BR, Hodge E, et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. PLoS One. 2014;9(7):e100776.

54. Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nature communications. 2015;6:8658.

55. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nature genetics. 2017;49(3):416-25.

56. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nature genetics. 2019;51(3):481-93.

57. Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nature genetics. 2017;49(3):426-32.

58. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. The Lancet Respiratory medicine. 2014;2(3):214-25.

59. Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, et al. A genome-wide association study of chronic obstructive pulmonary disease in Hispanics. Annals of the American Thoracic Society. 2015;12(3):340-8.

60. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Human molecular genetics. 2012;21(4):947-57.

61. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. American journal of respiratory and critical care medicine. 2015;192(5):559-69.

62. Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, et al. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. American journal of respiratory and critical care medicine. 2011;184(7):786-95.

63. Castaldi PJ, Cho MH, Litonjua AA, Bakke P, Gulsvik A, Lomas DA, et al. The association of genome-wide significant spirometric loci with chronic obstructive

pulmonary disease susceptibility. American journal of respiratory cell and molecular biology. 2011;45(6):1147-53.

64. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. American journal of respiratory and critical care medicine. 2012;186(7):622-32.

65. Hansel NN, Paré PD, Rafaels N, Sin DD, Sandford A, Daley D, et al. Genome-Wide Association Study Identification of Novel Loci Associated with Airway Responsiveness in Chronic Obstructive Pulmonary Disease. American Journal of Respiratory Cell and Molecular Biology. 2015;53(2):226-34.

66. Lee JH, Cho MH, Hersh CP, McDonald ML, Crapo JD, Bakke PS, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. Respiratory research. 2014;15:113.

67. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nature genetics. 2019;51(3):494-505.

68. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nature genetics. 2018;50(4):559-71.

69. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (New York, NY). 2012;337(6099):1190-5.

70. Obeidat M, Hao K, Bosse Y, Nickle DC, Nie Y, Postma DS, et al. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. The Lancet Respiratory medicine. 2015;3(10):782-95.

71. Cheng DT, Kim DK, Cockayne DA, Belousov A, Bitter H, Cho MH, et al. Systemic soluble receptor for advanced glycation endproducts is a biomarker of emphysema and associated with AGER genetic variants in patients with chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2013;188(8):948-57.

72. Yin NC, Lang XP, Wang XD, Liu W. AGER genetic polymorphisms increase risks of breast and lung cancers. Genetics and molecular research : GMR. 2015;14(4):17776-87.

73. Miller S, Henry AP, Hodge E, Kheirallah AK, Billington CK, Rimington TL, et al. The Ser82 RAGE Variant Affects Lung Function and Serum RAGE in Smokers and sRAGE Production In Vitro. PLoS ONE. 2016;11(10):e0164041.

74. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Human molecular genetics. 2012;21(6):1325-35.

75. Zhou X, Qiu W, Sathirapongsasuti JF, Cho MH, Mancini JD, Lao T, et al. Gene expression analysis uncovers novel Hedgehog interacting protein (HHIP) effects in human bronchial epithelial cells. Genomics. 2013;101(5):263-72.

76. Zerbino DR, Achuthan P, Akanni W, Amode M R, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic acids research. 2018;46(D1):D754-D61.

77. Moriguchi T, Haraguchi K, Ueda N, Okada M, Furuya T, Akiyama T. DREG, a developmentally regulated G protein-coupled receptor containing two conserved proteolytic cleavage sites. Genes to cells : devoted to molecular & cellular mechanisms. 2004;9(6):549-60.

78. Paavola KJ, Hall RA. Adhesion G Protein-Coupled Receptors: Signaling, Pharmacology, and Mechanisms of Activation. Molecular Pharmacology. 2012;82(5):777-83.

79. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. Nature Reviews Drug Discovery. 2017;16:829.

80. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol. 2003;63(6):1256-72.

81. Cui H, Wang Y, Huang H, Yu W, Bai M, Zhang L, et al. GPR126 Protein Regulates Developmental and Pathological Angiogenesis through Modulation of VEGFR2 Receptor Signaling. The Journal of biological chemistry. 2014;289(50):34871-85.

82. Glenn TD, Talbot WS. Analysis of Gpr126 function defines distinct mechanisms controlling the initiation and maturation of myelin. Development (Cambridge, England). 2013;140(15):3167-75.

83. Fimia GM, Sassone-Corsi P. Cyclic AMP signalling. Journal of Cell Science. 2001;114(11):1971.

84. Craven KB, Zagotta WN. CNG and HCN channels: two peas, one pod. Annual review of physiology. 2006;68:375-401.

85. de Rooij J, Zwartkruis FJ, Verheijen MH, Cool RH, Nijman SM, Wittinghofer A, et al. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. Nature. 1998;396(6710):474-7.

