THE ROLE OF GENETIC AND ENVIRONMENTAL VARIATION IN THE RESPIRATORY PHENOTYPE OF ALPHA-1-ANTITRYPSIN DEFICIENCY

BY

ALICE MARGARET WOOD

A Thesis submitted to
The University of Birmingham
for the degree of Doctor of Philosophy

Division of Medical and Dental Sciences
University of Birmingham
Date 24.08.2009

Dedication

This thesis is dedicated to my father John Kevin Wood

Acknowledgements

Many thanks go to Paul Newby, Darren Bayley and Ali Ahmad at the University of Birmingham for their direction regarding laboratory techniques, to Christopher McAloon of University Hospitals Birmingham for his assistance with *MMP* genotyping and to Patti Bharadwa at the University of Birmingham for her assistance with bacterial identification.

Thanks also to Peter Nightingale at the Wellcome Trust Clinical Research Facility and to Joanna Howson at the University of Cambridge for their advice on statistics, and to Professors Jon Ayres and Roy Harrison at the University of Birmingham for their advice regarding the outdoor air pollution and occupational analyses.

Finally, I would like to thank my supervisors Professor Robert Stockley, Dr Matthew Simmonds and Professor Stephen Gough at the University of Birmingham for their guidance and support throughout my PhD.

Funding for my studies was provided by a non-commercial grant from Talecris Biotherapeutics to my supervisor Professor Stockley.

Contents

List of Figures

List of Tables

Abstract

Alpha 1 antitrypsin deficiency (AATD) is the only established genetic predisposition to chronic obstructive pulmonary disease (COPD). The development of COPD in AATD is highly variable, probably relating to complex interactions between multiple genetic and environmental factors. This thesis will describe the COPD phenotypes observed in AATD and their inter-relationships, forming the basis for examining phenotypic associations with specific candidate genes and HLA class II type. Finally it examines the potential role of ambient air pollution.

Associations were seen for *TNFA* with chronic bronchitis, *SFTPB* with FEV1, *TGFB* with small airways disease and *GC* with bronchiectasis, consistent with the role of protein products in pathogenesis. Of four *MMP*s studied, association with gas transfer occurred in two. *HLA-DQA1**0301 and *HLA-DRB1**04 contributed significantly to gas transfer in regression models, and anti elastin antibodies were higher in *HLA-DRB1*04* and *HLA-DQA1**0301 homozygotes. Ozone levels contributed to the burden of disease in cross sectional and longitudinal models of pollution exposure, whilst PM_{10}, NO_2 and SO_2 were associated only in the longitudinal model.

In conclusion this thesis demonstrates the importance of genetic variation and environmental factors in determining respiratory phenotype in AATD. It also suggests a key role for adaptive immunity in pathogenesis of emphysema.

Chapter 1: Introduction

Respiratory diseases are a leading cause of morbidity and mortality in the United Kingdom (UK), with deaths due to respiratory disease being the third most common cause of death after diseases of the circulatory system and neoplastic diseases[1]. Within non-neoplastic respiratory diseases pneumonia remains the leading cause of death, with 1151 deaths per million members of the UK population, followed by chronic obstructive pulmonary disease (COPD) which is the registered cause of death in 920 cases per million of the population[1]. In the UK it is estimated that 1.5 million people may have COPD, with up to 1 in 8 hospital admissions related to the condition[2]. These simple statistics highlight the importance of studying lung diseases, and COPD in particular.

1.1 Chronic obstructive pulmonary disease (COPD)

COPD is characterised by airflow obstruction, together with an abnormal inflammatory response to noxious particles or gases[3]. Airflow obstruction is diagnosed by spirometry, such that the ratio between the forced expiratory volume in the first second of expiration (FEV1) and the the total volume expired (the forced vital capacity, FVC) is less than 0.7[4]. As with many lung diseases cigarette smoking is an important aetiological factor in COPD, and is widely accepted to be the main environmental risk factor for developing the disease. Despite this only about 15% of smokers develop clinically significant disease[5], suggesting that there are other influences on disease expression. Previous studies estimated that smoking contributes 15% to the variability of lung function[6], whilst genetic factors account for a further 40%[7]. The remainder of

the variability may be accounted for by other environmental influences, but formal studies have not been reported, possibly in part because of the methodological issues inherent in studies of environmental exposures, which are covered in more detail in chapter 7.

1.1.1 Clinical features

COPD is defined by airflow obstruction that is not fully reversible after administration of short acting bronchodilators (such as salbutamol)[3]. There are two main classification systems worldwide, issued by national respiratory societies, based on the severity of reduction of FEV1 compared to that predicted by age, gender and height (see Table 1.1). The most widely accepted is the American/European system, which differs from the British version only in that the lowest level of severity (very severe) is omitted in the UK[8]. Typically subjects present in middle age and are (or have been) tobacco smokers.

Several pathologies may contribute to the impairment of lung function, and it is likely that COPD in its current definition comprises overlapping syndromes (see Figure 1.1). Subjects may exhibit a range of pathologies, including chronic bronchitis[9], emphysema[10], bronchiectasis or a combination of these. Although the majority of subjects with such pathology will exhibit the fixed airflow obstruction that characterises COPD, there may be some that do not. In addition there may be an element of bronchodilator reversibility, defined by the change in FEV1 relative to baseline, or predicted normal value, after administration of short acting bronchodilators. Physiological tests of lung function can also identify impairment of small airway

function, which is measured using forced mid expiratory flow (FEF25-75%), and impairment of gas transfer, usually reported relative to lung volume (KCO). The small airways have been of great interest in COPD[11], as changes in their function may be the earliest sign of disease. Impairment of gas transfer usually indicates disease of the lung parenchyma, which in the case of COPD predominantly relates to emphysema. In addition subjects may exhibit systemic manifestations of the disease, including cachexia and co-morbid diseases, such as cardiovascular disease[12], diabetes[13] and osteoporosis[14].

	British Thoracic Society	American & European Thoracic Societies
	FEV1 % predicted	
Mild	50-80	80-100
Moderate	>30<50	>50<80
Severe	<30	>30<50
Very severe	-	<30

Table 1.1: Classification of severity of COPD by spirometry

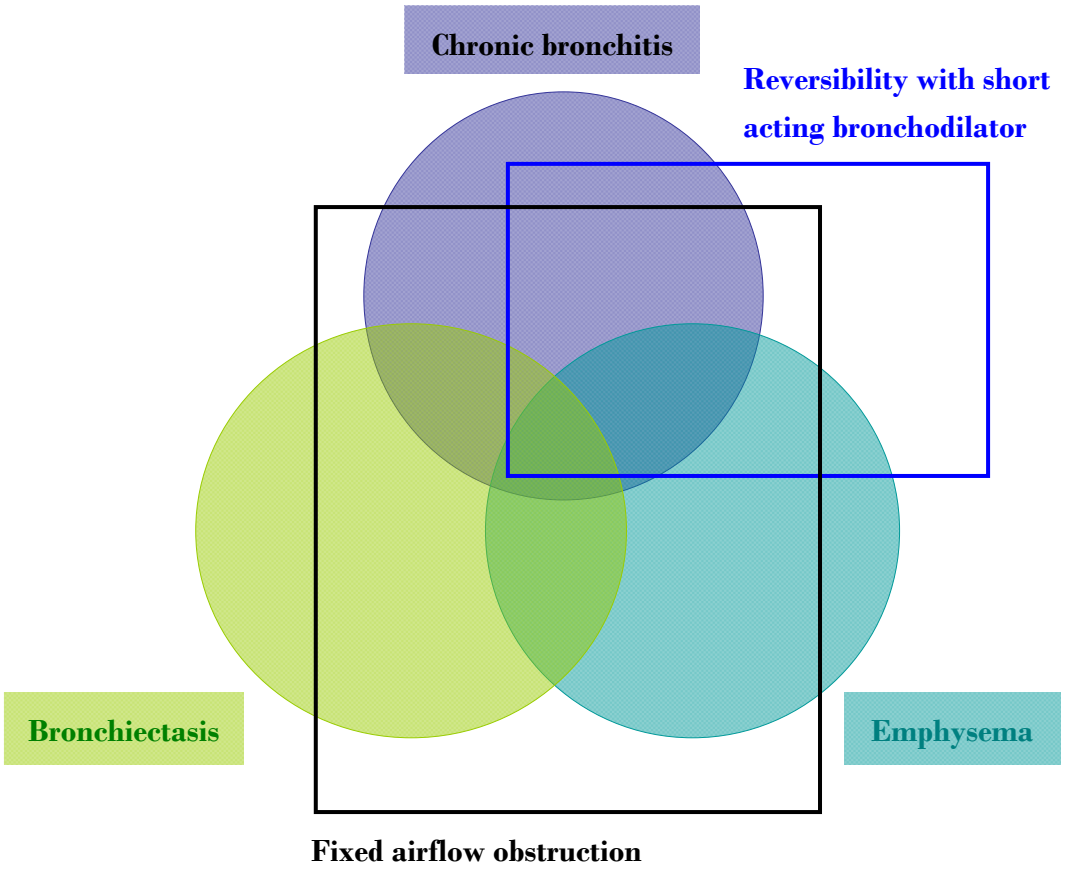
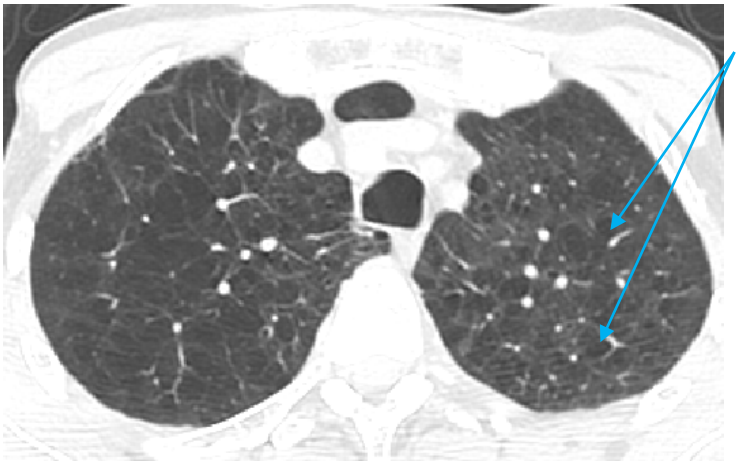


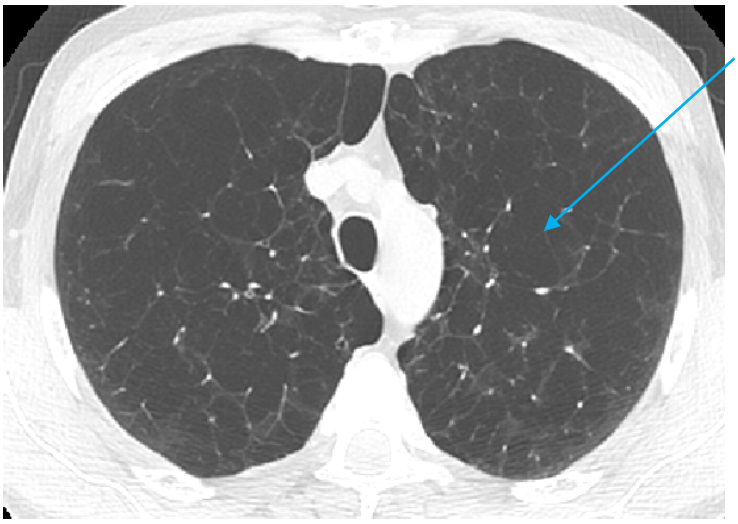
Figure 1.1: Phenotypic heterogeneity in COPD
The majority of subjects with chronic bronchitis, bronchiectasis or emphysema will have airflow obstruction diagnostic of COPD. Within those with COPD these pathologies may overlap such that subjects may have 1, 2 or 3 conditions present. In addition bronchodilator reversibility may exist, sufficient to diagnose co-existent asthma. The diagram illustrates the concept of these overlapping phenotypes.

Chronic bronchitis is defined as sputum production on most days for at least three months of at least two consecutive years[15]. Emphysema results from destruction of the lung parenchyma (the functional tissues of the lung, predominantly the alveoli) and may be diagnosed histologically[9], or more recently by computed tomographic (CT) scanning[16]. Its presence is also suggested by a pattern of raised residual volume (RV) and (to a lesser extent) total lung capacity (TLC) alongside impairment of KCO on pulmonary function testing[9]. Emphysema may be further subdivided according to its appearance and distribution on high resolution CT scan (HRCT). Typical radiological appearances of emphysema are shown in Figure 1.2. Emphysema in usual COPD is typically centrilobular in appearance, and predominantly in the upper zone. If it is due to alpha 1 antitrypsin deficiency (AATD), the only widely accepted genetic predisposition to emphysema, it is classically panacinar and lower zone dominant[17]. Differences in physiology have been seen according to emphysema zone[18], though this aspect has not been studied in detail in COPD unrelated to AATD. This, together with genetic associations of emphysema zone[19, 20], suggest that upper and lower zone emphysema may also be distinct pathologies.

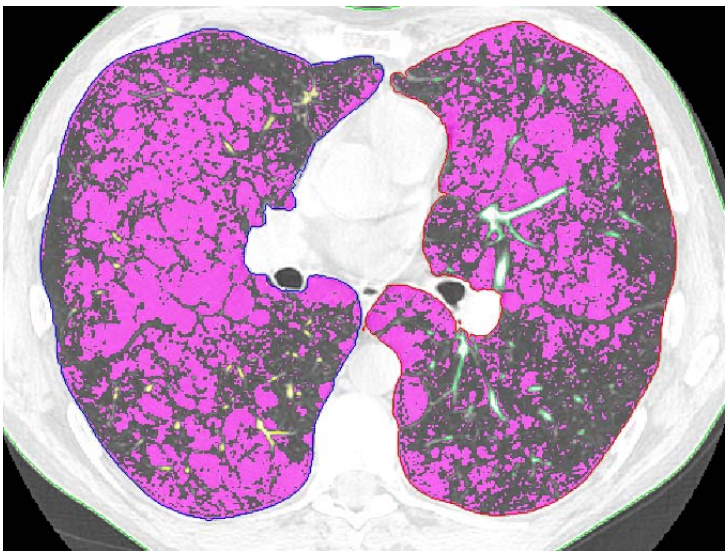
Bronchiectasis is usually diagnosed by CT scanning[21], the classical features being thickening of the bronchial wall, lack of tapering of the bronchi, bronchi visible within 1cm of the pleura and nodular densities due to mucous filled bronchi. Various patterns of bronchial dilatation are recognised, described by their shape, including cylindrical, varicose and saccular forms[21], illustrated in Figure 1.3. Subjects typically expectorate large volumes of sputum daily, though this may not always be the case where the condition exists alongside emphysema[22].



Centrilobular emphysema
The diseased areas, some of which are highlighted by arrows here, appear low in density, and are related to central bronchioles. This produces a patchy appearance to the scan image.

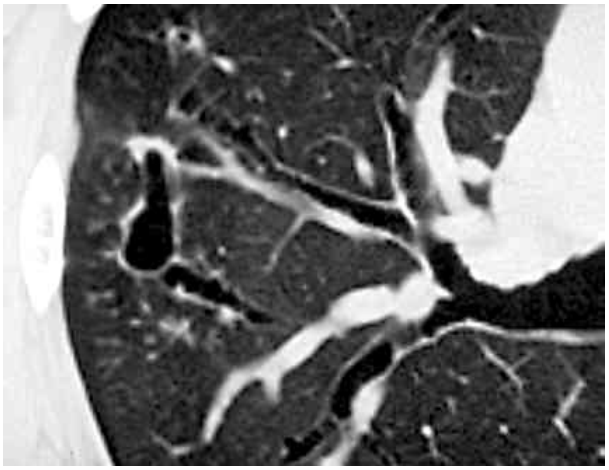


Panacinar emphysema
Disease is uniform and widespread within this image.

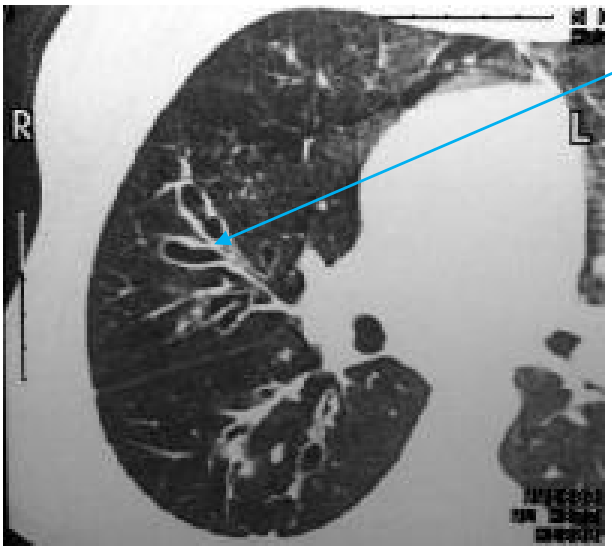


CT densitometry
Low density areas below a given threshold of Hounsfield Units are identified by quantitative CT. This is an example generated by Pulmo software (Leiden, Netherlands) using an image from the same patient as the panacinar image above, where the emphysematous areas are highlighted in pink.

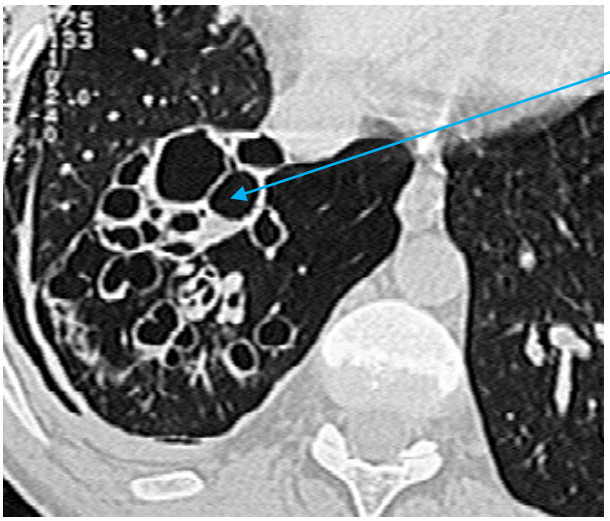
Figure 1.2: The radiological appearance of emphysema subtypes
The pictures show the two main subtypes of emphysema.



Tubular bronchiectasis
Lack of tapering of the bronchi and bronchial wall thickening can be seen.



Varicose bronchiectasis
The classical 'tree in bud' nodular appearance of mucus filled bronchi is indicated by the arrow, alongside irregular variceal bronchial dilatation.



Saccular, or cystic, bronchiectasis
Large dilated airways are shown in the right lower lobe.

Figure 1.3: The radiological appearance of bronchiectasis on CT scan
The three main subtypes, which may exist alone or together in an individual, are shown here, each being described by the appearance of the bronchi.

1.1.2 Management

Therapy in COPD is mainly directed at airflow obstruction and inflammation, with additional treatment for exacerbations, which may be infective or non-infective. Thus short and long acting bronchodilators, acting via beta adrenoceptors (β2 agonists) and anticholinergic pathways are recommended to be used in a stepwise manner, with the addition of inhaled steroids later in the disease. Current British Thoracic Society Guidelines suggest that they should be added combined with a long acting β2 agonist when FEV1 is below 60% of the predicted normal value, and when the patient is experiencing regular exacerbations[8]. Many of the newer treatments for COPD have been directed at individual components of inflammation, given its' importance in pathogenesis (discussed in section 1.1.3). However most, such as anti-TNFα, have been disappointing[23]. Some of the newer treatment strategies and their effects in clinical trials are shown in Table 1.2.

It may also be important to note that variability of response to established treatments (such as β2 agonists) and to newer ones (such as anti-TNF) may occur due to genetic variation[24-26] though it is not yet known if this is relevant in COPD. It is also possible that treatment response could vary between subphenotypes of COPD, such as those described in section 1.1.1. The importance of genetic variation is discussed further in the context of anti-TNFα therapy in chapter 3. Non-pharmacological interventions, such as pulmonary rehabilitation, which aims to improve patient fitness, are also effective[27].

Treatment	Mechanism	Clinical effects	Refs.
Cilomilast	PDE4 inhibitor	Improvement in FEV1 and quality of life. reduced FEV1 decline, fewer exacerbations	[28, 29]
Roflumilast	PDE4 inhibitor	Improvement in FEV1	[30]
BAYx1005	LTB4 synthesis inhibitor	Reduced bronchial inflammation	[31]
ABX-IL8	Monoclonal AB specific to IL8	Improves dyspnoea and FEV1 early in treatment, but no sustained improvement in lung function by the end of the trial	[32]
N-acetylcysteine	Antioxidant	No improvement in lung function or exacerbation frequency	[33]
Infliximab	Anti-TNFα	No benefit except in cachectic subjects whose 6MWT distance and frequency of hospital admissions improved	[23]
Marimastat	MMP inhibitor	Tested in asthma; reduced airway hyper-responsiveness	[34]
Al-trans-retinoic acid	Repairs elastase/smoke induced lung damage	Clinical trials in progress. Pilot studies confirm safety.	[35]
Montelukast	Leukotriene receptor antagonist	Improved FEV1 and quality of life, observational study suggested reduced hospital admissions & medication usage	[36, 37]

Table 1.2: Potential new medical treatments for COPD, their mechanism of action and reported clinical effects

PDE4 = phosphodiesterase 4, LTB4 = leukotriene B4, IL8 = interleukin-8, AB = antibody, TNFα = tumour necrosis factor-α, MMP = matrix metalloprotease, 6MWT = 6 minute walk test

1.1.3 Pathogenesis

A number of pathological processes have been deemed important in COPD, within which three main themes have emerged:

- Protease-antiprotease imbalance
- Oxidant-antioxidant imbalance
- Inflammation

The first theme concerns imbalance between proteases that digest elastin and other components of the extra-cellular matrix in the lung parenchyma, and anti-proteases that protect[38, 39]. The origin of this theory comes from the observation that patients with AATD develop early onset emphysema[40]. Alpha one antitrypsin (AAT) is an anti-protease, which acts predominantly to block the action of neutrophil elastase (NE), a protease released by neutrophils. Neutrophil elastase is a serine protease, the first of three classes of protease important in COPD. The remaining two classes are the cysteine proteases, such as cathepsin-B, and the matrix metalloproteases (MMP's) [41]. In general the serine and cysteine proteases are capable of degrading elastin and some forms of collagen [41], whilst MMP's have more of an effect on collagen, gelatin and laminin [39]. Each protease is inhibited by one or more anti-proteases, may inactivate other anti-proteases, or activate pro-inflammatory cytokines via interaction with proteinase activated receptors (PARs) [42]. A key inflammatory cytokine activated in this way is tumour necrosis factor alpha (TNFα). The proteases function to clear debris and damaged tissue, but if their action is not effectively controlled they may produce excessive lung damage.

Second, disparity between levels of harmful oxidants and protective antioxidants may lead to dominant oxidative stress. Cigarette smoke is a major source of oxidants (mainly free radicals and nitric oxide), though reactive oxygen species (ROS), such as O_2^-, are also produced by interaction of cigarette smoke with the epithelial cells, and released by inflammatory leukocytes, which accumulate in the lungs of smokers[43]. Certain pollutants, predominantly particulate matter, may also affect oxidant-antioxidant balance[44] and could contribute to disease by this mechanism, a concept discussed further in chapter 7. Antioxidant enzymes present in the airway include glutathione-S-transferase, superoxide dismutase and catalase[45]. Free radicals and ROS have direct toxic effects on the respiratory epithelium, leading to epithelial cell detachment, and increased cell lysis[46]. These processes also enhance pulmonary inflammation via up-regulation of genes encoding pro-inflammatory cytokines[46], exacerbating the damage to the lung.

Inflammation is the third key concept in COPD pathogenesis, and may also be stimulated by cigarette smoke[43], ozone[47] and particulate matter pollution[44]. The processes by which pollution may affect the lung are discussed further in chapter 7. The importance of inflammation in pathogenesis is supported by transgenic mouse models, such as that for IL13 – when this pro-inflammatory cytokine is over-expressed the mice exhibit induction of MMPs and develop emphysema[48], whilst TNFα knockout mice are relatively protected from developing these features after exposure to cigarette smoke[49]. In man the role of inflammation is supported by the observation of infiltration of the airways by inflammatory cells in increasing numbers as disease progresses[11] and of correlations between airway inflammatory cytokine levels, such as intereukin 8 (IL8),

and progression of respiratory disease[50]. Furthermore, subjects with COPD have higher circulating levels of several inflammatory markers[51], though the relationship between pulmonary and systemic inflammation is not yet clear. The hypothesis that systemic inflammation results from 'overspill' of inflammatory mediators from the lung is not consistent with the lack of correlation between inflammatory markers in sputum and plasma[52], although recent data suggests some signal communication between the 2 components when repeated sputum measures are used[53]. Inflammation is, however, a significant driver of proteolytic and oxidant damage, as shown in Figure 1.4, with many mediators implicated in pathogenesis [54], as summarised in Figure 1.5.

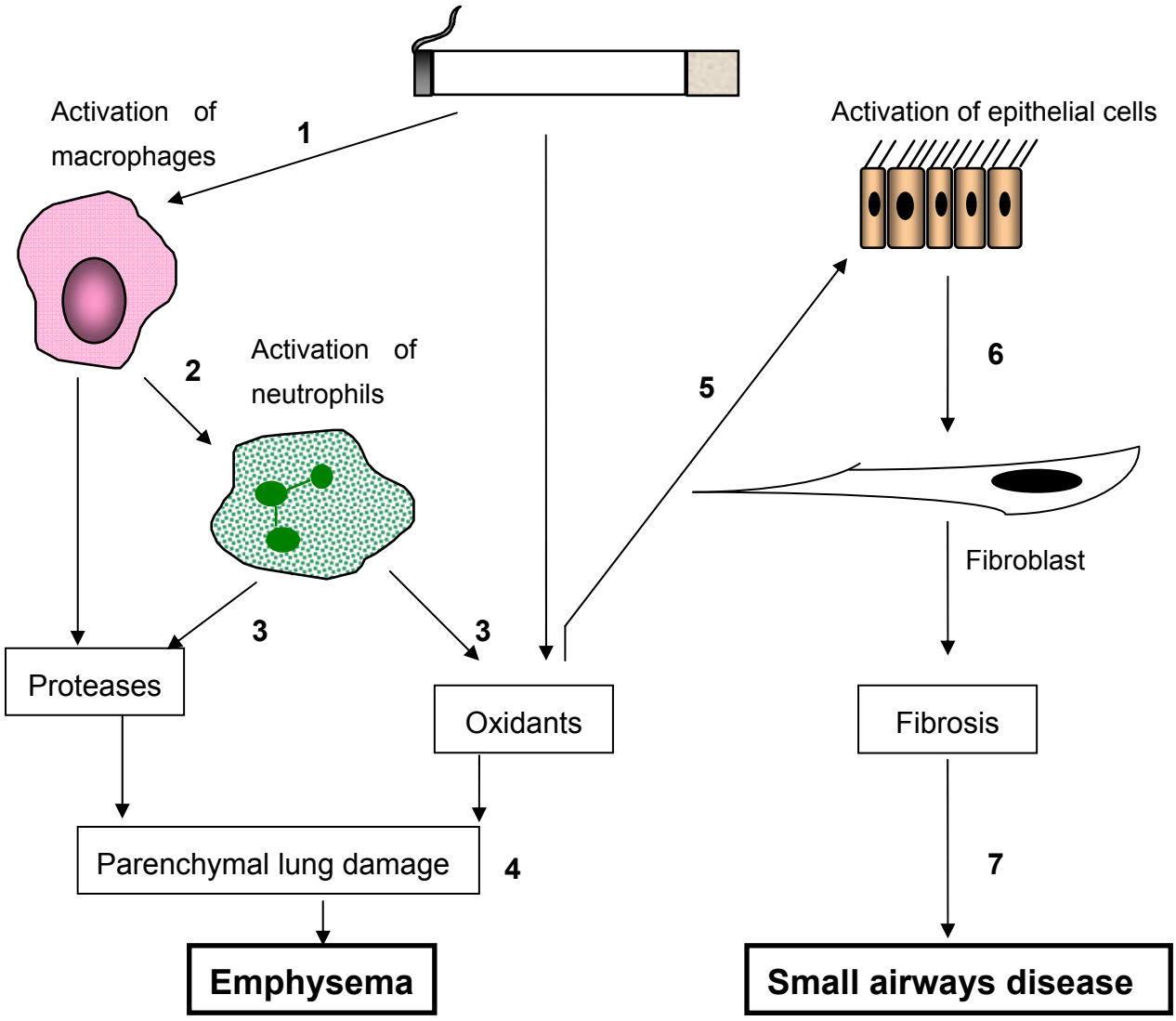


Figure 1.4: The pathogenesis of COPD

Cigarette smoke and other inhaled pro-inflammatory agents (such as ozone [47] or particulate matter [44]) activate macrophages (1), and neutrophils (2) leading to the direct release of proteases or neutrophil chemotracctants, together with the release of oxidants (3) resulting in subsequent breakdown of connective tissue in the lung (4), causing emphysema. Epithelial cell stimulation (5) by oxidants, such as oxygen free radicals, promotes fibroblast activity (6), eventually leading to fibrosis and small airways disease (7).

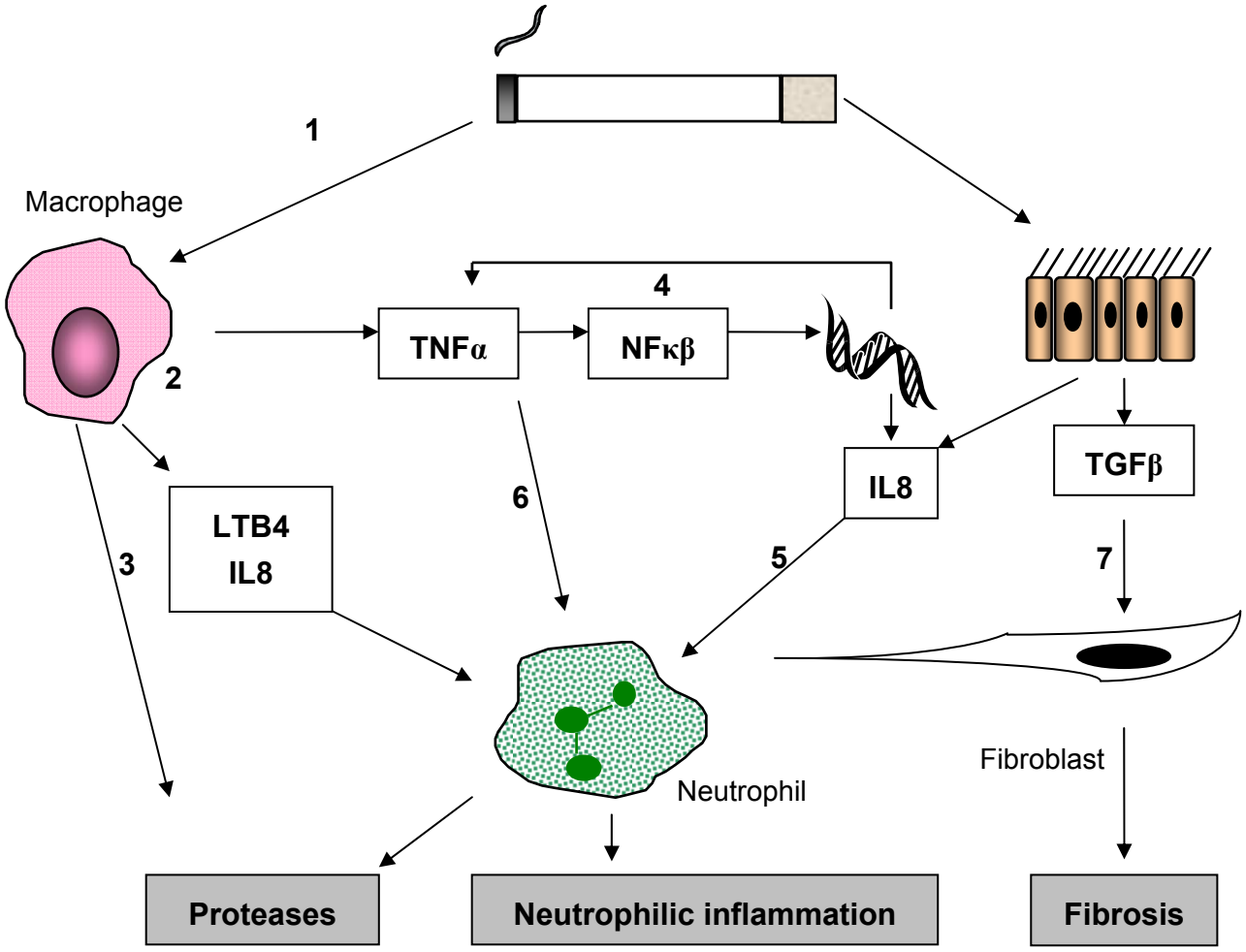


Figure 1.5: Inflammatory mediators in COPD

There are many pro-inflammatory mediators involved in COPD, some of which are illustrated here. Cigarette smoke, or another inhaled pro-inflammatory agent, activates macrophages (1) to release TNFα, leukotriene B4 (LTB4), interleukin 8 (IL8) and other neutrophil chemotactic factors (2), as well as proteases (3). TNFα promotes further IL8 release from other cells in the respiratory tract by nuclear factor kappa beta (NFκβ) mediated effects on gene transcription (4). This in turn increases local neutrophilic inflammation (5), and hence the release of proteases. TNFα itself may directly increase neutrophilic inflammation by upregulation of endothelial adhesion molecules (6). Epithelial cells also stimulate fibroblasts via transforming growth factor beta (TGFβ), leading to fibrosis (7).

1.1.4 Genetic aspects

Family based studies support a role for genetic susceptibility in COPD by demonstrating ancestral aggregation of spirometric measures in the general population [55, 56] and higher rates of airflow obstruction in first-degree relatives of subjects with COPD[57]. This suggests that lung function runs in families, whether or not they have developed any respiratory disease. This in turn is indicative of a genetic component, although assumes that all environmental influences have been taken into account prior to the analysis. Accounting for environmental factors is key in COPD since smoking also tends to aggregate in families; the more recent family studies in the area have accounted for this by regression analysis. Aggregation of pulmonary function parameters has been shown in families with AATD[58] and in the last year both CT diagnosed emphysema and airway wall thickness have been shown to aggregate in families with usual COPD[59], suggesting a genetic contribution to lung architecture, as well as to physiology.

Continued smoking is generally considered to be a risk factor for more rapid decline of lung function over time, compared to the general population, but major differences in the rate of decline of lung function occurs between smokers[60]. This suggests there may also be an interaction between genetic and environmental influences.

Approaches to genetic analysis

In diseases where multiple genetic influences are likely to be present, like COPD, several approaches have been used to identify susceptibility genes. These include linkage

analysis and population based case-control studies. Variation, or polymorphism, within genes can be classified in different ways[61], which describe the resultant change in the DNA sequence or protein. Common changes in the DNA sequence include microsatellites - multiple repeats of a short segment of DNA, and single nucleotide polymorphisms (SNPs) - a change of a single base, being the most common type of polymorphism in the human genome [62]. Such changes may occur in coding or non-coding regions of DNA. Changes in coding regions may be synonymous or non-synonymous, depending whether they affect the amino acid sequence of the protein product - synonymous changes do not alter the product, non-synonymous changes do. Non-synonymous changes in coding regions were thought to be most important for many years, since intuitively a change in the amino acid seuquence may alter the function of a protein[63]. However it is now recognised that alterations in non-coding regions may be equally important in complex disease.

Copy number variation (CNV) is a newer concept of interest for genetic association studies. CNV means that differing numbers of copies of a segment of DNA occur in a given population, this segment being anything from 1 kilobase to several megabases in size. The CNV can be caused by deletions, duplications, inversions or translocations of genetic material and might involve a candidate gene (or indeed many genes) thus has great potential to influence pathogenesis. CNVs are very common in the human genome [64], this again being a good reason to examine further for associations with disease. An example in lung disease is CNV of the endothelial growth factor receptor gene (*EGFR*), where higher copy numbers are seen in lung cancer and relate to response to treatment [65].

Linkage analysis

This approach identifies susceptibility areas within the genome by seeking markers that co-segregate with disease; if a marker is inherited together with a disease, and this is significant across many families then the marker must be located close to a susceptibility gene, and the two are said to be linked. Linkage analysis does not rely on prior knowledge of the gene product, or its role in pathogenesis. Microsatellites are the most commonly used markers in such analyses.

Parametric linkage analysis requires specification of a genetic model, including parameters such as penetrance, allele frequency and mode of inheritance. Results are expressed as a logarithm of the odds (LOD) score[66], which is a likelihood ratio derived from the recombination fraction between the marker and the proposed locus of the disease-causing gene. The threshold level of LOD score needed for genome-wide significance varies depending on the study design, from 3.3 for parametric linkage analyses to 3.6 in sibling pair studies[67]. The latter is a commonly used non-parametric linkage study design, which works on the principle that there will be an excess of allele sharing between sibling pairs with disease, compared to that expected by chance. Unlike parametric designs such methods do not require a mode of inheritance and are therefore more useful in complex diseases (like COPD) where multiple influences exist, than in single gene disorders where the mode of inheritance may be inferred from the pattern of disease observed clinically.

Linkage studies allowed relatively crude whole genome screening, but have been disappointing in their results in common complex diseases, in that they have either failed to deliver novel loci, or have failed to replicate well in independent populations.

Developments in genotyping technology mean that is now possible to screen the whole genome using hundreds of thousands of SNPs, whose genotypes can be detected by fluorescence on a chip. Sequencing technology is also progressing rapidly and it is now possible to ascertain the exact nucleotide sequence of an individual's genome using high throughput methods. Since linkage studies can only report 'linkage' of disease to a genomic region, which may be very large, they required follow up by other work, usually case-control association analyses, to locate the gene responsible for disease association. Historically false positive studies have been a problem[68]. Furthermore, follow up work has tended to concentrate on candidate genes suggested by current knowledge of pathogenesis in the regions identified, limiting the discovery of new processes relevant to disease.

Linkage disequilibrium

Linkage disequilibrium (LD) is a relevant concept to consider when discussing linkage studies, and prior to considering other study designs. This is the term used to describe how closely genetic markers are related to one another, such that two markers exhibiting a high degree of LD will usually be inherited together, whilst those with a low degree of LD may not. The degree of LD between two markers can be ascertained from the correlation coefficient between genotypes (r^2) or an alternative measure known as D-prime, which is calculated by dividing the observed degree of linkage between the two markers by the theoretical maximum. A higher r^2 indicates a higher degree of LD between the markers. Effectively linkage studies rely on this concept to narrow the region of genetic association.

Case-control association studies

International projects, such as the SNP consortium[69], which catalogues common SNPs in the human genome, and HapMap[70, 71], which has genotyped SNPs in Caucasian, African, Chinese and Japanese subjects, have contributed to the many databases available on genetic variation. Potential disease causing polymorphisms, or markers within genes and their likely population frequencies can be identified using such resources. This allows the design of case-control association studies, either following up a linkage study, or as a primary study of one or more candidate genes. This is a powerful and widely used approach, though sometimes producing inconsistent results[72]. This may be because of variation in the definition of cases, underpowered studies, racial differences and population heterogeneity.

The issues of case definition and power are particularly important when examining a complex disease such as COPD, as each gene may contribute to a particular sub-phenotype, or contribute only a small amount to the clinical phenotype. For instance, cases of COPD might be defined by standard spirometric measures, or sub-phenotypes might be chosen based on CT parameters or symptoms. If different genes contribute to each of these disease features, then studies which have used differing case definitions might not replicate each others' results. Similarly, if a number of genes contribute to each specific clinical feature as part of a cascade, as is likely given the complexity of pathogenesis (section 1.1.3), then each one may only lead to a small additional risk of that feature. The power of a genetic association study is dependent on the odds ratio (OR) of disease conferred by a specific polymorphism and the sample size, such that to detect the effect of a SNP conferring a small OR of disease a larger sample size will be

needed to have adequate power to detect it. Power calculations for case control association studies are usually based around an OR of disease of 1.5. Other problems can be limited by choosing subjects from only one racial group and genotyping markers specifically to assess population heterogeneity.

Genetic associations

During the years that the work contained in this thesis was performed the level of knowledge of the genetic associations of COPD has increased exponentially. At commencement of the studies contained herein the majority of COPD genetic association work was in small cohorts, phenotyped in many different ways, or were linkage studies. Since then the problems of these approaches have been recognised and the technology has improved to the extent that large, genome-wide studies replicated in similar sized groups have recently been reported [73]. It is now recognized that both airways and parenchymal disease are likely to have a genetic component [59], and that the gold standard for genetic association work should be to report associations in two COPD populations, phenotyped in the same way, in the same paper. Alternatively, confidence in the truth of the association might be gained by meta-analysis of the studies of any given polymorphism.

A range of genetic associations with recognised COPD phenotypes have been reported and are summarised in Table 1.3. These are predominantly in candidate genes suggested by the three pathogenetic themes of protease-antiprotease imbalance, oxidant-antioxidant imbalance and inflammation. Major studies of genes in each of these areas are discussed below, concentrating on the pre genome-wide era, with further detail on

those studied in this thesis in the relevant chapters. The results of genome-wide association studies (GWAs) in COPD, and the future of genetic association studies in the post-genome era will be considered in chapter 8, in order to put this thesis into context of the published work available at its inception.

Protease-antiprotease pathways

AATD is the most widely accepted genetic susceptibility to COPD, and exerts its influence by decreased antiprotease protection, as discussed in section 1.2. Other candidate genes relating to increased proteolytic destruction, or decreased antiprotease protection have been studied in COPD, predominantly in matrix metalloprotease genes (*MMP*s) or their inhibitors. This family of genes and protein products is discussed in more detail in chapter 6.

The most widely studied *MMP* polymorphism is a promoter region SNP in *MMP9* (C-1562T, rs3918242), located on chromosome 20q11. This increases *MMP9* activity via higher levels of gene transcription resulting from preferential binding of a transcription repressor protein to the C allele[74], and has been linked to COPD in both Chinese[75] and Japanese populations[76]. Both of these studies were relatively small, having 100 and 45 cases respectively, together with a similar number of controls. A later study in 84 Japanese subjects reported that the association was specific to upper zone dominant emphysema (UZDE)[20].

SNPs in *MMP1* and *MMP12* have also been studied in COPD. The single G allele variant of a promoter region SNP in *MMP1* (rs1799750, G-1607GG) has been associated with predisposition to rapid decline of lung function, as were combinations of this and a

non-synonymous SNP in *MMP12,* though the latter SNP was not associated when examined alone[77]. This suggests that the true variant linked to decline may be rs1799750, however the associated variant results in decreased *MMP1* expression[77] thus the association is not consistent with the role of MMP1 in COPD pathogenesis. The true association may, therefore, be in LD with these SNPs. The same SNP (rs1799750) was associated with UZDE in Caucasians in the National Emphysema Treatment Trial (NETT) genetics study[19], but not with features of lung function (such as FEV1)[72] or functional status[78]. It is of note that the region in which both genes lie has been associated with emphysema in a recent genome wide association study[79], detailed results from which will be discussed in chapter 8. Two SNPs in a gene coding for an MMP inhibitor (TIMP2) - tissue inhibitor of metalloprotease-2 (*TIMP2)* - have also been associated with COPD in Japan[80]. This result has not yet been reproduced in other ethnic groups, and the functional consequences of each SNP remain theoretical.

Alpha-1 antichymotrypsin (also known as SERPINA3) inhibits cathepsin G and mast cell chymase in a reversible fashion. Two SNPs in *SERPINA3*, associated with low α-1 antichymotrypsin levels (rs17473, rs1800463), have been associated with COPD in Swedish subjects [81, 82], though their cases were defined by a measure of airway resistance, rather than standard spirometric parameters. These associations were not reproduced in Japan[83, 84], though a non-synonymous SNP affecting the signal peptide region (rs4934) was found more commonly in the Japanese COPD group. None of these variants showed association in Italian subjects with airflow obstruction, though this included those with bronchiectasis as well as COPD[84]. The lack of replication could be explained on racial grounds or by the different diagnostic criteria used by each group.