86. Bos JL. Epac proteins: multi-purpose cAMP targets. Trends in biochemical sciences. 2006;31(12):680-6.

87. Murray AJ. Pharmacological PKA inhibition: all may not be what it seems. Sci Signal. 2008;1(22):re4.

88. Bjarnadottir TK, Fredriksson R, Schioth HB. The adhesion GPCRs: a unique family of G protein-coupled receptors with important roles in both central and peripheral tissues. Cellular and molecular life sciences : CMLS. 2007;64(16):2104-19.

89. Yona S, Lin HH, Siu WO, Gordon S, Stacey M. Adhesion-GPCRs: emerging roles for novel receptors. Trends in biochemical sciences. 2008;33(10):491-500.

90. Purcell RH, Hall RA. Adhesion G Protein–Coupled Receptors as Drug Targets. Annual Review of Pharmacology and Toxicology. 2018;58(1):429-49.

91. Monk KR, Hamann J, Langenhan T, Nijmeijer S, Schoneberg T, Liebscher I. Adhesion G Protein-Coupled Receptors: From In Vitro Pharmacology to In Vivo Mechanisms. Mol Pharmacol. 2015;88(3):617-23.

92. Krasnoperov V, Lu Y, Buryanovsky L, Neubert TA, Ichtchenko K, Petrenko AG. Post-translational proteolytic processing of the calcium-independent receptor of alpha-latrotoxin (CIRL), a natural chimera of the cell adhesion protein and the G protein-coupled receptor. Role of the G protein-coupled receptor proteolysis site (GPS) motif. The Journal of biological chemistry. 2002;277(48):46518-26.

93. Arac D, Boucard AA, Bolliger MF, Nguyen J, Soltis SM, Sudhof TC, et al. A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. The EMBO journal. 2012;31(6):1364-78.

94. Paavola KJ, Stephenson JR, Ritter SL, Alter SP, Hall RA. The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. The Journal of biological chemistry. 2011;286(33):28914-21.

95. Lin HH, Stacey M, Yona S, Chang GW. GPS proteolytic cleavage of adhesion-GPCRs. Advances in experimental medicine and biology. 2010;706:49-58.

96. Liebscher I, Schon J, Petersen SC, Fischer L, Auerbach N, Demberg LM, et al. A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. Cell reports. 2014;9(6):2018-26.

97. Petersen SC, Luo R, Liebscher I, Giera S, Jeong SJ, Mogha A, et al. The adhesion GPCR GPR126 has distinct, domain-dependent functions in Schwann cell development mediated by interaction with laminin-211. Neuron. 2015;85(4):755-69.

98. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. Science signaling. 2014;7(338):ra76-ra.

99. Salzman GS, Zhang S, Gupta A, Koide A, Koide S, Arac D. Stachel-independent modulation of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(38):10095-100.

100. Miller S, Melen E, Merid SK, Hall IP, Sayers I. Genes associated with polymorphic variants predicting lung function are differentially expressed during human lung development. Respiratory research. 2016;17(1):95.

101. Monk KR, Naylor SG, Glenn TD, Mercurio S, Perlin JR, Dominguez C, et al. A G protein-coupled receptor is essential for Schwann cells to initiate myelination. Science (New York, NY). 2009;325(5946):1402-5.

102. Monk KR, Oshima K, Jors S, Heller S, Talbot WS. Gpr126 is essential for peripheral nerve development and myelination in mammals. Development (Cambridge, England). 2011;138(13):2673-80.

103. Waller-Evans H, Promel S, Langenhan T, Dixon J, Zahn D, Colledge WH, et al. The orphan adhesion-GPCR GPR126 is required for embryonic development in the mouse. PLoS One. 2010;5(11):e14047.

104. Mogha A, Benesh AE, Patra C, Engel FB, Schoneberg T, Liebscher I, et al. Gpr126 functions in Schwann cells to control differentiation and myelination via G-

protein activation. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2013;33(46):17976-85.

105. Geng FS, Abbas L, Baxendale S, Holdsworth CJ, Swanson AG, Slanchev K, et al. Semicircular canal morphogenesis in the zebrafish inner ear requires the function of gpr126 (lauscher), an adhesion class G protein-coupled receptor gene. Development (Cambridge, England). 2013;140(21):4362-74.

106. Bok J, Chang W, Wu DK. Patterning and morphogenesis of the vertebrate inner ear. The International journal of developmental biology. 2007;51(6-7):521-33.

107. Poschl E, Schlotzer-Schrehardt U, Brachvogel B, Saito K, Ninomiya Y, Mayer U. Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. Development (Cambridge, England). 2004;131(7):1619-28.

108. Nguyen NM, Senior RM. Laminin isoforms and lung development: All isoforms are not equal. Developmental Biology. 2006;294(2):271-9.

109. Miner JH, Yurchenco PD. LAMININ FUNCTIONS IN TISSUE MORPHOGENESIS. Annual Review of Cell and Developmental Biology. 2004;20(1):255-84.

110. Relan NK, Schuger L. Basement membranes in development. Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society. 1999;2(2):103-18.

111. Relan NK, Yang Y, Beqaj S, Miner JH, Schuger L. Cell Elongation Induces Laminin α2 Chain Expression in Mouse Embryonic Mesenchymal Cells. Role in Visceral Myogenesis. 1999;147(6):1341-50.

112. Loscertales M, Nicolaou F, Jeanne M, Longoni M, Gould DB, Sun Y, et al. Type IV collagen drives alveolar epithelial-endothelial association and the morphogenetic movements of septation. BMC biology. 2016;14:59.

113. Soranzo N, Rivadeneira F, Chinappen-Horsley U, Malkina I, Richards JB, Hammond N, et al. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. PLoS genetics. 2009;5(4):e1000445.

114. Zhao J, Li M, Bradfield JP, Zhang H, Mentch FD, Wang K, et al. The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. BMC medical genetics. 2010;11:96.

115. Kou I, Takahashi Y, Johnson TA, Takahashi A, Guo L, Dai J, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. Nature genetics. 2013;45(6):676-9.

116. Xu JF, Yang GH, Pan XH, Zhang SJ, Zhao C, Qiu BS, et al. Association of GPR126 gene polymorphism with adolescent idiopathic scoliosis in Chinese populations. Genomics. 2015;105(2):101-7.

117. Ravenscroft G, Nolent F, Rajagopalan S, Meireles AM, Paavola KJ, Gaillard D, et al. Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. American journal of human genetics. 2015;96(6):955-61.

118. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology. 1977;36(1):59-74.

119. Hill SJ, Baker JG, Rees S. Reporter-gene systems for the study of G-protein-coupled receptors. Current opinion in pharmacology. 2001;1(5):526-32.

120. Aileen Paguio BA, Frank Fan, Pete Stecha, Denise Garvin, Monika Wood and Keith Wood. pGL4 Vectors: A New Generation of Luciferase Reporter Vectors. Promega Notes. 2005(89):7-10.

121. Le Jeune IR, Shepherd M, Van Heeke G, Houslay MD, Hall IP. Cyclic AMP-dependent transcriptional up-regulation of phosphodiesterase 4D5 in human airway smooth muscle cells. Identification and characterization of a novel PDE4D5 promoter. The Journal of biological chemistry. 2002;277(39):35980-9.

122. Rayner RE, Makena P, Prasad GL, Cormet-Boyaka E. Optimization of Normal Human Bronchial Epithelial (NHBE) Cell 3D Cultures for in vitro Lung Model Studies. Sci Rep. 2019;9(1):500.

123. Uyttersprot N, Costagliola S, Miot F. A new tool for efficient transfection of dog and human thyrocytes in primary culture. Molecular and cellular endocrinology. 1998;142(1-2):35-9.

124. Billington CK, Pascual RM, Hawkins ML, Penn RB, Hall IP. Interleukin-1beta and rhinovirus sensitize adenylyl cyclase in human airway smooth-muscle cells. American journal of respiratory cell and molecular biology. 2001;24(5):633-9.

125. A. White A. Separation and purification of cyclic nucleotides by alumina chromatography1974. 41-6 p.

126. Deng MY, Wang H, Ward GB, Beckham TR, McKenna TS. Comparison of six RNA extraction methods for the detection of classical swine fever virus by real-time and conventional reverse transcription-PCR. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc. 2005;17(6):574-8.

127. Bartlett JM, Stirling D. A short history of the polymerase chain reaction. Methods in molecular biology (Clifton, NJ). 2003;226:3-6.

128. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science (New York, NY). 1988;239(4839):487-91.

129. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. J Vis Exp. 2012(62).

130. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nature protocols. 2006;1(3):1559-82.

131. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques. 2008;44(5):619-26.

132. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif). 2001;25(4):402-8.

133. Scott MG, Swan C, Jobson TM, Rees S, Hall IP. Effects of a range of beta2 adrenoceptor agonists on changes in intracellular cyclic AMP and on cyclic AMP driven gene expression in cultured human airway smooth muscle cells. British journal of pharmacology. 1999;128(3):721-9.

134. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC molecular biology. 2006;7:3.

135. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 1977;74(12):5463-7.

136. Carr IM, Camm N, Taylor GR, Charlton R, Ellard S, Sheridan EG, et al. GeneScreen: a program for high-throughput mutation detection in DNA sequence electropherograms. Journal of medical genetics. 2011;48(2):123-30.

137. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One. 2013;8(12):e85024.

138. Kheirallah AK, de Moor CH, Faiz A, Sayers I, Hall IP. Lung function associated gene Integrator Complex subunit 12 regulates protein synthesis pathways. BMC genomics. 2017;18(1):248.

139. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols. 2012;7:562.

140. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research. 2016;44(D1):D877-81.

141. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555-7.

142. The GC. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.

143. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nature genetics. 2013;45(10):1238-43.

144. Kitagaki J, Miyauchi S, Asano Y, Imai A, Kawai S, Michikami I, et al. A Putative Association of a Single Nucleotide Polymorphism in GPR126 with Aggressive Periodontitis in a Japanese Population. PLoS One. 2016;11(8):e0160765.

145. Shaffer JR, Li J, Lee MK, Roosenboom J, Orlova E, Adhikari K, et al. Multiethnic GWAS Reveals Polygenic Architecture of Earlobe Attachment. American journal of human genetics. 2017;101(6):913-24.

146. Terzikhan N, Sun F, Verhamme FM, Adams HHH, Loth D, Bracke KR, et al. Heritability and genome-wide association study of diffusing capacity of the lung. European Respiratory Journal. 2018;52(3):1800647.

147. Stehlik C, Kroismayr R, Dorfleutner A, Binder BR, Lipp J. VIGR--a novel inducible adhesion family G-protein coupled receptor in endothelial cells. FEBS letters. 2004;569(1-3):149-55.

148. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome research. 2012;22(9):1790-7.

149. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS genetics. 2012;8(11):e1003029.

150. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.

151. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.

152. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11(4):361-2.

153. Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. Int J Genomics. 2015;2015:206937.

154. Voraphani N, Gladwin MT, Contreras AU, Kaminski N, Tedrow JR, Milosevic J, et al. An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. Mucosal immunology. 2014;7(5):1175-85.

155. Singhania A, Wallington JC, Smith CG, Horowitz D, Staples KJ, Howarth PH, et al. Multitissue Transcriptomics Delineates the Diversity of Airway T Cell Functions in Asthma. American journal of respiratory cell and molecular biology. 2018;58(2):261-70.

156. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science (New York, NY). 2015;347(6220):1260419.

157. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in biosciences = Theorie in den Biowissenschaften. 2012;131(4):281-5.

158. Lindberg S, Khan R, Runer T. The effects of formoterol, a long-acting beta 2-adrenoceptor agonist, on mucociliary activity. European journal of pharmacology. 1995;285(3):275-80.

159. Reigstad LJ, Varhaug JE, Lillehaug JR. Structural and functional specificities of PDGF-C and PDGF-D, the novel members of the platelet-derived growth factors family. The FEBS journal. 2005;272(22):5723-41.

160. Neufeld G, Cohen T, Shraga N, Lange T, Kessler O, Herzog Y. The Neuropilins: Multifunctional Semaphorin and VEGF Receptors that Modulate Axon Guidance and Angiogenesis. Trends in Cardiovascular Medicine. 2002;12(1):13-9.

161. Gaboriaud C, Thielens NM, Gregory LA, Rossi V, Fontecilla-Camps JC, Arlaud GJ. Structure and activation of the C1 complex of complement: unraveling the puzzle. Trends in immunology. 2004;25(7):368-73.

162. Milner CM, Day AJ. TSG-6: a multifunctional protein associated with inflammation. Journal of Cell Science. 2003;116(10):1863.

163. Perry SE, Robinson P, Melcher A, Quirke P, Buhring HJ, Cook GP, et al. Expression of the CUB domain containing protein 1 (CDCP1) gene in colorectal tumour cells. FEBS letters. 2007;581(6):1137-42.

164. Blanc G, Font B, Eichenberger D, Moreau C, Ricard-Blum S, Hulmes DJ, et al. Insights into how CUB domains can exert specific functions while sharing a common fold: conserved and specific features of the CUB1 domain contribute to the molecular basis of procollagen C-proteinase enhancer-1 activity. The Journal of biological chemistry. 2007;282(23):16924-33.

165. Bidan CM, Veldsink AC, Meurs H, Gosens R. Airway and Extracellular Matrix Mechanics in COPD. Frontiers in Physiology. 2015;6:346.