SERPINE2 is an inhibitor of trypsin-like serine proteases, related to AAT[85], though its major function is in coagulation and fibrinolysis[86]. *SERPINE2* was identified as a potential candidate gene in COPD by integration of an area of linkage on chromosome 2q33 from the Boston early-onset COPD cohort[87, 88] with knowledge of gene expression during murine lung development and human lung microarray datasets[89]. This showed *SERPINE2* to be within the area of linkage (chromosome 2q33) and to be highly expressed in the lung, making it a logical candidate gene. Several studies now support a role for this gene in COPD, though the precise location of functional variants has yet to be determined. In the NETT cohort many SNPs were significantly associated[89] (the most associated being rs6734100 with FEV1/FVC, p=0.00004) most of which were replicated by the International COPD Genetics Network (ICGN) and in a Norwegian case-control group[90]. Conversely, a large European case-control study did not find any association with COPD[91] and questioned the validity of some of the results reported in the original study since SNPs in complete LD in the European cohort had different p values for association from one another in the original study. Over-expression of *SERPINE2* is associated with an increased risk of COPD[89], which is not consistent with its protective anti-protease actions. The mechanism of association may be through its role in coagulation, since enhanced prothrombotic markers are associated with decline of FEV1 in COPD[92] though the role of such pathways in COPD pathogenesis has not been widely investigated. Alternatively SERPINE2 mediated inhibition of plasminogen activator urokinase (PLAU) may be important as PLAU is involved in activation of TGFβ and MMPs, and is over-expressed in COPD [93].

Gene	Markers/alleles		Function	COPD phenotype	N (cases: controls)	Effect size	p	Refs.
MMP9	rs3918242	C-1562T	↑gene transcription	UZDE, COPD	84:85 110:94 100:98	6.51 2.69 6.17	0.04 0.02 <0.01	[20, 75, 76]
MMP1	rs1799750	G-1607GG	↑gene transcription	UZDE, FEV1 decline	*282* 284:306	*4.1* *0.62*	0.02 0.02	[19, 77]
TIMP2	rs2277698 *	G853A G-418C	Altered protein ↓protein level	COPD	88:40	5.4 2.3	<0.0001 0.05	[80]
SERPINA3	rs4934 rs17473 rs1800463	Ala-15Thr Pro227Ala Leu55Pro	Altered protein ↓protein level ↓protein level	Altered airway resistance	100:100	Not given	0.04 0.02 0.04	[81, 82]
SERPINE2	Various§	N/A	N/A	COPD	*949* & 204:441 *1910* & 975:956	Not given *-0.113*	$4x10^{-5}$ 0.008	[89, 90]
GSTP1	rs947894	Ile105Val	↑enzyme activity	UZDE, COPD, FEV1 decline	*282* 286:308 *1940* 53:50 1098:544	-5.0 2.83 *-0.21* 3.5 1.69	0.001 0.03 0.03 0.01 0.02	[19, 94-97]
GSTM1	Gene deletion	Null	No protein	Emphysema, CB	1098:544 168:384 87:162	3.49 1.36 1.50	0.005 0.03 0.002	[97-99]
SOD3	rs1799895	Arg213Gly	↑ protein level	COPD	230:210	4.3	0.02	[45]
EPHX1	rs1051740 rs2234922	Tyr213His His139Arg	↓enzyme activity ↑enzyme activity	Emphysema, UZDE, FEV1 decline	*282* 304:441 *304* & 127 families 94:203 40:358 283:308	*Slow -5.9* Fast 0.73 *Slow 1.75* Slow 4.1 Slow 2.9 Slow 4.9	0.005 0.03 0.03 & ns 0.01 0.02 0.04	[19, 72, 78, 100-102]

HMOX1	Microsatellite in promoter	GT(n)	↑gene transcription ↓enzyme activity	Emphysema, COPD, FEV1 decline	*749* 101:100	*-62.0* 1.62	0.009 0.01	[103, 104]
TNFA	rs1800629	A-308G	↑ protein level	Emphysema, CB, COPD	106:110 44:40 42:42	2.58 2.15 11.1	<0.01 0.09 <0.01	[105-107]
TGFB	rs1800469 rs1982073	C-509T C613T	↑ protein level ↑ protein level	COPD, dyspnoea	*304* 304:441	*5.4 & 5.7* Not given	0.0002 & 0.0005 Both 0.02	[78, 108]
GC	rs4588 rs7041	Thr436Lys Asp432Glu	↓conversion to MAF	Emphysema, COPD, FEV1 decline	63:82 104:413 75:64	*GC1F* 2.2 *GC2* 0.8 *GC2* 0.17	0.01 0.04 <0.01	[109-111]
IL13	rs1800925	C-1055T	↑gene transcription	COPD	151:99	Not given	0.02	[112]
SFTPB	rs1130866 rs2118177	Thr131Ile	Altered protein	COPD, dyspnoea, exacerbations	304:441 127 families & *304* 101:81	1.58 Not given 3.65	0.05 0.03 & ns 0.03	[72, 78] [113]

Table 1.3: Some genetic polymorphisms associated with COPD

The table lists some studies of genetic association in COPD. Details of studies are given in the order that they appear in the references listed. Where the studies were case: control in nature the effect size is shown as an OR; where they were quantitative analyses in one group of cases the B value from the regression analysis is shown, and the data given in italics. Where quantitative studies concern FEV1 or FEV1 decline a negative B value is equivalent to a higher risk of disease, in contrast to most emphysema studies where this is indicated by a positive B value. Where studies have computed statistics by different risk alleles this is indicated in the effect size column with the allele name. If more than one cohort was studied results

*SNP unidentifiable in dbSNP[114]. §16 SNPs were associated with either FEV1 or FEV1/FVC; p value and effect size of most associated is shown. See abbreviations list for definitions of genes used in this table; common allelic variant descriptors used in published work are included for ease of cross-referencing to such studies. Clinical abbreviations used here are: UZDE = upper zone dominant emphysema, CB=chronic bronchitis, COPD=airflow obstruction with low FEV1, equivalent to BTS mild COPD or greater (table 1.1).

Oxidant-antioxidant pathways

Polymorphisms affecting the function of antioxidants in the airway might influence oxidative stress, and are thus logical candidates. The most widely studied class of antioxidant genes are the glutathione S transferases, *GSTP1* and *GSTM1*, thats protein products detoxify some of the harmful contents of tobacco smoke [115]. A number of polymorphisms with functional consequences have been examined in COPD[19, 94, 96, 97].

A non-synonymous SNP in *GSTP1* (rs947894, Ile105Val) which increases the metabolism of carcinogenic aromatic epoxides[116] has been widely studied in COPD. Studies of its relationship to lung disease have varied in their findings. The 105Ile variant encodes a change in protein structure which alters the substrate binding site[116], resulting in less anti-oxidant activity. It has been associated with UZDE in the NETT cohort[19], with airflow obstruction in a Japanese population[96] and with a rapid decline of FEV1 in the Lung Health Study (LHS)[94]. Conversely, it has also been associated with low baseline lung function[97] and 105Val homozygous children exhibit reduced annual growth of FEV1[95]. Gene-smoking or gene-gene interactions have been considered as causes for these disparate results. No evidence of a gene-smoking effect was seen in the LHS[97], but there was an additive effect of polymorphisms in *GSTP1* and other *GST*s[94], suggesting a role for gene-gene interaction.

GSTM1 has a null allele, such that null homozygotes have no GSTM1 activity. This genotype has been associated with emphysema[98] and chronic bronchitis[99], with conflicting results concerning its role in lung cancer[98, 117]. It was not associated with

decline of FEV1 in the LHS[97]. In common with most other genetic studies in COPD the positive results have been difficult to replicate[72], though this may be because of differences in case definition. The studies of *GSTM1* have varied in their dataset size, from 111 emphysematous subjects in the smallest[98] to over 1000 participants in the LHS[97].

Microsomal epoxide hydrolase (EPHX1) is an enzyme expressed in bronchial epithelial cells and metabolises highly reactive epoxide intermediates in cigarette smoke [118, 119]. There are 2 known SNPs in *EPHX1* that affect enzyme activity; the first is in exon 3 (rs1051740, Tyr113His), the second in exon 4 (rs2234932, His139Arg). In both cases the His variant is associated with lower levels of enzyme activity [120, 121]. Each polymorphism only accounts for a modest change in activity[121], and are insufficient on their own to explain the association with COPD. However, variation in the gene's regulatory regions can account for up to 30% of gene activity[122], such that SNPs here could also be important in COPD. Subjects carrying both His variants were at the highest risk of developing COPD and emphysema in a Scottish population[100]; the result being replicated in subjects with advanced COPD in Japan[101]. The LHS also demonstrated a relationship with rapid decline in FEV1 for subjects with the same His-His combination[102]. The His139 variant alone was associated with a spirometric diagnosis of COPD in the Boston early-onset COPD cohort[72] and with UZDE in NETT genetics[19], consistent with the observation of better gas transfer in those with the Arg139 variant[78].

Extracellular superoxide dismutase (SOD3) is present at high concentrations in areas of the lung containing large amounts of type 1 collagen, especially around large airways and adjacent to alveoli[123] and protects the lung, particularly during episodes of

increased inflammation [124]. A SNP (rs1799895, Arg213Gly) in *SOD3* that increases plasma enzyme levels was found to be protective against COPD in New Zealand[45], and in the Copenhagen City Heart Study[125].This effect was not observed in non-smokers suggesting a gene-smoking interaction, although robust statistical evidence of this was not reported[125].

Heme oxygenase-1 (HMOX1) catalyses the oxidative cleavage of haem - a molecule consisting of an iron atom within a porphyrin ring, of which there are four main types in man. This results in the release of carbon monoxide, biliverdin and iron[126]. Biliverdin is then broken down into bilirubin, which scavenges local ROS such that HMOX1 contributes to a pathway which generates antioxidants. It is present at higher concentrations in the lungs of smokers than non-smokers[127] suggesting up-regulation in response to increased ROS. A microsatellite (GT)n repeat in the 5' region of *HMOX1* alters transcription[128] such that with greater numbers of repeats it is not induced as effectively by ROS[104]. Thus in the presence of a long GT repeat ($n \geq 30$) smokers would be unable to protect their lungs from ROS induced damage, and would be susceptible to emphysema[103]. A French study associated GT repeats ≥ 33 in length with airflow obstruction and rapidly declining lung function[103], whilst Japanese subjects with GT repeats ≥ 30 in length were more likely to have emphysema[104]. The effect on decline was not reproduced in the LHS, despite the fact that this was a larger study in a similar racial group[94]. The association with emphysema has not been replicated to date.

Inflammation and inflammatory mediators

Since inflammation is central to the pathogenesis of COPD many mediators have been examined as candidate genes for association with disease. A key driver of inflammation in COPD is TNFα, that's gene has been studied relatively widely in COPD. A SNP in its' promoter (G-308A, rs1800629) directly affects gene regulation, and is associated with high TNFα production[129]. This has been studied in several COPD phenotypes with differing results. An initial case-control study in Taiwanese subjects with chronic bronchitis found an association with disease, consistent with greater airways inflammation[107]. It has also been linked to airflow obstruction without chronic bronchitis, and severity of emphysema in Japanese subjects [105, 106]. Studies in Caucasians have not been able to reproduce these results [72, 102].

A number of other genes associated with pro-inflammatory pathways have also been examined in COPD. The four surfactant proteins contribute to regulation of alveolar surface tension and have a role in host defence, both of which are processes relevant to lung homeostasis. Surfactant proteins B and C are hydrophobic membrane proteins that increase the rate at which surfactant spreads across the alveolar surface; both are required for normal respiratory function because of their effect on alveolar surface tension[130]. Surfactant proteins A and D have carbohydrate recognition domains that allow them to coat bacteria, to promote phagocytosis by macrophages[131]. Variations within surfactant protein B (*SFTPB*) have been associated with a variety of COPD phenotypes in Caucasian [72, 78, 132] and Hispanic subjects [133], and more recently a genome-wide study has reported association near to the gene for its binding protein[73]

The Boston early-onset COPD study showed linkage between FEV1 and chromosome

19q13, which contains the transforming growth factor beta (*TGFB*)[108]. The role of its protein product (TGFB) in COPD has been outlined in section 1.1.3. Subsequent studies have supported a role for variation within *TGFB* in COPD susceptibility[78, 134, 135]. Studies of the Group specific component gene (*GC*) (also known as Vitamin D binding protein) in COPD have concentrated on known allelic variants; such as the *GC2* and *GC1S* alleles, caused by non-synonymous SNPs (rs4588 and rs7041). The *GC2* allele appears to be protective in Caucasians [110, 111], whilst the *GC1S* allele has not been associated with COPD[72]. The *GC1F* allele has been consistently associated with a range of COPD phenotypes in Japanese subjects [109, 136] but results in Caucasians have been inconsistent [102, 110, 111]. This is probably because Caucasians have a lower frequency of *GC1F* and studies were therefore underpowered (section 4.1.3 and appendix 4).

Another polymorphism in an inflammatory pathway gene which has been less well replicated is in the promoter region of *IL13* (rs1800925, C-1055T). This variant is associated with increased IL13 production[137]. It is more common in COPD[112] though this may only be important in the presence of a gene-smoking interaction[138]. Studies in transgenic mice have shown that when *IL13* is over expressed, it results in cathepsin and MMP dependent, rather than NE driven, emphysema [48, 139]. In mice *IL13* induced emphysema is characterised by excessive pulmonary mucus production, but this feature has not been examined in genetic association studies in man.

1.2 Alpha-1-antitrypsin deficiency

Alpha one antitrypsin deficiency constitutes the only widely accepted genetic susceptibility factor to COPD. It was first described in 1963 by Laurell and Eriksson, who reported an absence of the α1 band on protein electrophoresis of human serum[140]. Serum electrophoresis separates serum proteins into five major bands: albumin, alpha-1-globulins, alpha-2-globulins, beta globulins and gamma globulins. Alpha 1 antitrypsin constitutes the majority of alpha-1-globulin band, thus in AATD the band is not present. The absence of this band was originally noted in conjunction with early onset emphysema and COPD[40], suggesting a role for pathways involving AAT in pathogenesis.

Alpha one antitrypsin is an acute phase glycoprotein, synthesised in hepatocytes[141], and subsequently secreted into the plasma. It is also produced in small quantities by alveolar macrophages, circulating monocytes[142] and lung epithelial cells[143, 144]. Figure 1.6 shows the structure of AAT, highlighting some of the features that classify it into the serpin family of proteins. It is comprised of 394 amino acids with an active site residue of methionine at position 358, and a predominant target protease of neutrophil elastase (NE), though it can also bind irreversibly to proteinase 3 and cathepsin G. Point mutations in its mobile domain can lead to protein polymerisation within the liver[145], such that it is not secreted effectively into the plasma, resulting in low pulmonary levels and vulnerability to NE induced damage. Mutations like these underlie the most common forms of AATD.

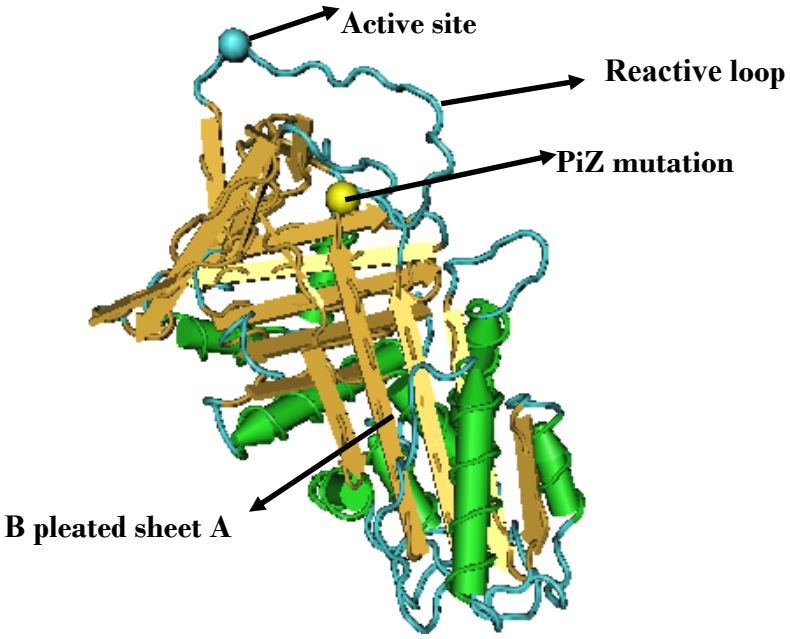


Figure 1.6: The structure of alpha 1 antitrypsin, including the PiZ mutation
AAT is part of the serpin family of proteins, thats structure is based around three beta pleated sheets and a reactive loop[146, 147], as shown. The function of AAT is determined by the active site, where NE binds irreversibly, inhibiting its function. Genetic variants influencing the shape, or movement, of the reactive loop where the active site is located are often associated with AATD, because alterations tend to allow the molecule to polymerise within the liver, such that there is a relative serum deficiency.

1.2.1 Clinical features

AATD is associated with COPD, emphysema, chronic bronchitis, bronchiectasis, neonatal jaundice, liver cirrhosis, vasculitis and panniculitis[17]. Although AATD represents a predisposing factor for development of COPD phenotypes, including emphysema, not all subjects develop pulmonary disease. Such variation in clinical phenotype may be due to environmental factors, such as cigarette smoking, but may also be influenced by other genetic factors. The evidence for genetic influences on clinical disease expression is discussed later in this chapter (section 1.2.3) and is the subject of the majority of studies contained in this thesis.

The diagnosis is usually made after investigation of pulmonary or liver disease, but is influenced by local and national practice, particularly with regard to neonatal screening. In countries where a screening programme exists subjects may be diagnosed at birth, rather than when symptoms of disease develop. Many of the symptoms of AATD are non-specific, with subjects experiencing cough, wheeze and chest infections, contributing to the delay in diagnosis experienced by many, of approximately 5.6 years from onset of symptoms[148]. The classical presentation is with progressive dyspnoea (breathlessness), with or without chronic bronchitis, similar to usual COPD, though often at a younger age or with less smoke exposure[17]. Regression modelling in the UK AATD population suggests that features of emphysema may occur in late teenage years[149], in contrast to the usual COPD presentation in middle age (>45). In the UK current national guidelines suggest testing for the condition only in people under the age of 45 or with a disease severity inconsistent with their level of cigarette smoke exposure[8]. Cigarette smoke

exposure is usually quantified in pack years, one pack year being equivalent to one pack of cigarettes a day for a year (i.e. 20 per day). A ten pack year or more exposure history is seen in most subjects with usual COPD, but this is likely to be far less in AATD[17].

Lung function tests usually show features typical of usual COPD (section 1.1.1), such as airflow obstruction with or without low gas transfer, and increased static lung volumes. However it is not uncommon for changes more usually associated with asthma, such as reversibility after bronchodilators, to be seen[150]. In the NHLBI AATD registry reversibility of at least 12% of baseline FEV1 was present in 28% of subjects at their initial assessment after administration of a bronchodilating agent[150]. Normal lung function may also be seen, most often but not exclusively in asymptomatic subjects. Smoke exposure is likely to be the most important determinant of lung function, with differences detectable between smokers and non-smokers even at age 18 years[151], and a significant relationship between smoking and lung function decline in adults, with a mean decline per year of 70ml in FEV1 in current smokers compared to 47ml in those who had never smoked[152]. Other determinants of rapidly declining lung function include male sex, low body mass index, frequent exacerbations and the severity of upper zone emphysema[153].

HRCT scanning is now being used more widely in the assessment of COPD and AATD because of advances in the interpretation of such images[154]. Scans can be used not only to diagnose emphysema and bronchiectasis, but also their type, distribution and severity[22, 154]. Emphysema distribution in AATD relates to lung function[18]. Most subjects with AATD have lower zone predominant emphysema on HRCT, in contrast to usual COPD where upper zone disease is more common. Bronchiectasis occurs in the

majority of subjects, often alongside emphysema, but is the predominant pathology in a small subgroup[22].

Specific tests for AATD include measuring the serum level of AAT, with a level of less than 11μM considered pathophysiologically important, and phenotyping (protein electrophoresis) or genotyping for the common disease associated variants, such as the Z allele – these are discussed further in section 1.2.3.

1.2.2 Management

At present treatment for AATD depends on the individual presentation: for most subjects this will mean interventions targeted at COPD (section 1.1.2). Treatment approaches targeting the molecular basis of disease, rather than its consequences, have the potential to treat all aspects of the condition, and are the subject of much current research. Such approaches include AAT augmentation[155], blocking the polymerisation process within hepatocytes[156, 157] and gene therapy[158]. A recent clinical trial has shown AAT replacement to be beneficial in reducing emphysema progression as quantified by CT scanning [159]. Gene therapy is in the early phase of clinical trials, but is not yet of proven efficacy. Blocking polymerisation, and hence enhancing secretion into the circulation, remains at the *in vitro* stage of development at present but has the current drawback of undertreating the problem. Alternatively liver transplantation from an individual homozygous for the normal allele returns AAT levels to normal. If this is performed prior to the development of lung disease it offers likely protection, but can only be used in those with end stage liver disease at present due to the shortage of suitable

organs, and the implications of subsequent life-long immunosuppression.

1.2.3 Genetic aspects

Alpha one antitrypsin is encoded by the protease inhibitor (Pi) locus on chromosome 14q32[160-162] which is 12.2kb in length with four coding exons (II, III, IV, V), three non coding exons (IA, IB, IC) and six introns (Figure 1.7). The region coding for the reactive loop, where the binding site for NE is located, is within exon V. It has been fully sequenced and cloned[163] and shows a co-dominant pattern of inheritance[40]. It is a highly pleomorphic gene with approximately 125 single SNPs (genotypes) reported in public databases[164], some of which have an effect on AAT level or function.

Traditionally the protein product is identified by its speed of migration on gel electrophoresis, the most common isoforms being F, M , S and Z (fast, medium, slow and very slow respectively)[165]. Changes in protein charge, resulting from alterations in amino acid composition[166] are the cause of differential migration speeds. The protein isoform is often termed 'phenotype', though this term will be avoided in this thesis to minimise confusion with clinical phenotypes. Theoretically this isoformic classification means that different genotypes resulting in a similar protein product would be grouped together. Nevertheless, it is usual to refer to isoforms identified in this way so that they seem analogous to the underlying genotype. Thus alleles encoding a protein of medium speed are termed PiM, and those encoding slow isoforms PiZ. Variants may also be classified by their effect on AAT level and function- normal, deficient, null (nil detectable) or dysfunctional. This classification is seen alongside the isoform and genotypic

classifications, since each describe a different aspect of the genetic variation. Serum deficiency or absence of AAT occurs because of alterations in gene expression or translation, or abnormal intracellular processing[167]. Dysfunctional variants are abnormal forms of AAT, which whilst present at a detectable level, do not function normally. The F isoform is an example of a dysfunctional variant[168]. Amongst the dysfunctional proteins described, those associated with poor inhibition of NE may be associated with lung disease. An example of this is PiMmineralsprings[169] caused by a non-synonymous SNP (rs28931568, Gly67Glu) which also contributes to disease by serum deficiency due to intracellular aggregation of AAT. A different amino acid substitution at the active site (unidentifiable in dbSNP, Met358Arg), known as PiPittsburgh, not only leads to reduced anti-NE activity, but also to an increased anticoagulant effect by way of inhibition of factor IXa, kallikrein and factor XIIf[170]. However, the majority of clinical disease due to AAT is from plasma deficiency and null variants, some of which are discussed below.

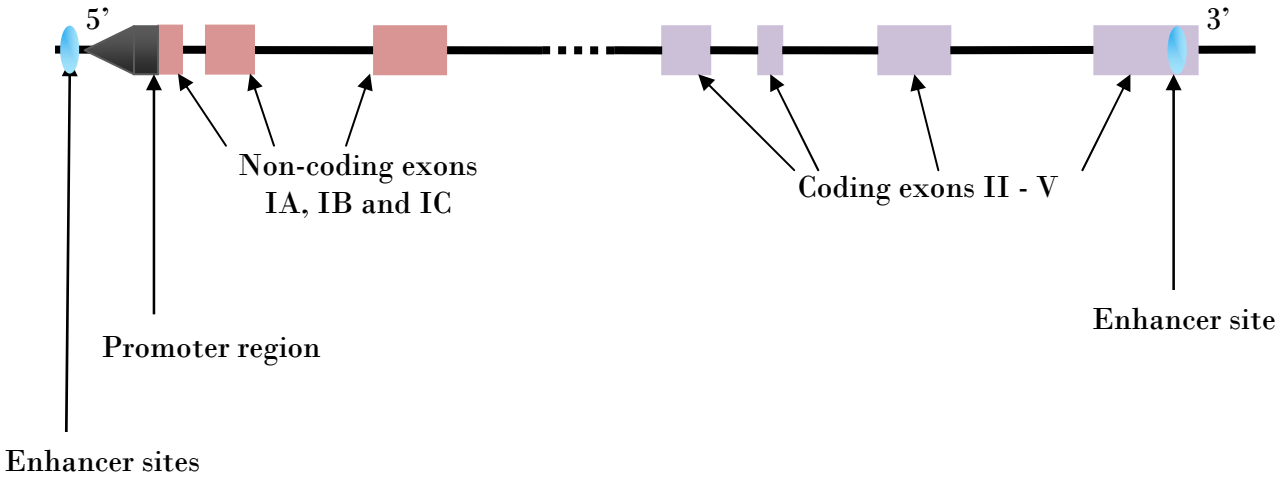


Figure 1.7: The structure of the alpha 1 antitrypsin gene

The gene encoding AAT is composed of 3 non coding exons and 4 coding exons, and is located on chromosome 14q32. There is a large intron between the non-coding and coding exons, and a small promoter (approximately 15kb) at the 5'end of the DNA sequence. The start of the transcription initiation site differs by 9bp between lung tissue and monocytes, suggesting a role for local control in the lung[171]. AAT expression systemically is regulated by synergistic action of two hepatocyte nuclear factors (1α and 4), and enhanced by sites induced by cytokines. Upstream oncostatin-M and IL-1 inducible enhancer sites are located 100bp and 13kb respectivelyfrom the 5' end, whilst an enhancer site 1.2kb from the 3' end of the final exon is induced by oncostatin-M and IL-6[172].

The PiZ mutation is a non-synonymous SNP (rs28929474, Glu342Lys) which causes a conformational change in the reactive loop, opens up β sheet A and allows the reactive loop of a second molecule to insert at this point[147]. This process extends to form AAT polymers as inclusion bodies in hepatocytes, preventing secretion into the circulation, a process which has been reproduced *in vitro* under physiological conditions[145]. This process causes hepatocyte death and ultimately clinical liver disease in some subjects, though the precise mechanisms for this are not well understood. Additional polymorphisms within the AAT gene may play a role in determining susceptibility [173].

In the UK the PiZZ genotype occurs in about 1 in 2000 individuals and the PiMZ genotype in 1 in 25 individuals[174]. In general the PiZZ genotype is that accepted to constitute a risk factor for both lung and liver disease[167], whilst the PiMZ genotype (a combination of a deficiency allele and a normal allele) probably does not in non-smokers. Nevertheless amongst patients with a diagnosis of COPD the PiMZ genotype has a prevalence up to 17.8%[175], greater than would be expected if it has no influence on disease expression. This might be because of interaction between the genotype and exposure to cigarette smoke. A meta-analysis of studies assessing risk of COPD due to the PiMZ genotype concluded that it represents a small riak factor, but was unable to account for smoking effects, or assess for gene-smoke interaction[176]. Further studies of this would certainly be interesting. In PiZZ subjects the expression of liver disease varies widely, from jaundice only at birth, through to cirrhosis, perhaps because of other genetic influences on this phenotype[173, 177]. Similarly variation in respiratory phenotype is observed even after standardisation by Pi genotype[17].

Polymerisation of AAT also underlies the PiMmalton (no dbSNP[114]reference number

available, 52Phe deleted)[178] and Siiyama (rs55819880, Ser53Phe)[179-181] forms of AATD, common in Sardinia and Japan respectively. As a result the clinical effect of these variants is very similar to the PiZ allele. A slower rate of polymer formation occurs with the PiS allele (rs17580, Glu264Val), because the structural change in β sheet A is not as radical[182, 183], resulting in a milder serum deficiency and little evidence of clinical disease. The PiS deficiency variant is more common in South-West Europe than the PiZ variant, which is more common in Northern Europe[184].If an individual has the PiSZ genotype their clinical phenotype, for liver disease[183] and lung disease in smokers[185] is usually intermediate between PiZZ and PiSS subjects.

Many null allelic variants have been described which result in an absence of AAT, and are denoted as QO rather than Pi. These are caused a variety of alterations in the DNA sequence, generally resulting in the formation of premature stop codons. A deletion of a single base pair (bp) in exon 1, which would usually contribute to the code for a tyrosine residue at position 160 in the protein, leads to the QOgranite falls genotype, resulting in a premature stop codon and unstable mRNA[186]. Deletion of 2bp in exon IV characterises QOhongkong, which also has a premature stop codon, hence translates to a truncated protein[187], which is then retained in the endoplasmic reticulum of the liver, thus accounting for its absence in serum. Three other null alleles have been described, which also result from the formation of stop codons due to either substitution, deletion or addition of base(s) in coding regions[188-190]. A larger 17 bp deletion is seen in the QOisola di procida allele, which is so substantial that no detectable mRNA or protein is produced[191]. Since such genotypes do not result in accumulation of polymers within the

liver subjects with null alleles are not susceptible to liver disease but may be more susceptible to lung disease than PiZZ individuals[192].

Some of the polymorphisms discussed here are summarised in Table 1.4.

Allele	Genetic defect	Cellular defect	Disease association
Z	1bp substitution	IC aggregation	Lung, liver
S	1bp substitution	IC aggregation	Lung*
Mmalton	3bp deletion	IC aggregation	Lung, liver
Siiyama	1bp substitution	IC aggregation	Lung
Mprocida	1bp substitution	IC degradation	Lung
Mheerlen	1bp substitution	IC degradation	Lung
QOgranite falls	1bp deletion	Unstable mRNA	Lung
QOhongkong	2bp deletion	Truncated protein	Lung
QOisola di procida	Deletion exons II-IV	No mRNA	Lung
Pimineral springs	1bp substitution	Poor inhibition of NE	Lung
PiPittsburgh	1bp substitution	Antithrombin III activity	Bleeding diathesis

Table 1.4: Effects of genetic variation within the Pi locus

The table shows some of the variants within the Pi locus that are associated with disease. All are listed by their usual allelic decriptors, rather than dbSNP references numbers, since this is the more usual nomenclature used.

*The risk of lung disease due to the S allele has been demonstrated only in smokers, and has not been studied in SS homozygotes[193].

IC=intracellular.

There is some evidence that other genes contribute to the clinical phenotype of AATD, as there is familial clustering of spirometric abnormalities in PiZZ and PiMZ subjects[58, 194] and declining lung function in PiMZ individuals[102]. The majority of genetic association work in AATD has concentrated on genes thought to have an influence in subjects with usual COPD, working on the premise that additive genetic risk would occur. Case-control studies examining candidate genes in AATD have not been widely published, but have implicated polymorphisms in endothelial nitric oxide synthase (*NOS3*) and *GSTP1*[195]. NOS3 regulates platelet aggregation, leukocyte adhesion, monocyte chemotaxis and smooth muscle tone[196]. The *NOS3* polymorphism was shown to have a significantly correlated with severity of lung disease, defined by FEV1 [197]. A lack of association with a functional variant in NOS3 suggests the presence of LD with an effect in a neighbouring gene.

GSTP1 is an antioxidant (section 1.1.3). The associated *GSTP1* SNP (rs947894, Ile105Val) [198] increases the metabolism of carcinogenic aromatic epoxides[116], a type of oxidant contained in cigarette smoke. Association studies of this variant with COPD have produced inconsistent results; the 105Ile variant is less active against oxidants and thus intuitively would be associated with higher levels of lung damage, but in AATD the opposite was found[195]. The study was small making a false positive result possible. A larger family based study has also suggested a role for variation in *IL10, TNFA, SERPINE2* and *EPHX1*[199], genes which were chosen for study based on their suspected role in either asthma or COPD. Evidence of a functional effect was seen for *IL10*, thats plasma levels varied with genotype[199].

1.3 Aims of this thesis

There is growing evidence that COPD is associated with a number of genetic predispositions, and exhibits a number of distinct sub-phenotypes. Subjects with AATD exhibit considerable clinical phenotypic variability, which cannot be explained by differences in smoke exposure. This suggests that AATD is part of the spectrum of genetic influences upon respiratory phenotype that interact with environmental factors to cause COPD. Genetic association studies in COPD, prior to the commencement of the work contained herein, had often been poorly replicated largely because of inadequate phenotyping and small datasets. Consequently, the only widely accepted genetic risk factor is AATD. Genetic associations in AATD have been less widely studied, being limited by the rarity of the condition, making recruitment of large datasets difficult. This thesis aims to examine selected candidate genes for association with COPD phenotypes in AATD, and dissect these from other clinical and environmental influences.

- Chapter 2 describes the methods used for clinical phenotyping of subjects, DNA extraction, quantification and genotyping, and the statistical methods used for analysis of clinical and genetic data. It also describes the use of public genetic resources in study design and interpretation.

- Inflammation, particularly that driven by TNFα, is key in COPD pathogenesis[200]. In AATD inflammation is more marked than in usual COPD[201] perhaps because interaction between AAT and TNFα[202] may

amplify TNFα effects. Chapter 3 investigates the relationship of variation within *TNFA* to clinical phenotypes in AATD, and the relationship of plasma and sputum TNFα levels to genotype and clinical phenotype.

- A number of other genetic associations in pathways relating to inflammation have also been identified in usual COPD, raising the question: do the same genes influence phenotype in AATD? The aim of chapter 4 is to examine phenotypic associations for three genes that had relatively well replicated associations in usual COPD, namely *SFTPB, TGFB* and *GC*.

- Protease driven lung damage is another key area of COPD pathogenesis. Matrix metalloproteases have been linked to two specific phenotypes in COPD: rapid decline of lung function and UZDE. Lung function decline tends to be faster in AATD[152], whilst UZDE is relatively uncommon. These phenotypes might, therefore, be of particular interest with regard to *MMP* genotype in AATD. Furthermore, a cluster of *MMP*s lie in an area of linkage found by a European study of AAT deficient discordant sibling pairs [203]. Chapter 5 details studies examining selected *MMP* variants to lung function, decline and UZDE in AATD.

- The HLA region was one of the first to be studied in COPD genetic association work[204], but only one gene in the HLA class III region has been associated (*TNFA*). The *TNFA* promoter shows strong LD with HLA class II alleles, but these have not been investigated, despite the observation of HLA class II

dependent cellular responses in emphysema[205]. HLA class II and III genes show association with autoimmune diseases, and emphysema also shows some features consistent with an autoimmune response in the lung[205, 206]. Chapter 6 describes phenotypic associations of HLA type in AATD, and evidence of autoimmunity in disease pathogenesis.

- Pollution, like cigarette smoke, is associated with airway inflammation and respiratory disease[207]. AAT provides the majority of protection from some pollutants, thus pollution could influence phenotype in AATD. The aim of this chapter was to test the hypothesis that pollution also plays a role in the pathophysiology of lung disease in AATD. Consideration was also given to the role of inhaled agents in the workplace and social deprivation.

- Chapter 8 concludes the thesis.

Chapter 2: Methods

2.1 Clinical phenotyping

2.1.1 Subject selection

Subjects were recruited from the UK national registry for AATD, which is based in Birmingham, UK. This is part of ADAPT (Antitrypsin Deficiency and Assessment Programme for Treatment), a programme started in 1996 and funded by a non-commercial grant from Bayer Biological (USA). Subjects were referred to the registry by health care professionals, including chest physicians and general practitioners, or identified through family screening of those related to someone with AATD. ADAPT has ethical approval to conduct clinical and genetic studies on registered participants (South Birmingham Local Ethics Committee 3359 and 3359a). Informed, written consent is obtained from all participants at registration, the consent forms being included in Appendix 1. All subjects registered with ADAPT have their diagnosis confirmed by genotyping (Heredilab, Salt Lake City, USA). Only Caucasians with a confirmed PiZZ genotype, and a stored sample suitable for DNA extraction, were selected for the studies included in this thesis, giving a total of 548 subjects. Subjects were divided into an unrelated group, for the case-control studies (n=444), and a familial group (n=104), acting as a replication dataset. Subdivisions in Chapter 7 differ from this, and are specified there.

The procedures described in this section are those for all subjects attending ADAPT.

Clinical data was collected by all research fellows in the department, including me, as well as research nurses. Extraction of data from the ADAPT records and calculation of specific phenotypes for use in studies contained in this thesis was performed solely by me.

2.1.2 Demographic data collection

At the baseline assessment a full clinical history was taken, including the symptoms necessary to diagnose chronic bronchitis[15] and frequency of exacerbations[208]. All basic demographic details (gender, date of birth, height, weight and home address) were recorded, alongside a full smoking history, including age started, age stopped (if relevant) and amount smoked per day, allowing calculation of pack year exposure (calculated by dividing the number cigarettes per day by 20 and multiplying by number of years smoked). Subjects were classified as index cases if they were diagnosed with AATD after investigation of their symptoms and non-index if they were diagnosed through family screening. This classification was of importance only in the familial replication dataset for the studies contained herein, but was available for all subjects.

2.1.4 Lung function

All tests were performed in the stable state by trained respiratory physiologists within the Lung Investigation Unit at University Hospitals Birmingham, UK, and compared to predicted values derived from standard reference equations for Caucasian adults (Appendix 2). At the baseline visit spirometry was performed before and after nebulised short acting bronchodilators were given[209]. Positive bronchodilator reversibility is

defined as an increase of greater than 200ml and greater than 12% predicted of the forced expiratory volume in 1 second (FEV1)[4] and was calculated for all subjects based on this definition from the raw data recorded in ADAPT databases. At subsequent annual visits spirometry was performed only after bronchodilatation, such that changes in lung function over time were derived from post-bronchodilator values. Annual follow up has been offered throughout the period that ADAPT has been established, such that those registered in the early years of the programme may have up to 12 years data available, with correspondingly fewer years for those registered later.

Static lung volume measurements including total lung capacity (TLC) and residual volume (RV) were assessed using helium dilution[210]. Gas transfer measurements (DLCO) were obtained by the single breath carbon monoxide method[211]. Together these values were used to calculate the uptake of gas per unit of effective alveolar volume (KCO). All these values wre determined only after bronchodilatation. The equation used to calculate this, and the reference equations for predicted values are in Appendix 2.

2.1.5 Computed tomography

High resolution computed tomography (HRCT) scans of the chest to detect emphysema and bronchiectasis were acquired using Prospeed and Lightspeed scanners (General Electrical Systems, Milwaukee, USA) at full inspiration, with the patient lying prone, as described elsewhere[212]. The presence of emphysema[213] or bronchiectasis[21] was documented according to published guidelines and reported by a single radiologist. Emphysema was quantified by the same radiologist, and by previous research fellows

within the department, using density mask analysis at a threshold of -910Hounsfield Units (HU), as described previously[209, 212]. The upper zone was assessed using a slice at the level of the aortic arch and the lower zone using a slice at the level of the inferior pulmonary veins. These regions were chosen in earlier studies as they were easily identifiable structures, giving greater reproducibility[209]. Density mask analysis of this type results in upper and lower zone voxel indices (UZVI and LZVI) at -910HU. Normal values for this have been derived from scans of healthy non-smokers[214] and are listed in Appendix 2. Relative predominance of emphysema in the upper zone (UZDE) was defined as a scan having a visual appearance of emphysema and UZVI>LZVI, whilst lower zone dominant emphysema had a similar radiological appearance but with LZVI>UZVI.

For the studies contained in this thesis the radiological diagnoses and density mask analyses were extracted from medical records and added to the ADAPT databases.
The clinical phenotypes obtained using the above data are shown in Table 2.1.

2.1.6 ADAPT databases

The ADAPT programme comprise 2 clinical databases, the older of the two was used to collect data from 1996, the newer one started in 2000 to improve both the quality and quantity of data collected. There is also a microbiological database and a laboratory database containing information on the location and type of stored samples. Data was extracted from all four sources, and unified with a project database, set up in Access 2000 and employing referential integrity to link all data types to individuals. Additional data

from CT scan reports was extracted from medical notes and intranet based hospital records.

Phenotype	Data type	Derived from
Chronic bronchitis	History	*
Exacerbation frequency per year	History	Symptoms using Anthonisen criteria[208]
FEV1 %predicted	Lung function	Raw ADAPT data and reference equations
FEV1/FVC	Lung function	*
KCO %predicted	Lung function	Raw ADAPT data and reference equations
FEV1 decline over 3 years	Lung function	Linear regression of raw ADAPT data
KCO decline over 3 years	Lung function	Linear regression of raw ADAPT data
Emphysema	Radiological	HRCT reports
Bronchiectasis	Radiological	HRCT reports
UZVI	Radiological	HRCT reports
LZVI	Radiological	HRCT reports
Emphysema zone	Radiological	LZVI-UZVI

Table 2.1: COPD phenotypes derived from clinical data

*Data regarding chronic bronchitis and FEV1/FVC was extracted directly from ADAPT databases and did not require any calculation or transformation for the studies contained within this thesis.

2.2 Laboratory methods

This section gives a general overview of methods used for DNA extraction, quantification, genotyping and other laboratory procedures. Specific details appear in the relevant chapters, and the composition of solutions used is in Appendix 3.

2.2.1 DNA extraction

DNA precipitation method

This method is a modified form of a commercially available kit (Nucleon Bacc II, Tepnel Life Sciences, UK) and was used for all subjects where at least 5ml whole blood, collected into an EDTA tube (Greiner Bio-one, UK) to prevent clotting, was available. Blood was mixed by inversion and transferred to a 50ml Falcon tube (BDH, UK), where 20ml of Reagent A (Appendix 3) was added to lyse the red blood cells. This was rotary mixed at 170rpm for 4 minutes at room temperature, followed by centrifugation at 2400rpm for 4 minutes. The supernatant was discarded without disturbing the white cell pellet and the centrifugation step repeated after addition of a further 10 ml of Reagent A. The supernatant was again discarded and the final cell pellet was then resuspended in 2 ml of Reagent B (Appendix 3) and vortexed briefly to lyse it. The cell suspension was transferred to a 4.5ml centrifuge tube (Alpha labs, UK), 500µl of 5M sodium perchlorate (Fisher Scientific, UK) added and contents mixed by inversion. Two ml of chloroform (at -20°C) was then added and the tube was inverted 10 times by hand. The tube was centrifuged at 3000rpm for 10 minutes and the resulting supernatant transferred to a new

4.5ml cryovial (alpha labs, UK) using a Pasteur pipette, without disturbing the lower pink protein layer. One hundred percent ethanol (Fisher Scientific, UK) at -20°C was added to fill the tube, which was then inverted several times to precipitate the DNA. If the DNA was not visible at this point the sample was incubated at -80°C for 30 minutes prior to the next step. The sample was then centrifuged at 3000rpm for 10 minutes to pellet the DNA, and the ethanol discarded. The DNA pellet was then washed with 1ml of 70% ethanol, and centrifuged again at 3000rpm for 5 minutes. The supernatant was discarded and the pellet left to air dry for at least 3 hours before resuspension in 200µl TE buffer (Appendix 3).

Spin column method

This method using a commercially available kit (DNEasy, Qiagen, UK) was employed for subjects where only a small blood sample (<1ml) was available. All reagents and sample tubes required are provided with the kit. Twenty µl Proteinase K was added to a microcentrifuge tube, together with 100µl of whole blood. The solution was made up to 220µl using phosphate buffered saline (PBS, appendix 3) prior to addition of the first buffer solution (Buffer AL), vortexing and incubation for 10 minutes at 70°C. This process lyses the cells to allow DNA extraction.

Subsequent steps were undertaken to bind the DNA to the filter in the spin column, and wash away other cell contents (such as proteins). Two hundred µl of ethanol was mixed with the sample and this mixture pipetted into a spin column, within a microcentrifuge tube, and spun at 8000rpm for 1 minute. The flow through and collection

tube were discarded and the process repeated using 500µl of buffer AW1, spinning at 8000rpm for 1 minute, followed by 500µl of buffer AW2 and further centrifugation at 14000rpm for 3 minutes. The DNA was then eluted from the spin column using 200µl of buffer AE. The column was incubated at room temperature for 1 minute and centrifuged at 8000rpm for 1 minute on 2 consecutive occasions to maximise the yield.

2.2.2 DNA quantification

Picogreen® method

The Picogreen® double stranded DNA (dsDNA) quantitation reagent is a fluorescent nucleic acid stain which integrates into the DNA and allows measurement of DNA concentrations down to 25ρg/ml using a standard spectrofluorimeter. The method relies on a standard curve prepared using calf thymus DNA (Sigma, UK) as shown in Table 2.2, and is unaffected by the presence of single stranded nucleic acids or RNA.

The DNA quantitation process was carried out over 2 days. On the first day 1:100 DNA dilutions were prepared in 1.5ml sterile eppendorfs (Alpha labs, UK) using 2µl of DNA and 198µl TE buffer (Appendix 3). Samples were mixed by vortexing and left overnight at 4°C to resuspend the DNA. On the next day the process was completed, using 96 well plates set up as shown in Figure 2.1, where rows A & B contain duplicates of the standard curve and rows C to H contain duplicate DNA samples to be measured. A 1:1200 dilution of Picogreen was made by mixing 1194µl of TE buffer with 6µl Picogreen in a 1.5ml eppendorf, which is referred to as Picogreen Solution A. This was vortexed and added to the blanks and standards in rows A and B. Picogreen solution B is a 1:280

dilution, made by mixing 7580µl of TE buffer with 20µl of Picogreen in a 30ml sterile foil wrapped Falcon tube (BDH, UK). This was vortexed prior to adding 95µl of Picogreen Solution B to each of the samples in rows C-H.

Plates were read at the standard fluorescein wavelength (excitation 480nm, emission 520nm) in a spectrofluorimeter (Perkin Elmer (Wallac 1420), UK). The fluorescence value of the blank was subtracted from the each of the sample values, and the DNA concentration of each sample was derived from the standard curve. If the standard curve duplicates deviated by more than 20% from one another the plate was rejected and repeated with new standard curve dilutions. If individual paired sample readings deviated by more than 20% from one another they were repeated.

	Calf thymus DNA	TE (µl)	Concentration after addition of Picogreen® (ng/ml)
1	4µl of 1mg/ml stock	1246	1600
2	600µl of 1	600	800
3	600µl of 2	600	400
4	600µl of 3	600	200
5	600µl of 4	600	100
6	600µl of 5	600	50
7	600µl of 6	600	25
8	600µl of 7	600	12.5
9	600µl of 8	600	6.25

Table 2.2: Preparation of a standard curve for the Picogreen® method of DNA quantitation

	1	2	3	4	5	6	7	8	9	10	11	12
A	BLANK	800	400	200	100	50	25	12.5	6.25			
B	BLANK	800	400	200	100	50	25	12.5	6.25			
C	S1	S1	S2	S2	S3	S3	S4	S4	S5	S5	S6	S6
D	S7	S7	S8	S8	S9	S9	S10	S10	S11	S11	S12	S12
E	S13	S13	S14	S14	S15	S15	S16	S16	S17	S17	S18	S18
F	S19	S19	S20	S20	S21	S21	S22	S22	S23	S23	S24	S24
G	S25	S25	S26	S26	S27	S27	S28	S28	S29	S29	S30	S30
H	S31	S31	S32	S32	S33	S33	S34	S34	S35	S35	S36	S36

Figure 2.1: Arrangement of 96 well plate for Picogreen® DNA quantitation
Blank = 50µl TE, Rows A-B 2-9 = 50µl of prepared standards, at the given concentration (ng/ml), Rows C-H = 5µl of each sample in duplicate.

2.2.3 Polymerase chain reaction

This section provides an overview of the principles of the technique, with specific details of the compositions of each reaction, specific methods and cycling conditions appearing in relevant chapters.

The polymerase chain reaction (PCR) amplifies a segment of DNA between regions of known nucleotide sequence. Two oligonucleotides are used as primers, which are different in sequence and complementary to the sequence on the opposite strand of the DNA template flanking the segment of interest. This allows amplification of the area of interest which lies between the primers. The other components of the reaction are deoxyribonucleotide triphosphates (dNTPs), *Taq* polymerase, magnesium chloride ($MgCl_2$) and a buffer. Template DNA is denatured at 95°C, and cooled to allow annealing of the primers to the template DNA. The action of *Taq* polymerase then amplifies the segment of interest through addition of complementary dNTPs, and the cycle is repeated many times, with the product of each cycle acting as the template for the next. This process is illustrated in Figure 2.2. Each PCR reaction should be optimised both in terms of the amount and concentration of its components and the thermal cycling conditions.

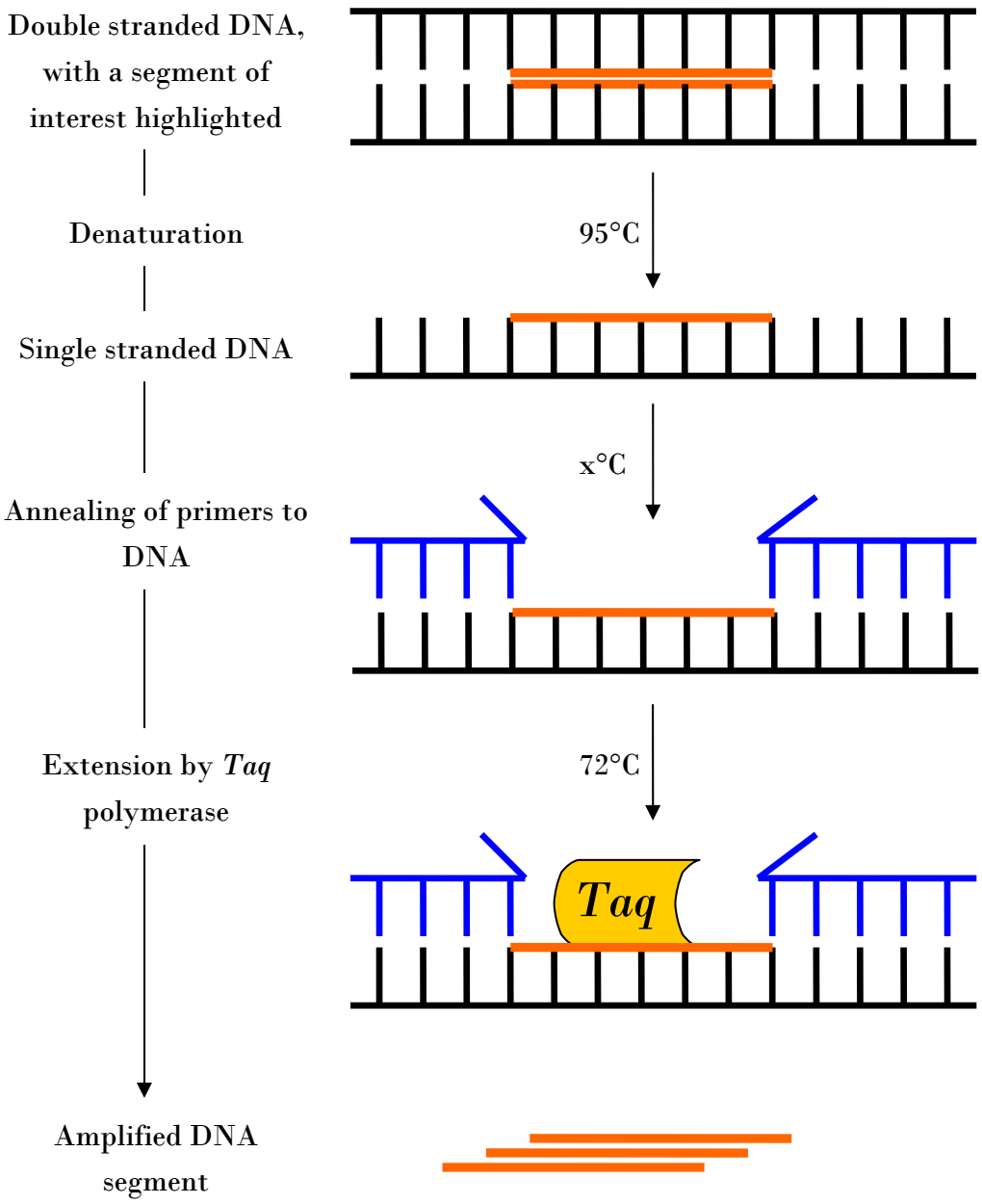


Figure 2.2: The polymerase chain reaction
Double stranded DNA is denatured at 95°C, to form single stranded DNA. The oligonucleotide primers can anneal to the exposed areas where their sequence is complementary, the temperature at which this occurs being dependent on the oligonucelotide sequence. When *Taq* polymerase is present it amplifies the region between the primers, the optimum temperature for being 72°C. The end result is an amplified amount of the DNA of interest.

Oligonucleotide primers

Primers are designed so that they are complementary to the template DNA flanking the segment of interest. The technique, therefore, requires knowledge of the nucleotide sequence. Access to relevant information has become easier with the advent of a number of online public genetic databases that catalogue sequences and known SNPs within the human genome, including services hosted by the National Centre for Biotechnology Information (NCBI), such as Entrez SNP[114], and Ensembl[215], hosted by the Wellcome Trust and the Sanger Institute. Both of these resources were used for studies within this thesis.

In general primers are between 20 and 30 base pairs in length, and should have <50% GC content in order to avoid high annealing temperatures (Tm). Primer pairs should have a similar GC content, whilst being non-complementary, to decrease the formation of primer dimers. For the HLA studies contained in Chapter 6 the primer sequences had already been determined and published elsewhere[216]. In fluorescence based PCR techniques, such as TaqMan®, the primers have a fluorescent marker attached which is cleaved during extension by *Taq* polymerase. This then emits a signal detectable and quantifiable by fluorimetry. This technique was used for most of the SNP genotyping studies within this thesis, and as such there were 2 fluorescent signals per reaction to detect the two possible alleles present within a diallelic SNP - these being named VIC and FAM, and allocated by the manufacturer of the primers. The highest VIC signal represents homozygotes for one allele, the highest FAM signal homozygotes of the other allele and those of an intermediate level represent heterozygote DNA. Both signals are

read by fluorimetry based techniques which produces plots on which the signal from each subject can be visualised (see results in chapters 3, 4 and 5). This system works well with low DNA concentrations, such that all samples were diluted to a fixed concentration of 20ng/µl for the studies reported in this thesis. The manufacturers of TaqMan® genotyping technologies (Applied Biosystems) both prepared and optimise the primers based on sequence supplied to them by me, and the precise primer sequence was not released to me.

$MgCl_2$ concentration

Buffers used in PCR are at the optimum at pH (8.3) for *Taq* polymerase activity, however the pH of the reaction falls to 7.2 when incubated at the extension temperature (72°C) such that the presence of divalent cations is critical. Magnesium cations are the most effective, although the concentration required varies between reactions, dependent on the primer pairs used. Optimisation can be achieved by performing an assay at concentrations from 0.5-4.0mM. This step was not necessary for the studies contained herein, as all reactions worked well under the published conditions for HLA[216] and for the TaqMan® studies the PCR reagents were supplied pre-mixed by Applied Biosystems.

dNTP concentration

Deoxyribonucleotide triphosphates (dNTPs) are present to allow synthesis of the new DNA during PCR. Usually they are used at a concentration of 200µM, which is sufficient to synthesise 12.5µg of DNA when half the dNTPs are incorporated. An increased concentration may increase the efficiency of the PCR, but can also increase the error rate

of *Taq* polymerase, or even inhibit it. Alterations in dNTP concentration also affect the optimal $MgCl_2$ concentration required. No alteration of published conditions was required for HLA genotyping[216] and for the TaqMan® studies the PCR reagents were supplied pre-mixed by Applied Biosystems.

Taq DNA polymerase

This enzyme catalyses the PCR and amplifies the product. As such increasing concentrations increase the efficiency of the reaction, although this must be balanced against the risk of producing non-specific PCR products alongside the area of interest. No alteration of published conditions was required for HLA genotyping[216] and for the TaqMan® studies the PCR reagents were supplied pre-mixed by Applied Biosystems.

Reagent purity

A blank sample containing all components except DNA was run alongside each PCR to confirm the lack of PCR product, adding validity to positive results. If a PCR product was detected in the blank it indicated that contamination of the reagents had occurred; results from the same reaction were discarded and PCR repeated with fresh reagents.

Thermal cycling conditions

Denaturation

This separates the dsDNA template into single strands and was achieved by incubating the PCR mixture at 94°C for between 30 and 60 seconds (sections 3.2.2 and 6.3.2). Higher temperatures can decrease the activity of *Taq* polymerase, which must be taken into consideration when designing an assay.

Annealing

This step was used prior to extension and took 30 to 60 seconds to enable primers to anneal completely and stabilise the template DNA. The temperature at which the primer anneals can be estimated from the following equation

Tm =4(GC) + 2(AT)

Tm = annealing temperature
A, T, G and C = adenine, thymine, guanine and cytosine, the nucleotide bases

Tm, therefore, varies directly with the GC content of the primer. In the case of PCR reactions containing multiple primers, like those for HLA typing (section 6.3.2), primers had similar annealing temperatures.

Extension

This occurs at 72°C, which is close to the optimal temperature for *Taq* polymerase action (75°C) but prevents detachment of the primers from the template. The duration depends on the length of the sequence to be amplified, though 1 minute per 1000bp is usually sufficient. The TaqMan® PCR technique combines the annealing and extension steps (section 3.2.2).

2.2.4 Gel electrophoresis

Agarose gel electrophoresis was used to visualise the PCR products for the non-TaqMan® genotyping techniques, namely HLA typing. Gels were prepared at a concentration appropriate to the size of the product generated using electrophoresis grade agarose (Sigma, UK). All gels were made with an appropriate buffer (TBE, Appendix 3) and run in an electrophoresis tank filled with the same. The agarose was dissolved in the buffer by warming in a microwave and allowed to cool prior to addition of ethidium bromide (Sigma, UK) to aid visualisation, as it cross links DNA and fluoresces under UV light, which is used to view the results. The gel was then poured into a casting tray with combs and allowed to set. Once set the gel was placed in the electrophoresis tank, samples were mixed with a loading dye (see appendix 3), electrophoresed and visualised under UV light. Photographs were taken to record results.

2.2.5 Enzyme linked immunoadsorbent assay (ELISA)

This technique was used to measure TNFα and anti-elastin antibody levels. This section describes general methods for an ELISA; the specific process for the TNFα assay and anti-elastin antibody assay are given in the relevant chapters (sections 3.3.2 and 6.3.4 respectively)

A quantitative sandwich enzyme linked immunosorbent assay (ELISA) is illustrated in Figure 2.3. The amount of measured substance is quantified by means of comparison to a standard curve, this being set up using samples of known concentration, such that unknown samples are expressed relative to this. The final colour generating step must be stopped at a given time point (specific to the ELISA performed) prior to reading the plate in a standard microplate reader. All ELISA measurements were run in duplicate on 96 well plates, the set-up of the plates being similar to that shown for Picogreen measurements of DNA concentration (Figure 2.1).

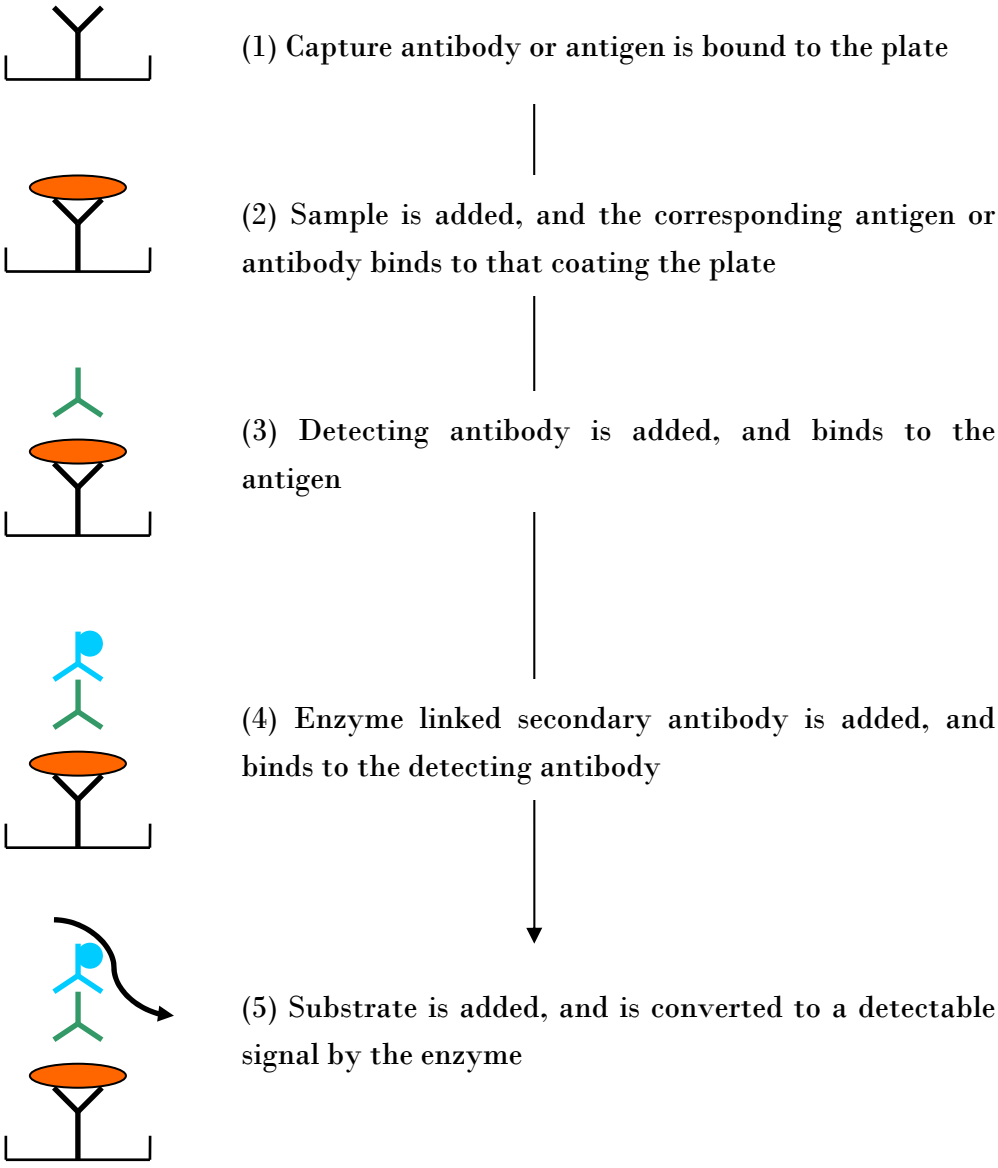


Figure 2.3: The process of a sandwich ELISA
Microplates are coated with a specific antibody or protein (1) to bind the corresponding protein present in the samples (2). Unbound substances are washed away, and a polyclonal antibody specific to the protein being measured is added to the plate (3). A second wash removes unbound antibody and an enzyme linked antibody specific for the primary antibody is added (4), followed by an amplifier solution which develops colour in proportion to the amount of measured substance present (5). The reaction is stopped by addition of a stop solution (usually an acid); plates are then read using a microplate reader.

2.2.6 Quantitative sputum culture

Bacterial colonisation status was required for studies reported in chapters 3 and 4 of this thesis. Some quantitative sputum culture results were already recorded in ADAPT databases. Fresh stable state sputum samples were also obtained from subjects who spontaneously produced sputum and attended ADAPT for assessment. Bacterial colonisation was taken to be >1 x 10^5 colony forming units/ml (cfu/ml) of recognised respiratory pathogens in a stable state sample.

Quantity of organisms

One gram of sputum was taken from the samples and homogenised by addition of an equal volume of dithiothreitol (Sputasol, Oxoid, UK) and vortexing for 60 seconds. Serial dilution of the sample was carried out using a pipette to remove 100µl, which was added to an appropriate volume of distilled water (Baxter, UK) to create dilutions starting at 1:10 and continuing to 1:1000000. Each dilution was vortexed prior to use and 10µl of neat or diluted sample was spread onto chocolate agar plates (Oxoid, UK) using a sterile plastic 'hockey stick' (Appleton Woods, UK) in the 4 zone streak method, illustrated in figure 2.4. The 1:10000 dilution was also added to a blood plate (Oxoid, UK) to aid subsequent species identification. Plates were incubated at 37°C for 24-48 hours and the number of colonies counted from the most appropriate dilution (a plate with no more than 300 colonies). The number of colony forming units per ml (cfu/ml) was then calculated based on the dilution of sample on the plate.

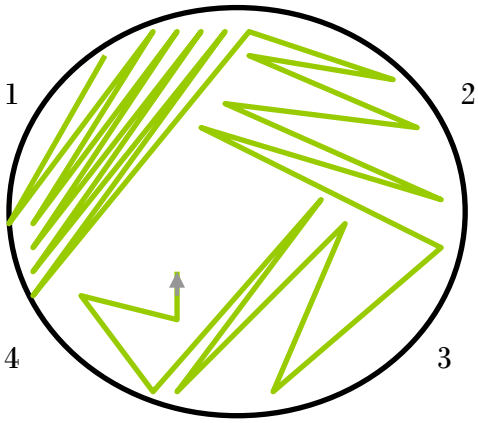


Figure 2.4: Plating of sputum samples
The four zone method ensures even spreading of the sample across the plate, such that most colonies are in zone 1 and progressively fewer down to zone 4.

Identification of organisms

The predominant organisms were identified using the appearance of the colonies, microscopy (identifying bacterial morphology), Gram stain and a series of functional tests, usually performed in the order in which they are listed here. Further physiological testing was carried out by a microbiologist if needed. The features seen with some common respiratory pathogens are shown in Table 2.3.

Gram stain identification of bacteria

Gram positive bacteria have a thick cell wall containing a high proportion of peptidoglycans (50-90%), whilst Gram negative bacteria have a thinner cell wall containing much less of these substances (around 10%). A small amount of sputum from the sample was smeared onto a clean microscope slide (Appleton Woods, UK) and allowed to air dry for a few minutes. This was stained by addition of crystal violet (Biomerieux, UK), which was allowed to act for 1 minute and rinsed with tap water prior to addition of Lugol's solution (Biomerieux, UK). This was left for 1 minute then washed off prior to decolourisation with ethanol (Fisher Scientific, UK). Finally a counter stain (Safranin, Biomerieux, UK) was added. Gram positive organisms retain the crystal violet based complexes within the cell due to the high peptidoglycan content of the cell wall, thus appear purple. The violet coloured complexes are washed out of Gram negative cells by the ethanol, and then appear pink due to the counterstain. This was visualised using a microscope and the results noted into the ADAPT microbiology database. The microscope did not have the facility to photograph magnified samples.

Functional tests

The following list is not an exhaustive list of the functional tests that are performed but contains those needed to identify the common respiratory pathogens. The results of the oxidase and catalase tests, together with colony appearance, with or without a Gram stain result (see also Table 2.3) guided the specific tests of bacterial species listed by the bacterial name.

Oxidase test

This identifies bacterial species able to oxidase tetramethyl-sp-phenylene-diamine-dihydrochloride (BDH, UK), such that when a colony was added to a drop of this solution the colour changed from colourless to purple.

Catalase test

This identifies bacterial species able to reduce hydrogen peroxide (BDH, UK) such that bubbles of hydrogen were seen if a colony was added to a drop of this solution.

Characteristic	Growth on agar	Colony appearance	Gram stain	Cell morphology	Oxidase test	Catalase test	Other tests
Haemophilus influenzae	Chocolate	Small, grey, translucent	Negative	Small coccobacilli	Negative	Negative	X & V
Moraxella catarrhalis	Chocolate Blood	Opaque, small, white, dry, pushes along agar surface	Negative	Large diplococci	Positive	Positive	Tributyrin
Pseudomonas aeruginosa	Chocolate Blood	Mucoid or rough, usually colourless, green haemolysis of blood agar	Negative	Bacilli	Positive	Positive	Cetrimide plate
Streptococcus pneumoniae	Chocolate Blood	Dark green, green haemolysis, draughtsman colonies	Positive	Diplococci	Negative	Negative	Optochin
Staphylococcus aureus	Chocolate Blood	Large, yellow, relatively small area of haemolysis	Positive	Clusters of cocci	Negative	Positive	Agglutination
β-haemolytic *Streptococci*	Chocolate Blood	Small, dry, white, large clear haemolytic zone	Positive	Chains of cocci	Negative	Negative	-

Table 2.3: Identification of common respiratory pathogens by their appearance and functional testing

The table shows the preferred growth medium, colony appearance, microscopic appearance after Gram staining and basic functional test results for selected respiratory pathogens.

***Haemophilus* spp**

X and V factor testing was used to identify *Haemophilus* spp. X factor is comprised of protoporphyrin IX, haemin or other iron containing porphyrins. X dependent strains are unable to convert d-aminolaevulinic acid to protoporphyrin, thus require this factor for growth. V factor comprises nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP)[2]. Both factors are present in blood, so care must be taken to avoid carryover of blood from the original plates when picking the colony for testing. Testing was performed by placing discs containing either or both factors (Oxoid, UK) onto agar, to which the colony to be identified was applied. *H.influenzae* (a common respiratory pathogen) requires both factors for growth, thus colonies will appear only around the disc containing both.

Streptococcus pneumoniae

The optochin test was used for *Streptococcus pneumoniae*. This detects susceptibility to ethylhydrocupreine hydrochloride (optochin) and differentiates *S. pneumoniae* from other α haemolytic *Streptococci*. The bacterial cell membrane of *S.pneumoniae* lyses in the presence of optochin due to changes in surface tension. The test was performed by applying a paper disc impregnated with optochin (Oxoid, UK) to an inoculated agar plate, and if positive a zone of inhibition was seen around the disc.

Pseudomonas spp

Suspected *Pseudomonas* spp were grown on cetrimide plates (Biomerieux, UK). Most non-*Pseudomonas* species will not grow on this medium. These plates identify *P.aeruginosa* because only this species of *Pseudomonas* produces pyocyanin, a blue-green pigment, and fluoresces under UV light.

Staphylococcus aureus

Specific agglutination kits (Biomerieux, UK) were used to identify *Staphylococcus aureus* (Figure 2.6) using colonies emulsified in normal saline (added to the two reagents provided.

***Moraxella* spp**

The tributyrin hydrolysis test was used to identify *Moraxella* spp. Oxidase and catalase positive colonies of suggestive morphology can be distinguished from *Neisseria* spp by this test which detects the enzyme butyrate esterase, present in *Moraxella* spp, but not *Neisseria* spp. Some colonies were emulsified in water, sufficient to create a cloudy solution, to which a tributyrin tablet (Bioconnections, UK) was added. Initially this causes the solution to go peachy-white in colour. If *Moraxella* spp are present it will turn yellow, usually within twenty minutes, though samples were left for several hours to avoid false negatives.

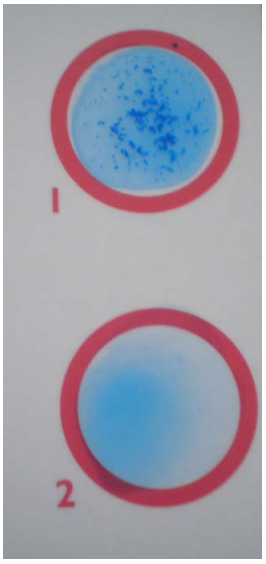


Figure 2.5: Positive agglutination test
A positive result is indicted by the appearance of agglutination with solution 1. A negative result would have the same appearance with both solutions as shown here for solution 2.

2.3 Statistical analysis

All data with the exception of that for familial genetic association, haplotype associations and HLA type associations were analysed using SPSS version 12 (SPSS Inc, Chicago, USA).

2.3.1 General statistical tests

Studies reported have generated data pertaining to a variety of COPD sub-phenotypes (described in section 1.1.1), TNFα levels in both plasma and sputum, and anti-elastin antibody levels, as well as genetic polymorphisms. Thus, for each chapter it has been necessary to describe the cohort studied, and then compare subgroups. In each case data normality was assessed using the Kolmogorov-Smirnov test. Parametric data has been displayed as mean with standard error (SE) and non-parametric data as median with inter-quartile range (IQR). Subgroup comparisons between clinical phenotypes and genotypes were made using the t-test, Mann-Whitney or Kruskal-Wallis test as appropriate. Spearman's test was used for continuous correlations where the distribution of data was non-normal and Pearson's for normally distributed data. Frequency comparisons were made using the Chi squared (χ^2) test.

Linear regression was used to calculate decline in lung function. Subjects were included for the genetic studies if they had at least 4 data points in the regression: a baseline and 3 subsequent values. Other methods of considering decline for pollution analyses are described in Chapter 7 (section 7.3.2). The slope of the resultant regression line was used to indicate decline in the given lung function parameter per year. Those variables that were

non-normally distributed within the group were transformed appropriately prior to use in subsequent regression models assessing influences on this phenotype.

Statistical significance was taken as $p<0.05$. Where correction for multiple statistical testing was appropriate a Bonferroni correction was used. However, all values of genetic studies are reported uncorrected, since this type of correction may be too conservative in this type of study[68]. Genome wide significance was accepted as $p<5 \times 10^{-5}$ [68]; though more stringent cut off values have been proposed by some groups[217]. Where genome wide significance was not reached in the case-control genetic association analyses the studies were repeated in a replication dataset of AATD sibling pairs. This is a recognised method of confirming the significance of a genetic association in complex disease[68].

2.3.2 Genetic association studies

For the genetic studies reported within this thesis the majority of analyses were carried out using haplotype tagging SNPs. Methods relating to their selection, prior to the actual analyses performed are discussed below.

Choice of haplotype tagging SNPs

The HapMap is a catalogue of common genetic variation in different ethnic groups (originally, and on the build of HapMap used herein, in just four races: Caucasians, Japanese, Chinese and Africans)[70]. Haplotype blocks are present throughout the human genome because of the limited number of recombination sites present in human DNA. During meiosis chromosomal strands can cross over such that pieces of DNA are swapped,

resulting in a hybrid chromosome composed of parts of each original chromosome. Genetic variants, such as SNPs, vary in their frequency between racial groups, as does the structure of resulting haplotype blocks. Genes can be broken down into component blocks, each of which will contain a number of SNPs, at known frequencies. Since these SNPs will be inherited together (in LD) if they are in the same haplotype block, it is only necessary to genotype 1 SNP per block. This is illustrated in Figure 2.6. A SNP that marks a block in this way is known as a tag SNP. Tag SNPs can be chosen using programmes within HapMap, for the specified racial group, setting limits for the minor allele frequency (MAF) you wish to detect and the level of correlation between SNP pairs. For tags in strong LD the r^2 between them will be at least 0.8 (see section 1.1.4) and it is this value that identifies the blocks. Conversely values of r^2 between 0.2 and 0.8 indicate incomplete LD and values <0.2 no LD. This may be visualised by the use of LD plots which show areas of high r^2 in red and low LD in white. Examples of such plots obtained for studies within this thesis can be seen in chapters 3-5.

For all the tag SNP studies in this thesis a cut off of MAF = 5% (0.05) and r^2 >0.8 were chosen and tags identified in HapMap build 35. HapMap output files detail the reference SNP[114] number of the tag, and those that it captures within the haplotype block. Using this algorithm a total of 38 tags were chosen to cover the following candidate genes: *GC, SFTPB, TGFB, MGC4093, CCDC97, MMP1, MMP3* and *MMP12*. For each block the tag with the highest MAF was chosen, where possible, in order to improve the power of the study (see appendix 4). At the time of the studies reported here HapMap had not fully reported all genes, such that some known SNPs within the literature were not reported there. This was most notable for *TNFA* and thus other published resources were used for

studies of this gene[218]. Details of the tags explored in this series of studies, and the number of captured SNPs, are shown in Table 2.4.

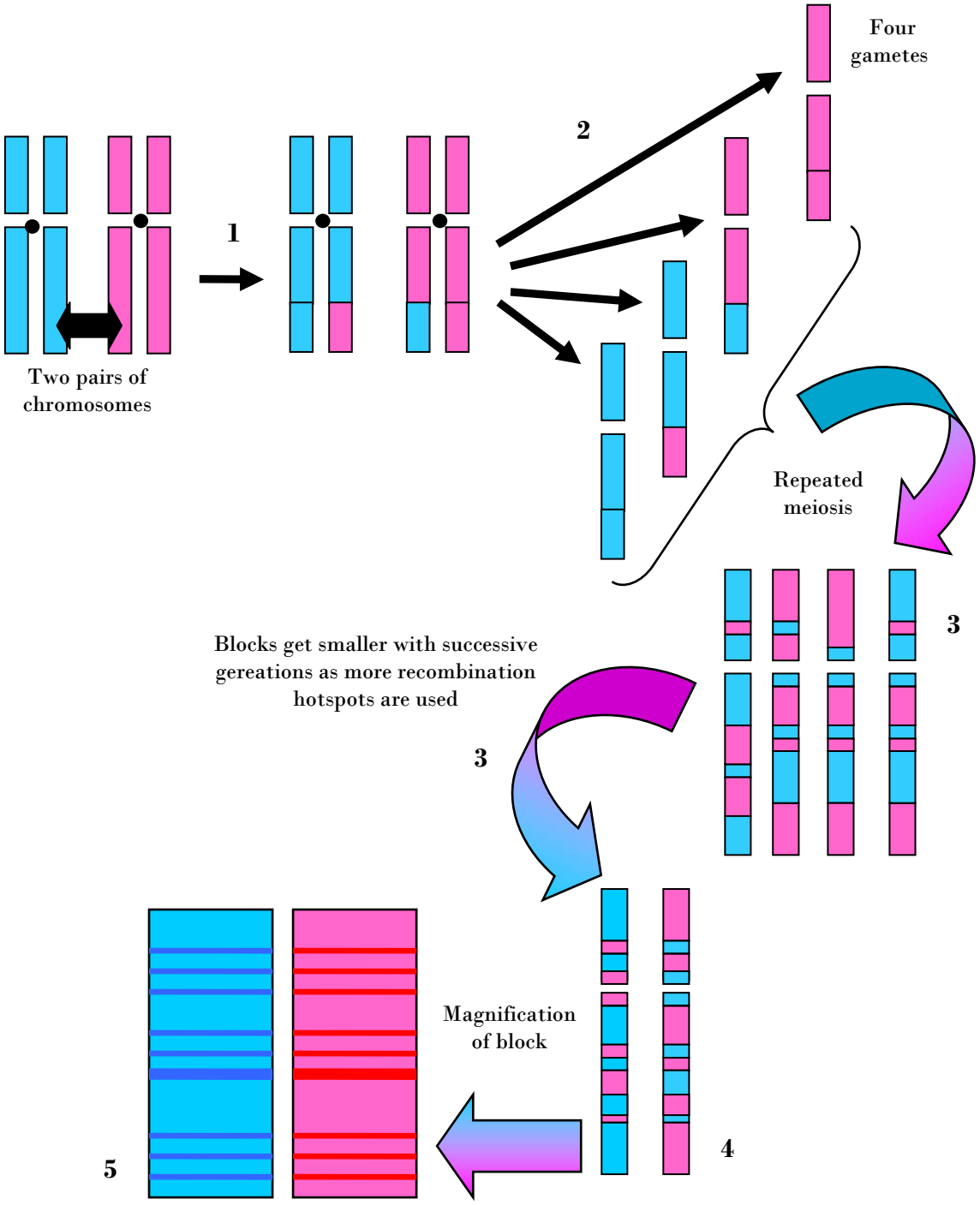


Figure 2.6: The formation of haplotype blocks and concept of tagging a block
During meiosis chromosomal segments may cross over (1) resulting in the formation of gametes with different combinations of genotypes (2). This process is repeated over many generations (3). HapMap provides information about the current haplotype blocks present for any given chromosome (4), by genotyping of SNPs within them. For any given block there will be a number of SNPs present, illustrated by lines (5), the colour indicating the allele present, any of which might be chosen as a tag. In this case the tag SNP chosen has been marked by showing the SNP with a thicker line.

Gene	Tag SNPs		Number of captured SNPs
GC	rs222035	rs2070741	28
	rs1491718	rs705117	
	rs222023	rs1352844	
	rs2298849	rs3733359	
	rs222029	rs3755967	
SFTPB	rs2118177	rs1130866	3
TGFB	rs8110090	rs11466338	6
	rs4803455	rs11466345	
	rs8179181	rs1800469	
MGC4093	rs11083616		3
CCDC97	rs8109627	rs1041794	9
	rs8109167		
MMP1	rs470747	rs7125062	13
	rs996999	rs3213460	
	rs7125320	rs5031036	
	rs470358		
MMP3	rs678815	rs3025066	14
	rs476762	rs650108	
MMP12	rs476185	rs11225442	13
	rs652438	rs17368659	
	rs632009		
TNFA	rs1800629	rs1799964	8
	rs361525	rs3093662	

Table 2.4: Tag SNPs covering selected candidate genes

Genetic association analyses

Deviation from Hardy-Weinberg equilibrium (HWE) was checked for all SNPs (Appendix 4). This is a crude measure of genotyping error, as well as being an indicator of the stability of a SNP between generations. A SNP not in HWE may represent a genotyping error, particularly if this is seen in control subjects. It is possible for deviation from HWE to occur in cases if the SNP adversely affects survival, and procreation, but this is less common than problems of genotyping. Those that were not in HWE were excluded from further analysis in the studies contained herein.

The power of the study to detect phenotypic differences between groups was calculated for qualitative phenotypes, prior to adjustment for covariates, based on the genotype frequencies obtained (Appendix 4).

Primary comparisons were made against AATD subjects without the phenotype for which association was sought. All AATD subjects were UK Caucasian, thus controlling for racial genetic variation.

Case-control multiple regression analyses

Binary logistic regression was used to test for independent predictors of qualitative phenotypes, with genotypes coded as a categorical variable for genotypic association, effectively being a general genetic model. A dominant model was also constructed with coding of 1 or 0 to indicate presence or absence of the minor allele. The OR of disease conferred by the associated allele or genotype was indicated by the B value in these regressions. Stepwise linear regression was used to test for independent predictors of

quantitative phenotypes. The B value from the regression indicates the amount of change in the outcome variable if the predictor is increased by 1 unit and all other features remain constant, whilst the r^2 value indicates the proportion of variation in the outcome variable attributable to the predictor. It should be noted that in a regression the r^2 is not the same as that derived from correlations, such as those used to ascertain LD, even though the annotation is identical, so this has been clarified where needed.

Haplotype analyses

Haplotype analyses were carried out in Haploview[219] for qualitative phenotypes only, between affected and unaffected AATD subjects. Where they were required for quantitative outcomes the haplotypes were identified in Haploview and the association analyses performed in SPSS using coding of each haplotype in a general model in the same manner as the regression analyses above.

Familial genetic association

Theory

Genetic data from related individuals cannot be analysed in the same way as a standard case-control association study, since affected and unaffected family members will share a proportion of alleles due to common ancestry. In the UK ADAPT databases there are a number of sibling pairs, who comprise the replication dataset for the larger case-control studies contained within this thesis. Sibling pair genetic studies are 'model free';

the rationale being that an excess of disease causing alleles will be seen in affected family members regardless of the mode of inheritance. The first step is to calculate the proportion of alleles identical by descent (IBD). This can be used to assess qualitative or quantitative phenotypes, the latter having been chosen where possible for the studies in this thesis. The basis of this is to regress the squared difference in the trait value between siblings on the proportion of alleles IBD[220]. Two siblings that share more alleles IBD would be expected to have more similar trait values if the alleles were influencing the phenotype. Consequently there would be a negative relationship between the squared trait differences and the IBD sharing. Since this is a linear regression covariate effects can be taken into account as they would in any stepwise regression, and interactions between covariates (including genes) can be specified with most software packages.

Sibling pair trait analyses

All analyses were carried using the software package Statistical Applications in Genetic Epidemiology (S.A.G.E.)[221]. The GENIBD function was used to generate IBD sharing distributions, the output file indicating the probability of each pair sharing 0, 1 or 2 alleles IBD. This file was then used for trait analysis using the SIBPAL function. The output file generated indicated significance for each predictor in the analysis and the direction of association, the latter being determined by the t statistic.

HLA region analyses

HLA class II loci are multiallelic; as such the methods described for coding genetic data in SPSS for regressions could not be used. Analyses were carried out using stata version 10

(www.stata.com) specifying a multiplicative model, and using routines available from www-gene.cimr.cam.ac.uk/clayton/software/stata to test for deviations from HWE. Linear regression was used for pulmonary function (FEV1, FEV1/ FVC and KCO) and logistic regression for emphysema, adjusting for smoking, age and gender. Locus associations were tested first. The alleles were then categorised within the regression to obtain individual allelic associations, in each case setting the most frequent allele as the comparator against which the likelihood ratio was calculated and grouping at the 1% frequency level. Further detail is given in Chapter 6.

Chapter 3: Phenotypic associations of TNFα and *TNFA* in AATD

3.1 Aims of this chapter

This chapter will detail the COPD phenotypes observed in AATD and the phenotypic associations *TNFA* variation, and of circulating and sputum TNFα levels in such subjects. Independent effects of *TNFA* variants on circulating and airway levels of TNFα will then be described.

The introduction, methods and results for *TNFA* and TNFα will be presented in sub-sections specific to genetic association, and protein associations with clinical phenotype, prior to an analysis of protein level associations with genotype which utilises results from both preceding sub-sections, prior to an overall conclusion for the chapter. Much of the data presented here has been published in a peer reviewed journal (Appendix 7).

3.2 TNFA

3.2.1 Introduction

TNFA, which encodes TNFα, is located on chromosome 6p21.3, and is situated within the HLA class III region, illustrated in Figure 3.1. *TNFA* is composed of 4 exons arranged over approximately 3kb of DNA[222], as shown in Figure 3.2. Regulation of TNFα production at the transcriptional level is controlled by several regulatory sequences at the 5' end of the gene, including sequences for the activator proteins 1 and 2 sites and a cyclic adenosine monophosphate (cAMP) responsive element[223]. The 3' untranslated region also contains a sequence which affects mRNA stability and therefore subsequent translation efficiency[223]. SNPs affecting these sequences have the potential to alter the production of TNFα, which if present in excess might drive the inflammation seen in COPD.

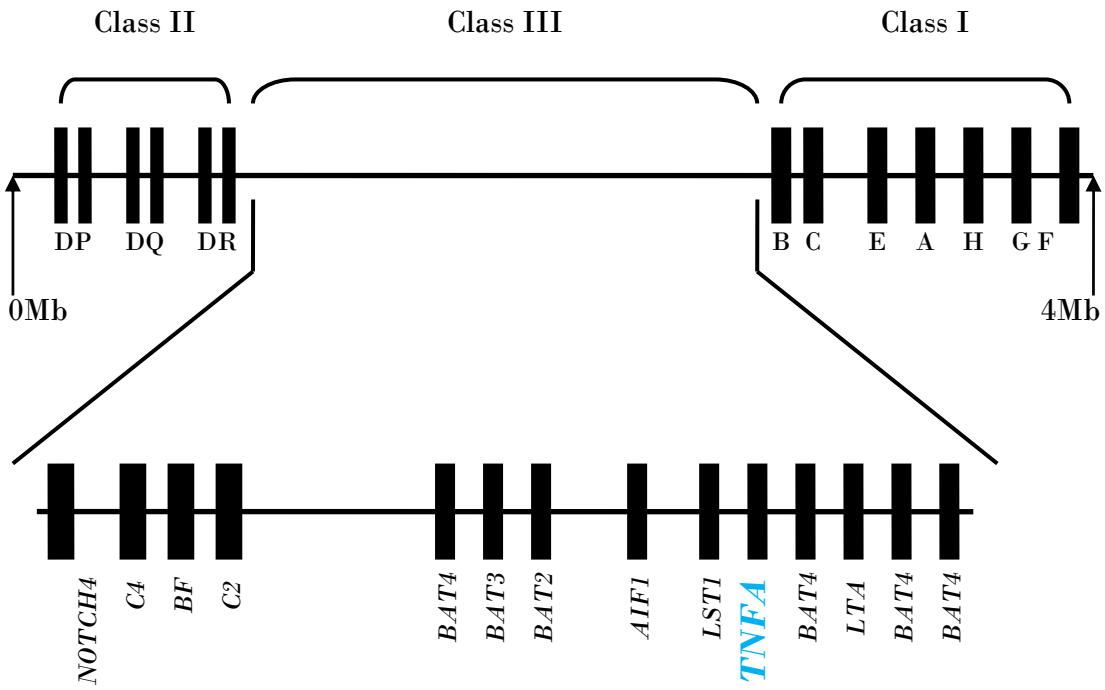


Figure 3.1: The HLA region
This is a simplified diagram of the HLA region on chromosome 6p21. Detail of key genes in each class is shown, with class III expanded.