166. Hirota N, Martin JG. Mechanisms of airway remodeling. Chest. 2013;144(3):1026-32.

167. Hamano Y, Kalluri R. Tumstatin, the NC1 domain of α3 chain of type IV collagen, is an endogenous inhibitor of pathological angiogenesis and suppresses tumor growth. Biochemical and Biophysical Research Communications. 2005;333(2):292-8.

168. Jones RL, Noble PB, Elliot JG, James AL. Airway remodelling in COPD: It's not asthma! Respirology (Carlton, Vic). 2016;21(8):1347-56.

169. Langenhan T, Aust G, Hamann J. Sticky signaling--adhesion class G protein-coupled receptors take the stage. Sci Signal. 2013;6(276):re3.

170. Patra C, van Amerongen MJ, Ghosh S, Ricciardi F, Sajjad A, Novoyatleva T, et al. Organ-specific function of adhesion G protein-coupled receptor GPR126 is domain-dependent. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(42):16898-903.

171. Kuffer A, Lakkaraju AK, Mogha A, Petersen SC, Airich K, Doucerain C, et al. The prion protein is an agonistic ligand of the G protein-coupled receptor Adgrg6. Nature. 2016;536(7617):464-8.

172. Demberg LM, Winkler J, Wilde C, Simon KU, Schon J, Rothemund S, et al. Activation of Adhesion G Protein-coupled Receptors: AGONIST SPECIFICITY OF STACHEL SEQUENCE-DERIVED PEPTIDES. The Journal of biological chemistry. 2017;292(11):4383-94.

173. Billington CK, Hall IP, Mundell SJ, Parent JL, Panettieri RA, Jr., Benovic JL, et al. Inflammatory and contractile agents sensitize specific adenylyl cyclase isoforms in human airway smooth muscle. American journal of respiratory cell and molecular biology. 1999;21(5):597-606.

174. Philp CJ. The Potential Therapeutic Effect of Manipulating the Extracellular Matrix in Idiopathic Pulmonary Fibrosis: University of Nottingham; 2016.

175. Lu H, Hoshiba T, Kawazoe N, Chen G. Comparison of decellularization techniques for preparation of extracellular matrix scaffolds derived from three-dimensional cell culture. Journal of biomedical materials research Part A. 2012;100(9):2507-16.

176. Ngoka LC. Sample prep for proteomics of breast cancer: proteomics and gene ontology reveal dramatic differences in protein solubilization preferences of radioimmunoprecipitation assay and urea lysis buffers. Proteome science. 2008;6:30.

177. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7:539.

178. Amrani Y, Panettieri RA. Airway smooth muscle: contraction and beyond. The international journal of biochemistry & cell biology. 2003;35(3):272-6.

179. Doeing DC, Solway J. Airway smooth muscle in the pathophysiology and treatment of asthma. Journal of applied physiology (Bethesda, Md : 1985). 2013;114(7):834-43.

180. Yan F, Gao H, Zhao H, Bhatia M, Zeng Y. Roles of airway smooth muscle dysfunction in chronic obstructive pulmonary disease. Journal of translational medicine. 2018;16(1):262-.

181. Panettieri RA, Jr., Kotlikoff MI, Gerthoffer WT, Hershenson MB, Woodruff PG, Hall IP, et al. Airway smooth muscle in bronchial tone, inflammation, and remodeling: basic knowledge to clinical relevance. American journal of respiratory and critical care medicine. 2008;177(3):248-52.

182. Hackett TL, Singhera GK, Shaheen F, Hayden P, Jackson GR, Hegele RG, et al. Intrinsic phenotypic differences of asthmatic epithelium and its inflammatory responses to respiratory syncytial virus and air pollution. American journal of respiratory cell and molecular biology. 2011;45(5):1090-100.

183. Aghapour M, Raee P, Moghaddam SJ, Hiemstra PS, Heijink IH. Airway Epithelial Barrier Dysfunction in Chronic Obstructive Pulmonary Disease: Role of Cigarette Smoke Exposure. American journal of respiratory cell and molecular biology. 2018;58(2):157-69.

184. Georas SN, Rezaee F. Epithelial barrier function: at the front line of asthma immunology and allergic airway inflammation. The Journal of allergy and clinical immunology. 2014;134(3):509-20.

185. Lambrecht BN, Hammad H. The airway epithelium in asthma. Nature medicine. 2012;18(5):684-92.

186. Lambrecht BN, Hammad H. Biology of lung dendritic cells at the origin of asthma. Immunity. 2009;31(3):412-24.

187. Heijink IH, Noordhoek JA, Timens W, van Oosterhout AJ, Postma DS. Abnormalities in airway epithelial junction formation in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2014;189(11):1439-42.