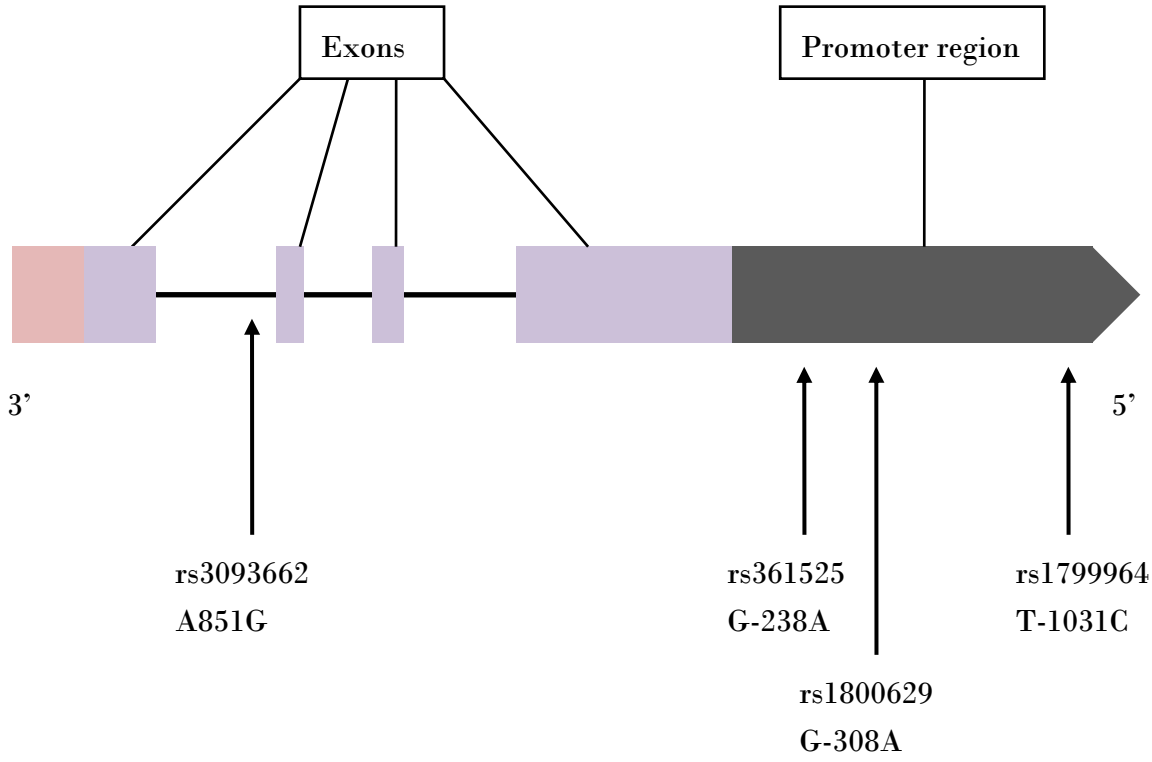


Figure 3.2: The structure of *TNFA*
The structure of the gene is shown, together with the SNPs studied in this chapter. The original SNP descriptors are shown beneath the current reference SNP numbers, for ease of comparison with prior published work.

TNFA in COPD and AATD

TNFA has been investigated in several COPD phenotypes[72, 105-107, 224, 225], predominantly with selected SNP work[105-107, 224, 225], though one study used tag SNPs to screen the whole gene for association[72]. These studies have defined their cases by FEV1 and FEV1/FVC diagnosed COPD[72, 105, 224], presence of emphysema[106, 225] and chronic bronchitis[107]. The associations have been poorly replicated, perhaps because of these differences in patient characterisation, or because of low numbers of patients in each study. LD with HLA alleles[226] and racial variation may also have contributed[227]. If an association is actually due to an allele with which it exhibits a degree of LD some studies may show the association and others, perhaps where the population have a different degree of LD, may not. Similarly racial differences may affect the distribution of SNP genotypes, a fact which prompted the genotyping of four racial groups in the studies underying HapMap. A recent meta-analysis concluded that a promoter region SNP (rs1800629, G-308A) is significantly associated with COPD[228], but the validity of this may be questioned as it pooled results from studies that defined their COPD subjects using differing criteria.

In AATD there is significant variation in the clinical phenotype. Both familial and non-familial studies suggest that, as in usual COPD, other genetic influences may be important in determining clinical features[194, 195, 197, 199], but no studies have examined the role of candidate genes in the full range of COPD phenotypes previously associated with *TNFA* polymorphisms. Furthermore, the interaction between AAT and TNFα might magnify any effect of *TNFA* polymorphisms in AATD. Epistatic effects

explicable by cytokine interactions, such as this, are also seen in asthma when multiple polymorphisms in genes involved in the Th2 immune response magnify the risk of asthma phenotypes[229]. For instance, both interleukin-13 (IL13) and interleukin-4 (IL4), produced by Th2 cells, are capable of inducing B lymphocytes to produce IgE after allergen exposure and share a common receptor component, interleukin-4 -receptor alpha (IL4Rα)[229]. The presence of associated alleles of single SNPs with asthma phenotypes (such as bronchial hyperresponsiveness) in *IL4RA* and *IL13* together in an individual was found to increase the risk of asthma relative to either associated allele being present alone[229].

3.2.2 Methods

Subjects

Four hundred and twenty four unrelated UK Caucasian subjects with the PiZZ genotype of AATD were studied in the initial case-control study. Although phenotypic information and blood had been stored on a further 20 subjects, the samples of blood on these subjects were so small as to yield insufficient DNA for genotyping of all SNPs and were therefore excluded from further analyses. Subjects' characteristics, with additional stratification by the phenotypes described in chapter 2 (section 2.1 and Table 2.1) are shown in Table 3.1. The frequency and degree of phenotypic overlap for the qualitative phenotypes is illustrated in Figure 3.3. It should be noted that only 82% of subjects underwent a CT to formally diagnose emphysema, and as such the number of subjects for whom CT derived diagnostic information is available is less than that for lung function.

Similarly annual lung function data with 4 data points over 3 years, providing data for annual decline, was only available for a sub-set of subjects, so the number for this analysis is again smaller.

Subjects with emphysema had worse lung function and greater smoke exposure compared to those without emphysema (all $p<0.001$). There was no difference in smoke exposure, lung function or UZVI between those with and without bronchiectasis, though affected subjects were older ($p<0.001$). Bronchiectatics had more severe lower zone emphysema, ($p=0.026$), than those without bronchiectasis. Subjects with chronic bronchitis had lower FEV1 ($p=0.013$) but no other demographic or clinical feature differed from those without bronchitis. There were no gender differences between those with and without each of the qualitative phenotypes.

Seventy two sibling pairs were studied as a replication dataset; samples suitable for DNA extraction were only available on 52 pairs, thus comparison of genotypes was only possible in this smaller group. The characteristics of the siblings are shown in Table 3.2. Significant differences were observed between index and non-index siblings in that index siblings tended to have worse lung function and a higher incidence and severity of emphysema. There was no difference in the incidence of chronic bronchitis between siblings.

	Whole group n=424	Emphysema		Bronchiectasis		Chronic bronchitis	
		Yes n=279	No n=70	Yes n=83	No n=266	Yes n=159	No n=265
% of subjects of male gender	59.91	64.16	54.29	57.83	63.67	64.78	57.03
Age in years	50.01 (0.52)	51.76 (0.78)	41.52 (33.77-49.27)	54.18 (1.03)	48.90 (0.63)	50.79 (0.77)	49.48 (0.69)
Pack years	14.00 (2.35-25.65)	17.50 (7.65-27.35)	0 (0-4.18)	14.63 (4.33-25.93)	13.50 (1.40-24.60)	14.63 (2.55-26.71)	12.75 (0.75-24.75)
FEV1 %predicted	37.15 (18.30-56.04)	32.08 (18.94-45.22)	102.88 (80.93-124.83)	36.64 (14.59-58.69)	35.54 (17.39-53.69)	32.76 (15.49-50.03)	38.17 (14.18-62.16)
FEV1/FVC	38.40 (25.95-50.85)	35.00 (26.00-44.00)	73.60 (63.10-84.10)	40.10 (24.35-55.85)	37.50 (25.00-48.00)	37.85 (26.85-48.85)	39.40 (26.40-52.40)
KCO %predicted	69.82 (1.18)	63.18 (1.38)	98.66 (2.55)	67.85 (2.75)	70.57 (1.54)	67.62 (2.31)	71.34 (1.89)
UZVI	31.09 (1.05)	32.71 (1.02)	2.31 (0.39)	33.45 (1.89)	30.63 (1.27)	32.98 (1.69)	29.88 (1.36)
LZVI	44.59 (1.21)	46.78 (1.14)	5.85 (1.01)	54.15 (39.50-68.80)	47.95 (32.15-63.75)	48.85 (34.60-63.10)	49.50 (31.90-67.10)
FEF25-75 (l)	0.38 (0.18-0.58)	-	-	-	-	-	-

Table 3.1: Clinical features of the AATD case-control dataset

Data is presented as mean (standard error of the mean) or median (interquartile range) dependent on its distribution.

FEV1 = forced expiratory volume in one second, FVC = forced vital capacity, KCO = gas transfer corrected for alveolar volume, FEF25-75 = forced expiratory flow between 25 and 75% of total, UZVI = upper zone voxel index, LZVI = lower zone voxel index.

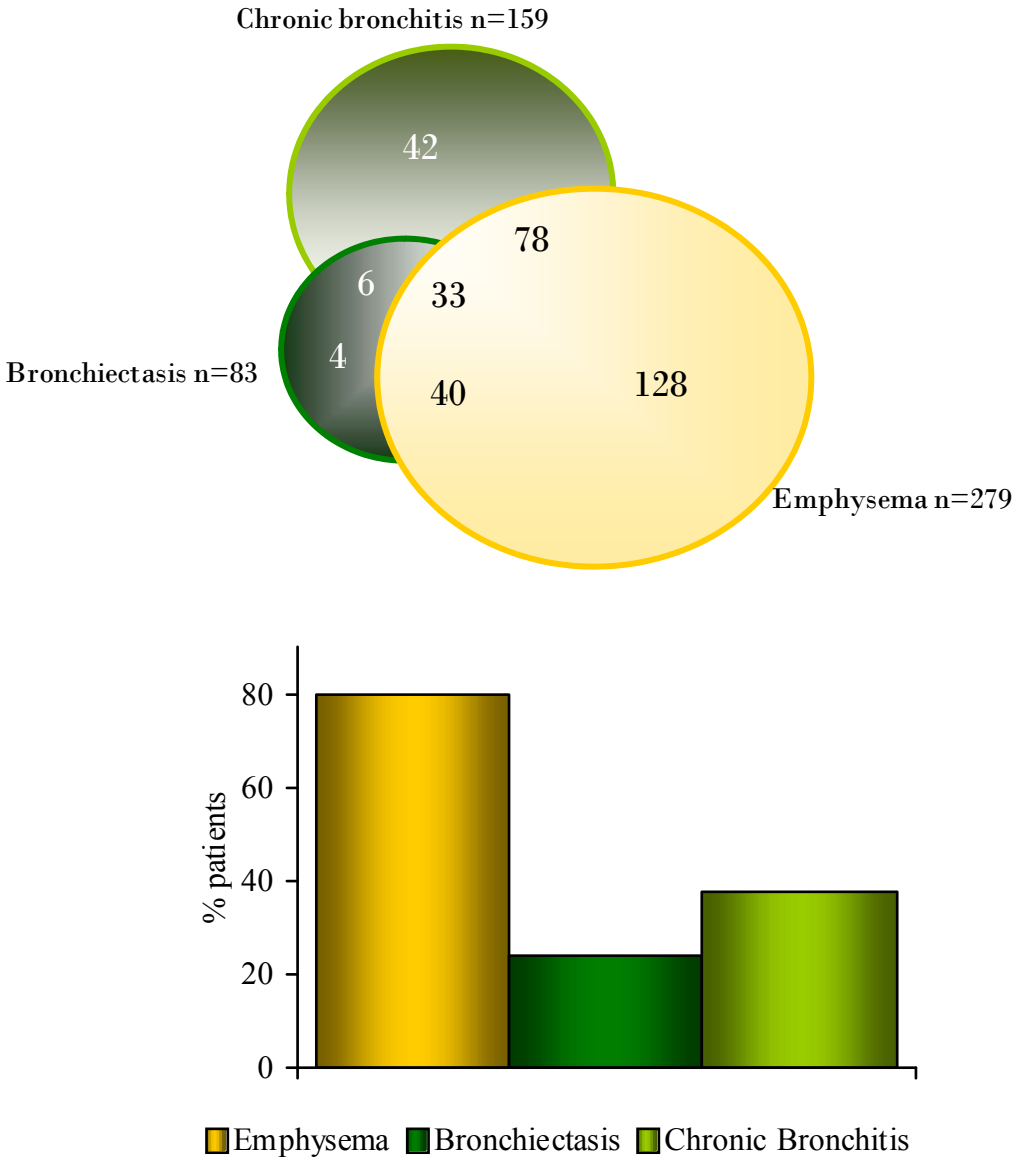


Figure 3.3: Phenotypic overlap and frequency of qualitative phenotypes and in the AATD case-control dataset

The Venn diagram shows the number of subjects with each of the three qualitative phenotypes studied, and the degree of overlap between them. The bar chart shows the proportion of each phenotype within the group (%).

	Index	Non-index	p value
Gender (% male)	59.68	53.18	NS
Age in years	47.91 (1.20)	46.00(1.22)	NS
Smoking in pack years	12.50 (1.5-23.00)	8.50 (0-18.50)	NS
FEV1 %pred	31.00 (21.00-55.00)	68.00 (46.00-107.00)	<0.001
FEV1/FVC %	37.00 (30.0-47.0)	59.00 (43.00-75.00)	<0.001
KCO %pred	69.10 (3.21)	82.70 (2.51)	0.023
Chronic bronchitis (%)	46.20	42.60	NS
Emphysema (%)	82.50	55.70	0.001
UZVI % voxels <-910HU	35.35 (23.32-47.39)	16.02 (5.32-26.72)	<0.001
LZVI % voxels <-910HU	53.90 (40.45-67.35)	27.92 (12.80-43.05)	<0.001

Table 3.2: Characteristics of the familial replication dataset

Data is presented as mean (standard error of the mean) or median (interquartile range) dependent on its distribution.

FEV1 = forced expiratory volume in one second, FVC = forced vital capacity, KCO = gas transfer corrected for alveolar volume, UZVI = upper zone voxel index, LZVI = lower zone voxel index, NS= non-significant

TNFA genotyping

Tag SNPs to cover *TNFA* were chosen using the algorithm described in section 2.3.2 and published data pertaining to *TNFA* variation in UK Caucasian subjects[218]. This resource was chosen as coverage of *TNFA* was relatively poor in HapMap at the time of commencement of the studies herein, and the population in this resource was the most closely geographically matched group for the subjects studied in this thesis. Population matching is important as *TNFA* lies within the HLA region, which is known to show geographical variation[217]. The number of SNPs captured is shown in Table 2.4 and their position relative to *TNFA* structure is shown in Figure 3.1.

All tags were genotyped using TaqMan® genotyping technologies. This uses primers designed and supplied by Applied Biosystems (ABI), based on the sequence surrounding the SNP of interest. The exact primer sequence, therefore, remains the property of the company, hence only context sequences used for design can be given in this thesis. After primers have been designed and optimised by ABI for a particular SNP it will be listed on their website (using the reference SNP number) to indicate that a pre-designed genotyping assay is available. Assays for other SNPs can be designed upon request, using sequence data supplied by the customer. For the tags chosen within *TNFA,* assays had already been designed by ABI (www.appliedbiosystems.com), so no further sequence data was necessary. The context sequence associated with the primers for each tag is shown in Table 3.3.

The PCR reactions were set up on 384 well plates (Greiner Bio-one, UK), using 4 blanks as negative controls per plate. For each plate 1050µl of genotyping master mix (supplied

ready mixed, containing dNTPs, $MgCl_2$ and *Taq* polymerase and optimised for use with TaqMan® genotyping assays by Applied Biosystems, UK), 70µl of genotyping assay at a 20x concentration (ABI, UK) and 35µl of MillQ water were mixed in an opaque eppendorf (Alpha labs, UK), since the assay is light sensitive. For assays supplied at a 40x concentration the amount of assay was halved and the amount of water correspondingly increased, such that the total volume was unchanged.

A volume of 2.25µl of DNA at 20ng/µl was pipetted into each well of a 384 well plate, leaving 4 wells free to act as blanks, to which the same volume of MillQ water was added in order to standardise the reaction volume. The master mix (2.75µl) was then added to each well of the plate, the plate sealed with an optically clear seal (Alpha Labs, UK), and covered with foil in order to protect the light sensitve genotyping assay. The plate was then centrifuged at 1000rpm for 1 minute to ensure that all reagents were at the bottom of the plate and once completed the plates were placed in a thermal cycler (MJ Research (PTC-225) UK) and PCR performed under the following conditions:

50°C	2 minutes	
95°C	10 minutes	
92°C	15 seconds	} 40 cycles
60°C	1 minute	

After completion of the PCR the plates were again covered with foil to prevent degradation of the light sensitive assay and stored at 4°C until they could be read. This step was carried out using an ABI-7900HT (ABI, UK), which is able to read the fluorescent signals emitted by the VIC and FAM markers within each assay. For each tag ABI supplied information indicating which allele the VIC and FAM markers detect, such

that this can be specified prior to reading the plate. The instrument collects fluorescence data on the samples which are analysed using SDS software (version 2.2.2, ABI, UK). The software presents the results as a scatterplot. There are four potential clusters of points, which correspond to the two homozygous genotypes, heterozygotes or no amplification. Parameters may be set to indicate cut-off values for allocation of genotypes within the software, and was set at 95% for the studies contained herein (as recommended by the manufacturers). Each plot was reviewed for the position and relative signal quality of samples where a genotype could not be allocated by the software, and a genotype assigned manually if possible. If the signal quality was <80% or the position was such that the genotype was unclear such samples were reanalysed.

SNP	Context sequence	Design strand
rs1800269	GAGGCAATAGGTTTTGAGGGGCATG[A/G]GGACGGGGTTCAGCCTCCAGGGTCC	Reverse
rs361525	GGCCCAGAAGACCCCCCTCGGAATC[A/G]GAGCAGGGAGGATGGGGAGTGTGAG	Reverse
rs1799964	GGAAGCAAAGGAGAAGCTGAGAAGA[C/T]GAAGGAAAAGTCAGGGTCTGGAGGG	Reverse
rs3093662	TTGAATGCCTGGAAGGTGAATACAC[A/G]GATGAATGGAGAGAGAAAACCAGAC	Reverse

Table 3.3: *TNFA* sequences for primer design

The 2 alleles of each SNP are highlighted in blue,within square brackets, at the point where it appears within the primer sequence.

Statistical analysis

Each tag SNP genotype was assessed for its contribution to COPD phenotypes by regression analysis. In the case-control dataset phenotypic associations of each SNP genotype were sought using the qualitative phenotypes of emphysema, bronchiectasis and chronic bronchitis, and quantitative analyses using FEV1, KCO and HRCT densitometry. Associations seen were then sought in the familial replication dataset.

3.2.3 Results

Case control dataset

Genotyping was successful, on average, in 95.3% (n=404) of cases. An example of the genotyping plots is shown in Figure 3.4 for rs1799964. All SNPs were in Hardy Weinberg equilibrium (all $p>0.05$ for deviation from HWE). Allele and genotype frequencies are shown in Table 3.4. The dataset had 80% power to detect an OR of between 1.62 and 2.26 of developing a given disease phenotype (Table 3.5), this being calculated retrospectively based on the genotype frequencies obtained. Lower power was exhibited where the minor allele frequency was low.

In the regression analyses the influence of age, gender and smoking on each phenotype was sought prior to addition of genetic data. In the logisitic regression models for qualitative phenotypes, age was a significant predictor of both emphysema and bronchiectasis (both $p<0.0001$), whilst pack years smoked was associated with emphysema ($p<0.0001$). No demographic or clinical features predicted the development of

chronic bronchitis. In the linear regressions both FEV1 and FEV1/FVC required log transformation due to their non-normal distribution. Age, gender and smoking were significant predictors of logFEV1, logFEV1/FVC, KCO, UZVI and LZVI (all $p<0.05$). Gender was the only significant predictor of FEV1 decline ($p=0.03$). This aspect of the regressions is summarised in Table 3.6.

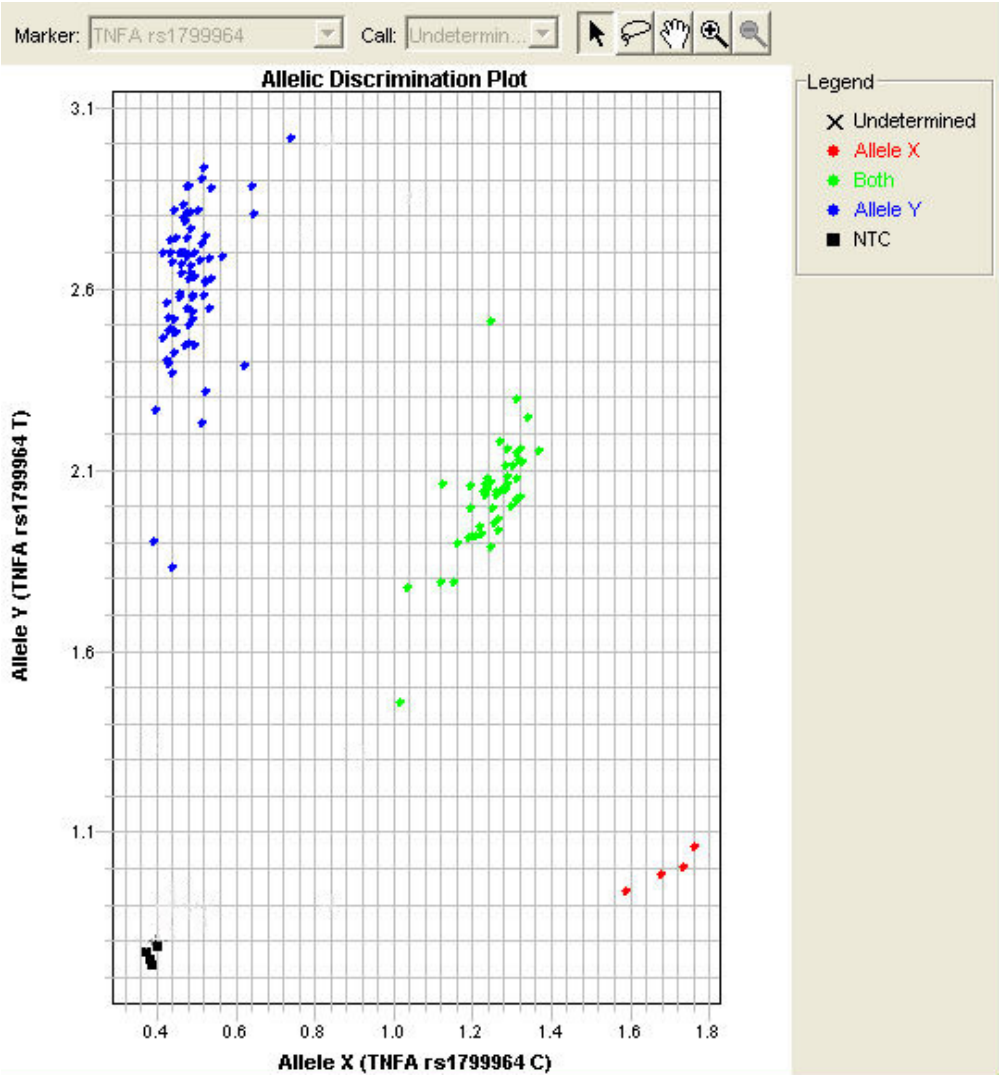


Figure 3.4: TaqMan® genotyping plot for *TNFA* rs1799964

CC homozygotes give the strongest signal along the X axis, whilst TT homozygotes have a strong signal from the fluorophore allocated to Y axis. Heterozygotes emit signal from both markers, thus lie between the other groups. Three clear groups are visualised for this SNP. A cut off of 95% was chosen on the software, with subsequent manual allocations based on the appearance of the plot. The plot was generated using SDS version 2.2.2 (ABI, UK).

SNP	All AATD	Chronic bronchitis	Emphysema	Bronchiectasis
rs1800629 [G/A]				
A allele	21.91	24.84	21.85	26.58
GG	61.63 (249)	56.60 (90)	62.22 (168)	54.43 (43)
GA	32.92 (133)	37.75 (60)	31.85 (86)	37.97 (30)
AA	5.45 (22)	5.66 (9)	5.96 (16)	7.59 (6)
rs361525 [G/A]				
A allele	8.54	11.64	8.52	5.13
GG	83.67 (338)	77.99 (124)	84.07 (227)	91.03 (71)
GA	15.56 (63)	20.75 (33)	14.81 (40)	7.69 (6)
AA	0.74 (3)	1.31 (2)	1.11 (3)	1.28 (1)
rs1799964 [T/C]				
C allele	25.62	26.89	25.19	26.58
TT	57.32 (231)	55.34 (88)	57.78 (156)	55.70 (44)
TC	34.49 (139)	35.85 (57)	34.07 (92)	35.44 (28)
CC	8.19 (33)	8.80 (14)	8.15 (22)	8.86 (7)
rs3093662 [A/G]				
G allele	8.81	11.39	7.99	4.49
AA	82.63 (333)	77.36 (123)	84.39 (227)	91.02 (71)
AG	17.12 (69)	22.64 (36)	15.24 (41)	8.97 (7)
GG	0.25 (1)	-	0.37 (1)	-

Table 3.4: Allele and genotype frequencies of *TNFA* tag SNPs in the case control dataset
SNP alleles are listed [major, minor] and the minor allele frequency given as a percentage, together with all genotype frequencies, where the raw data is also shown in parentheses. Statistically significant differences contributing to a COPD phenotype were ascertained by regression and are detailed in the text (section 3.2.3).

SNP	Phenotype	Minimum odds ratio detectable with 80% power
rs1800629	Emphysema	1.83
	Bronchiectasis	1.82
	Chronic bronchitis	1.62
rs361525	Emphysema	2.26
	Bronchiectasis	2.18
	Chronic bronchitis	2.05
rs1799974	Emphysema	1.79
	Bronchiectasis	1.75
	Chronic bronchitis	1.58
rs3093662	Emphysema	2.16
	Bronchiectasis	2.17
	Chronic bronchitis	2.00

Table 3.5: Power of the case-control dataset to detect differences in phenotypes for *TNFA* tag SNPs

The table shows the odds ratio of developing each phenotype, prior to adjustment for covariates, calculated using the formulae detailed in Appendix 4.

Regression type	Phenotype	Predictor	B	SE (B)	r2 change	p value
Logistic	Emphysema	Age	0.094	0.016	NA	<0.0001
		Gender	-	-	-	NS
		Pack years	0.095	0.017	NA	<0.0001
	Bronchiectasis	Age	0.05	0.013	NA	<0.0001
		Gender	-	-	-	NS
		Pack years	-	-	-	NS
	Chronic bronchitis	Age	-	-	-	NS
		Gender	-	-	-	NS
		Pack years	-	-	-	NS
Linear	logFEV1	Age	-0.003	0.001	0.015	0.006
		Gender	-0.086	0.025	0.024	0.001
		Pack years	-0.007	0.001	0.153	<0.0001
	logFEV1/FVC	Age	-0.004	0.001	0.066	<0.0001
		Gender	-0.082	0.016	0.077	<0.0001
		Pack years	-0.004	0.001	0.074	<0.0001
	KCO	Age	-0.625	0.103	0.077	<0.0001
		Gender	-2.602	2.294	0.012	0.029
		Pack years	-0.387	0.078	0.046	<0.0001
	UZVI	Age	0.514	0.092	0.090	<0.0001
		Gender	8.896	1.961	0.082	0.016
		Pack years	0.192	0.065	0.019	<0.0001
	LZVI	Age	0.813	0.099	0.170	<0.0001
		Gender	5.968	2.11	0.043	<0.0001
		Pack years	0.445	0.070	0.084	<0.0001
	FEV1 decline	Age	-	-	-	NS
		Gender	-0.702	0.336	0.032	0.030
		Pack years	-	-	-	NS

Table 3.6: Clinical predictors of phenotypes used in regression models

The table shows the regression models constructed and the predictors assessed. B is the regression coefficient and represents the change in the outcome variable if the predictor is increased by 1. In all logistic regressions the phenotype was coded 1 if present and 0 if absent, thus a positive B value indicates an increased risk of developing the phenotype. Gender was coded male=1 and female=0 in all regressions, thus a positive value indicates an increased risk of developing a phenotype (logistic models) or an increase in the quantitative outcome variable (linear models). Where the significance of the F change was >0.05 this is shown as non-significant (NS), and no regression coefficient is given as a result. The level of r^2 change in the linear regressions indicates the amount of variability of the measure explicable by variation in the predictor, such that if r^2 change =0.084 then the amount of variability explicable is 8.4%. FEV1 decline was calculated using for each individual using raw data, but a log value of this figure was used for the regressions, as the distribution of decline within the group was non-normal. NS= not applicable.

Genetic data was added to the regression models by forced entry subsequent to adjustment for clinical covariates; in each case the genotype was coded 0, 1 or 2 and analysed as a categorical variable, such that each genotype was treated independently. This is effectively a general model. A second model was constructed to assess allelic effects, with the number of minor alleles present being coded 0, 1 or 2 in a quantitative manner. This is equivalent to a co-dominant model.

A significant difference in allele and genotype frequency was seen between subjects with and without chronic bronchitis for rs361525 (both $p=0.01$), with the A allele conferring an OR of 1.99 (95% confidence interval (CI) 1.18-3.67) in the general model regression analysis. No other tag SNPs predicted this phenotype.

There were no tag SNP associations with emphysema, bronchiectasis, lung function or emphysema severity (as measured by UZVI and LZVI), all models showing $p>0.05$ for the step where genetic data was added.

When haplotype analysis was performed, no haplotypes were observed for analysis; the LD plot is shown in Figure 3.5.

Familial dataset

The hypothesis that rs361525 is associated with chronic bronchitis could not be tested in the replication dataset because this phenotype did not differ significantly between family members (Table 3.2).

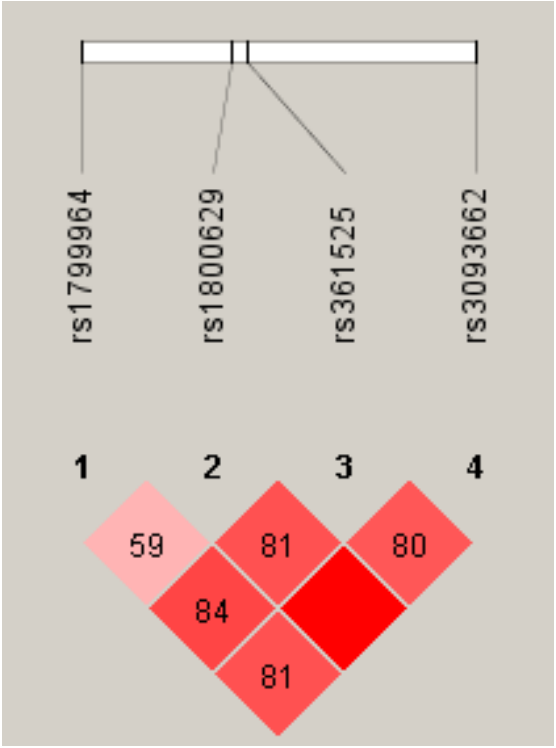


Figure 3.5: Linkage disequilibrium plot for *TNFA* tag SNPs

The linkage disequilibrium plot shows the r^2 values between pairs of SNPs, identified by following the lines formed by the blocks within the plot, such that the value between rs1799964 and rs1800629 is 59, and that between rs1799964 and rs361525 is 84. The software displays the r^2 value as a whole number by default, such that these figures are actually 0.59 and 0.84 respectively. Higher values are indicative of greater LD, since the SNP alleles are correlating more closely. The Haploview software highlights any pairs or groups of SNPs whose r^2 values indicate significant LD ($p<0.05$). No haplotype blocks were observed in Haploview for further analysis, as shown above.

3.3 TNFα

3.3.1 Introduction

TNFα in COPD

Chronic obstructive pulmonary disease is associated with an abnormal inflammatory response within the lungs and the circulation (section 1.1.3). Tumour necrosis factor-α (TNFα) is an inflammatory cytokine which is elevated in the sputum[230], bronchial biopsies[231] and circulation[232] of COPD subjects. It is central to the inflammatory cascade in COPD (Figure 1.5), and probably also influences co-morbidities such as vascular disease and osteoporosis, as shown in Figures 3.6 and 3.7. A murine model also confirms a role for TNFα in COPD pathophysiology, with TNFα receptor (*TNFAR*) knock out mice failing to develop the inflammatory response in the lung typical of COPD after either short or long term cigarette smoke exposure[49]. This was the rationale for testing of anti-TNFα therapy in COPD[23]. However, a clinical trial of a drug with this action (infliximab), already in use in other inflammatory conditions, did not show any significant effect on symptoms or lung function when given regularly intravenously over two years[23]. A subgroup of cachectic subjects, however, showed an improvement in exercise capacity (as measured by the six minute walk test) after infliximab treatment[23], perhaps lending support to the theory that phenotypic subgroups within COPD respond differently to treatment, and should be considered as different disease entitities.

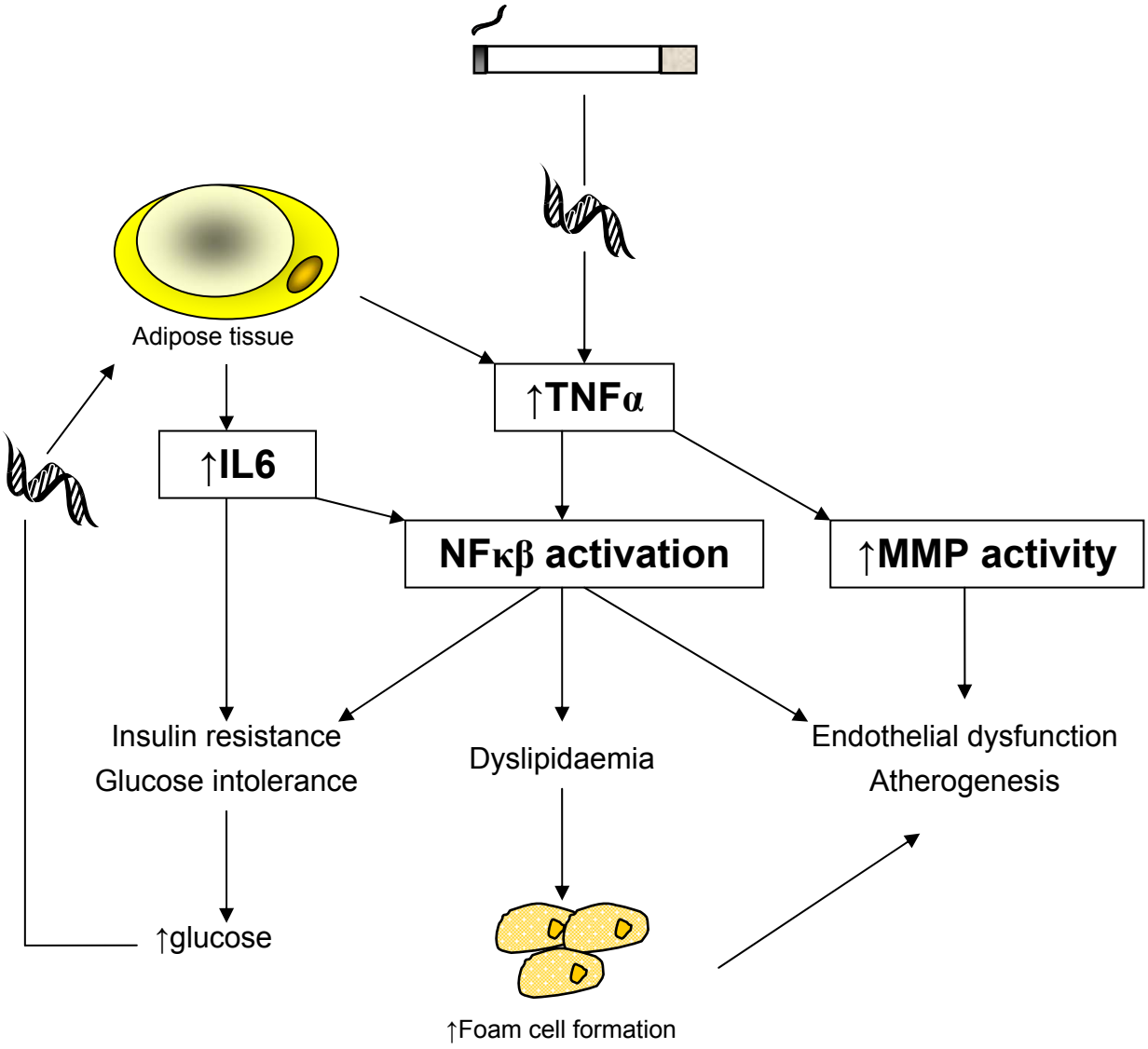


Figure 3.6: The TNFα mediated inflammatory cascade
Cigarette smoking is associated with greater gene expression of TNFα [200] and subsequent systemic inflammation. Increased adiposity also contributes to TNFα and interleukin-6 (IL6) production, both of which activate nuclear factor kappa beta (NFκβ), leading to insulin resistance, dyslipidaemia and atherogenesis. The latter is enhanced by the activity of MMPs, which is increased by TNFα. Raised glucose levels also affect gene expression, predominantly in adipose tissue, elevating inflammatory cytokine levels further.

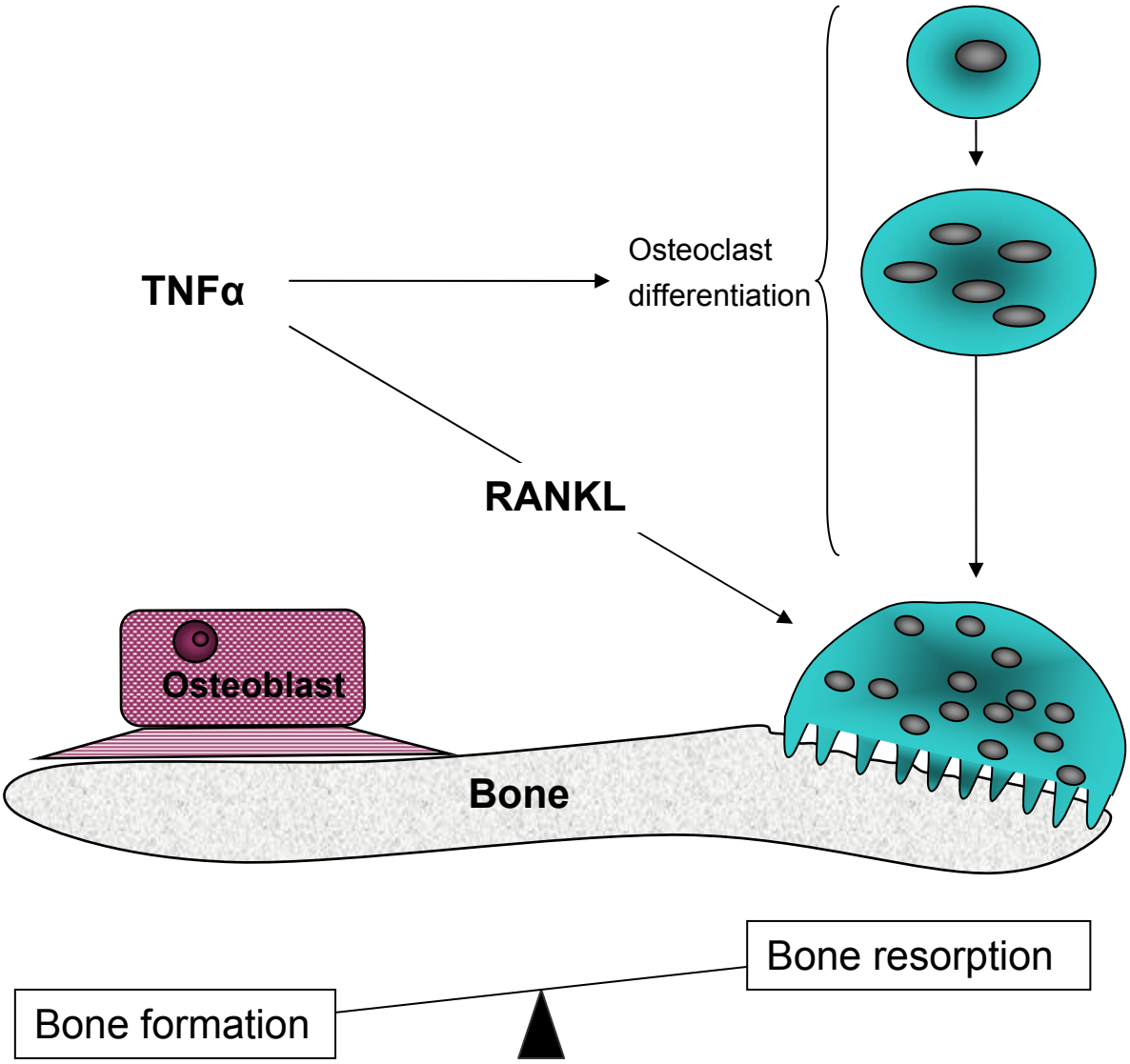


Figure 3.7: The role of TNFα in osteoporosis
TNFα both upregulates receptor activator for the NFκβ ligand (RANKL) and enhances osteoclast differentiation. This promotes bone resorption by osteoclasts, causing an imbalance between this and bone formation by osteoblasts, ultimately resulting in osteoporosis.

TNFα in AATD

Alpha one antitrypsin (AAT) and TNFα show a degree of interaction[202], such that TNFα release is suppressed by AAT *in vitro*[202] and in animal models[233]. This is thought to occur because AAT inhibits thrombin and plasmin activity, which prevents proteinase activated receptor-1 activation, a process which stimulates TNFα release[202] (Figure 3.8). In AATD this suppression would be reduced thus any TNFα effects are likely to be more pronounced, perhaps contributing to the observation that inflammation is more marked in AATD than usual COPD[234]. The relationship of TNFα levels to clinical phenotype has not been reported in AATD.

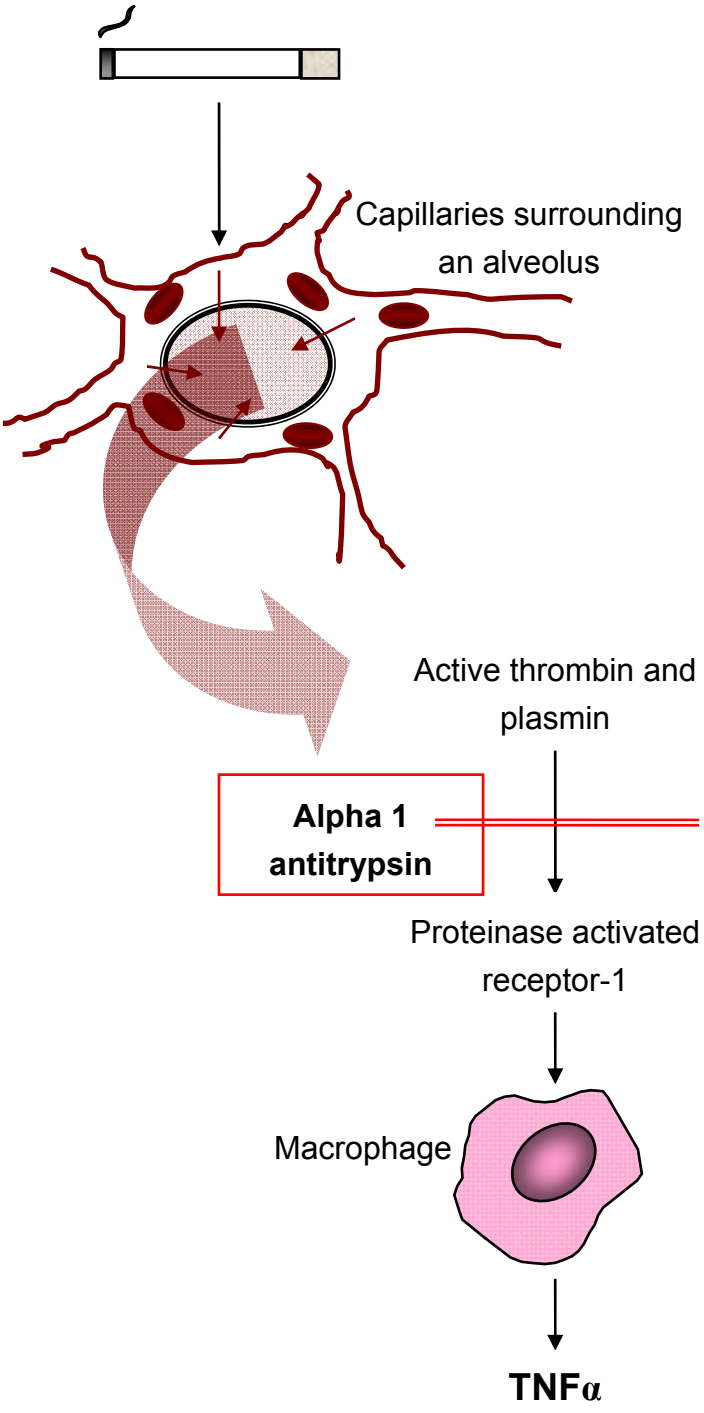


Figure 3.8: The mechanism by which AAT suppresses TNFα release in the lung
Cigarette smoke increases the permeability of the alveolus, such that protein leakage occurs. In the lung thrombin and plasmin cause release of TNFα from macrophages, acting via proteinase activated receptor-1, a process which is blocked by AAT.

3.3.2 Methods

Subjects

The subjects on whom circulating and airway levels of TNFα were measured comprised all those in the case-control dataset described in section 3.2.2 with appropriate stored samples available. Plasma concentrations were measured on all PiZZ subjects for whom a sample was available (n=237), though this was less than the number genotyped, as some genotyped subjects did not have a stored plasma sample (n=97). Sputum concentrations were measured in patients with chronic bronchitis (the only group who can spontaneously expectorate sputum when stable; even subjects without lung disease may do so in the presence of clinical features of infection). The number of suitable samples was, therefore, fewer than plasma (n=60). Many factors influence airway inflammation in COPD including bacterial colonisation[235] and cachexia[232] so any study of airway inflammatory markers requires matching of subjects for these features. In this study a subset of subjects were matched for age, gender, FEV1, body mass index (BMI) and bacterial colonisation status (section 2.2.6).

Quantitation of TNFα

TNFα concentrations were measured using a pre-validated high sensitivity (HS) commercially available ELISA kit (R&D systems, UK). Circulating TNFα concentrations were measured using stored plasma samples, and airway concentrations using stored sputum sol phase samples. Blood samples were collected from stable subjects in EDTA tubes (Greiner Bio-one, UK) and centrifuged at 3000rpm at 4°C for 15 minutes prior to

storage of plasma in aliquots of 200µl at -80°C. Sputum samples, again from stable subjects, had been centrifuged at 2000rpm at 4°C for 90 minutes and the resultant sol phase stored in aliquots of 50-200µl at -80°C.

All solutions required for the ELISA were provided in the kit. The plates provided are pre-coated by the manufacturer with an antibody specific to TNFα. A standard curve was set up by serial dilution of the standard provided (concentration 32pg/ml) in the calibrator diluent to 16, 8, 4, 2, 1 and 0.5pg/ml. Standards and samples were run in duplicate on 96 well plates. Fifty µl of assay diluent and 200µl of plasma or standard was added to each well, and the plate incubated at room temperature (~18°C) for 3 hours. It was then washed 6 times using 400µl wash buffer per well for each wash. Two hundred µl of TNFα HS conjugate was then added to each well, followed by incubation for 2 hours and washing as before. Fifty µl of substrate solution was added next, followed by an hour of incubation. Amplifier solution (50µl) was then added prior to a final 30 minute incubation period, again at room temperature. Before reading the plate, 50µl of stop solution was applied to each well. Colour readings were taken at 490nm, with wavelength correction set to 650nm, using a microplate reader (Biotek plate reader with Gen5 software, both Northstar Scientific, UK) and compared to the standard curve.

Sputum sol phase samples had higher TNFα concentrations, and hence required dilution to obtain an accurate reading within the range of the standard curve. Samples were diluted using the calibrator diluent from the kit to 1:50, and a volume of 200µl of this dilution added to each well of the plate, in a similar manner to the plasma. If the result fell outside the range of the standard curve at this dilution the measurement was repeated at lower or higher concentrations as needed.

Statistical analysis

Plasma and sputum concentrations of TNFα were compared with demographic and clinical features. Phenotypic associations of circulating TNFα levels were sought in order to ascertain any associations thats presence should be corrected for in subsequent regression models involving genetic data.

3.3.3 Results

Clinical associations of circulating TNFα

Plasma TNFα levels were obtained on 230 subjects out of the total case-control cohort of 424. There was no significant difference in plasma level between male and female subjects, nor did it vary with smoke exposure. Older subjects tended to have higher TNFα levels, as shown by the correlation in Figure 3.9 ($p=0.029$, $r=0.145$). Plasma TNFα showed no significant correlations with FEV1, KCO or lung function decline, nor were there any significant differences between those with and without emphysema, chronic bronchitis or bronchiectasis (Table 3.7).

Clinical associations of airway TNFα

Sputum TNFα levels were measured on 58 subjects, showing a mean of 31.14pM with a standard error of 9.84, reflecting the wide range (0.06-390.00pM). No samples were available from subjects of the AA genotype for analysis.

Three subjects whose values lay more than 3 standard deviations from the mean were

excluded from further analysis because examination of their clinical records showed that samples had been obtained at the time of an infective exacerbation. Exacerbations are associated with increased airway inflammation[234], such that inclusion of these samples would have biased the results.

There was no significant difference in sputum level between male and female subjects, nor did it vary with age, smoke exposure, FEV1 or KCO (all $p>0.1$). There was no correlation between sputum and plasma TNFα levels ($p=0.836$). Contemporaneous quantitative sputum microbiology was available in 35 subjects; of whom 51% were colonised with recognised airway pathogens. Sputum TNFα tended to be higher in samples colonised with bacteria (mean 43.78pM; SE=23.91pM v 14.99pM; SE=5.21pM), although the difference was not statistically significant ($p=0.310$). There was no correlation with bacterial numbers in colonised subjects ($p=0.991$).

The analysis was repeated in a subset of subjects matched for age, gender and severity of COPD (see also section 1.1.1 and Table 1.1), comprising 11 male subjects, with normal body mass index, mean age of 53.09 years (SE=1.16) and FEV1 23.65%predicted (SE=2.19). One subject, whose concentration was vastly different from the rest of the cohort (594pM), was excluded as the stored sample had been obtained whilst he was participating in a clinical trial. Again, there was a wide range of TNFα concentrations, with a mean of 67.12pM (range 0.20-389.00pM). There was no difference in sputum TNFα between those with and without stable state bacterial colonisation of the airway ($p=0.441$).

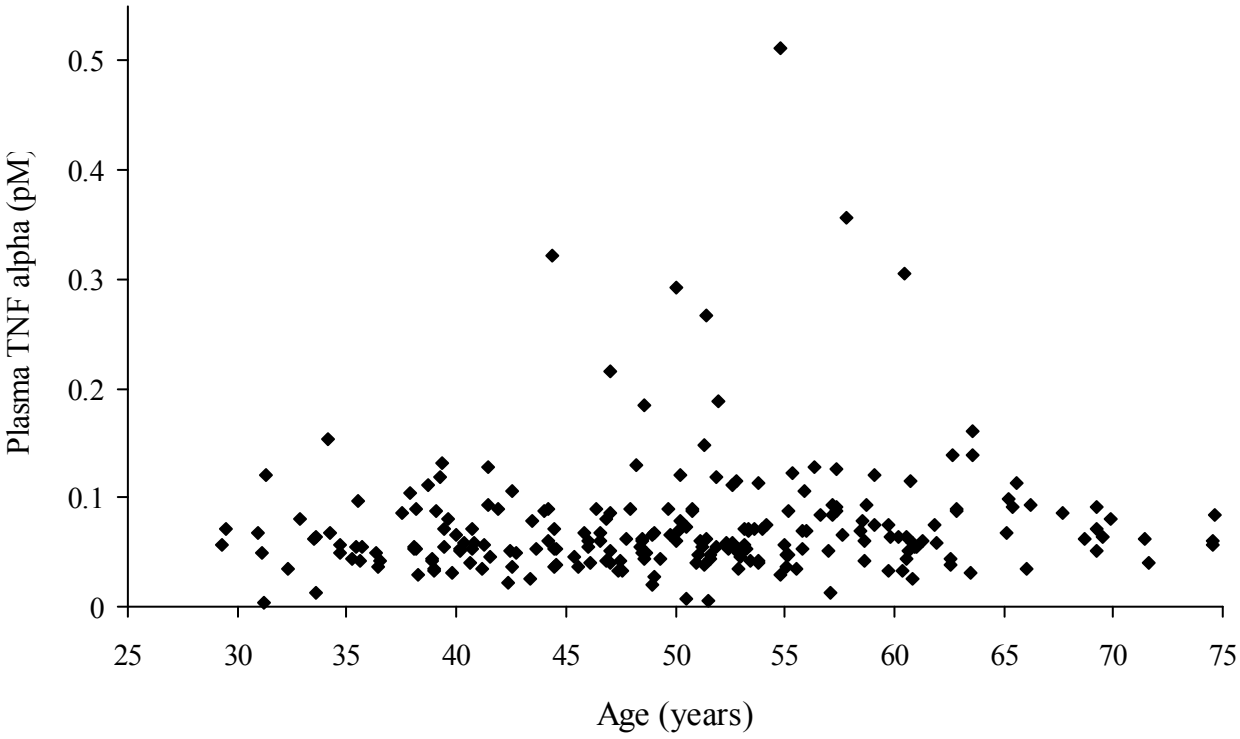


Figure 3.9: Plasma TNFα correlates with age
The graph shows levels of circulating TNFα for each subject relative to their age. There was a weak but positive correlation (r=0.145, p=0.029) indicating a general increase in older subjects. Since the data was non-parametric the correlation line has been omitted from the graph.

	Plasma TNFα (pM)	
All subjects	0.061 (0.051-0.071)	
Phenotype	**Affected**	**Unaffected**
Emphysema	0.079 (0.061-0.098)	0.065 (0.005)
Bronchiectasis	0.065 (0.045-0.085)	0.061 (0.044-0.079)
Chronic bronchitis	0.062 (0.040-0.082)	0.061 (0.040-0.081)

Table 3.7: Plasma TNFα levels in AATD subjects
The table shows mean (standard error of the mean) or median (interquartile range) plasma TNFα dependent on its distribution. The mean circulating level of TNFα was 0.061pM in the cohort; the data is then substratified into groups affected or unaffected by the 3 phenotypes listed in the first column (emphysema, bronchiectasis or chronic bronchitis), all of which are defined in section 1.1.1 and the first two illustrated in figures 1.2 and 1.3. There were no significant differences between subjects with and without these phenotypes (all p>0.05).

3.4 Associations between TNFA and TNFα

3.4.1 Introduction

A functional effect for a SNP may be inferred if an association of genotype and protein level is seen, together with a phenotypic association[236]. The analyses in the preceding sections (*TNFA* and TNFα associations with phenotype in AATD) highlight many factors that need to be considered when assessing whether genotype-phenotype associations are reflective of a causative association mediated by the protein product.

3.4.2 Methods

Differences in circulating and airway levels of TNFα reported in section 3.3.3 were sought for associated genotypes identified in section 3.2.3, controlling for associated clinical features by regression and matching of subjects respectively. Since only rs361525 was associated with any phenotype, this was the only SNP studied here.

3.4.3 Results

Circulating TNFα stratified by genotype

Plasma TNFα levels stratified for rs361525 genotype are shown in Figure 3.10. There was a trend suggesting a higher level in those with at least one A allele, indicating it may have an effect on circulating levels of the protein. This might imply that circulating TNFα levels are important in pathogenesis, since this allele was associated with the chronic

bronchitis phenotype. However, since age was a significant predictor of plasma TNFα (section 3.3.3), a regression model to control for age prior to seeking genetic association with protein levels was constructed. The distribution of plasma TNFα concentrations deviated sufficiently from normality (Kolmogorov-Smirnov statistic 0.203, $p<0.0001$) such that regression models to assess the influence of genotype, corrected for age, could not be constructed prior to data transformation. Using log values the regression model did not show any significant effect of rs361525 genotype on plasma TNFα, either in the whole group, or in those with chronic bronchitis (associated with the SNP in the case control dataset). Although there were 4 cases whose standard residuals lay more than 3 standard deviations from the mean (indicating outliers), none influenced the model unduly (maximum Cooks distance 0.115). The regression results are summarised in Table 3.8.

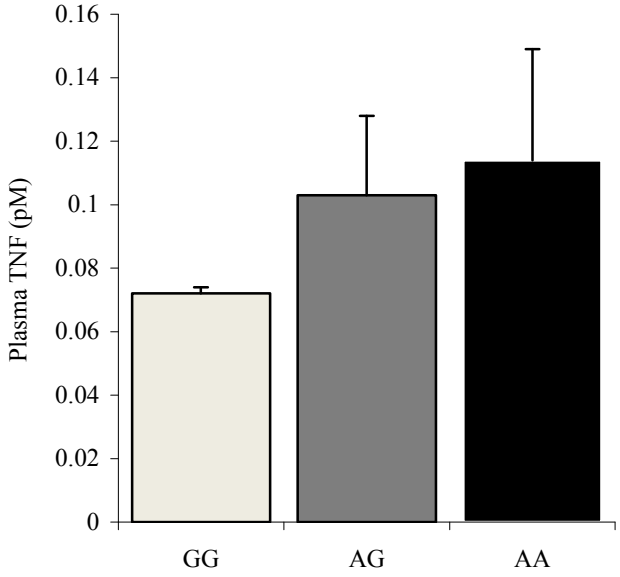


Figure 3.10: Plasma TNFα stratified for rs361525 genotype
The bar chart shows the mean (SE) of circulating TNFα levels for each of the three observed genotypes of rs361525 in AATD subjects. Although levels appear higher in subjects with one or more A alleles the difference was not statistically significant.

Predictor	B	SE (B)	r^2 change	p value
rs361525 A allele	0.056	0.052	0.005	0.285
Age	0.003	0.002	0.013	0.087

Table 3.8: Linear regression results assessing independent predictors of plasma TNFα
The table shows the regression coefficient and change in r^2 value observed with addition of each predictor. Neither achieved statistical significance.

Sputum TNFα stratified by genotype

Neither gender, age, smoke exposure or any other recorded demographic feature were associated with sputum TNFα levels in this study (section3.3.3). Since no feature affected the outcome measure (sputum TNFα level) no adjustment by regression modelling was needed and as such the levels were compared directly between those with and without the associated allele of rs361525. However, since prior studies have shown several features associated with airway inflammation (sections 1.1.3 and 3.3.2) the statistics for association of sputum TNFα levels were carried out in the matched subset described in the methods (section 3.3.2), as well as the larger group.

Sputum TNFα levels were markedly higher in those with the A allele of rs361525, the difference being more marked in the matched subset, as shown in Table 3.9 and Figures 3.11 and 3.12. Although all subjects in the smaller matched group had significant numbers of airway bacteria (i.e. diagnostic of airway bacterial colonisation), four subjects had only those deemed mixed normal flora, rather than potential pathogens, leaving 3 of each genotype matched for pathogen colonisation status. Since this factor might influence airway inflammation it was appropriate to sub-stratify the results for this feature.

	rs361525 genotype		p value
	AG	GG	
All subjects (AG, n=11; GG, n=43)	100.02 (45.76)	18.33 (6.85)	0.028
Matched subjects (AG, n=4;GG, n=6)	165.11 (76.12)	1.80 (0.69)	0.011
Matched subjects with airway pathogens (AG & GG, n=3)	90.14 (18.65)	0.54 (0.30)	0.050

Table 3.9: Sputum TNFα between subjects stratified for rs361525 genotype
The mean concentration (SE) in pM for each group is shown in the table. In all comparisons the levels are significantly higher in subjects with the A allele.

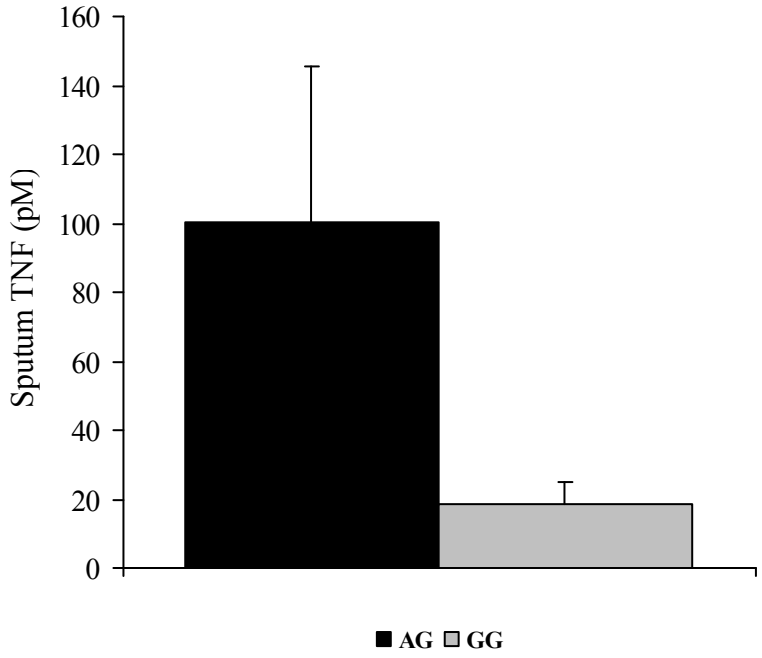


Figure 3.11: Sputum TNFα stratified for rs361525 genotype in AATD subjects
The bar chart shows the mean (SE) sputum concentrations of TNFα in the unmatched AATD cohort (all subjects); the difference is significant (p=0.028).

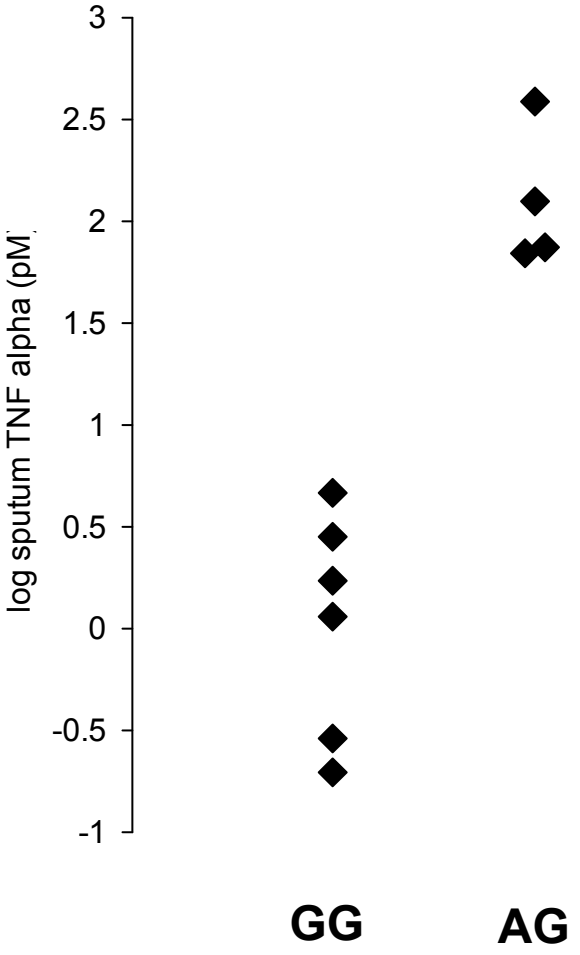


Figure 3.12: Scatterplot of sputum TNFα concentrations stratified for rs361525 genotype in subjects matched for age, gender, BMI and FEV1

Sputum TNFα concentrations are shown as log values for ease of visualisation on the graph, since the raw values differed so widely between the genotypes. The difference is statistically significant (p=0.01).

3.5 Discussion

The results presented in this chapter suggest an association between the A allele of rs361525, a SNP in the promoter region of *TNFA*, and chronic bronchitis in AATD. Furthermore, a functional effect of the SNP is suggested by the observation of higher airway levels of TNFα in subjects with the A allele. The presence of an association of genotype with protein levels, localised to the airway, and an airway related phenotype strongly suggests that this polymorphism has a role in the pathophysiology of chronic bronchitis.

Chronic bronchitis is a feature of approximately 40% of the AATD subjects studied here, and was not predicted by any demographic feature in the regression models (unlike other COPD phenotypes) suggesting that genetic or other unidentified environmental influences might be important in its pathogenesis. Its' importance as a phenotypic outcome measure has been recognised by other COPD genetic association studies[99, 107], perhaps for this reason. Clinically it is an important phenotype to ascertain, as it is associated with more frequent exacerbations [237], which in turn are associated with a tendency to disease progression [238, 239]. In some patients, difficulty in expectorating may also occur, leading to increased cough. Mucolytic therapy may be more beneficial in the presence of a productive cough[8], facilitating airway clearance and reducing exacerbation frequency [240]. Association of chronic bronchitis with both *TNFA* genotype and airway TNFα levels suggests that treatment strategies aimed at blocking TNFα effects might be of benefit in such patients. Anti-TNFα therapy is available, and has undergone clinical trials in COPD[23], but was not assessed specifically in subjects

with chronic bronchitis despite the recognised importance of TNFα in the pathophysiology [231]. Furthermore, response to anti-TNFα treatment may be dependent on *TNFA* genotype [26], perhaps indicating a need for further clinical trials directed to clinical sub-phenotype or genotype.

Although there is a degree of overlap of COPD phenotypes, these results are specific to chronic bronchitis, concurring with previous work on TNFα in airways disease[231]. TNFα is known to be elevated in the sputum[230] and bronchial biopsies[231] of subjects with COPD, and is generally accepted to be important in the pathogenesis of both pulmonary and systemic features of COPD[241]. Such an association is lacking for presence or severity of parenchymal disease, defined by emphysema on HRCT. This suggests different genetic factors influence the development of each clinical sub-phenotype and supports the view that inadequate patient characterisation may underlie poor replication of previous COPD genetic association studies. It is also of interest that the influence of age and smoke exposure differed between COPD phenotypes in the regression models, further emphasising that the pathophysiology of individual phenotypes has a different basis.

Despite the genetic association observed, there was no difference in circulating TNFα between sub-phenotypes. It was not possible to test differences in airway TNFα between sub-phenotypes, as only those with chronic bronchitis produce sputum for testing. This is an inherent problem in studies of airways disease; whilst sputum induction may allow collection of airway secretions from healthy individuals factors relating to the process of sputum induction could confound results[242].

Genetic association with another promoter region SNP (rs1800629, G-308A) in *TNFA*

is accepted as important in COPD, following a supportive meta-analysis[228]. However, previous studies included in the meta-analysis have defined their cases using FEV1 and FEV1/FVC [72, 105, 224], CT diagnosed and pulmonary function diagnosed emphysema[106, 225] and chronic bronchitis[107]. It is, therefore, debatable whether such studies can be merged in a single analysis of their outcome, as they may include different, albeit related, disease phenotypes. Furthermore many of the supportive publications included studied less than 100 subjects with a COPD phenotype [106, 107, 224, 225], and as such are small for determining genetic association work and may produce false positive results. It will be of importance to study this SNP, or at least the region in which it is located, in the large genome wide association studies in COPD underway at present, with particular reference to the bronchitis phenotype.

The rs361525 (G-238A) SNP has been less comprehensively studied in COPD. One previous study has tagged *TNFA*[72] but failed to find an association. However, subjects included a family cohort with airflow obstruction, and cases with emphysema; no previous studies in COPD have looked for association of chronic bronchitis with rs361525. Another study by the same group in subjects with AATD tagged several candidate genes and found associations with lung function for three SNPs in *TNFA*, including one in LD with rs361525[236]. Interaction of TNFα with AAT may make the relationship exclusive to those with AATD or alternatively the specificity of the phenotypic association might explain the lack of previous association of this SNP in COPD.

The rs361525 SNP is located in the promoter region of *TNFA* and has been studied extensively in many different disease states, although little evidence exists for a functional effect. Cellular studies of *TNFA* promoter region SNPs suggest that the A allele of

rs361525 increases gene transcription[243]. Conversely gene expression in patients with psoriasis (a disease with which the A allele of rs361525 is strongly associated[244]) appears reduced[245]. Further functional work clarifying the role of the rs361525 SNP would be indicated if fine mapping of the region surrounding it confirmed it to be the functional variant associated with chronic bronchitis, rather than merely a polymorphism in LD with another such SNP. Such studies are expensive, and alternative surrogate support for a functional effect may be gained by assessing TNFα levels[236], which was the strategy chosen for the studies in this chapter.

The data did not show an association between rs361525 genotype and plasma TNFα levels. This may be because of the specific association with airways disease since airway and plasma inflammatory markers do not correlate in COPD [52], as confirmed here. It was only possible to obtain airway secretion samples from a relatively small cohort for measurement of airway TNFα, the results show a significant difference between those with and without the associated allele, and no variation with any other clinical feature. Whilst it is appropriate to view the results with caution because of the small sample size for airway TNFα measurements, they support a functional effect of the A allele in airways disease. Further study involving a larger cohort of patients is therefore indicated.

The genetic association results reported here do not include adjustment for multiple statistical testing as such corrections (for instance the Bonferroni) appear overly conservative in genetic association studies[68], and hence are generally inappropriate. This is particularly true when several theories are being tested, as in this study in which the hypothesis was that each SNP could be driving any of three qualitative COPD phenotypes, or affecting disease severity via several quantitative phenotypes. Statistical

significance is usually accepted if it reaches the level needed for a genome wide study (generally considered to be $p \leq 5 \times 10^{-5}$[68]), or by replication in an independent population. The studies reported in this thesis have considered the latter approach, using a small familial dataset to act as a replication group. However, within this dataset there was no difference in the prevalence of chronic bronchitis between family members. As such it was not possible to test the hypothesis of association with this phenotype in this replication cohort. These problems are inherent in studying rare conditions, such as AATD, as recruitment of datasets powered for genome wide significance using healthy AATD controls is currently impossible, because of the high rates of disease phenotypes (Figure 3.3). International collaborations to create larger datasets could help overcome this problem although geographical variation of genes, particularly those within the HLA region such as *TNFA*, may be a further confounding factor [217].

3.6 Conclusions

This study suggests an association between rs361525 (G-238A) and chronic bronchitis in AATD, but requires further independent replication. No functional effect of the SNP was demonstrated in plasma, and plasma TNFα did not differentiate between COPD phenotypes or correlate with disease severity. A functional effect in the airway is suggested by the difference in sputum TNFα levels seen between those carrying the associated allele and those without, but must be viewed with caution because of the small sample size.

Chapter 4: The role of selected genes which encode proteins relating to inflammation in the respiratory phenotype of AATD

4.1 Introduction

Inflammation is more marked in AATD than usual COPD[234], and whilst the interaction between AAT and TNFα[202] (see also section 3.3.1) may contribute to this, polymers resulting from the PiZ mutation are also thought to be pro-inflammatory[246], though the mechanisms underlying this are not yet fully elucidated. This suggests that inflammation will be present in all PiZZ subjects, including those with no clinical evidence of disease and who do not smoke. In this situation an additive effect of different variations in a series of pro-inflammatory proteins, or inhibitors of inflammation, resulting from genetic polymorphism, could be clinically important.

A number of genes relating to inflammation have associations in usual COPD (section 1.1.4); for the studies contained in this chapter 3 genes, each of which have at least 4 prior studies supporting a role in usual COPD, were examined for association with clinical phenotype in the cohort of PiZZ subjects described in chapters 2 and 3. These genes are Surfactant protein B (*SFTPB*), transforming growth factor β (*TGFB*) and Group specific component (*GC*), which is also known as Vitamin D binding protein. These genes, their protein products and prior evidence for a role in COPD will be discussed in sub-sections of this introduction, prior to results on each gene in turn.

Data presented here has been published in a peer reviewed journal (Appendix 7).

4.1.1 Surfactant protein B

The surfactant protein B gene *(SFTPB)* lies on chromosome 2p12 and is composed of 11 exons spread over approximately 10kb of DNA, with an 823 nucleotide 3' untranslated region[247]. The protein it encodes (SFTPB) is derived from extensive proteolytic digestion of a 381 residue precursor within alveolar type II cells, to form a final protein 79 residues in size[248]. As a consequence of this, intronic polymorphisms resulting in changes in protein function, associated with disease, are relatively well recognised and described[249].

Surfactant protein B is thought to inhibit pulmonary inflammation by reducing nitric oxide release from alveolar macrophages[250] and reducing L selectin expression, with a consequent reduction in neutrophil migration to the lungs[251]. The role of SFTPB in COPD pathogenesis has not been widely studied, but low levels are observed in animals with emphysema[252]. Mathematical models of emphysema [253] suggest that surfactant related changes in surface tension would contribute to disease although this has yet to be studied *in vivo*. Pathways relating to this protein may be of even greater interest in AAT deficient subjects with COPD because surfactant proteins interact with AAT to decrease anti-elastase activity[254] which could be critical in AATD as anti-elastase activity is already profoundly reduced.

Variation within *SFTPB* has been positively associated with COPD in Hispanic subjects compared to smokers with normal spirometry, such that the C allele of rs1130866

(Thr131Ile) conferred an OR of 3.65 of disease [133]. This is a high OR of disease for a single SNP, given that COPD is a disease where many genetic factors are thought to influence outcome, and it is of note that the 95% confidence interval for the OR was very wide (1.19-12.12), possibly because of relatively low numbers of cases in the study (n=94)[133]. The same polymorphism was associated with emphysema[72] and a flanking microsatellite with exercise capacity[78] in the NETT cohort. In the first of these studies[72] the SNP was studied in a case-control group, and replication attempted in a familial group. In the case control study it was associated with emphysema in the presence of a gene by smoking interaction term, and replicated in those of the family group with severe disease, but not in those with milder disease. It is of note that the defining characteristics of the cases in the 2 groups differed slightly, though it is likely both had emphysema, data pertaining to this specific phenotype was not reported for the familial group. The second of these studies was performed by the same research group, utilising different phenotypic data from the case-control dataset, and the same familial replication group [78]. It is perhaps not surprising, therefore, that the results were largely concordant with the earlier study. A German study, examining only microsatellite lengths, found a significant association with respiratory failure in COPD (OR=4.9), though no association with COPD per se[132]. Again this was a relatively small study, comprising 118 subjects with COPD, an identical number of matched control subjects and 110 population based control subjects. However only 24 of the subjects with COPD also had respiratory failure, which is reflected in the wide 95% confidence interval reported (1.76-13.60)[132].

The functional SNP in this gene (rs1130866, Thr131Ile) has been studied in AATD more recently, and was not associated with lung function[236]. No phenotypic data other

than lung function was reported, and the mean FEV1 (%predicted) was 65.9%, indicating that most subjects had only mild COPD according to British standards (Table 1.1). Although the studies supporting a role for *SFTPB* in COPD have some weaknesses their results suggest that if the gene is of importance it may be so in defining those with more severe lung disease, reflected by the association seen with this phenotype in the study by Hersh *et al*[72] and with respiratory failure[132]. The phenotypic data available in the UK AATD population, reported in this thesis, is of sufficient detail to ascertain if the lack of association seen in the first study of this gene in AATD[236] is truly reflective of its role in disease or simply reflects the relatively mild disease seen in the population studied.

4.1.2 Transforming growth factor β

The transforming growth factor β gene (*TGFB)* lies on chromosome 19q13, and is composed of 7 exons with 6 very large introns, giving a total gene size of approximately 23kb[255]. The final protein, TGFβ1, consists of 112 amino acids from the C terminal end of the precursor, formed into 2 polypeptide chains linked by disulphide bridges[256]. Mice who are unable to activate latent TGFβ1 develop emphysema, suggesting that disordered activation of TGFβ1 relates to the pathogenesis of COPD[257]. This occurs via alterated expression of *MMP12*, such that increased proteolysis occurs because of increased levels of its protein product (MMP12), which is a matrix metalloprotease (MMP)[257]. The role of MMPs in emphysema pathogenesis is discussed in the introduction and in chapter 5 (sections 1.1.3 and 5.1). TGFβ1 also regulates extra-cellular matrix production, cell growth and differentiation, tissue repair and some immune responses [258], many of which

are thought to be important in COPD pathogenesis. It is capable of conversion of naïve T cells into both regulatory T cells (Treg), and pro-inflammatory Th17 cells, with a shift to the latter being observed in the presence of IL6[259] - a cytokine known to be important in COPD[260]. Retinoic acid, a vitamin A metabolite, regulates this process such that it promotes anti-inflammatory change[259], and provides some of the rationale for clinical trials of drugs targeting these pathways in COPD by reducing the inflammation driving the disease process (section 1.1.3)[261].

Three SNPs in *TGFB* (rs2241712, rs1800469 and rs1982073) were associated with severe COPD in the NETT cohort[108], two (rs2241712 and rs1800469) being linked to subjective measures of dyspnoea (though not objective measures of exercise capacity) in a later study by the same group[78]. Amongst these, functional roles are known for rs1800469 (C-509T), which enhances promoter function, thus increasing levels of TGFβ1[262], whilst rs1982073 (T869C) leads to an amino acid substitution (Leu10Pro) and is associated with higher levels of mRNA and protein[263]. The direction of association in prior work suggests a potential protective role for TGFβ1: a case-control study examining rs1982073 in subjects with COPD, resistant smokers and healthy controls found the T allele less commonly in COPD subjects[135]. This association was replicated in a larger (1390 subjects) European study[134]. Although an animal model (described in paragraph 1 of this section) suggests this association has pathogenetic credibility[257] some clinical studies contradict this hypothesis, showing that TGFβ1 is highly expressed in COPD airways[264, 265], and that this correlates with lung function[265] suggesting that the protein may contributes to pathophysiology.

Two SNPs in the region surrounding *TGFB* have also been associated with COPD -

rs2241718 and rs6957[108]. These SNPs lie in a protein coding region, and it is not yet known whether their association simply reflects proximity to an associated SNP in *TGFB*, whether one of them is the major associated variant in the area (such that those in *TGFB* are only associated because of their proximity) or represent independently associated variants, in addition to those reported in *TGFB*. Both rs2241718 and rs6957 lie in the coiled-coil domain-containing protein 97 gene (*CCDC97*), at the 3' end of *TGFB*. Although it is known that this gene codes for a protein its function and hence any possible role in respiratory pathology is unknown. However, because of the previously published association with COPD [108] it was felt that this gene should be studied alongside *TGFB* for the studies contained in this chapter. The version of HapMap used to choose tag SNPs for this study (build 36)[71] showed that the block containing the promoter of *TGFB* also spans some of the adjacent hypothetical protein coding gene *MGC4093*, and as such this gene has been examined in the same way as *TGFB*, in the studies reported in this chapter in order to aid ascertainment of the true associated variant(s) in the region.

4.1.3 Group specific component

The Group-specific component gene (*GC)*, lies on chromosome 4q12, is approximately 42kb in size, and is comprised of 13 exons, 1 of which is entirely untranslated, and 2 only partially translated[266]. Its protein product, Gc globulin (GC), also known as Vitamin D binding protein, is a precursor of macrophage activating factor[267] and enhances the neutrophil chemotactic properties of C5 derived peptides[268]. This is consistent with a role for GC in COPD pathogenesis, since activated macrophages and neutrophilic

inflammation are believed to be important in COPD pathogenesis (section 1.1.3). The latter process is prevented by neutrophil elastase (NE) inhibitors[269], suggesting a relationship between GC dependent inflammation and proteolytic pathways, which could be of importance in AATD. It is thought that this is because NE cleaves the GC binding site on neutrophils, such that GC is freed into the circulation resulting in more neutrophil chemotaxis[269].

The *GC2* allele (rs4588), resulting from a non-synonymous SNP, is protective against COPD in Caucasians[110, 111]conferring an OR of 0.5-0.7 of disease in heterozygotes (varying depending which allele it is combined with)[110] and an OR of 0.17[111] to 0.5[110] in homozygotes. This is consistent with the fact that only 10% of this form of the protein can be converted to macrophage activating factor[270], though it is not known to alter neutrophil chemotaxis[111]. The *GC1F* allele was linked to an increased risk of developing airflow obstruction, emphysema and a rapid decline of FEV1 in Japanese subjects[109, 136], but studies in Caucasians have varied in their results[102, 110, 111]. It is also of note that *GC* variants have been associated with diffuse panbronchiolitis[109], a condition seen almost exclusively in Japan, which leads to bronchiectasis[271].

4.2 Aims of this chapter

This chapter describes specific phenotypic associations of variation in *SFTPB*, *TGFB*, its surrounding region, and *GC*. Associations were sought with the COPD phenotypes described in chapter 2, with the addition of a measure of pulmonary function relevant to small airways disease for *TGFB* (and surrounding region), and of airway bacterial

colonisation for *GC* because of its potential influence on host defences. Data pertaining to frequency and type of airway bacterial colonisation in the study population is therefore presented here, prior to this analysis.

4.3 Methods

The subjects studied have been described in chapter 3 (section 3.2.2), consisting of a case-control population of 424 subjects and a familial replication dataset of 52 sibling pairs. Tag SNPs to cover *SFTPB*, *TGFB, GC, MGC4093* and *CCDC97* were chosen using HapMap, as described in chapter 2 (section 2.3.2). The latter 2 genes lie immediately adjacent to *TGFB* in a region where association with COPD has been reported[108]; the rationale for studying them has been described in sections 1.1.4 and 4.1.2. All tags were genotyped using TaqMan® genotyping technologies on an ABI7900HT, using the methods described in chapter 3 (section 3.2.2). The locations of tags within the genes are shown in Figure 4.1, and their context sequences in Table 4.1.

Each tag was examined for qualitative and quantitative phenotypic associations as before, with additional analyses for small airways disease for the *TGFB* region SNPs, using FEF25-75, and airway bacterial colonisation for *GC* SNPs. FEF25-75 is a measure of pulmonary function most specific to small airways disease, found in COPD (section 1.1.1). It is of particular relevance to study for *TGFB* since its protein product (TGFβ1) is highly expressed in the small airways of subjects with COPD[264], suggesting a sub-phenotype in which this gene/protein is of most relevance. Similarly since GC is known to

have a role in defence of the airway, by means of neutrophil recruitment[268], and is associated with DPB[109], a condition in which airway infection is of particular prognostic importance[271], it was relevant to examine the phenotype of airway bacterial colonisation (section 2.2.6) for this gene alone.

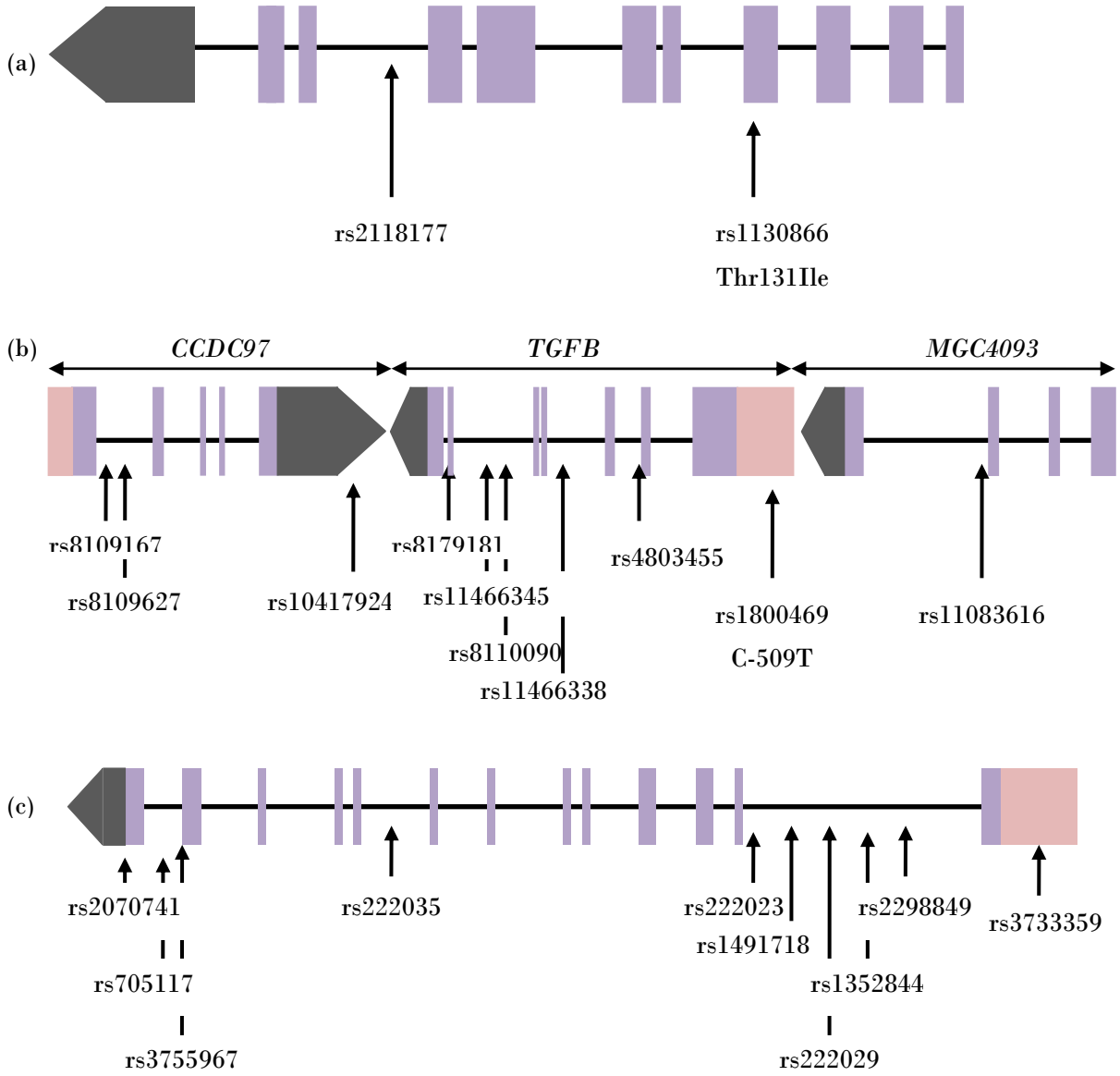


Figure 4.1: Location of tag SNPs within *SFTPB*, the *TGFB* region and *GC*

In each picture the mauve blocks represent exons, the promoter is represented by a grey pentagon, other regulatory regions or non-coding areas by pink blocks and introns by a black line. In each diagram the 3' end is on the left of the page. All SNPs are denoted by their reference SNP number, with additional decriptors added only where these are commonly used in past literature.

(a) *SFTPB*

(b) *TGFB* region

(c) *GC*

Gene	SNP	Context sequence	Design strand
SFTPB	rs2118177	CTTTGTCCAGAGGCGACTCCTAACC[C/T]TTAGCAGGCTCTGCCCTAACTTACA	Reverse
	rs1130866	AGGAGGTGAGCTTGCAGCCCTCACA[A/G]TCTGGTTCTGGAAGTAGTCGATGAC	Reverse
CCDC97	rs8109167	CAGAAGCCTTCTGAACAAGACACAC[G/T]GTGAAAACCATGAAGGAAAAAAAAG	Forward
	rs8109627	GAATAGATCAGAATTTAGTAGATGT[C/T]AGAAGTGGCATTTCAGATTGGTAGG	Reverse
	rs10417924	AGCTAACACGTGCTGGGGACACACA[C/T]AGACAAGACCATCCATACCAAACAG	Reverse
TGFB	rs8179181	GTGGAGCTGCAGGCAGGAGAGACGC[A/G]TCAGGGGCAGGGAGGGGCTACCACC	Reverse
	rs11466345	CCTCAGCCTCCCAAAGTGCTGGGATTACGGG[T/C]GTGAGCTACCATACCTGGCCTCTCTTAG GAC	Reverse
	rs811090	CCCAGAGACAGGACCAAAATATTTT[A/G]AGCAAGGGAGGAATAAGGTCAGATG	Reverse
	rs11466338	GATGGCGCTAAGATCTGGTGGTGAAGGAGCC[T/C]GGGAGCAGGGCTGATTTTCTGGCA	Reverse
	rs4803455	CCTGAATTCTCAGTAACTTAGAAGT[A/C]ATTTCTAATGATTCCGGCTGGGCAC	Forward
	rs1800469	GAGGAGGGGGCAACAGGACACCTGA[A/G]GGATGGAAGGGTCAGGAGGCAGACA	Forward
MGC4093	rs11083616	AGAAAAGAAAAAAAGAAAATGCACC[A/G]AGTACTAACTAAATCCCAGATGCTA	Forward
GC	rs2070741	TGCTCAACAGACGCATATAAACTCA[T/G]GCTATTTTAATATTTATGCCAAAAA	Forward
	rs705117	TAGTTATGTCACAGTTCTATTTGCA[C/T]GGTGTAAAAAATTCCATGTTTCATT	Reverse
	rs3755967	TTGTGGTCTCATTATTTCTGGATAT[C/T]AAGTACCTCAGACCTATAACTTTGT	Reverse
	rs222035	GCTTCAAACTATTTGTGAAATATTT[G/T]TATGCCATCATTATGCATACACATC	Reverse
	rs222023	TTTTTTTTTAAAGATTCTCTTTAGG[A/G]CTCATCAGATTAATGTCACCTGTTT	Forward
	rs1491718	GCTGTTAAACTAATACAAGAAAACA[C/T]AGGGGAAAAGCTCCATGATGTTGGT	Reverse
	rs222029	TTCACAAGGCTTGCATTCTATATTT[A/G]TTACAATAAGCAAATGAACAACCAG	Reverse
	rs1352844	GTAGTGTTTCTCAGAACCTTTAGGA[C/T]GACAAAGGCATGAATTGTGAAACCA	Reverse
	rs2298849	AAACACATTACTATTATGAATAATA[A/G]TGTACGAATTAGGTGAGGCTGTTAA	Reverse
	rs3733359	AGAGAGTCTTGCAGCACCTCCTCTCTCCT[G/A]TAGGTGACCATGTAAAAGTGGTAGCC	Forward

Table 4.1: Tag SNP sequences for primer design in *SFTPB*, *TGFB* region and *GC*

The two possible alleles of each SNP are highlighted in blue between square brackets indicating the position of the SNP within the gene.