188. Brown K, Filuta A, Ludwig MG, Seuwen K, Jaros J, Vidal S, et al. Epithelial Gpr116 regulates pulmonary alveolar homeostasis via Gq/11 signaling. JCI insight. 2017;2(11).

189. Kubo F, Ariestanti DM, Oki S, Fukuzawa T, Demizu R, Sato T, et al. Loss of the adhesion G-protein coupled receptor ADGRF5 in mice induces airway inflammation and the expression of CCL2 in lung endothelial cells. Respiratory research. 2019;20(1):11.

190. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature reviews Genetics. 2012;13(4):227-32.

191. Strutt D, Schnabel R, Fiedler F, Promel S. Adhesion GPCRs Govern Polarity of Epithelia and Cell Migration. Handb Exp Pharmacol. 2016;234:249-74.

192. Yates LL, Schnatwinkel C, Murdoch JN, Bogani D, Formstone CJ, Townsend S, et al. The PCP genes Celsr1 and Vangl2 are required for normal lung branching morphogenesis. Human molecular genetics. 2010;19(11):2251-67.

193. Frantz C, Stewart KM, Weaver VM. The extracellular matrix at a glance. J Cell Sci. 2010;123(Pt 24):4195-200.

194. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. Advanced Drug Delivery Reviews. 2016;97:4-27.

195. Clause KC, Barker TH. Extracellular matrix signaling in morphogenesis and repair. Current Opinion in Biotechnology. 2013;24(5):830-3.

196. Hidaka H, Inagaki M, Kawamoto S, Sasaki Y. Isoquinolinesulfonamides, novel and potent inhibitors of cyclic nucleotide dependent protein kinase and protein kinase C. Biochemistry. 1984;23(21):5036-41.

197. Engh RA, Girod A, Kinzel V, Huber R, Bossemeyer D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89. Structural implications for selectivity. The Journal of biological chemistry. 1996;271(42):26157-64.

198. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10(1):57-63.

199. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. Bioinformatics and biology insights. 2015;9(Suppl 1):29-46.

200. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.

201. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov Jill P, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems. 2015;1(6):417-25.

202. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007-2015.

203. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic acids research. 2005;33(Database issue):D54-D8.

204. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current protocols in bioinformatics. 2016;54:1.30.1-1..3.

205. Hu H, Zhao Y, Xiao Y, Zhang R, Song H. Disruption of plasminogen activator inhibitor-1 gene enhances spontaneous enlargement of mouse airspace with increasing age. The Tohoku journal of experimental medicine. 2010;222(4):291-6.

206. Lee MJ, Van Brocklyn JR, Thangada S, Liu CH, Hand AR, Menzeleev R, et al. Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. Science (New York, NY). 1998;279(5356):1552-5.

207. Suto F, Ito K, Uemura M, Shimizu M, Shinkawa Y, Sanbo M, et al. Plexin-a4 mediates axon-repulsive activities of both secreted and transmembrane semaphorins and plays roles in nerve fiber guidance. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2005;25(14):3628-37.

208. Fuller RW, Dixon CM, Cuss FM, Barnes PJ. Bradykinin-induced bronchoconstriction in humans. Mode of action. The American review of respiratory disease. 1987;135(1):176-80.

209. Ikezoe K, Oga T, Honda T, Hara-Chikuma M, Ma X, Tsuruyama T, et al. Aquaporin-3 potentiates allergic airway inflammation in ovalbumin-induced murine asthma. Sci Rep. 2016;6:25781.

210. Crane R, Gadea B, Littlepage L, Wu H, Ruderman JV. Aurora A, meiosis and mitosis. Biology of the cell. 2004;96(3):215-29.

211. Singh M, Cowell L, Seo S, O'Neill G, Golemis E. Molecular basis for HEF1/NEDD9/Cas-L action as a multifunctional co-ordinator of invasion, apoptosis and cell cycle. Cell Biochem Biophys. 2007;48(1):54-72.

212. Tikhmyanova N, Little JL, Golemis EA. CAS proteins in normal and pathological cell growth control. Cellular and molecular life sciences : CMLS. 2010;67(7):1025-48.

213. Pasterkamp RJ, Peschon JJ, Spriggs MK, Kolodkin AL. Semaphorin 7A promotes axon outgrowth through integrins and MAPKs. Nature. 2003;424(6947):398-405.

214. Suzuki K, Okuno T, Yamamoto M, Pasterkamp RJ, Takegahara N, Takamatsu H, et al. Semaphorin 7A initiates T-cell-mediated inflammatory responses through alpha1beta1 integrin. Nature. 2007;446(7136):680-4.