4.4 Results

4.4.1 SFTPB

Initially the genotyping success rate was 85.7%, which was unacceptably low, but the plots suggested that this was because of many failures on the plates (Figure 4.2). The plates had been aliquoted entirely by hand, a potentially error prone technique given the small volumes and large number of samples per plate. When a robotic system (Biomek 2000, UK) was used for DNA aliquoting the success rate improved to over 95% (n=413). Both SNPs were in Hardy Weinberg equilibrium (p=0.18 and 0.22 for deviation from HWE). Allele and genotype frequencies are shown in Table 4.2.

The case-control dataset had 80% power to detect a risk allele OR of 1.71-1.83 of developing each phenotype prior to adjustment for covariates, as shown in Table 4.3. This is equivalent for a protective allele of 1/(risk OR), in this case a protective variant with OR 0.54-0.56. There were no allelic associations of either tag with any phenotype, but minor allele homozygotes (CC) of rs2118177 exhibited a lower risk of COPD (defined as FEV1/FVC<0.7 with FEV1<80% predicted) in the general regression model for this phenotype, producing an OR of 0.28 (95% CI 0.11-0.73) for this phenotype (p=0.01, Figure 4.3). This OR is for the minor allele homozygote only, and was obtained after adjustment for covariates. No allele OR was significant. Quantitative analyses in the case-control dataset were non-significant (p>0.1). Quantitative analysis of genotypes in the replication dataset showed a significant association of FEV1 with the same SNP genotype, the t statistic indicating that FEV1 tended to be higher (t=2.96, p=0.002) which is consistent with the result in the case-control group.

When haplotype analysis was performed no haplotype was observed (Figure 4.4).

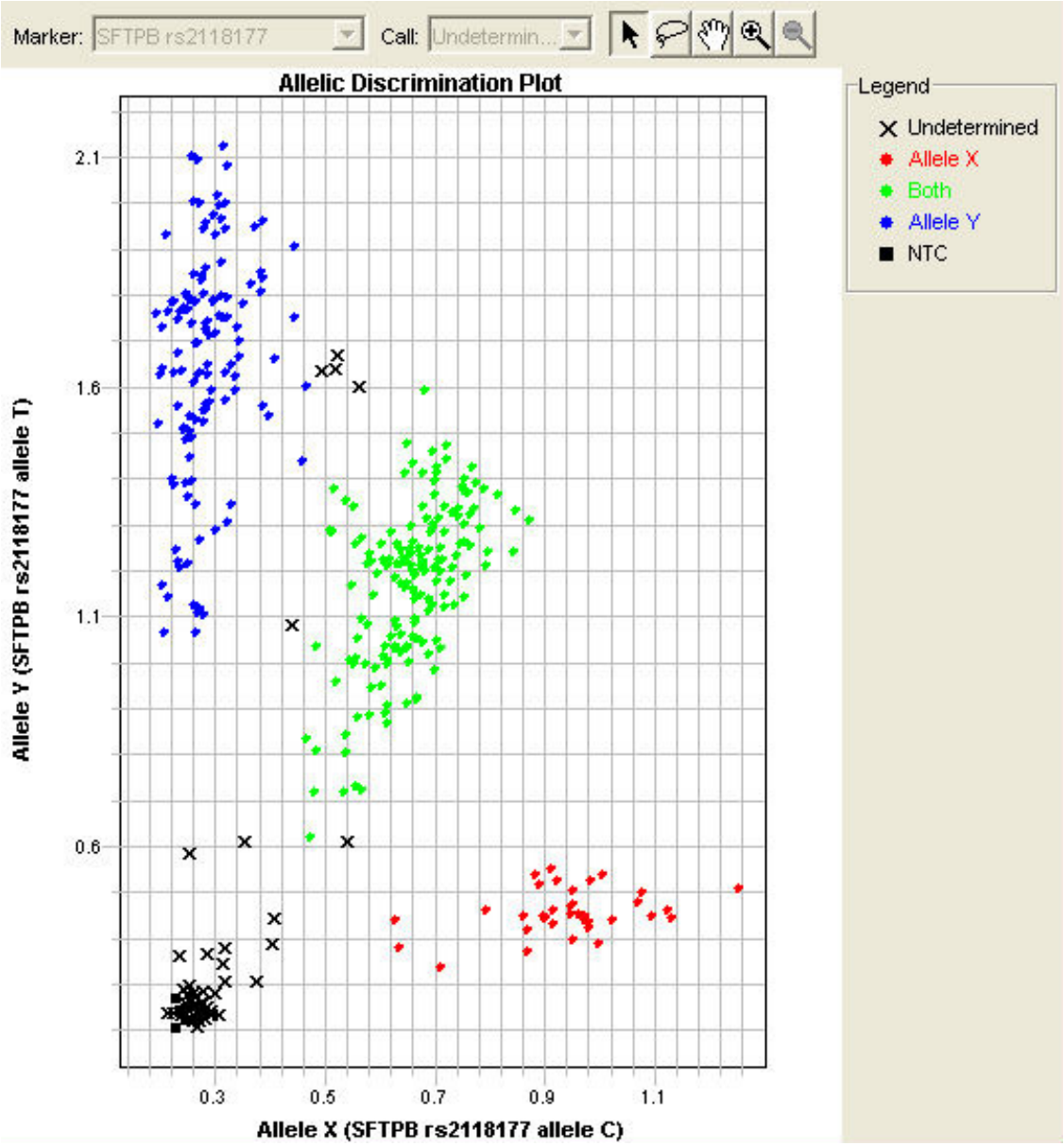


Figure 4.2: Example of TaqMan® plot with high failure rate for *SFTPB* rs2188177
The plot (SDS version 2.2.2, ABI, UK) shows clustering of failed samples adjacent to the negative controls suggesting that there was no DNA present in the wells. Use of a robotic aliquoting system improved this, indicating that human error in repeated aliquoting of small volumes may have been the cause.

SNP	All AATD	Chronic bronchitis	Emphysema	Bronchiectasis
rs2118177 [T/C]				
C allele	36.92	33.02	35.71	32.23
TT	38.26 (158)	42.14 (67)	38.83 (106)	43.42 (33)
TC	49.63 (205)	49.69 (79)	51.65 (141)	48.69 (37)
CC	12.11 (50)	8.18 (13)	9.52 (27)	7.89 (6)
rs1130866 [A/G]				
G allele	49.76	54.09	49.08	45.39
AA	23.73 (98)	18.86 (30)	24.91 (68)	30.26 (23)
AG	53.03 (219)	54.08 (86)	52.01 (142)	48.68 (37)
GG	23.25 (96)	27.04 (43)	23.08 (63)	21.05 (16)

Table 4.2: Allele and genotype frequencies of *SFTPB* tag SNPs

SNP alleles are listed [major, minor] and the minor allele frequency given as a percentage, together with all genotype frequencies as a percentage, with the raw value in brackets.

SNP	Phenotype	Minimum odds ratio detectable with 80% power
rs2118177	Emphysema	1.81
	Bronchiectasis	1.76
	Chronic bronchitis	1.57
rs1130866	Emphysema	1.86
	Bronchiectasis	1.74
	Chronic bronchitis	1.54

Table 4.3: Power of the *SFTPB* studies to detect qualitative phenotypes in the case-control dataset

Power calculations were carried out as before (Appendix 4), and the ORs shown are prior to adjustment for covariates.

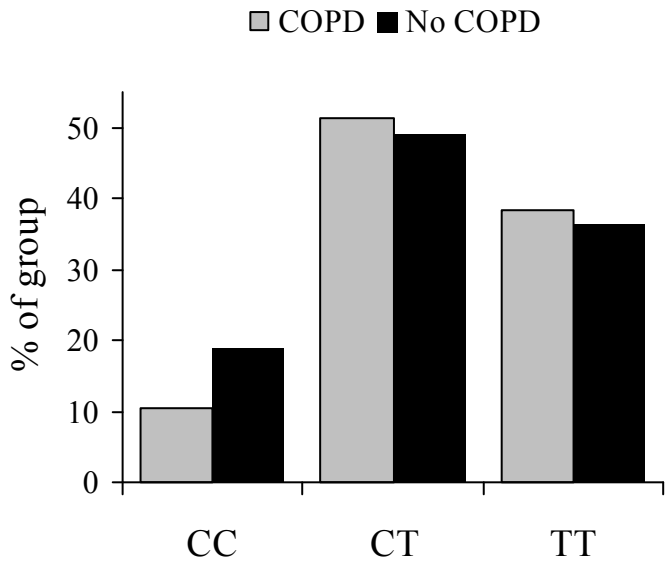


Figure 4.3: Distribution of rs2118177 genotypes in subjects with and without COPD
The CC genotype, being the minor allele homozygote, was less common in those with COPD, suggesting a protective effect of this genotype. The difference in genotype frequency was statistically significant, with p=0.01 in the regression model after adjustment for smoking, gender and age.

Figure 4.4: Linkage disequilibrium plot for *SFTPB* tag SNPs
No haplotype block was observed in Haploview for further analysis, as shown by the low r^2 value (17) between the 2 SNPs in this LD plot. See Figure 3.5 for further information about interpretation of LD plots.

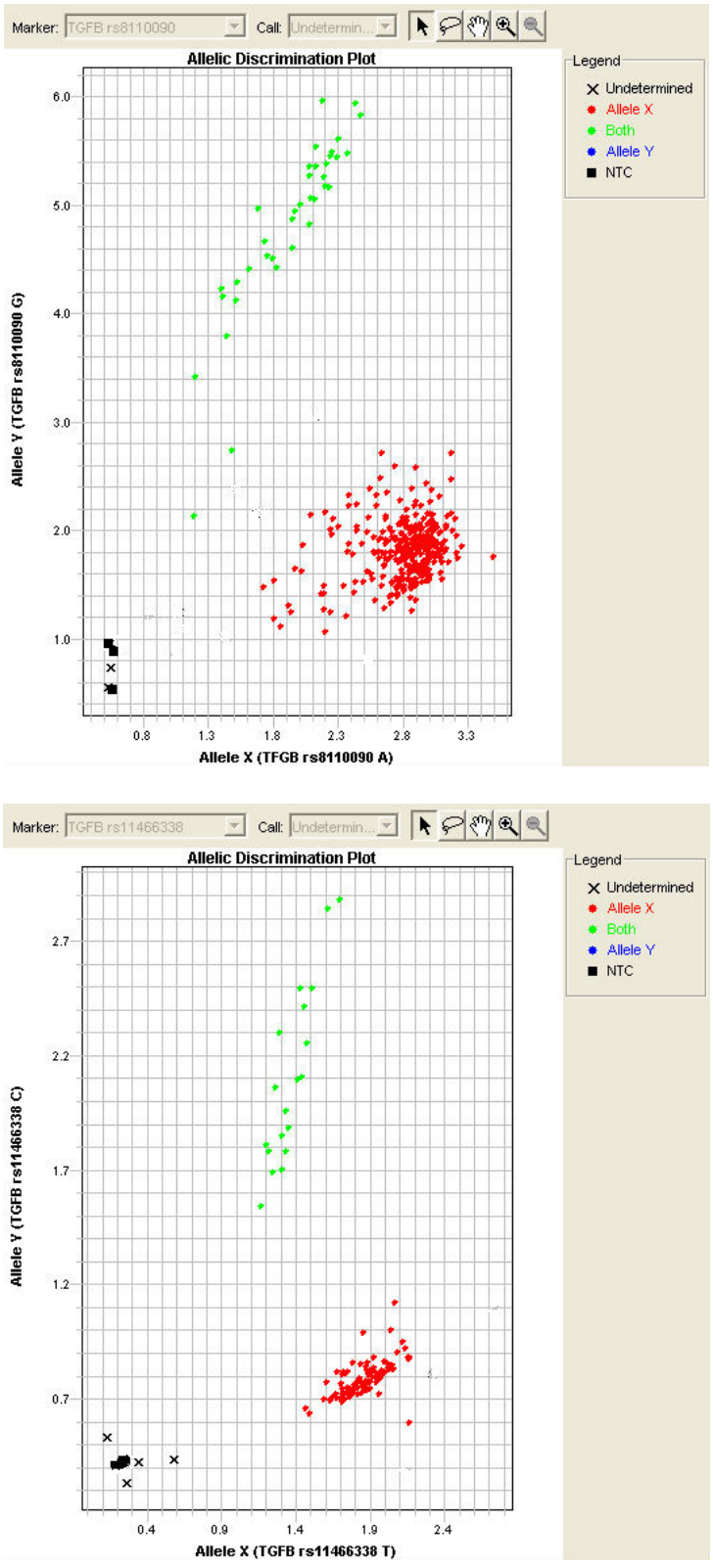
4.4.2 TGFB region

Genotyping was successful in 93.2% of cases (n=395). All SNPs were in HWE (p>0.05). Two cluster plots were observed for several *TGFB* SNPs (Figure 4.5) indicating absence of the minor allele homozygote. As a consequence power lessened, such that for the least common minor allele (rs8110090) the study had 80% power to detect an OR of 2.96 in the worst affected phenotypic analysis, compared to the best OR of 1.55 achieved with rs4803455. Allele and genotype frequencies are shown in Table 4.4, and power prior to adjustment for covariates in Table 4.5.

There were no allelic or genotypic associations with any qualitative phenotype in the case-control dataset. One SNP was associated with FEV1 in the quantitative analyses (rs4803455, p=0.043), with a trend towards association being observed for the same SNP with FEV1/FVC (p=0.076). SNPs tagging the haplotype blocks upstream of rs4803455 also showed a trend towards an association with FEV1/FVC although this also failed to reach significance (rs1800469, p=0.052; rs11083616, p=0.100). Abnormalities of this alone may occur with small airways disease, and the lung function measure most specific to this phenotype (FEF25-75) was associated with rs1800469 (p=0.017). The most associated SNP genotype was the TT genotype of rs1800469; the difference in FEF25-75 observed is illustrated in Figure 4.6. Significant tag SNP associations, and supporting trends, from the regressions are shown in Table 4.6.

In the replication dataset quantitative analyses using FEV1/FVC were not significant ($p>0.05$), but there was insufficient information available pertaining to FEF25-75 to assess this phenotype.

Haplotype analysis revealed one haplotype block, which showed no qualitative phenotypic associations. It is of note that the 3 SNPs (rs4803455, rs1800469 and rs11083616) showing associations and trends with FEV1/FVC and FEF25-75 tended to form a haplotype (Figure 4.7), but the Haploview software did not show significant correlations between these SNPs ($p>0.05$) so the group was not analysed as a block.



<u>Figure 4.5: Examples of two cluster TaqMan® plot for rs8110090 & rs1146638 in *TGFB*</u>
Two cluster plots were seen for two SNPs within *TGFB*, examples of the plots from SDS (version 2.2.2, ABI, UK) being shown above.

Gene	SNP	All AATD	Chronic bronchitis	Emphysema	Bronchiectasis
CCDC97	rs8109167 [G/T]				
	T allele	20.00	21.48	21.27	19.18
	GG	63.44 (251)	61.07 (91)	62.42 (167)	65.75 (48)
	GT	32.91 (130)	34.90 (52)	32.84 (88)	30.14 (22)
	TT	3.54 (14)	4.03 (6)	4.85 (13)	4.11 (3)
	rs8109627 [T/C]				
	C allele	31.01	29.28	29.86	29.45
	TT	47.34 (187)	45.64 (68)	47.69 (128)	49.32 (36)
	TC	43.29 (171)	44.97 (67)	44.62 (120)	42.47 (31)
	CC	9.37 (37)	9.39 (14)	7.46 (20)	8.22 (6)
	rs10417924 [C/T]				
	T allele	19.12	17.45	16.60	19.18
	CC	65.06 (257)	68.46 (102)	68.28 (183)	64.38 (47)
	CT	31.65 (125)	28.19 (42)	30.22 (81)	32.88 (24)
	TT	3.29 (13)	3.36 (5)	1.49 (4)	2.74 (2)
TGFB	rs8179181 [G/A]				
	A allele	23.80	24.49	26.87	24.66
	GG	56.96 (225)	57.72 (86)	48.13 (129)	56.16 (41)
	GA	38.48 (152)	37.58 (56)	50.00 (134)	38.36 (28)
	AA	4.56 (18)	4.70 (7)	1.87 (5)	5.48 (4)
	rs11466345[A/G]				
	G allele	9.75	11.07	10.07	9.59
	AA	81.01 (320)	78.52 (117)	79.85 (214)	82.19 (60)
	AG	18.48 (73)	20.81 (31)	20.15 (54)	16.44 (12)
	GG	0.51 (2)	0.67 (1)	-	1.37 (1)
	rs8110090 [A/G]				
	G allele	5.32	5.37	4.29	6.16
	AA	89.36 (353)	89.26 (133)	91.42 (245)	87.67 (64)
	AG	10.63 (42)	10.74 (16)	8.58 (23)	12.33 (9)
	GG	-	-	-	-
	rs11466338[A/G]				
	G allele	8.48	7.72	7.09	7.54
	AA	83.04 (328)	84.78 (126)	85.82 (230)	84.93 (62)
	AG	16.96 (67)	15.44 (23)	14.18 (38)	15.07 (11)
	GG	-	-	-	-
	rs4803455 [A/C]				
	C allele	42.53	43.29	41.60	43.84
	AA	33.42 (132)	32.21 (48)	35.45 (95)	32.88 (24)
	AC	48.10 (190)	48.99 (73)	45.90 (123)	46.58 (34)
	CC	18.48 (73)	18.89 (28)	18.66 (50)	20.55 (15)

	rs1800469 [C/T]				
	T allele	26.96	27.18	22.01	26.03
	CC	52.91 (209)	52.35 (78)	60.82 (163)	54.79 (40)
	CT	40.25 (159)	40.94 (61)	34.43 (92)	38.36 (28)
	TT	6.84 (27)	6.71 (10)	4.85 (13)	6.85 (5)
MGC4093	rs11083616[A/G]				
	G allele	32.78	34.23	30.78	33.56
	AA	44.57 (176)	42.28 (63)	47.76 (128)	42.47 (31)
	AG	45.32 (179)	46.98 (70)	42.91 (115)	47.95 (35)
	GG	10.13 (40)	10.74 (16)	9.33 (25)	9.59 (7)

Table 4.4: Allele and genotype frequencies of *TGFB* region tag SNPs
SNP alleles are listed [major, minor] and the minor allele frequency given as a percentage, together with all genotype frequencies as a percentage, the raw number being given in brackets.

SNP	Phenotype	Minimum odds ratio detectable with 80% power
rs8109167	Emphysema	1.86
	Bronchiectasis	1.86
	Chronic bronchitis	1.62
rs8109627	Emphysema	1.77
	Bronchiectasis	1.72
	Chronic bronchitis	1.53
rs10417924	Emphysema	1.92
	Bronchiectasis	1.84
	Chronic bronchitis	1.67
rs8179181	Emphysema	1.83
	Bronchiectasis	1.83
	Chronic bronchitis	1.61
rs11466345	Emphysema	2.23
	Bronchiectasis	2.32
	Chronic bronchitis	1.85
rs8110090	Emphysema	2.96
	Bronchiectasis	2.60
	Chronic bronchitis	2.24
rs11466338	Emphysema	2.50
	Bronchiectasis	2.40
	Chronic bronchitis	2.03
rs4803455	Emphysema	1.79
	Bronchiectasis	1.74
	Chronic bronchitis	1.55
rs1800469	Emphysema	1.87
	Bronchiectasis	1.76
	Chronic bronchitis	1.58
rs11083616	Emphysema	1.77
	Bronchiectasis	1.71
	Chronic bronchitis	1.53

Table 4.5: Power of the *TGFB* region studies to detect qualitative phenotypes in the case-control dataset

Power calculations were carried out as before (Appendix 4), and the ORs shown are prior to adjustment for covariates.

	Tag SNP	Phenotype	B	SE (B)	r^2 change	p value
*	rs4803455	logFEV1	0.037	0.018	0.009	0.043
		logFEV1/FVC	0.022	0.012	0.007	0.076
§	rs1800469	logFEV1/FVC	-0.041	0.033	0.007	0.052
		FEF25-75	-1.501	0.629	0.041	0.019
*	rs11083616	logFEV1/FVC	0.20	0.013	0.006	0.100

Table 4.6: *TGFB* region tag SNP associations in linear regression models
The table shows the regression parameters models from those models where p≤0.1 Co-dominant models are denoted *, in each case the heterozygote was coded 1, and the minor allele homozygote 2, such that when the B value is positive it indicates a higher level of the outcome variable in minor allele homozygotes. § indicates a recessive model.

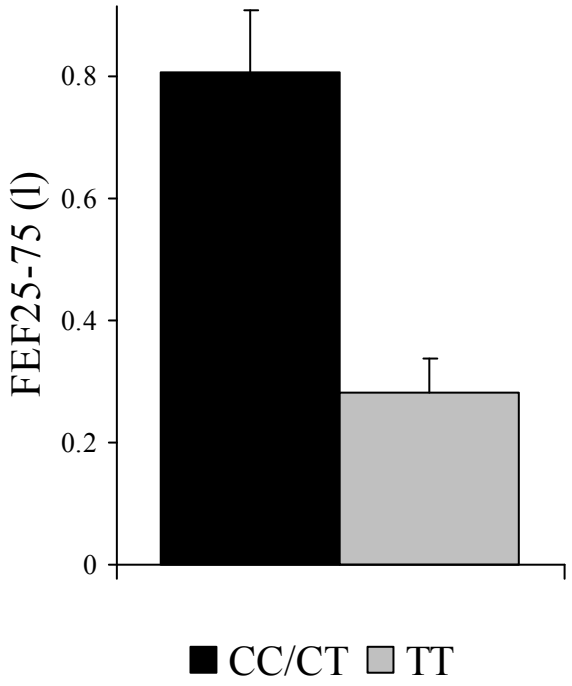


Figure 4.6: FEF25-75 relates to rs1800469 genotype
The bar chart shows the mean (SE) prior to adjustment for covariates, stratified by genotype. The minor allele homozygote is shown separately as this genotype was only significant in a recessive model. The difference is significant, with p=0.019 in the regression model.

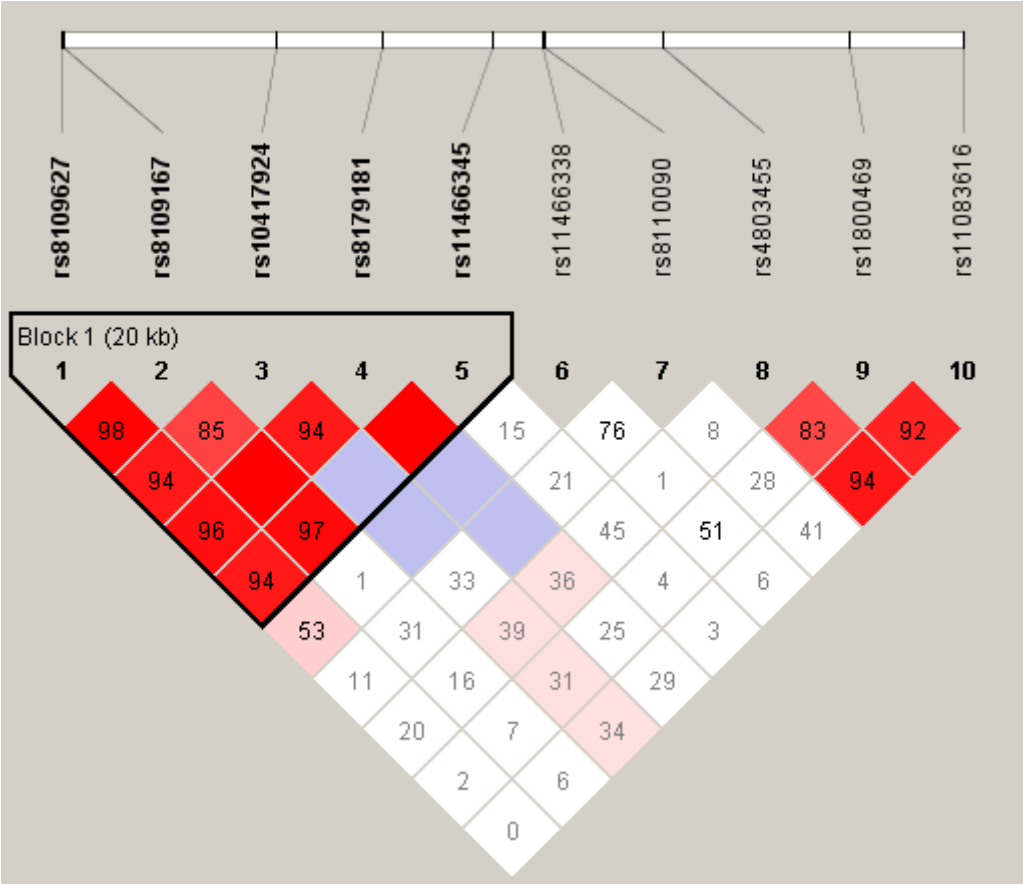


Figure 4.7: Linkage disequilibrium plot for tag SNPs in the *TGFB* region

This output from Haploview shows the r^2 values between SNPs, and identified one haplotype block containing all tags within *CCDC97* and 2 within *TGFB*. The 3 tags showing a tendency to association with small airways disease (rs4803455, rs1800469 and rs11083616) have a propensity to form a haplotype, as can be seen from the r^2 values, and colour of the blocks, although this was not statistically significant.

4.4.3 *GC*

Genotyping was successful in 90.1% of cases (n=382). This is a little lower than some of the other genes contained in this thesis, although above the acceptable threshold of 90%. This occurred predominantly because of initial problems with the assay for rs705117, which required alterations of the PCR reaction mixture before adequate genotype discrimination was achieved. This process is summarised in Figure 4.8. Despite this improvement, a genotype could not be confidently allocated in 42 cases for this SNP. All SNPs were in Hardy Weinberg equilibrium ($p>0.05$ for deviation from HWE), except rs3755967 ($p=0.04$), which was therefore excluded from further anlayses. Allele and genotype frequencies are shown in Table 4.7. In the qualitative analyses the mean OR the study had 80% power to detect was 2.01; full details of power are shown in Table 4.8.

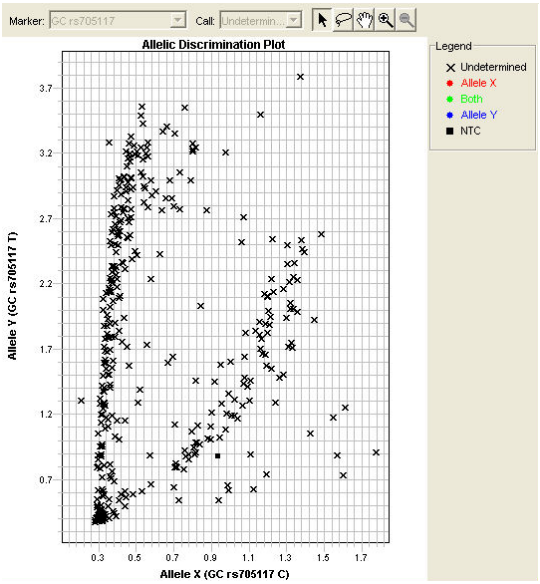
One SNP (rs2070741) showed both allele and genotype association with bronchiectasis, with the minor allele [C] conferring an OR of 1.80, as shown in Table 4.9. A second SNP showed an association with the same phenotype in the minor allele homozygote, but its significance was lost in an allelic analysis (Table 4.9). No other associations were seen with the qualitative or quantitative phenotypes described earlier ($p>0.1$ in regression models).

Further analyses were undertaken based on the existence of bacterial colonisation of the airway (section 2.2.6), to give an indication of the mechanism behind the phenotypic association. One hundred sputum samples were analysed; the frequency of colonisation and bacterial species identified are shown in Table 4.10. Some examples of the growth plates used are shown in Figure 4.9. Subjects with airway colonisation exhibited no significant clinical differences from those without (Table 4.11), but the phenotype was

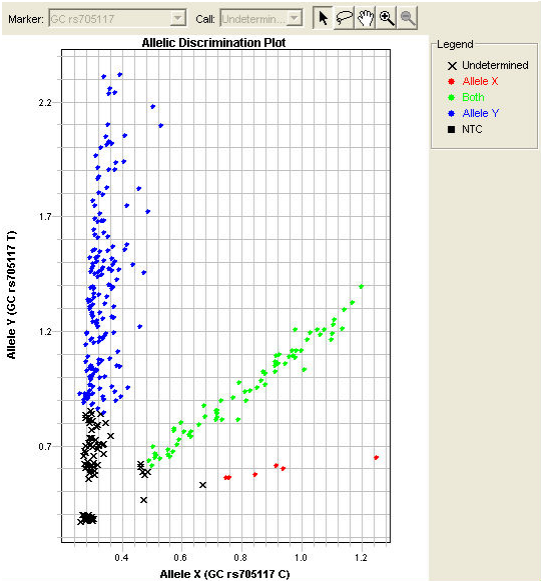
more likely to occur in carriers of the C allele of rs2070741. The allele conferred an OR of 3.83 for this phenotype (p=0.04). There was insufficient data relating to bronchiectasis and bacterial colonisation in the family dataset to test replication.

Three haplotype blocks were observed in the case-control dataset (Figure 4.10), but none showed any qualitative phenotypic associations (all p>0.1 in each analysis).

(a) Original PCR reaction



(b) Double amount of fluorescent assay



(c) Increased PCR cycles to 50, with assay amount from (b)

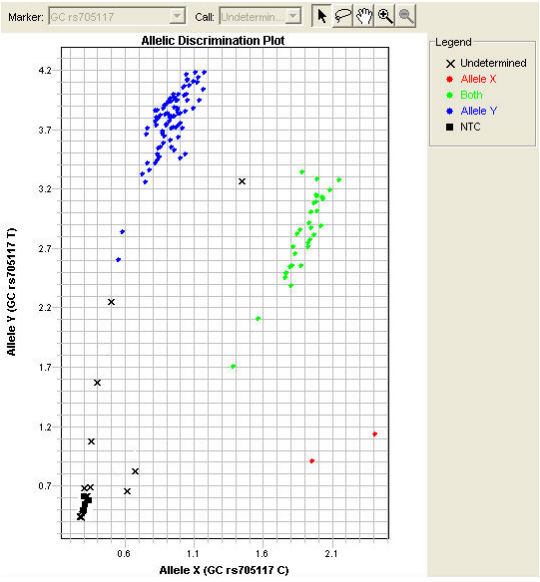


Figure 4.8: TaqMan® plots for selected rs705117 genotyping reactions

(a) No genotypic groups could be distinguished by fluorescence signals in the original PCR reaction (b) Increasing the amount of fluorophore allowed genotypic groups to be distinguished by the software, but there was a linear pattern; clear clusters were not seen until the PCR cycles were also extended (c).

SNP	All AATD	Chronic bronchitis	Emphysema	Bronchiectasis
rs2070741 [A/C]				
C allele	9.29	8.78	9.55	14.00
AA	81.41 (311)	83.11 (123)	81.65 (218)	74.67 (56)
AC	18.11 (69)	16.22 (24)	17.77 (47)	22.67 (17)
CC	0.52 (2)	0.68 (1)	0.75 (2)	2.67 (2)
rs705117 [T/C]				
C allele	17.54	16.89	17.23	17.33
TT	68.85 (263)	70.95 (105)	70.04 (187)	69.33 (52)
TC	27.23 (104)	24.33 (36)	25.47 (68)	26.67 (20)
CC	3.93 (15)	4.73 (7)	4.49 (12)	4.00 (3)
rs3755967 [C/T]				
T allele	27.62	28.04	29.03	28.00
CC	50.26 (192)	51.35 (76)	47.19 (167)	49.33 (37)
CT	44.24 (169)	41.22 (61)	47.57 (127)	45.33 (34)
TT	5.50 (21)	7.43 (11)	5.24 (14)	5.33 (4)
rs222035 [G/T]				
G allele	42.93	44.59	44.76	46.00
GG	33.25 (127)	34.46 (51)	32.21 (86)	36.00 (27)
GT	47.64 (182)	42.14 (62)	46.07 (123)	36.00 (27)
TT	19.12 (73)	23.65 (35)	21.72 (58)	28.00 (21)
rs222023 [G/A]				
A allele	5.89	4.05	5.43	7.33
GG	88.74 (339)	92.57 (137)	89.89 (240)	88.00 (66)
GA	10.73 (41)	6.76 (10)	9.36 (25)	9.33 (7)
AA	0.52 (2)	0.68 (1)	0.75 (2)	2.67 (2)
rs1491718 [T/C]				
C allele	13.22	13.18	12.55	13.33
TT	75.13 (287)	74.32 (110)	76.40 (204)	74.67 (56)
TC	23.30 (89)	25.00 (37)	22.10 (59)	24.00 (18)
CC	1.57 (6)	0.68 (1)	1.50 (4)	1.35 (1)
rs222029 [A/G]				
G allele	18.32	16.22	17.98	20.40
AA	67.02 (256)	71.62 (106)	67.79 (181)	61.84
AG	29.32 (112)	24.32 (36)	28.46 (76)	31.53
GG	3.67 (14)	4.05 (6)	3.91 (10)	2.63
rs1352844 [C/T]				
T allele	12.83	11.15	13.30	15.33
CC	75.13 (287)	77.71 (115)	73.78 (197)	70.67 (53)
CT	24.08 (92)	22.30 (33)	25.84 (69)	28.00 (21)
TT	0.79 (3)	-	0.37 (1)	1.35 (1)
rs2298849 [C/T]				
T allele	19.50	17.23	20.22	24.00
CC	64.92 (248)	69.59 (103)	63.67 (170)	54.67 (41)

CT	31.15 (119)	26.35 (39)	32.21 (86)	42.67 (32)
TT	3.93 (15)	4.05 (6)	4.12 (11)	2.70 (2)
rs3733359 [G/A]				
A allele	9.55	9.80	8.80	8.67
GG	82.03 (313)	81.38 (120)	83.15 (222)	82.67 (62)
GA	17.02 (65)	18.24 (27)	16.10 (43)	17.33 (13)
AA	1.05 (4)	0.68 (1)	0.75 (2)	-

Table 4.7: Allele and genotype frequencies of *GC* tag SNPs

SNP alleles are listed [major, minor] and the minor allele frequency given as a percentage, together with all genotype frequencies as a percentage, with the raw value in brackets.

SNP	Phenotype	Minimum odds ratio detectable with 80% power
rs2070741	Emphysema	2.22
	Bronchiectasis	2.33
	Chronic bronchitis	1.89
rs705117	Emphysema	2.13
	Bronchiectasis	2.13
	Chronic bronchitis	1.82
rs3755967	Emphysema	1.82
	Bronchiectasis	1.75
	Chronic bronchitis	1.56
rs222035	Emphysema	1.77
	Bronchiectasis	1.70
	Chronic bronchitis	1.54
rs222023	Emphysema	2.93
	Bronchiectasis	2.92
	Chronic bronchitis	2.08
rs1491718	Emphysema	2.05
	Bronchiectasis	2.05
	Chronic bronchitis	1.76
rs222029	Emphysema	2.00
	Bronchiectasis	1.91
	Chronic bronchitis	1.64
rs1352844	Emphysema	2.21
	Bronchiectasis	2.05
	Chronic bronchitis	1.74
rs2298849	Emphysema	2.06
	Bronchiectasis	1.90
	Chronic bronchitis	1.61
rs3733359	Emphysema	2.06
	Bronchiectasis	2.18
	Chronic bronchitis	1.90

Table 4.8: Power of the *GC* studies to detect qualitative phenotypes in the case-control dataset

Power calculations were carried out as before (Appendix 4), and the ORs shown are prior to adjustment for covariates.

	Minor allele homozygote	Minor allele	Allele OR	95% CI
rs2070741 [A/C]	0.031	0.046	1.80	1.02-3.19
rs2298849 [C/T]	0.013	NS	-	-

Table 4.9: *GC* tag SNP associations with bronchiectasis

The table shows the p values for associated SNPs only. SNPs are listed showing [major/minor] alleles, together with the OR of bronchiectasis conferred by genotypes containing the minor allele.

	Frequency (n)
Not colonised	30
Mixed normal flora	20
Haemophilus influenzae	23
Moraxella catarrhalis	11
Haemophilus parahaemolyticus	8
Staphylococcus aureus	6
Pseudomonas aeruginosa	5
Streptococcus pneumoniae	4
Haemophilus parainfluenzae	3
Pseudomonas species	2
Proteus mirabilis	2
Enterobacteriaceae	2
Haemophilus haemolyticus	1
Klebsiella oxytoca	1
Haemophilus species	1
Bacterial numbers in cfu/ml	8.0×10^7 (1.2×10^7-1.3×10^9)

Table 4.10: Microbiology of stable state sputum samples in AATD subjects

The majority of subjects had at least one bacterial species in the airway, with only 30 samples having none present.

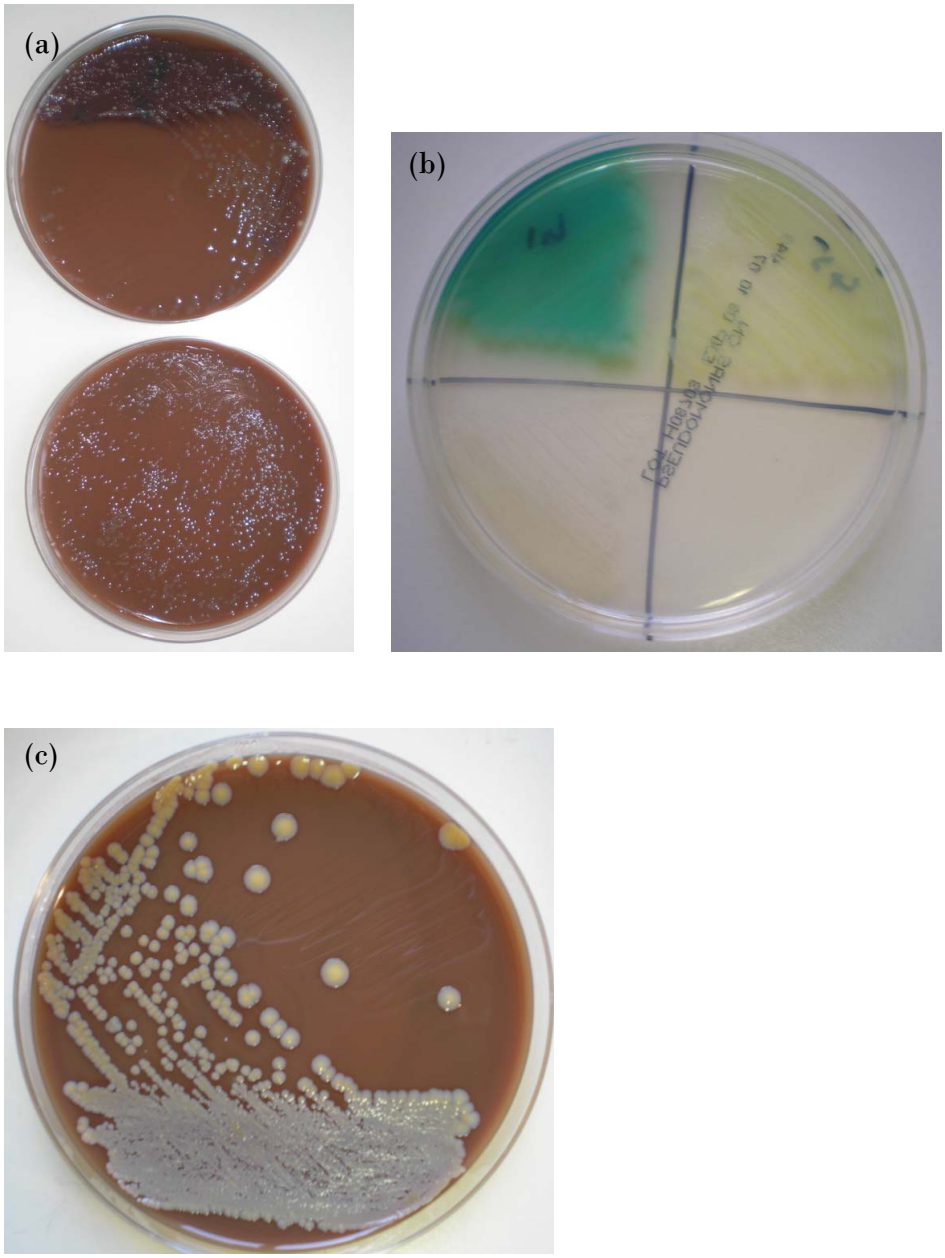


Figure 4.9: Sputum culture plates
(a) *Pseudomonas* colonies: small, white colonies, shown on primary plate and a subsequent dilution and a specific cetrimide plate (b). *P.aeruginosa* appears green and will fluoresce under UV light when grown on this medium.
(c) *Staphylococcus aureus* colonies: larger, yellow colonies, with dilution across the plate

	Airway bacterial colonisation	No colonisation
Age in years	51.37 (47.24-55.31)	50.74 (42.48-59.05)
Smoking in pack years	16.00 (5.70-26.30)	18.00 (11.00-22.00)
FEV1%predicted	32.71 (21.14-56.18)	28.15 (19.57-38.21)
FEV1/FVC %	37.70 (26.10-47.00)	33.00 (25.00-45.00)
KCO %predicted	66.31 (3.32)	56.57 (5.07)
Bronchiectasis %	35.56	34.78
Emphysema %	86.67	86.96
UZVI % voxels <-910HU	30.96 (2.53)	44.50 (4.53)
LZVI % voxels <-910HU	49.80 (29.68-59.43)	59.40 (52.10-65.50)

Table 4.11: Characteristics of subjects with and without airway bacterial colonisation
Data is shown as mean (SE) and median (IQR), as described in chapter 2 (section 2.3.1). There were no statistically significant differences.

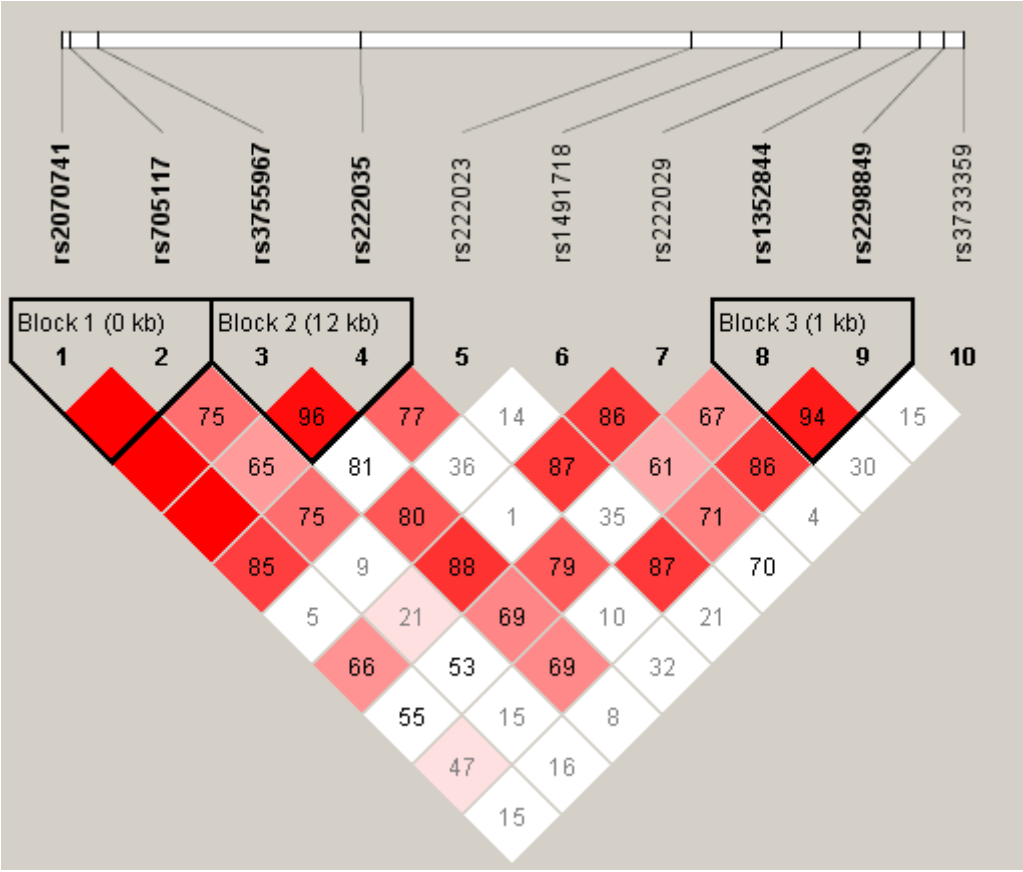


Figure 4.10: Linkage disequilibrium plot for *GC* tag SNPs
The output above from Haploview, using r^2 values as before, shows that 3 haplotype blocks were observed. No phenotypic associations were seen with any of these blocks.

4.5 Discussion

The results presented in this chapter illustrate the concept that detailed phenotypic information is key when examining genetic associations, since *SFTPB*, the *TGFB* region and *GC* show different phenotypic associations from one another, yet have all been associated with the broader diagnosis of COPD in previous published work. It is interesting that the associations seen are consistent with their protein products' putative roles in pathogenesis, though further studies would be needed to endorse this. However the strength of these associations and the degree of confidence in the results is also variable, and will now be discussed.

The association of rs2118177 SNP in *SFTPB* with COPD, defined by lung function, in the case-control dataset and with FEV1 in the family dataset indicate a potential protective effect of the minor allele homozygote. It is an intronic SNP of no known function that tags the promoter region of the gene in HapMap[71], though coverage here is relatively sparse. In usual COPD this SNP has been associated with exacerbation frequency[113]; a number of SNPs within the gene were associated with this outcome (defined by health care utilisation only) in a study population comprising participants in NETT. The most associated SNP was in the promoter region of the gene (rs3024791, p=0.007) but rs2118177 showed a weaker association. This phenotype has not been examined in other studies, so the result should be viewed with caution until it has been replicated. Other genetic association studies derived from NETT have not examined this SNP [19, 72, 78] and the only similar study in AATD did not genotype this *SFTPB* SNP [236]. The AAT genetic modifier study only genotyped rs1130866 (Thr131Ile), since this

had previously been the most associated SNP in *SFTPB* with COPD, and was only able to examine association with spirometric measures[236]. These two facts might explain the lack of association seen.

Evidence from the study reported in this chapter alone would be insufficient to justify the expense of fine mapping the haplotype block to find the actual disease associated variation, but should encourage further study of the role of surfactant proteins in emphysema and AATD, since a genetic association implies a role for the protein product in pathogenesis. Current evidence would suggest that it is most likely to be involved in parenchymal disease, since levels are low in an animal model of emphysema[252] and mathematical models suggest that the effect of this in the lung would be to alter alveolar surface tension in such a way that emphysema could result[253]. We were unable to prove such an association here, perhaps because the number included in the analyses based on a CT diagnosis of emphysema was lower, as CT data was not available on all subjects. If SFTPB contributes to emphysema it might show potential in treating the condition, especially as preparations of surfactant suitable for delivery to the lung are already in use[272].

In the *TGFB* region 3 tag SNPs showed a tendency to associate with lung function, consistent with small airways disease in the case-control dataset. However, since FEF25-75 (the lung function measure most specific for small airways disease) was not available in the familial dataset, it was not possible to replicate the association. The associations with other lung function measures were tested in the familial dataset but showed no association. The most associated SNP was rs1800469 (C-509T) in which the T allele elevates circulating TGFβl by increasing gene transcription[262]; this suggests that the

protein contributes to small airways disease, and is consistent with prior clinical work examining small airway biopsies which show elevated levels of the protein product (TGFβ1) in this location[264]. In usual COPD genetic association work on rs1800469 has shown conflicting results - in American Caucasian subjects the T allele has been associated with both COPD[108] and increased dyspnoea[78], but C allele containing haplotypes are also associated with COPD in European Caucasians[134] and in Chinese subjects[273]. If the association is specific to small airways disease, as our results suggest, then inadequate phenotyping might explain these inconsistencies. It is also possible that the true association is with a SNP in LD with rs1800469. The two SNPs tagging haplotype blocks adjacent to rs1800469 (rs4803455 and rs11083616) were associated with FEV1 and showed a slight trend to association with FEV1/FVC respectively, both being phenotypes consistent with early disease, and the haplotype analysis suggested that these tended to form a block with rs1800469 (Figure 4.8) though this did not reach statistical significance. The rs4803455 SNP is an intronic SNP in *TGFB* of no known functional consequence, and with no known disease associations. rs11083616 is an intronic SNP, again with no known functional consequence, and lies in *MGC4093*, a gene which codes for a hypothetical protein. Expression of *MGC4093* is increased in the small airway epithelium of smokers compared to those who have never smoked[274], and correlated with emphysema severity in subjects in NETT[275]. This suggests a role for the protein in small airway disease, and emphysema, such that genetic variation might also be associated with disease. This gene, its putative protein product, and their roles in COPD are, therefore, worthy of further study. Since TGFβ1 is a multifunctional protein, if a specific association were proven with *TGFB* it is unlikely to be a target for drug design, as blocking or altering its actions would

probably have widespread, potentially harmful effects.

Two *GC* SNPs showed association with bronchiectasis and/or bacterial colonisation of the airway in the case-control dataset. Information pertaining to these phenotypes was not available in the familial replication dataset, so the association could not be replicated. Both are intronic SNPs of no known function, do not tag known functional variants, and were at low frequency in the cases studied. The latter fact in particular means that the result must be viewed with caution as the possibility of false positive results is higher in small datasets. However, given that GC is involved in both neutrophil chemotaxis and macrophage activation, by enhancement of the chemotactic properties of complement derived peptides[268] and because it is a precursor of macrophage activating factor[267], the observation of an apparent genetic association with airway defence against infection is interesting. Airway bacterial colonisation is associated with increased pulmonary inflammation[235], and with an increased frequency of exacerbations[276], both of which may influence disease progression. The cause of bronchiectasis in AATD is unknown; though it tends to co-locate with emphysema in the majority of subjects, a bronchiectasis predominant phenotype is also recognised[22]. This genetic association generates a hypothesis that structural damage to the airway from repeated exacerbations, after bacterial colonisation of the airway, could lead to bronchiectasis. It will therefore also be of importance to examine the role of *GC* polymorphisms in bronchiectasis unrelated to AATD, and to study the protein's role further in airway disease.

The level of p-value to determine significance in genetic studies is controversial, as discussed previously (section 3.5). The studies herein adopted the well recognised strategy of replication in an independent dataset. Unfortunately there was insufficient phenotypic

data to test for replication of all observed associations in *TGFB* and *GC* in the family dataset, thus the association in *SFTPB* is the only one that can be viewed with confidence. In diseases where a number of genes influence outcome, as is likely to be the case with COPD, each polymorphism may confer only a small increased risk of disease, thus large numbers are needed to detect differences between populations. In this respect the studies in this thesis are limited, and underpowered. Power was a particular issue with the studies contained in this chapter where many tag SNPs exhibited low minor allele frequencies. This means that an effect cannot be excluded in all SNPs, except where the allele effect size exceeds that reported in Table 4.8. These are somewhat higher than many of the OR's now being reported in other complex diseases with multiple genetic associations, which may be as low as 1.1[277]. The level of OR more typically observed when the association is within a locus whose protein product is deemed pathophysiologically important, such as the thyroid stimulating hormone receptor gene in Grave's disease, is closer to 1.5[278]. Nevertheless it does not diminish the significance of the result for rs2118177, within *SFTPB,* since this has replicated in both datasets.

4.6 Conclusions

Variation within the haplotype block tagged by rs2218177 in *SFTPB* is associated with COPD in AATD. Variation within the *TGFB* region and within *GC* associates with small airways disease and bronchiectasis respectively, but these results require verification in an independent patient population.

Chapter 5: The role of *MMP*s

5.1 Introduction

Matrix Metalloproteinases (MMP's) compromise a structurally and functionally related group of twenty proteolytic enzymes, which are capable of degrading components of the extracellular matrix[39]. The functions of some of those thought to be important in COPD are summarised in Table 5.1. It is this potential for destruction which is thought to contribute to the pathogenesis of COPD, consistent with the protease-antiprotease theme discussed in Chapter 1 (section 1.1.3)). Many clinical studies have reported the presence of MMPs in the lung, with elevated levels in subjects with COPD[279, 280], supporting this concept.

MMPs are synthesised as proenzymes which are maintained in an inactive form by interaction of a zinc residue in their catalytic domain with a cysteine residue within the proenzyme domain, as shown in Figure 5.1. Macrophages secrete many MMPs, including MMP1, MMP3, MMP7 and MMP12[39]. A separate group, known as membrane type MMPs, are localised to the cell surface (MMP14-17). The production of both MMP types is highly regulated at the transcriptional level by a variety of growth factors and cytokines[39]. A further two macrophage derived MMPs (MMP8 and MMP9) are not as well regulated, being stored in the granules of neutrophils for rapid release[39]. Subsequent MMP activity is regulated by tissue inhibitors of metalloproteases (TIMPs). Although all TIMP's are capable of inhibiting any MMP their affinity for each MMP varies[281].

	Interstitial collagens	Basement membrane components	Elastin
MMP1	III, I, II	Laminin, entactin, proteoglycan	-
MMP3	-	Laminin, entactin, proteoglycan, pepsinised type IV collagen	+
MMP9	Gelatin	Types IV and V collagen, entactin, proteoglycan	++
MMP12	-	Laminin, entactin, proteoglycan, pepsinised type IV collagen	++++

Table 5.1: Activity of selected MMP against components of the extracellular matrix
The table shows some of the MMPs most widely studied in relation to COPD, delineating their substrates in 3 columns. The first two list interstitial collagens and basement membrane components that the given MMP is capable of degrading. The third column indicates relative activity levels against elastin by the use of + and - signs, such that a single + indicates some activity, further + signs indicates higher activity and - no activity.

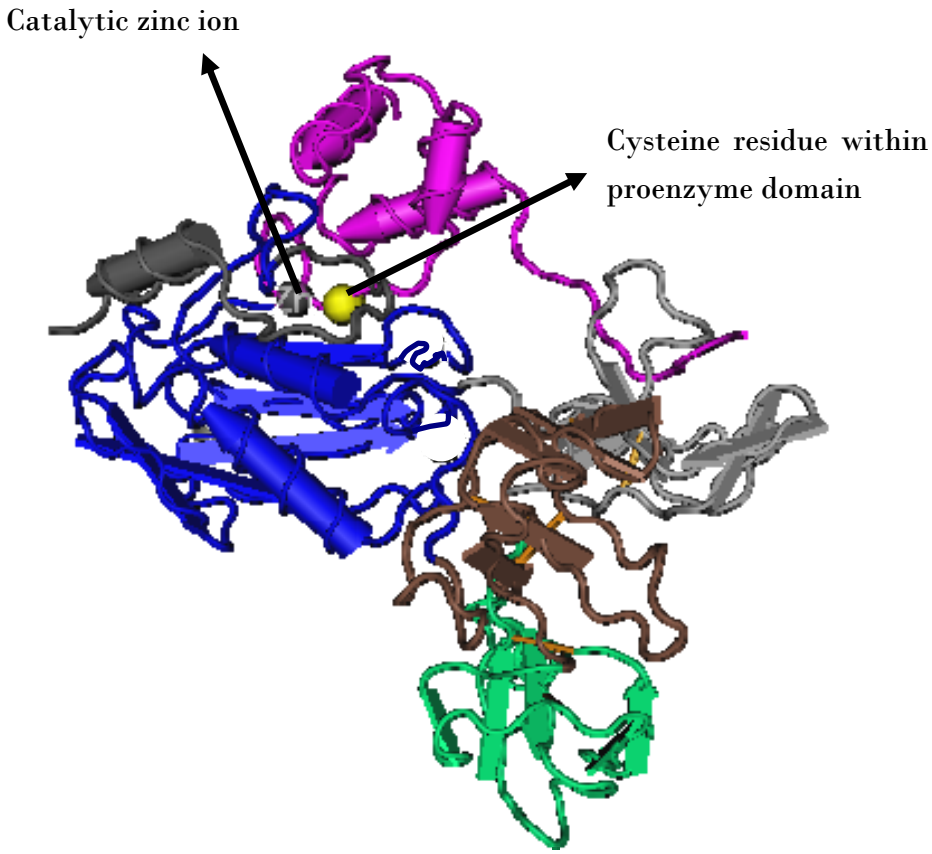


Figure 5.1: The structure of an MMP molecule

The picture shows the structure of MMP9, used as an example of the molecular structure of MMPs. All MMPs have a catalytic domain, highlighted here in pink and a proenzyme domain, shown in blue. Within these areas the zinc and cysteine residues that interact to keep the molecule in inactive conformation have been labelled.

MMP's are logical candidate genes for association with COPD, and have been the subject of a number of genetic studies, as outlined in Chapter 1 (section 1.1.4). *MMPs* have been mapped to chromosomes 11, 14, 16, 20 and 22 with several (including *MMP1*, *MMP3* and *MMP12*) being clustered on the long arm of chromosome 11[282]. This area was linked to both FEV1 and KCO in a European familial genetic association study in AATD[203]. For this reason this cluster of *MMP*s, rather than any other previously associated *MMP*, was chosen for the studies contained in this chapter.

5.1.1 MMP1, MMP3 and *MMP12*

MMP1, *MMP3* and *MMP12* lie adjacent to one another on chromosome 11q22-23. *MMP1* is approximately 17kb in size[283] and is comprised of 10 exons and 11 large introns. MMP1, also known as collagenase, is capable of breaking down type I, II and III collagen, which are ubiquitous in the connective tissues of the body, including the lung. *MMP3* is similar in structure to *MMP1*, such that their protein products share over half their amino acid sequences, suggesting a common evolutionary origin[284]. MMP3 is also known as stromelysin, and digests proteoglycan, laminin and type IV collagen[285] which also contribute to the extracellular matrix in the lung. *MMP12* is somewhat smaller, being 1.8kb in length, but codes for a similar sized protein (470 amino acids). It is known as macrophage metalloelastase, is produced by alveolar macrophages and is capable of degrading elastin[286].

In usual COPD a SNP in the promoter region of *MMP1* (rs1799750) was implicated in upper zone dominant emphysema[19] and in decline of FEV1[77], as detailed in chapter 1

(section 1.1.4). The rs1799750 SNP increases gene expression by creation of an extra transcription factor binding site (ETS)[287]. An animal model in which transgenic mice overexpress *MMP1,* results in the presence of emphysema [288]. The SNP associations seen must, however, be viewed with caution as neither has been replicated. The NETT genetics study[19] attempted replication of the association of rs1779750 with emphysema zone by examining both surgically and medically treated groups, but found no association in the latter. When the entire NETT cohort were compared to healthy subjects from the normative aging study no association with emphysema was seen, nor was there association in a family cohort with COPD[72]. No other studies have examined rs1799750 in relation to lung function decline. A SNP in *MMP12* (rs652438) was also implicated in lung function decline in the LHS, but only as part of a haplotype with rs1799750 (*MMP1*)[77]. Animal studies here show that *MMP12* null mice are protected against smoke related lung injury[289], thus making it a good candidate gene. Again, however, the association described in usual COPD has not been replicated, and should be viewed with caution. Important interactions of AAT and MMP12 have also been observed in *in vitro* studies, where smoke exposed macrophages fail to secrete MMP12 in the presence of AAT[202] providing further rationale for the study of this gene in subjects with AATD. Unlike *MMP1* and *MMP12, MMP3* had not previously been associated with COPD at the start of this project, nor are there are any specific animal models or *in vitro* work regarding its protein product in respiratory disease. However, *MMP3* lies between *MMP1* and *MMP12*, thus is still within the region of the linkage peak in AATD[203] and was worthy of further association analysis in AATD. Subsequent to the completion of this project a large genome wide study in usual COPD has reported association of a SNP within *MMP3*

(rs615098) with a quantitative emphysema phenotype [73].

5.2 Aims of this chapter

This chapter describes associations of *MMP1, MMP3* and *MMP12* with lung function and lung function decline in AATD, and forms a follow up study after the report of linkage of FEV1 and KCO to this region in AATD[203]. It, therefore, aims to provide initial evidence of replication of this association, and narrow the size of the block containing the associated variant(s) such that further fine mapping might be planned if appropriate.

5.3 Methods

The subjects studied have been described in Chapter 3 and comprise the unrelated case-control dataset only (n=424), since some members of the familial dataset had participated in the linkage study that reported an association which this study aims to replicate[203]. Specific phenotypes relating to lung function, lung function decline and emphysema zone were chosen because of prior evidence in AATD[203] and usual COPD[19, 77]. The group assessed for decline of lung function showed no difference in any demographic or clinical feature from the whole group (described in Chapter 3), except that they had smoked slightly more (median 18.75 v 14.00 pack years, p=0.030). Clinical characteristics of the genotyped decline group are shown in Table 5.2. Non-genetic influences upon decline of lung function have been assessed further and are described in Chapter 7, significant independent predictors being included in the regression models herein. Emphysema zone was analysed as a quantitative phenotype, defined as UZVI-LZVI in the presence of a visual appearance and densitometry consistent with emphysema.

Tag SNPs covering the 3 genes were chosen using HapMap, and genotyped using TaqMan® genotyping technologies, as described previously (Chapter 3, section 3.2.2). The location of tags is shown in Figure 5.2, and the context sequences for primer design in Table 5.3. One tag in *MMP12* (rs476185) failed design, probably due to a high GC content in the region. No other tags were available within the frequency limits (MAF>5%) thus this haplotype block could not be genotyped.

Haplotypes were identified in Haploview, and analyses of association with continuous outcome variables carried out in SPSS. This was done by coding each observed haplotype

as described in Chapter 2 (section 2.3.2) and using this code in the regression models for lung function and lung function decline.

	Decline group n=110
Gender (% male)	66.36
Age in years	51.48 (0.92)
Pack years	18.75 (9.01-28.49)
FEV1 %predicted	35.01(18.25-51.77)
FEV1/FVC	37.30(27.10-48.40)
KCO %predicted	68.09 (2.12)
UZVI	32.84 (1.69)
LZVI	52.10(38.53-65.68)
Decline of FEV1 in ml/yr	-42.00 (-87.00-3.00)
Decline of KCO in [mmol/min/kPa/l]/yr	-0.03 (0.004)

Table 5.2: Clinical characteristics of genotyped subjects for decline of lung function

The table shows demographic features, lung function and HRCT densitometry measures in the group assessed for decline of lung function. Data is shown as mean (SEM) or median (IQR) dependent on its distribution. FEV1 decline was non-normally distributed and required log transformation prior to subsequent linear regression modelling. KCO decline was normally distributed, and as such could be used in its raw form for subsequent regressions. The group showed no difference in any demographic or clinical feature from the whole cohort (Table 3.1), except that they had smoked slightly more ($p=0.03$).

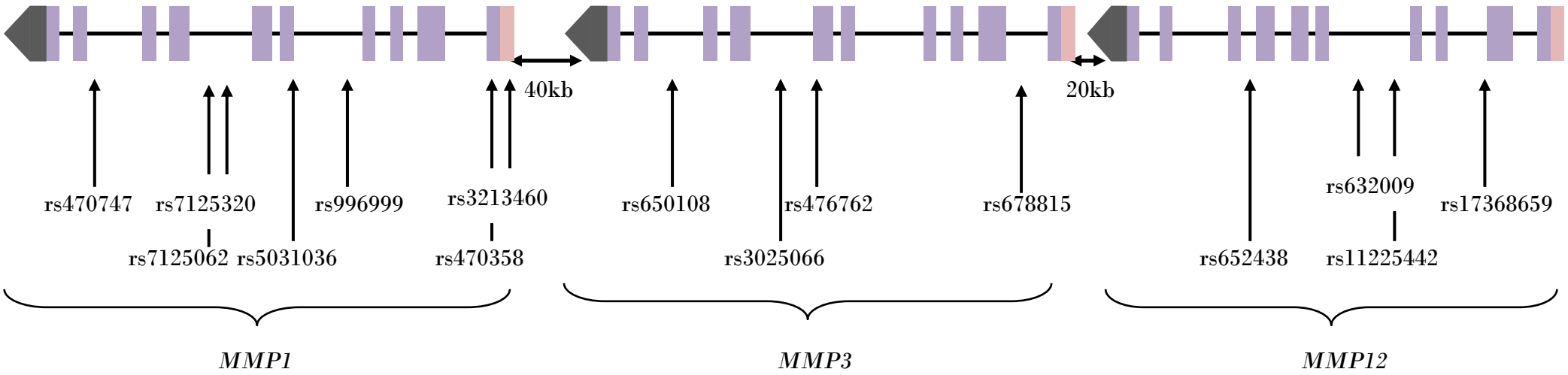


Figure 5.2: Location of tag SNPs in *MMP1*, *MMP3* and *MMP12*

The diagram shows the structure of the 3 *MMP*s on chromosome 11 chosen for further study. The mauve blocks represent exons, the promoter is represented by a grey pentagon, other regulatory regions by pink blocks and introns by a black line. All SNPs are denoted by their reference SNP number.

Gene	SNP	Context sequence	Design strand
MMP1	rs470747	TTTGCTAAACATGTGCACTCTGAAA[A/G]TCATTACAGAAAATACACAATGAGG	Reverse
	rs7125062	AAAAAAAGCAAGAAAAGTTTCCATC[C/T]AATTTCTGGTAATCTCTGGCAGATA	Forward
	rs7125320	ACCATTGTATAGCATACAAATAAAA[G/T]AGGCATAAGTAACTTTAAGAACATA	Reverse
	rs996999	GGAAAACACAAAAAGCCTTCCTTTT[C/T]TTTAATTTCCACATAAGTTTATGAC	Forward
	rs470358	ACTAAAAATTTACATTATTGTCACA[C/T]GCAATGCAGCATTTACCTGGACTAA	Forward
	rs3213460	AGTGATGGCTTCCCAGCCTCTTGCT[A/G]CTCCAATATCCCAGCTAGGAAGCTC	Forward
	rs5031036	TATCGCTAATGGCTGTTTNATTTGAAATACATATAAAGACCACCAGGCCCTTGTCCGTAA [T/C]GTTTTTCCCCATACTCACCATATATGGCTTGGATGCCATCAATGTCATCCTGAGCTAGC	Forward
MMP3	rs650108	GAGGTGACAAGTGGGTGAGGTTAGA[A/G]GTAGCCCTTGAAAATCATATAGTAG	Reverse
	rs3025066	TGTTTGTCATGTGCATATTGATCCA[C/T]CAGTAAGTAAATTACTTGATTAAAT	Reverse
	rs476762	AAATGTGGGAAAAGGCAGCACCAGA[A/T]CTAAATGATTCCAAGCTCAGAAGAA	Forward
	rs678815	ATTTCTCTAGCTTGCTGAAATAATG[C/G]CAAATTTTATAATATGATACTAGCA	Forward
MMP12	rs652438	GGGATAATTTGGCTCTGGTCTTAAA[C/T]TGCTAATTAACCAGTATTTGTCATC	Forward
	rs632009	CCTCATCTGTATAATGAAGATAATA[C/T]CTGTGTCACCTGCCACACAGGGTTC	Reverse
	rs11225442	GTCTTTTCGTAAGCCCAAATATAAG[A/G]CTATCATTAAGTTTCAGGTACTGTG	Forward
	rs17368659	TTACCACAATACTTCATTTTTGCTA[G/T]CACAATTGGAAAATGGATCATTTCA	Forward

Table 5.3: Context sequences for primer design in *MMP1*, *MMP3* and *MMP12*

The two possible alleles of each SNP are highlighted in blue between square brackets indicating the position of the SNP within the gene.

5.4 Results

Adequate genotype discrimination could not be achieved in AATD subjects for rs5031036 thus it could not be assessed further. No other tags were available for this block within the required allele frequency limits - that is, no other SNPs with MAF>5% were tagged by rs5031036. The genotyping success rate in the remaining SNPs was 92.5%. All SNPs were in HWE, except rs476762 (*MMP3*, $p<0.01$), which was excluded from further analyses. The allele and genotype frequencies of successfully genotyped SNPs are shown in Table 5.4.

In the dominant quantitative models of lung function, two SNPs (rs470358 and rs678815) were independent predictors of baseline KCO. The minor allele (T) of rs470358 in *MMP1* showed a protective effect in the regression model ($B=6.130$, $p=0.025$), the B value indicating that KCO is 6.13% higher in those carrying at least one T allele. The major allele (G) of rs678815 in *MMP3* was associated with a protective effect on KCO of similar magnitude ($B=7.766$, $p=0.025$). The B value from the regression shows that subjects who did not have the G allele exhibited KCO 7.77% lower on average than those with at least one copy of it. Mean KCO stratified for presence of these alleles is shown in Figure 5.3.

There were no genetic associations with FEV1, FEV1/FVC or emphysema zone (all $p>0.1$). In the models to assess decline no SNP genotype predicted FEV1 or KCO decline (all $p>0.1$).

One haplotype was observed in Haploview (Figure 5.4), which showed no associations (all $p>0.1$) with the quantitative phenotypes above.

Gene	Tag SNP	Genotype frequency % (n)			MAF (%)
		Major allele homozygote	Heterozygote	Minor allele homozygote	
MMP1	rs470747 [G,A]	40.55 (159)	50.39 (198)	9.06 (36)	34.25
	rs7125062 [T,C]	50.29 (197)	42.77 (168)	6.94 (28)	28.32
	rs7125320 [C,T]	83.75 (328)	14.95 (58)	1.27 (5)	8.82
	rs996999 [C,T]	64.65 (253)	30.21 (118)	5.14 (20)	20.24
	rs470358 [C,T]	34.40 (135)	50.71 (199)	14.89 (58)	40.25
	rs3213460 [G,A]	63.82 (250)	34.13 (134)	2.05 (8)	19.11
MMP3	rs650108 [A,G]	55.06 (216)	40.17 (157)	4.78 (19)	24.86
	rs3025066 [G,A]	85.52 (335)	13.45 (52)	1.03 (4)	7.76
	rs476762 [T,A]	72.97 (286)	19.77 (77)	7.27 (28)	17.15
	rs678815 [G,C]	27.80 (109)	52.12 (204)	20.08 (79)	46.14
MMP12	rs652438 [G,A]	87.89 (345)	12.11 (47)	-	6.06
	rs632009 [T,C]	44.97 (176)	42.59 (167)	12.43 (49)	33.73
	rs11225442 [G,A]	74.01 (290)	23.16 (91)	2.82 (11)	14.41
	rs17368659 [G,T]	77.38 (303)	21.31 (84)	1.31 (5)	11.97

Table 5.4: *MMP1*, *MMP3* and *MMP12* allele frequencies
The table shows the tag SNPs within the MMP cluster that were successfully genotyped. SNPs are listed as [major allele, minor allele], with the incidence of each genotype shown as a percentage, the raw data being in brackets, and the minor allele frequency (MAF) in the final column.

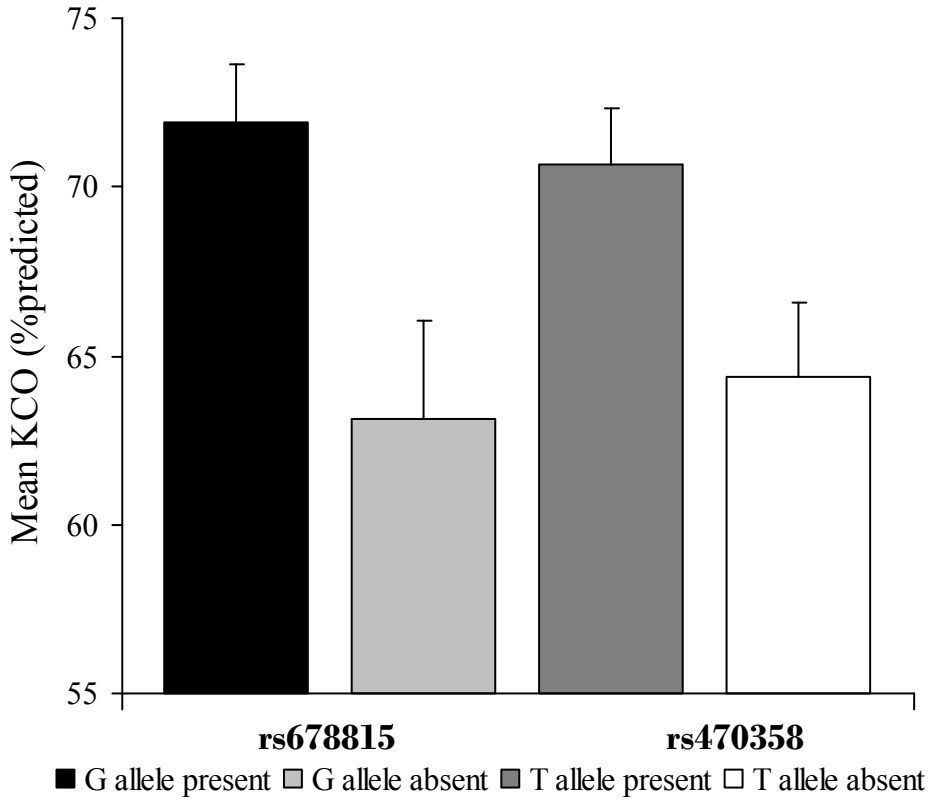


Figure 5.3: KCO stratified for associated *MMP* SNP genotype

The graph shows the mean and standard error of KCO from the regression models after adjustment for age, gender and smoking. The difference in KCO was significant for both SNPs, in each case p=0.025.

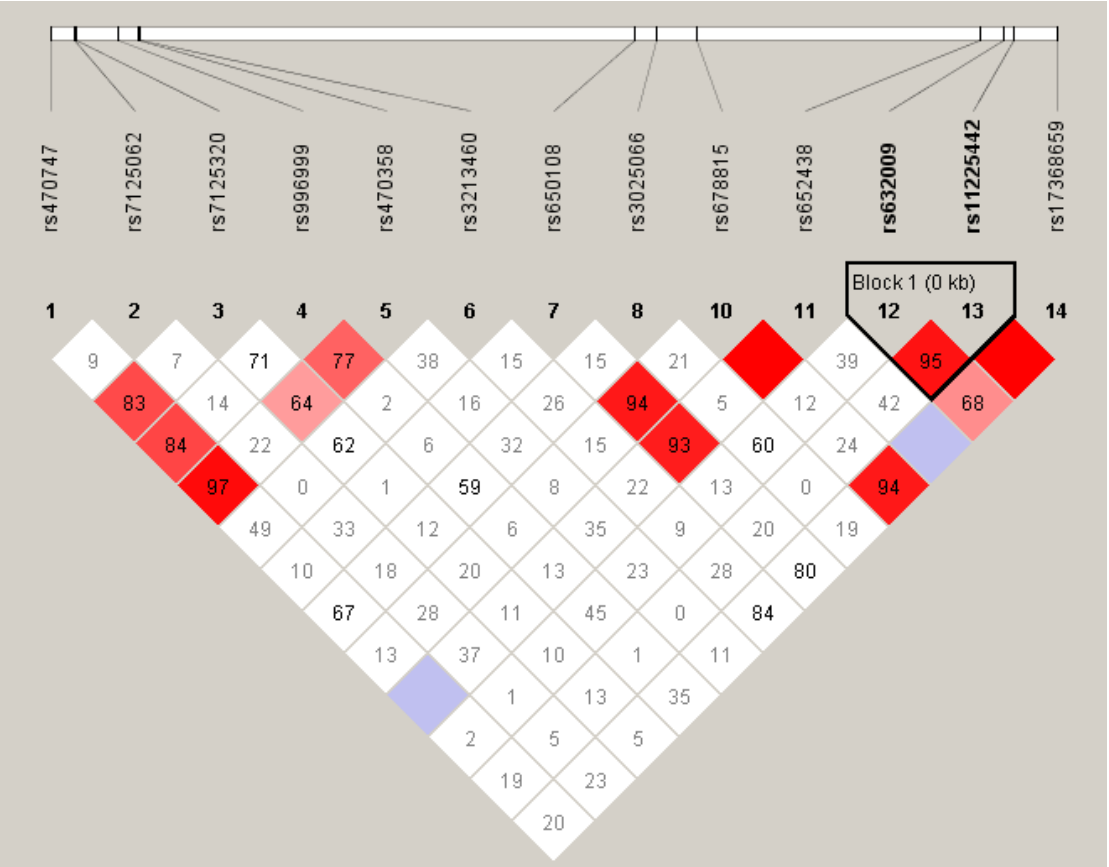


Figure 5.4: Linkage disequilibrium plot for *MMP1*, *MMP3* and *MMP12*

The LD plot shows the r2 values generated in Haploview. One haplotype block was identified, which did not show any significant phenotypic association.

5.5 Discussion

The main factor prompting the investigation of the 3 *MMPs* reported in this chapter was a linkage study in AATD, which suggested that the region of chromosome 11 containing them contained a protective variant[203]. Many of the siblings in the UK familial dataset participated in this study, which is why associations were not assessed in this group again. Two protective variants were identified in the unrelated dataset reported here; the T allele of rs470358 (*MMP1*) and the G allele of rs678815 (*MMP3*), both being associated with higher KCO. The results presented in this chapter, therefore, support prior work. Both SNPs are located in intronic regions of their respective genes and neither has an identified functional role.

Previous work in usual COPD implicated a SNP in the promoter region of *MMP1* (rs1799750) in UZDE[19]. In this study two tags cover the promoter region of *MMP1,* one of which is rs470358, though since rs1799750 allele frequencies were not reported in HapMap for this SNP the HapMap software did not report that it was tagged by rs470358. However, given that the MAF in previous studies has been greater than 5% it is likely that it is. Nevertheless we were unable to confirm association with emphysema zone, defined using the difference between LZVI and UZVI in an individual. This may have been because there were lower numbers of subjects in the CT based analyses since not all subjects underwent a local quantitative scan. In AATD emphysema zone relates to lung function with FEV1 tending to reflect LZDE better than UZDE and KCO the converse (i.e. reflecting UZDE better than LZDE)[18]. Thus the results for KCO are consistent with a tendency to protection from UZDE in those with the associated alleles.

This would also be in agreement with the association seen in NETT of rs1799750 with UZDE.

The results of this chapter did not show any associations within MMP12, but it should be noted that variation within *MMP12* alone has not usually been reported with any phenotype in either COPD or AATD previously, rather only as part of a haplotype. A SNP in *MMP12* (rs652438) was implicated in lung function decline in the LHS, but only as part of a haplotype with rs1799750[77]. Another study on two mixed European populations showed an enhanced risk of COPD in individuals with a haplotype of two *MMP12* SNPs (rs652438 and rs2276109)[164]. Our results do not show any haplotypes containing tags within *MMP1* and *MMP12* or any significant haplotypes within *MMP12*, though rs470747 and rs470358 approached significance as a pair (Figure 5.4). This might be because the population studied herein (although of the same racial group to the LHS, and indeed European) come from a different geographical region. The LHS was carried out in Denmark, and although the population there is predominantly Caucasian, as in the UK, the racial ancestry may differ. A similar difference might occur between a UK and a mixed mainland European population. This is because in recorded history the UK has been invaded by Danish Vikings, Nordic Vikings and Romans amongst others. Thus the UK population's genetic make up will reflect this[290], unlike Denmark, where the influences are different. HapMap exemplifies the differences in LD that occur between racial groups, though it has not studied racial ancestry within Caucasians. It is, therefore, possible that differences between the Caucasian cohorts might account for differences in haplotypes observed in the 2 populations. High degrees of geographical variation have been observed on both the long and short arm of chromosome 11 in a large genome-wide

study in the UK[217]. We might speculate that this is due to differences in the ancestry of populations residing in each area of the UK – for instance, areas in the South-East might show more markers of Danish Viking ancestry than those in the North-West, because of the site of Viking invasion in the South-East and relatively slow spread of the invading population North and West. This highlights the importance of geographical standardisation, as well as replication in different racial groups, when considering genetic associations with complex disease.

MMP3 had not previously been associated with COPD when this project started. However, recently a genome-wide study has reported association of a SNP (rs615098) very close to the *MMP3* promoter region with a quantitative emphysema phenotype[79]. Although the association did not replicate in the second population tested it remained significant at genome wide level in a meta-analysis of the two groups. This type of meta-analysis is relatively new and remains a little controversial in the field so will require further replication. There are no specific animal models or *in vitro* work regarding its protein product (MMP3) in COPD. A mouse model of inflammatory lung injury, performed by instilling a chemotactic cytokine into the trachea, showed less lung injury in MMP3 deficient mice than their normal counterparts[291]. This may indirectly support a role for the protein in inflammatory diseases of the lung, including COPD. A recent Japanese study identified a functional polymorphism in the *MMP3* promoter, associated with increased serum MMP3[292], supporting the existence of genetic variants that could change the risk of diseases where MMP3 plays a role. Elevated pulmonary levels of other MMPs have been observed in emphysema[279, 293] and may contribute to disease pathogenesis via increased proteolytic damage to the lung parenchyma. It is conceivable

that this is the same for MMP3, though the only study to report airway levels (in sputum) thus far found no difference between subjects with COPD and healthy controls[294]. However, the study was small and further investigation is warranted, which should include examining levels of the protein in airway biopsies, resected specimens of lung parenchyma, or airway secretions in subjects with and without COPD.

Genetic association implies a role for the protein product in pathogenesis, and hence a rationale for treatment. Our results suggest that treatments targeting MMPs might be appropriate in AATD. Endogenous anti-proteases (TIMPs) given to patients as a recombinant protein, or by gene therapy, and synthetic MMP inhibitors have been considered[295]. However, recombinant proteins are not yet cost effective and gene delivery is unlikely to produce enough protein to be effective[295]. A small study of a synthetic MMP inhibitor in asthma showed some promise in reducing pulmonary inflammation [34] but no trials have been conducted in COPD. Due to the associations of *MMP* genes with phenotype in COPD and AATD, such drugs might be of benefit in subjects with both conditions. However, the observation that pulmonary inflammation is more marked in AATD than usual COPD [234] might make the efficacy of the intervention more easily detectable in this group.

5.6 Conclusions

Variation within *MMP1* and *MMP3* is associated with KCO, a marker of parenchymal disease, in AATD. These results support a family based linkage study in AATD, and whilst not true replication, go some way to confirming the importance of the region.

Chapter 6: The role of HLA type

6.1 Introduction

Most previous work on COPD pathogenesis has concentrated on the role of proteases, oxidants and inflammation, as discussed in Chapter 1 (section 1.1.3). However, an autoimmune component to disease pathogenesis has also been proposed, in part because inflammation in the lungs persists after smoking cessation. This indicates a pathogenic process initiated by smoking but one that is subsequently self-perpetuating, which could be via an autoimmune process. Support for autoimmunity comes from the observation that a similar inflammatory cytokine and cellular profile ocuur in autoimmune conditions, such as RA, and COPD[296].

Characteristic features of autoimmune diseases include the presence of autoantibodies or autoreactive T lymphocytes, reproducibility in animals (or indeed in man) after transfer of either of these moieties and lymphocytic infiltration of the target organ[297]. These features are present in patients with COPD, and in particular in patients with emphysema. Elevated anti-elastin antibodies have been observed in the circulation of patients with emphysema[205] as well as autoantibodies to pulmonary epithelium[298] and the vascular endothelium[299]. Results from animal models also support an autoimmune basis to disease development: the transfer of pathogenic CD4+ T cells from rats with emphysema and anti-endothelial cell antibodies results in the development of emphysema in healthy animals [206] and mice injected with elastin peptides develop both anti-elastin antibodies and a bronchoalveolar lavage (BAL) fluid profile similar to that seen in COPD[279, 300]. In addition a lymphocytic infiltrate in the lung has been

reported, with pulmonary CD4+ and CD8+ T cell numbers associated with worsening disease [11, 301]. Furthermore, oligoclonal T cell populations have been demonstrated in the lung, suggesting accumulation due to antigenic stimulation [302, 303]. Thus far the most likely stimulus in COPD is an anti-elastin autoantibody[205, 300], but it remains possible that other antibodies against self antigens play a role, or that cigarette smoke itself is the main factor [303]. Finally, evidence of clustering of organ specific autoimmune diseases within individuals with COPD unrelated to smoking[304] may suggest a common genetic basis to disease development.

A number of loci exhibit association with autoimmune diseases, many of the most well replicated of which are in the HLA region[217]. The HLA region is located on chromosome 6p21, and is one of the most gene dense, polymorphic and LD rich regions in the human genome [305, 306]. It can be subdivided into class I, II and III regions (Figure 3.3). Molecules encoded by class I genes are involved in presentation of endogenous antigens, as well as those derived by intracellular processing of viruses, to CD8+ T cells, whilst those encoded by class II present exogenous antigens to CD4+ T cells[306]. The class III region contains other immune response genes, such as *TNFA*. Autoimmune diseases, such as type 1 diabetes have a complex pattern of HLA association, the strongest evidence being for class II alleles, with strong predisposing effects being seen for *HLA-DRB1**04[307]. Associations with HLA class I and III are also well known, though though there is significant LD between these areas and class II so it can be difficult to ascertain if a single locus or combination of these is responsible for disease association [308, 309]. Attempts have been made to dissect these effects in other autoimmune diseases. In diabetes, for instance, both HLA class I and class II alleles are associated with

disease but analysis shows that after taking account of HLA class II effects the influence upon disease from class I is smaller[310]. This suggests that HLA class II is the most important part of the region in autoimmune disease.

Genetic association of COPD with the HLA region, specifically HLA class I, has been reported in two studies [204, 311]. However, these studies considered different sub-phenotypes of COPD - FEV1[204] and chronic bronchitis[311] and results lacked consistency. Polymorphisms in *TNFA,* part of the HLA class III region, have also been associated with COPD phenotypes [105, 107], but association with class II has not been reported, despite the high degree of LD between HLA class II and *TNFA*[312]. The association with *TNFA*, togther with the evidence of autoimmunity described above, suggests that investigation of genetic association studies of HLA class II in COPD may be fruitful. Furthermore, cells from subjects with COPD exhibit an attenuated inflammatory response when blocking HLA class II *in vitro* suggesting an HLA class II dependent inflammatory pathway[205]. This leads to the hypothesis that HLA class II type might influence the development of emphysema.

6.2 Aims of this chapter

This chapter compares HLA class II genotype frequencies between AATD patients with and without emphysema, and assesses HLA class II association with lung function. It also describes phenotypic associations of anti-elastin antibodies in AATD, and assesses differences in anti-elastin antibody levels between genotypes with and without associated HLA class II alleles. This provides an indication of whether anti-elastin mediated pathways are dependent on or unrelated to genetic factors in the HLA class II region.

6.3 Methods

6.3.1 Subjects

Subjects were taken from the cohort previously described (Chapters 2 and 3)and comprised all subjects whose DNA samples (extracted by DNA precipitation, as previously described) exceeded 100ng/μl and was of sufficient volume to allow HLA class II typing. HLA phototyping requires DNA at this concentration, unlike TaqMan® based genotyping methods, which work well at lower DNA concentrations (down to 4ng/μl). This resulted in a total of 322 subjects suitable for study.

The characteristics of the subjects are shown in Table 6.1.

	Whole group n=322	Emphysema n=219	No emphysema n=48
Age in years	50.86 (0.74)	52.33 (0.77)	41.19 (33.53-48.86)
% of group of male gender	58.19	63.24	43.18
Pack years	13.75 (0.03-24.00)	16.00 (5.47-26.53)	0 (0-4.38)
FEV1 %predicted	36.45 (23.56-60.00)	32.44 (19.93-44.95)	102.44 (80.81-124.04)
FEV1/FVC	29.33 (56.63)	34.75 (26.33-43.18)	72.30 (2.32)
KCO %predicted	66.10 (54.60-85.95)	63.65 (51.95-73.35)	98.18 (3.11)
UZVI	31.75 (15.60-45.93)	33.00 (18.80-47.20)	2.61 (0.44)
LZVI	49.65 (30.62-61.90)	53.00 (39.93-66.06)	6.99 (0.66)

Table 6.1: Characteristics of subjects studied for HLA class II type

Data is shown as mean (SEM) or median (IQR) dependent on its distribution.

Age, gender, smoke exposure, lung function and CT densitometry differed significantly between emphysematous and non-emphysematous subjects (all $p<0.0001$ except gender $p=0.012$). Only 82% of subjects had a CT scan to diagnose emphysema, hence the number listed for the whole group exceeds the sum of the numbers of subjects with and without emphysema.

6.3.2 HLA class II typing

This was performed primarily by the phototyping method[216], with additional typing using commercial kits for samples which failed by this method. All the methods used contain internal control bands.

Phototyping

This technique uses PCR with sequence specific primers to identify class II alleles at a resolution equivalent to serological typing. Forty two reactions are required to assign a class II type for *DRB1, DQB1* and *DQA1*, the results grouping into 13 *DRB1* types, 8 *DQB1* and 8 *DQA1* types. The pattern of bands visible on subsequent gel electrophoresis identifies the HLA type because each allele will generate bands of a known size in specific reactions, as shown in Table 6.3. Two control bands are also seen, showing that the PCR was successful - the PCR product in this band is an amplified part of intron 3 of *DRB1* and a dimer derived from it. Where the control bands were not seen it suggested that amplification had not occurred and it was repeated. All primers have a melting point between 58-62°C, to enable them to be run in the same thermal cycling conditions (see section 2.2.3). The primers used are listed in Table 6.3.