215. Wiley SR, Cassiano L, Lofton T, Davis-Smith T, Winkles JA, Lindner V, et al. A novel TNF receptor family member binds TWEAK and is implicated in angiogenesis. Immunity. 2001;15(5):837-46.

216. Hamann J, Aust G, Arac D, Engel FB, Formstone C, Fredriksson R, et al. International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors. Pharmacological reviews. 2015;67(2):338-67.

217. Cheng X, Ji Z, Tsalkova T, Mei F. Epac and PKA: a tale of two intracellular cAMP receptors. Acta Biochim Biophys Sin (Shanghai). 2008;40(7):651-62.

218. Wang W, Qu X, Dang X, Shang D, Yang L, Li Y, et al. Human beta-defensin-3 induces IL-8 release and apoptosis in airway smooth muscle cells. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2017;47(9):1138-49.

219. Shang J, Zhao J, Wu X, Xu Y, Xie J, Zhao J. Interleukin-33 promotes inflammatory cytokine production in chronic airway inflammation. Biochemistry and cell biology = Biochimie et biologie cellulaire. 2015;93(4):359-66.

220. Xia J, Zhao J, Shang J, Li M, Zeng Z, Zhao J, et al. Increased IL-33 expression in chronic obstructive pulmonary disease. American journal of physiology Lung cellular and molecular physiology. 2015;308(7):L619-27.

221. Knolle MD, Owen CA. ADAM8: a new therapeutic target for asthma. Expert opinion on therapeutic targets. 2009;13(5):523-40.

222. Chen J, Jiang X, Duan Y, Long J, Bartsch JW, Deng L. ADAM8 in asthma. Friend or foe to airway inflammation? American journal of respiratory cell and molecular biology. 2013;49(6):875-84.

223. Polverino F, Rojas-Quintero J, Wang X, Petersen H, Zhang L, Gai X, et al. A Disintegrin and Metalloproteinase Domain-8: A Novel Protective Proteinase in Chronic Obstructive Pulmonary Disease. American journal of respiratory and critical care medicine. 2018;198(10):1254-67.

224. Lin CH, Shih CH, Tseng CC, Yu CC, Tsai YJ, Bien MY, et al. CXCL12 induces connective tissue growth factor expression in human lung fibroblasts through the Rac1/ERK, JNK, and AP-1 pathways. PLoS One. 2014;9(8):e104746.

225. Lipson KE, Wong C, Teng Y, Spong S. CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. Fibrogenesis & Tissue Repair. 2012;5(1):S24.

226. Ponticos M, Holmes AM, Shi-wen X, Leoni P, Khan K, Rajkumar VS, et al. Pivotal role of connective tissue growth factor in lung fibrosis: MAPK-dependent transcriptional activation of type I collagen. Arthritis and rheumatism. 2009;60(7):2142-55.

227. Cesari M, Pahor M, Incalzi RA. Plasminogen activator inhibitor-1 (PAI-1): a key factor linking fibrinolysis and age-related subclinical and clinical conditions. Cardiovascular therapeutics. 2010;28(5):e72-e91.

228. Savarimuthu Francis SM, Davidson MR, Tan ME, Wright CM, Clarke BE, Duhig EE, et al. MicroRNA-34c is associated with emphysema severity and modulates SERPINE1 expression. BMC genomics. 2014;15(1):88.

229. Muth M, Theophile K, Hussein K, Jacobi C, Kreipe H, Bock O. "Hypoxia-induced down-regulation of microRNA-449a/b impairs control over targeted SERPINE1 (PAI-1) mRNA - a mechanism involved in SERPINE1 (PAI-1) overexpression". J Transl Med. 2010;8:33.

230. Xu X, Wang H, Wang Z, Xiao W. Plasminogen activator inhibitor-1 promotes inflammatory process induced by cigarette smoke extraction or lipopolysaccharides in alveolar epithelial cells. Experimental lung research. 2009;35(9):795-805.

231. To M, Takagi D, Akashi K, Kano I, Haruki K, Barnes PJ, et al. Sputum plasminogen activator inhibitor-1 elevation by oxidative stress-dependent nuclear factor-kappaB activation in COPD. Chest. 2013;144(2):515-21.

232. Jiang Y, Xiao W, Zhang Y, Xing Y. Urokinase-type plasminogen activator system and human cationic antimicrobial protein 18 in serum and induced sputum of patients with chronic obstructive pulmonary disease. Respirology (Carlton, Vic). 2010;15(6):939-46.

233. Ghosh AK, Vaughan DE. PAI-1 in tissue fibrosis. Journal of cellular physiology. 2012;227(2):493-507.

234. Petersen SL, Intlekofer KA, Moura-Conlon PJ, Brewer DN, Del Pino Sans J, Lopez JA. Nonclassical progesterone signalling molecules in the nervous system. Journal of neuroendocrinology. 2013;25(11):991-1001.

235. Nutu M, Weijdegard B, Thomas P, Bergh C, Thurin-Kjellberg A, Pang Y, et al. Membrane progesterone receptor gamma: tissue distribution and expression in ciliated cells in the fallopian tube. Molecular reproduction and development. 2007;74(7):843-50.

236. Kunzmann S, Ottensmeier B, Speer CP, Fehrholz M. Effect of progesterone on Smad signaling and TGF-beta/Smad-regulated genes in lung epithelial cells. PLoS One. 2018;13(7):e0200661.

237. Bartram U, Speer CP. The role of transforming growth factor beta in lung development and disease. Chest. 2004;125(2):754-65.

238. Ruiz C, Huang W, Hegi ME, Lange K, Hamou MF, Fluri E, et al. Growth promoting signaling by tenascin-C [corrected]. Cancer research. 2004;64(20):7377-85.

239. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. Molecular systems biology. 2005;1:2005.0010-2005.0010.

240. Vallath S, Hynds RE, Succony L, Janes SM, Giangreco A. Targeting EGFR signalling in chronic lung disease: therapeutic challenges and opportunities. The European respiratory journal. 2014;44(2):513-22.

241. Saied EM, Bediwy AS. Expression of epidermal growth factor receptor (EGFR) in the bronchial epithelium of patients with chronic obstructive pulmonary disease (COPD). European Respiratory Journal. 2011;38(Suppl 55):p3823.

242. Corradetti MN, Guan KL. Upstream of the mammalian target of rapamycin: do all roads pass through mTOR? Oncogene. 2006;25(48):6347-60.

243. Kalma Y, Marash L, Lamed Y, Ginsberg D. Expression analysis using DNA microarrays demonstrates that E2F-1 up-regulates expression of DNA replication genes including replication protein A2. Oncogene. 2001;20(11):1379-87.

244. Hur MW, Yoon JH, Kim MY, Ko H, Jeon BN. Kr-POK (ZBTB7c) regulates cancer cell proliferation through glutamine metabolism. Biochimica et biophysica acta Gene regulatory mechanisms. 2017;1860(8):829-38.

245. Wang L, Li J-j, Guo L-y, Li P, Zhao Z, Zhou H, et al. Molecular link between glucose and glutamine consumption in cancer cells mediated by CtBP and SIRT4. Oncogenesis. 2018;7(3):26.

246. Zhong J, Baquiran JB, Bonakdar N, Lees J, Ching YW, Pugacheva E, et al. NEDD9 stabilizes focal adhesions, increases binding to the extra-cellular matrix and differentially effects 2D versus 3D cell migration. PLoS One. 2012;7(4):e35058.

247. Pugacheva EN, Golemis EA. The focal adhesion scaffolding protein HEF1 regulates activation of the Aurora-A and Nek2 kinases at the centrosome. Nature cell biology. 2005;7(10):937-46.

248. Clay H, Wilsbacher LD, Wilson SJ, Duong DN, McDonald M, Lam I, et al. Sphingosine 1-phosphate receptor-1 in cardiomyocytes is required for normal cardiac development. Dev Biol. 2016;418(1):157-65.

249. Mehta P, Piao X. Adhesion G-protein coupled receptors and extracellular matrix proteins: Roles in myelination and glial cell development. Developmental dynamics : an official publication of the American Association of Anatomists. 2017;246(4):275-84.

250. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. 2018:447367.

251. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, et al. Open Targets Platform: new developments and updates two years on. Nucleic acids research. 2018;47(D1):D1056-D65.

252. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Human molecular genetics. 2015;24(R1):R102-10.

253. van de Bunt M, Cortes A, Brown MA, Morris AP, McCarthy MI. Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci. PLoS genetics. 2015;11(9):e1005535.

254. van Meerloo J, Kaspers GJ, Cloos J. Cell sensitivity assays: the MTT assay. Methods in molecular biology (Clifton, NJ). 2011;731:237-45.

255. Cory G. Scratch-wound assay. Methods in molecular biology (Clifton, NJ). 2011;769:25-30.

256. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281-308.

257. Suzuki N, Yoshida A, Nakano Y. Quantitative analysis of multi-species oral biofilms by TaqMan Real-Time PCR. Clinical medicine & research. 2005;3(3):176-85.