In the PCR reactions required for HLA typing many different alleles are being sought, and a large number of primers are used in different combinations to amplify them[216]. The combinations were therefore mixed in advance and stored. Primers were purchased as crude nucleotides (Sigma, UK) and the amount required for dilution calculated using the following formulae

$$\text{Primer amount } (\mu M) = \frac{\text{Primer OD x 100}}{1.5A + 0.71C + 1.2G + 0.84T}$$

A, C, G and T are the number of times adenine, cytosine, thymine and guanine respectively appear in the primer sequence.

$$\text{Primer amount } (\mu l/ml) = \frac{\text{Required working primer concentration x 1000}}{\text{Primer amount } (\mu M)}$$

The amount of primer was calculated using the first formula, and the primer dilutions were then made at the required concentration[216] using the second formula to indicate the total volume of primer(s) required, the amount being made up to 1ml with MillQ water for later use. Primer plates were aliquoted in batches, using 5µl of the primer mixes and then 10µl mineral oil (Sigma, UK) over this in each well (to prevent evaporation during PCR), as shown in Figure 6.1, such that two subjects could be typed per 96 well plate.

For each DNA sample to be amplified 300µl TDMH (Appendix 3), 169µl MillQ water, 9µl DNA and 2.25µl *Taq* polymerase (Bioline, UK) were mixed in a 1.5ml eppendorf (Alpha labs, UK) by vortexing, and 8µl of this mixture added to each of the 42 wells needed for a the chosen HLA class II loci. This was then sealed with a transparent lid (Greiner bio-one, UK) and centrifuged at 1000rpm for 1 minute prior to cycling under the following conditions:

96°C 1 minute

96°C 20 seconds
70°C 45 seconds
72°C 25 seconds
} 5 cycles

96°C	25 seconds	21 cycles
65°C	50 seconds	
72°C	30 seconds	
96°C	30 seconds	4 cycles
55°C	60 seconds	
72°C	90 seconds	
20°C	5 minutes	

Thirteen µl of PCR product per well were run on a 1% agarose gel made with 0.5xTBE and 0.024µl ethidium bromide/ml (see also section 2.2.4). Each gel was produced from 500ml of TBE and 0.5g of agarose and allowed 6 individuals to be typed at the 3 loci being studied in this chapter and was electophoresed for 15 minutes at 350 Volts, 300 Amps and 150 Watts. Gels were photographed using UV light in a 'G-box' (Syngene, UK) and assigned a type accordingly (Appendix 6).

	Alleles amplified	Forward primer/s (5'-3')	Reverse primer/s (5'-3')	Size (bp)
DRB1	0101/2/4	TTGTGGCAGCTTAAGTTTGAA	CCGCCTCTGCTCCAGGAG	194
	0103	TTGTGGCAGCTTAAGTTTGAA	CCCGCTCGTCTTCCAGGAT	195
	1501-5	TCCTGTGGCAGCCTAAGAG	CCACCGCGGCCCGCGC	206
	1601-6	TCCTGTGGCAGCCTAAGAG	CTCCGTCACCGCCCGGT	137
	03011/12/2/3/4/5, 1107	GTTTCTTGGAGTACTCTACGTC	GTCCACCCGGCCCCGCT	211
	0301/4	GACGGAGCGGGTGCGGTA	CTGCACTGTGAAGCTCTCCA	216
	0302/5, 1302/3, 1109/20, 1402/3/9/13/19	TACTTCCATAACCAGGAGGAGA	CTGCACTGTGAAGCTCTCAC	188
	0401-22, 1410, 1122	GTTTCTTGGAGCAGGTTAAACA	(i)CTGCACTGTGAAGCTCTCAC (ii)CTGCACTGTGAAGCTCTCCA	259
	0701	CCTGTGGCAGGGTAAGTATA	CCCGTAGTTGTGTCTGCACAC	231
	0801-11, 1415	AGTACTCTACGGGTGAGTGTT	(i)CTGCAGTAGGTGTCCACCAG (ii)TGTTCCAGTACTCGGCGCT	162 213
	0901	GTTTCTTGAAGCAGGATAAGTTT	CCCGTAGTTGTGTCTGCACAC	235
	1001	CGGTTGCTGGAAAGACGCG	CTGCACTGTGAAGCTCTCAC	203
	1101-21	(i)GTTTCTTGGAGTACTCTACGTC (ii)TTCTTGGAGTACTCTACGGG	CTGGCTGTTCCAGTCCTCCT	174 176
	1201/2/3	AGTACTCTACGGGTGAGTGTT	CTGTTCCAGGACTCGGCGA	163
	1301/2/4/8/15-17/19/20/22, 1416, 1102/3/11/14/16/20/21	(i)GTTTCTTGGAGTACTCTACGT (ii)TTCTTGGAGTACTCTACGGG	TCCACCGCGGCCCGCTC	210
	1301/2/4/8/9/16/20, 1116/20, 1416	GTTCCTGGACAGATACTTCC	(i)TCCACCGCGGCCCGCTC (ii)CCACCGCGGCCGCGC	152
	1301/2/10/15/16, 1116/20, 1419/20	TACTTCCATAACCAGGAGGAGA	(i)CCCGCTCGTCTTCCAGGAT (ii)TCCACCGCGGCCCGCTT	129 138
	1305/6/10, 1109, 1419/21	TACTTCCATAACCAGGAGGAGA	(i)ACCGCGGCCCGCCTGCG (ii)TCCACCGCGGCCCGCTT	136 138
	1303/10, 1419/21	GTTTCTTGGAGTACTCTACGTC	TCCACCGCGGCCCGCTT	210
	1303/4/12/21, 1413	GTTTCTTGGAGTACTCTACGTC	TGTTCCAGTACTCGGCCT	170

	1305/7/11/14/18/21, 1101/4/6/9/10/(11)/12/15	GTTTCTTGGAGTACTCTACGTC	CCCGCCTGTCTTCCAGGAA	200
	1301/2/5-9/10/11/14-16/18/19/20/22, 1402/3/6/9/12/14/17/19-21, 0301-5	GTTTCTTGGAGTACTCTACGTC	CTGTTCCAGTACTCGGCATC	171
	1401/3-5/7/8/10-12/14/15/18, 1318, 0809, 1117	GTTCCTGGACAGATACTTCC GTTCCTGGAGAGATACTTCC	(i)TCTGCAATAGGTGTCCACCT (ii)CTGCAGTAGGTGTCCACCAG	164 165
	1401/7/16	GTTTCTTGGAGTACTCTACGTC	CTGTTCCAGTGCTCCGCAG	171
	1402-3/6/9/12/13/17, 1318, (1418/19/21), (0301/2/3/5)	TACTTCCATAACCAGGAGGAGA	(i)CTGCAGTAGGTGTCCACCAG (ii)CCGCCTCTGCTCCAGGAG	129 151
DQB1	0305	GCTACTTCACCAACGGGACC	TGCACACCGTGTCCAACT	195
	02	GTGCGTCTTGTGAGCAGAAG	CGTGCGGAGCTCCAACTG	198
	04	GCTACTTCACCAACGGGACC	TGGTAGTTGTGTCTGCATACG	208
	05	ACGGAGCGCGTGCGGGG	CCCGCGGTACGCCACCTC	207
	0601-3	(i)TTTCGTGCTCCAGTTTAAGGC (ii)GACGTGGGGGTGTACCGC	CCGCGGAACGCCACCTC	249 140
	0603-9	GGAGCGCGTGCGTCTTGTA	(i)TGCACACCGTGTCCAACTC (ii)TGCACACCCTGTCCACCG	176
	0301/4	GACGGAGCGCGTGCGTTA	CGTGCGGAGCTCCAACTG	207
	0302	GTGCGTCTTGTGACCAGATA	CTGTTCCAGTACTCGGCG	119
	03032	GACGGAGCGCGTGCGTCT	CTGTTCCAGTACTCGGCGT	129
DQA1	0101/4	CATGAATTTGATGGAGATGAGG	ATGATGTTCAAGTTGTGTTTTGC	149
	0102/3	CATGAATTTGATGGAGATGAGC	ATGATGTTCAAGTTGTGTTTTGC	149
	0201	ACGGTCCCTCTGGCCAGTT	CAGGATGTTCAAGTTATGTTTTAG	173
	03011/12	TTCACTCGTCAGCTGACCAT	CAAATTGCGGGTCAAATCTTCT	183
	0302	TTCACTCGTCAGCTGACCAC	CAAATTGCGGGTCAAATCTTCT	183
	0401	ACCCATGAATTTGATGGAGAC	CACATACCATTGGTAGCAGCA	194
	0501/2	ACGGTGCCTCTGGCCAGTA	AGTTGGAGCGTTTAATCAGAC	189

	0601	ACGGTCCCTCTGGCCAGTT	GGTCAAATCTAAATTGTCTGAGA	191
Control	Part of intron 3 of *DRB1*	TGCCAAGTGGAGCACCCAA	GCATCTTGCTCTGTGCAGAT	796

Table 6.2: Primers used for HLA class II typing

All primers were used at a concentration of 3.4mM. Where more than one primer was used in a mix this is denoted by (i) for the first primer (ii) for the second.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0101/2/4	0701	1301/2/10/15/16, 1116/20, 1419/20	1402-3/6/9/12/13/17, 1318	02	0101/4	A1	A2	A3	A4	A5	A6
B	0103	0801-11, 1415	1305/6/10, 1109, 1419/21		04	0102/3	B1	B2	B3	B4	B5	B6
C	1501-5	0901	1303/10, 1419/21		05	0201	C1	C2	C3	C4	C5	C6
D	1601-6	1001	1303/4/12/21, 1413		0601-3	03011/12	D1	D2	D3	D4	D5	D6
E	03011/12/2/3/4/5, 1107	1101-21	1305/7/11/14/18/21, 1101/4/6/9/10/(11)/12/15		0603-9	0302	E1	E2	E3	E4	E5	E6
F	0301/4	1201/2/3	1301/2/5-9/10/11/14-16/18/19/20/22, 1402/3/6/9/12/14/17/19-21, 0301-5		0301/4	0401	F1	F2	F3	F4	F5	F6
G	0302/5, 1302/3, 1109/20, 1402/3/9/13/19	1301/2/4/8/15-17/19/20/22, 1416, 1102/3/11/14/16/20/21	1401/3-5/7/8/10-12/14/15/18, 1318, 0809, 1117		0302	0501/2	G1	G2	G3	G4	G5	G6
H	0401-22, 1410, 1122	1301/2/4/8/9/16/20, 1116/20, 1416	1401/7/16	0305	03032	0601	H1	H2	H3	H4	H5	H6

Figure 6.1: Primer locations in 96 well plate for HLA Class II typing

The diagram shows the positions of primers for alleles at the 3 loci within HLA class II; *DRB1* alleles are in the grey boxes, *DQB1* in pink and *DQA1* in blue. Setting up in this way allowed repetition of the primer pattern in the second half of the plate (columns 7-12), as indicated by the well locations listed from columns 1-6 in these columns, such that 2 subjects could be typed on each plate.

Dynal

It is possible for phototyping to fail at any of the 3 class II loci if the allele present is not one adequately identified by the primer mixes, or if control bands cannot be adequately visualised in some reactions, despite repeated attempts. This necessitates more high resolution typing methods with different primer mixes, such as those made by commercial providers such as Dynal. In the studies reported here no phototyping failed at the *DQA1* locus, thus Dynal was only necessary for *DRB1* and *DQB1*.

DRB1

Samples where a *DRB1* type could not be allocated by phototyping were analysed using a *DRB1* specific high resolution PCR kit (SSP UniTray, Dynal, Invitrogen, UK). A *DRB1* type is assigned using 96 reactions, the primers for which are pre-aliquoted into plates by the suppliers. For each DNA sample to be amplified 1 vial (580µl) of the PCR buffer provided in the kit, 9.3µl *Taq* polymerase (Bioline, UK), 268µl MillQ water and 80µl DNA (at a concentration of 75-125ng/µl) were mixed in a sterile 1.5ml eppendorf (Alpha labs, UK), with 8µl of the mixture being aliquoted into each well of the plate. Ten µl of mineral oil (Sigma, UK) was added to each well, prior to sealing with a transparent lid (Greiner Bio-one, UK), centrifuging at 1000rpm for 1 minute and thermal cycling as follows:

96°C 1 minute

96°C 25 seconds
70°C 50 seconds
} 5 cycles

72°C 25 seconds

96°C 25 seconds
65°C 50 seconds
72°C 45 seconds
} 21 cycles

96°C 25 seconds
55°C 60 seconds
72°C 120 seconds
} 4 cycles

Thirteen µl of each PCR product was run on a 1% agarose gel as before (sections 2.2.4 and 6.3.2, phototyping). Gels were photographed and applied to a worksheet provided with the kit (Appendix 6) and results interpreted according to the manufacturer's instructions.

DQB1

Samples where a *DQB1* type could not be allocated by phototyping were analysed using specific low resolution PCR kits (AllSet, Dynal, Invitrogen, UK). A strip of 8 PCR tubes was used, containing primers aliquoted by the suppliers. Fewer primers were required than for *DRB1* typing because the *DQB1* locus exbibits fewer alleles[313]. For each DNA sample to be amplified 19.0µl of DNA (at 50ng/µl), 80µl of the provided PCR buffer and 0.80µl *Taq* polymerase (Bioline, UK) were mixed, and 10µl dispensed into each primer tube, followed by 10µl of mineral oil (Sigma, UK), sealing with a strip of lids provided, centrifuging at 1000rpm for 1 minute and thermal cycling as follows:

96°C 90 seconds

96°C 15 seconds
} 10 cycles

65°C 1 minute

96°C 10 seconds
61°C 50 seconds
72°C 30 seconds
} 20 cycles

Samples were run on a 1% agarose gel using the same volumes and conditions as previously, photographed, applied to a worksheet (Appendix 6) and interpreted according to the manufacturer's instructions.

6.3.3 Statistical analysis of genetic association

Hardy-Weinberg equilibrium was checked for the *HLA-DRB1, HLA-DQB1* and *HLA-DQA1* loci using an add-on package to the stata programme (available from www.gene.cimr.cam.ac.uk/clayton/software/stata). Linear regression models were constructed using FEV1, FEV/FVC and KCO, adjusting for smoke exposure, gender and age where these were significant predictors of the outcome variable. A logistic regression model to compare subjects with and without emphysema was constructed in a similar manner. Since not all subjects had a CT scan the number included in the lung function models was greater than in the emphysema model (n=306 v 248). The regression models initially included all the alleles for a given locus to obtain a p-value for the association of the locus. For each locus the model with covariates and genotypes was compared to a model with just covariates via a likelihood ratio test. The alleles were then categorised within the regression to obtain individual allelic associations, in each case setting the most frequent allele as the comparator against which the likelihood ratio was calculated. In

each model alleles were grouped at the 1% frequency level, such that *HLA-DRB1**09 and *HLA-DRB1**10 were combined.

6.3.4 Anti-elastin antibody ELISA

Anti-elastin antibodies were measured in 172 stored serum samples, prepared by collection of blood from subjects at the time of their baseline clinical assessment in plain tubes (Greiner Bio-one, UK), which was allowed to clot prior to centrifuging at 3000rpm for 15 minutes. The resultant supernatant was stored in aliquots of 200µl at -80°C in a freezer. A smaller number of subjects had serum stored at their baseline assessment than those that had DNA stored, such that the number of samples genotyped exceeded the number on which it was possible to measure the autoantibody. The method for this assay was adapted from that published by Lee *et al*[205].

Soluble human lung elastin (Sigma, UK) was diluted to 25µg/ml in a 0.01% solution of bovine soluble albumin (BSA; Sigma, UK) in phosphate buffered saline (PBS, Appendix 3). The addition of BSA to the solution aided solubilisation of the elastin. Forty µl of this solution was added to each of 96 wells of a maxisorb plate (Nunc, UK), which allow optimal binding of protein to the plate surface. The outer wells of the plate (i.e. the first and last column and row) were not used because of the risk of excess evaporation from these during subsequent 37°C incubations, which could confound results. The elastin was left to bind to the plate by incubation overnight at 4°C. The following morning a 0.01% solution of BSA (prepared as before) was added to the plate, and incubated at 37°C for 2 hours. The plate was then washed using PBS with 0.05% Tween (Sigma, UK) to remove

unbound elastin, using a 3 wash cycle with 400µl of the washing solution per well.

Serum samples were diluted in PBS to a 1:500 dilution. One hundred µl of each sample, run in duplicate as with previous ELISAs (section 2.2.5), was added to the wells of each plate. Plates were then incubated at 37°C for 2 hours. During this stage any anti-elastin antibodies present, bound to the elastin on the plate. Plates were then washed as before, to remove any unbound substances, prior to addition of 40µl of anti-human IgG (Sigma, UK) diluted to 1:40000 in PBS, and a further incubation at 37°C for 2 hours. After another wash 200µl of tetramethylbenzidine (Sigma, UK) was added to each well to allow the colour change necessary to detect the antibody level (Section 2.2.5), and the plate incubated at room temperature for 20 minutes. The reaction was stopped by addition of sulphuric acid, prior to reading at 490nm, with wavelength correction set to 570nm, using a standard microplate reader (Biotek plate reader with Gen5 software, both Northstar Scientific, UK).

The standard curve was generated by serial dilution of a pooled plasma sample from healthy volunteers from 1:100 to 1:1600. The patient samples could then be measured relative to the top standard, which was given an arbitrary value of 100 international units.

The ELISA was validated in three ways. Firstly the coefficient of variation was calculated using a plate made up only of standards at the given concentration, such that there were 10 repeats at each concentration. This step ensured that the assay was performing consistently, giving the same result for each sample of the same concentration. Secondly a check of the sensitivity of the assay to small changes in concentration was performed using a plate set up with 4 columns made up using 50µl standard plus 50µl of

the 1:800 dilution in each well of the first column, and 50µl standard plus 50µl of the 1:1600 dilution in each well of the next 2 columns. These columns allowed calculation of spike recovery, which is the difference between observed and expected concentration expressed as a percentage of the expected value. Finally, to ensure assay specificity for anti-elastin antibodies, 2 columns were set up using a 1:40000 dilution of an anti-CD11 antibody, rather than the anti-human IgG. This step confirmed assay specificity for anti-human IgG antibodies to the elastin bound to the plate.

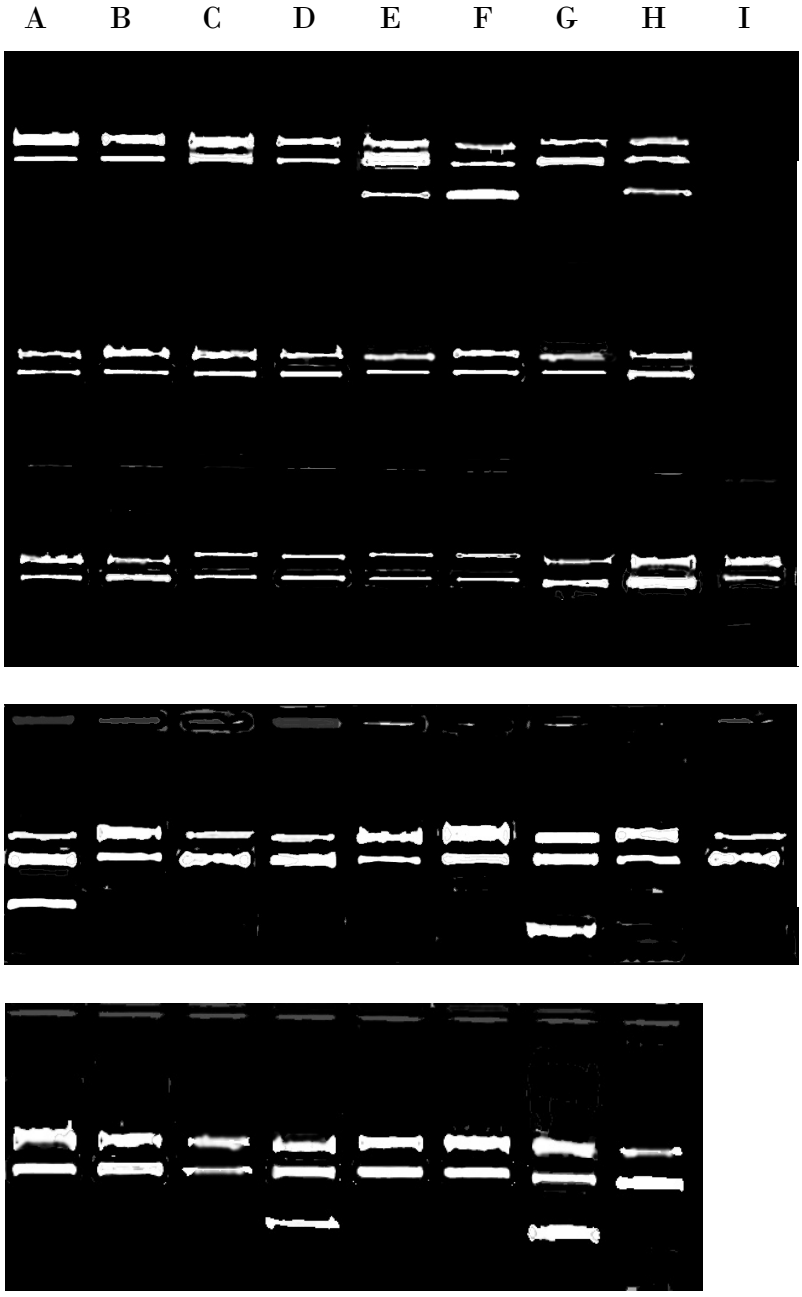
6.3.5 Statistical analysis of anti-elastin antibody levels

Anti-elastin antibody levels were compared to clinical variables using standard statistical methods outlined in chapter 2, prior to comparison between genotypes at associated loci using the Kruskal-Wallis test.

6.4 Results

6.4.1 HLA class II association with clinical features

Genotyping was successful in 95.3% of cases. Examples of the gels obtained by phototyping and Dynal methods are shown in Figures 6.2-6.4. Allele frequencies are shown in Table 6.3. The *HLA-DRB1*, *HLA-DQB1* and *HLA-DQA1* loci were in Hardy-Weinberg equilibrium in both emphysematous and non-emphysematous patients (all $p>0.05$).



(a) *DRB1* typing
25 sets of primers are used for this locus. This example shows bands below the controls in lanes E, F and H on the top row, the first two corresponding to *DRB1**03 and the last to *DRB1**04.

(b) *DQB1* typing
Bands are seen in lanes A and G corresponding to *DQB1**02 and *DQB1**0302 respectively.

(c) *DQA1* typing
Bands are seen in lanes D and G corresponding to *DQA1**0301 and *DQA1**0501 respectively.

Figure 6.2: Examples of gels from HLA phototyping
Forty two wells are required for one subject to obtain a class II type at the *DRB1, DQB1* and *DQA1* loci by this method. This series of pictures shows the gel as it runs from the well at the top of each row in the picture towards the bottom. Larger PCR products move more slowly through the gel, thus the control band appears first since it is much bigger than those that identify alleles (see Table 6.3 and Appendix 6 for further information). The above example, from a single subject, shows the following classical haplotypes.

*DRB1**03 *DQB1**02 *DQA1**0501

*DRB1**0401-22 *DQB1**0302 *DQA1**03011/12

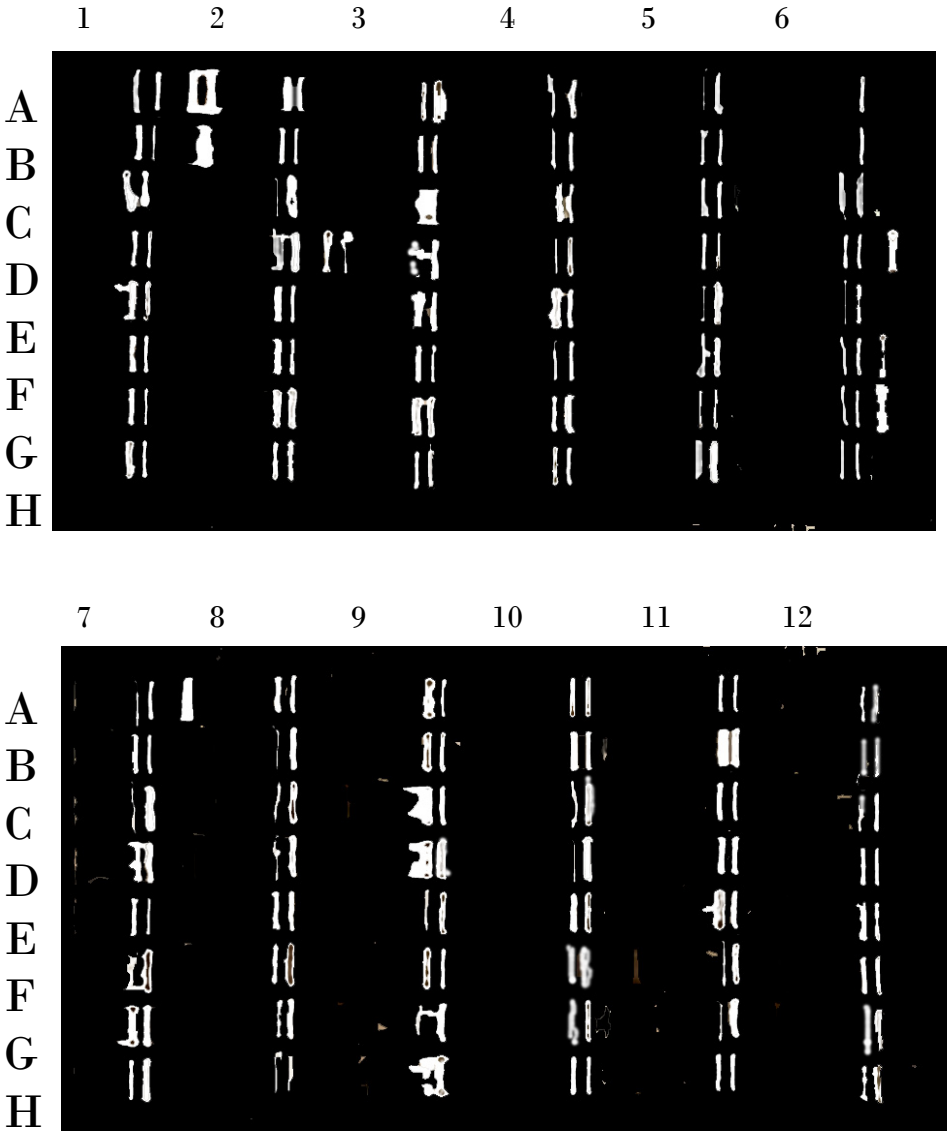


Figure 6.3: Examples of gels from high resolution *DRB1* typing
This requires 96 wells to obtain a single subject's *DRB1* type. The largest 1-2 bands in each lane are the control bands. The typing sheet (appendix 6) indicates the well numbers and sizes for each type. The resultant type is *DRB1**010101-03 and *DRB1**040101/02. This kit also identifies alleles at the *DRB3*, *DRB4* and *DRB5* loci, which were not required for the studies herein.

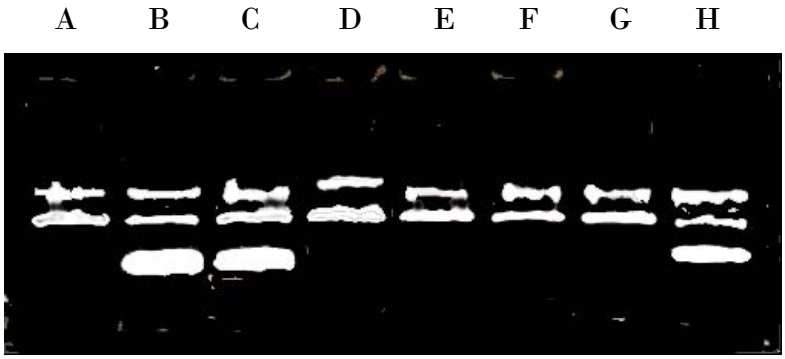


Figure 6.4: Example gel from low resolution *DQB1* typing
This requires 8 wells, similar to phototyping, to obtain a *DQB1* type, but the required DNA concentration is lower. The above example shows bands in lanes B, C and H, corresponding to *DQB1**050101-0504 and *030302-3/0306/0315/0317, as indicated by the interpretation sheets in Appendix 6.

Locus	Allele	All patients	Emphysema	No emphysema
HLA-DRB1	*01	12.8 (78)	13.2 (55)	15.6 (14)
	*03	17.0 (104)	16.7 (70)	20.0 (18)
	*04	16.6 (101)	15.8 (66)	21.1 (19)
	*07	12.8 (78)	12.0 (50)	13.3 (12)
	*08	2.6 (16)	2.9 (12)	1.1 (1)
	*09	0.2 (1)	0.2 (1)	0
	*10	0.8 (5)	0.7 (3)	0
	*11	6.4 (39)	6.7 (28)	5.6 (5)
	*12	2.3 (14)	2.2 (9)	3.3 (3)
	*13	12.0 (73)	13.6 (57)	5.6 (5)
	*14	3.6 (22)	3.1 (13)	2.2 (2)
	*15	11.8 (72)	12.0 (50)	12.2 (11)
	*16	1.1 (7)	1.0 (4)	0
HLA-DQB1	*02	20.0 (122)	18.4 (77)	24.4 (22)
	*0301	14.6 (89)	15.1 (63)	16.7 (15)
	*0302	9.0 (55)	8.4 (35)	11.1 (10)
	*03032	4.6 (28)	4.5 (19)	3.3 (3)
	*04	3.0 (18)	3.3 (14)	1.1 (1)
	*05	19.3 (118)	18.9 (79)	17.8 (16)
	*0601-9	29.5 (180)	31.3 (131)	25.6 (23)
HLA-DQA1	*0101	16.4 (100)	16.5 (69)	15.6 (14)
	*0102/3	23.3 (142)	24.4 (102)	17.8 (16)
	*0201	10.7 (65)	10.0 (42)	8.9 (8)
	*0301	19.4 (118)	18.7 (78)	23.3 (21)
	*0401	2.5 (15)	2.6 (11)	1.1 (1)
	*0501	27.2 (166)	27.3 (114)	33.3 (30)
	*0601	0.6 (4)	0.5 (2)	0

Table 6.3: HLA Class II allele frequencies according to respiratory phenotype
The table shows the allele frequency of each allele within the *HLA-DRB1*, *HLA-DQB1* and *HLA-DQA1* loci, shown as a percentage with the raw numbers in parentheses.

No locus associations were observed with emphysema, however pulmonary function showed association at the *HLA-DQA1* locus with KCO (p=0.034) and trends towards association with FEV1 and FEV1/FVC (Table 6.4, p=0.08 and 0.07 respectively). Since parenchymal disease (i.e. emphysema) is generally best represented by KCO, with FEV1 and FEV1/FVC impaired later in the disease process[149] these results are suggestive of an association with this phenotype, perhaps not detectable in the CT based analysis because of lower numbers in this group.

The allelic associations were examined next, and are detailed in Table 6.4. The most associated allele was *HLA-DQA1**0301, which predicted lower FEV1 and FEV1/FVC as shown by the negative regression coefficients (p=0.003 and 0.005 respectively), and tended to associate with lower KCO (trend; p=0.063). It is of note that the associations with FEV1 and FEV1/FVC would remain significant even after a Bonferroni correction for multiple statistical testing, which may be too conservative when considering areas of high LD, such as the HLA region. *HLA-DRB1**04, which was in LD with *HLA-DQA1**0301 (r2=0.92, p<0.0001) in this dataset (this relationship being consistent with current knowledge of regional LD[313]) showed association in the same direction with FEV1/FVC (p=0.046) and a trend to association with FEV1 and KCO (p=0.075 and 0.098 respectively). These were the only 2 alleles that showed evidence, or suggested evidence, of association with more than one measure of pulmonary function.

*DQB1**02 and *DQB1**03032 showed association with higher KCO and lower FEV1 respectively (p=0.014 and 0.035 respectively). Neither showed association with other features of pulmonary function. Alleles with which *DQB1**02 is commonly in LD, such as DRB1*03[313], showed no association with any feature of pulmonary function. This is

also the case for *DQB1**03032, and might indicate a *DQB1**02 or *DQB1**03032 specific effect, or a false positive result at these alleles. The latter is perhaps more likely, given the lack of suggestion of association in the locus analysis for *DQB1*.

Since a locus association was observed with KCO, and both *DQA1**0601 and *DQA1**0101 associate with this feature (p=0.011 and 0.015 respectively) these results deserve some further consideration. Alleles with which *DQA1**0601 and *DQA1**0101 are commonly in LD, such as *DRB1**01, showed no association with any feature of disease. Both alleles associate with lower KCO, but the confidence interval for the regression coefficient for *DQA1**0601 is very wide (Table 6.4), indicative of low numbers of subjects carrying this allele (Table 6.3). This result must, therefore, be viewed with great caution, as it may represent a false positive. The number of subjects carrying the *DQA1**0101 allele was much greater, so the same cannot be said of this allele, and it remains possible that this allele is having an effect on risk of disease. On balance, however, the strength of effect at *DQA1**0301 and consistency of effect through measures of pulmonary function, and with an allele with which it is commonly in LD (*DRB1**04) makes this a more likely candidate at the *DQA1* locus.

Locus/Allele	Phenotype	Regression coefficient	95% confidence interval	p value
DRB1	FEV1/FVC	-	-	0.416
	FEV1	-	-	0.122
	KCO	-	-	0.437
DQB1	FEV1/FVC	-	-	0.245
	FEV1	-	-	0.085
	KCO	-	-	0.191
DQA1	FEV1/FVC	-	-	0.070
	FEV1	-	-	0.080
	KCO	-	-	**0.034**
*DQA1*0301*	FEV1/FVC	-0.06	-0.09 - -0.02	**0.005**
	FEV1	-0.09	-0.15 - -0.03	**0.003**
	KCO	-4.96	-10.19 - 0.27	0.063
*DRB1*04*	FEV1/FVC	-0.05	-0.09 - -0.0008	**0.046**
	FEV1	-0.06	-0.12 - 0.01	0.075
	KCO	-5.32	-11.64 - 0.99	0.098
*DRB1*13*	KCO	-6.54	-14.07 - 0.99	0.088
*DQB1*02*	KCO	6.98	1.45 - 12.50	**0.014**
*DQB1*03032*	FEV1	-0.10	-0.20 - -0.01	**0.035**
*DQA1*0101*	KCO	-7.11	-12.59 - -1.63	**0.011**
*DQA1*0601*	KCO	-27.37	-49.49 - -5.26	**0.015**

Table 6.4: Association analyses for HLA class II and pulmonary function

For the three class II loci, the p values for likelihood ratio test results, with 13 degrees of freedom for *HLA-DRB1* and 8 for *HLA-DQB1* and *HLA-DQA1* are shown for each measure of pulmonary function. Details from the regressions are shown for alleles where $p \leq 0.1$ only without adjustment for multiple statistical testing; after a Bonferroni correction for the number of alleles tested the p value for significance at *HLA-DRB1* should be 0.004, and at *HLA-DQB1* or *HLA-DQA1* it should be 0.006. Results significant at $p<0.05$ are shown in bold type.

The only alleles to show consistent effects across several pulmonary function parameters were *DQA1*0301* and *DRB1*04*, which were in LD with one another ($r^2=0.92$, $p<0.0001$). Furthermore *HLA-DQA1*0301* would remain significantly associated with FEV1/FVC and FEV1 after a Bonferroni correction. The lack of consistent effects across pulmonary function measures for this and other alleles below *DRB1**04 in the table may mean these are less likely to be true associations, though further study will be required. The wide 95% CI of *DQA1**0601 reflects its low frequency within the group, and further reduces the confidence that this is a true association.

6.4.2 Anti-elastin antibody associations with clinical features

Anti-elastin antibody levels were obtained from samples contemporaneous to the clinical assessment for all 172 subjects, the mean value being shown in Table 6.5. The coefficient of variation for the assay over the working range was 7.13 - 9.47%. At the lowest concentration of the standard (6.25IU) it was higher (21.16%). Samples with anti-elastin antibody concentrations at this level were repeated at appropriate dilutions to ensure a result was obtained within the most reliable range of the assay. The spike recovery was 87.34%, and assay specificity for anti-human IgG was confirmed by a zero reading in the columns where control anti-CD11 antibody had been used.

No difference in level was seen between genders, and the level did not vary significantly with age ($p=0.64$ and 0.74 respectively). Antibody level did not differ between those who had never smoked and those who had ($p=0.07$, Figure 6.5), nor did they differ between those with normal FEV1/FVC, normal FEV1 or normal KCO and those where these parameters were abnormal ($p= 0.09$, 0.22 and 0.26 respectively). However there was a statistically significant correlation between antibody levels and FEV1/FVC, $p=0.03$, Figure 6.6). However, if those who had never smoked or those with a normal FEV1/FVC were excluded the significance of the relationship was lost ($p=0.84$ and 0.20 respectively). This suggests that the data is being influenced predominantly by these subjects. The level did not vary between emphysematous and non-emphysematous subjects ($p=0.074$, Table 6.5), nor did it correlate with any other feature of lung function (all $p>0.05$).

	Mean (SE) anti-elastin antibody titre (IU)
All subjects	39.54 (7.54)
Emphysema	29.30 (6.06)
No emphysema	51.12 (17.33)
p value	0.074

Table 6.5: Anti-elastin antibody levels in the AATD subjects

The table shows the mean anti-elastin antibody level in the AATD group, stratified by presence of emphysema. The differences were not statistically significant.

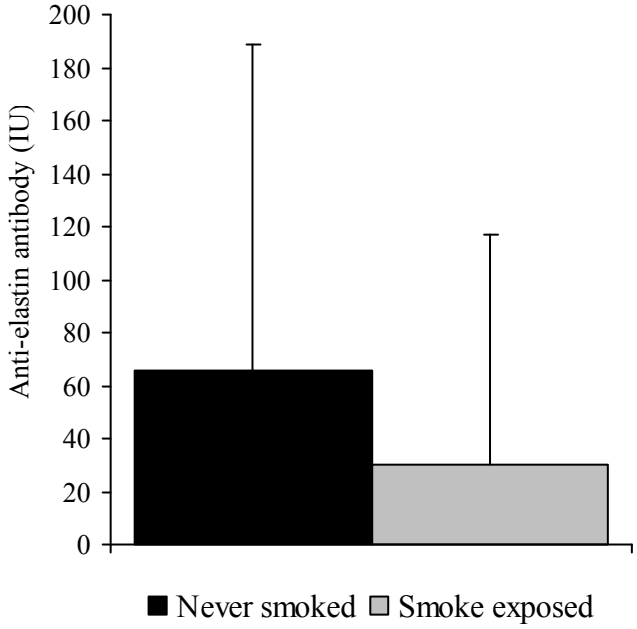


Figure 6.5: Anti-elastin antibody levels do not differ between those that have and have not smoked

The histogram shows the mean anti-elastin antibody level, with error bars representing SEM, stratified by smoke exposure. The difference was not statistically significant (p=0.068).

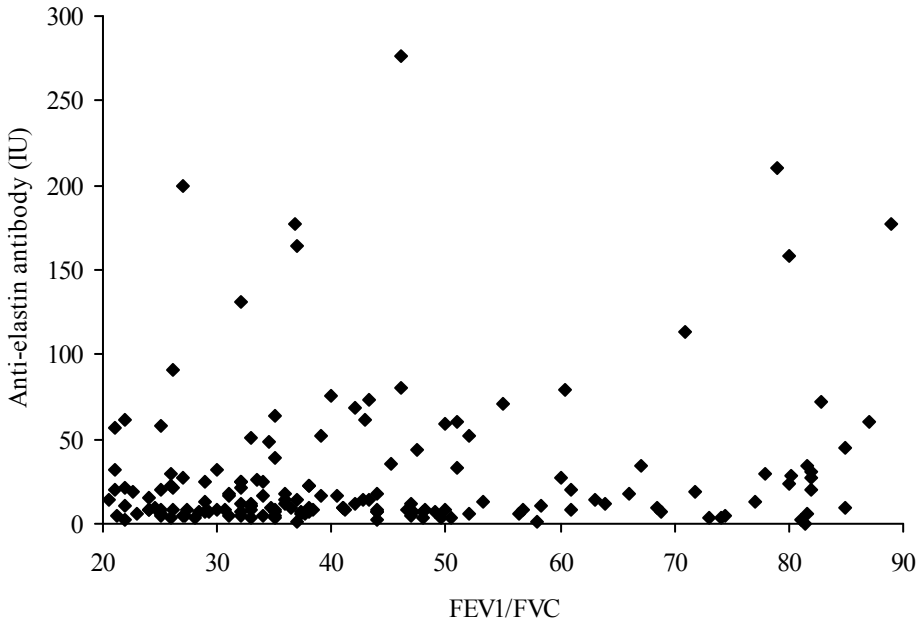


Figure 6.6: Anti-elastin antibody levels correlate with FEV1/FVC

The graph shows anti-elastin antibody levels, with each point representing an individual. Less airflow obstruction was observed in those with high anti-elastin antibody levels (r=0.16, p=0.03). The trendline has been omitted due to non-normal data distribution.

6.4.3 HLA class II associations with anti-elastin antibody levels

Anti-elastin autoantibody levels were compared between genotypes of phenotypically associated alleles only, in order to assess if anti-elastin autoimmunity was driving the phenotypic association with HLA type. This would indicate a potential mechanism for the observed genetic association.

Antibody levels were associated with *HLA-DQA1**0301 and *HLA DRB1**04 genotype (p=0.027 and 0.024 respectively), such that they were higher in those homozygous for the associated alleles. For the *HLA-DRB1*04* allele the mean (SEM) antibody level was 109.65 (38.96) IU/ml, whilst in heterozygotes it was 36.99 (15.63) IU/ml and in those with no *HLA-DRB1**04 alleles 38.73 (8.77) IU/ml. For the *HLA-DQA1**0301 allele the corresponding levels were identical for homozygotes, since these 3 subjects were homozygous for both *HLA-DRB1**04 and *HLA-DQA1**0301, whilst heterozgotes had a mean level of 37.36 (16.01) and those with no *HLA-DQA1**0301 alleles a level of 38.60 (8.70). This is consistent with the pattern of LD in the region. These results are illustrated only for the most associated allele (*HLA-DQA1**0301), since the results for *HLA-DRB1**04 are similar, in Figure 6.7.

This relationship was apparent only in smokers, with levels being higher in associated allele homozygotes (*HLA-DRB1**04; p=0.018, *HLA-DQA1**0301; p=0.021. Figure 6.8 illustrates this for the most associated allele (*HLA-DQA1**0301). The difference in antibody levels between non-smokers by genotype was non-significant (both p=0.86). This analysis was performed to determine if the generation of anti-elastin antibodies was both smoke and genotype dependent. Splitting of the dataset in this way does, however, limit

the size of the homozygous groups further and should, therefore, be viewed with caution.

Antibody levels did not differ significantly between genotypes for other alleles that exhibited a phenotypic association (all $p>0.75$).

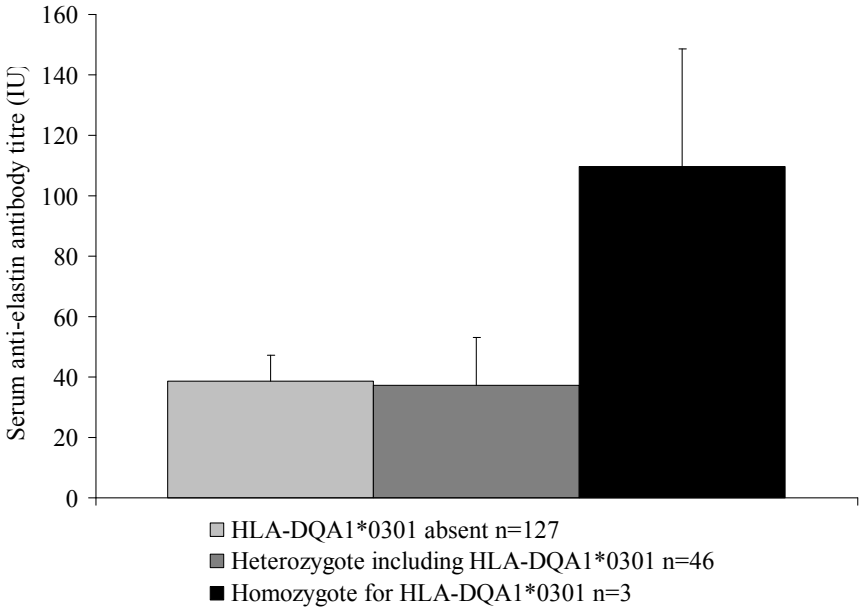


Figure 6.7: Anti-elastin antibody levels vary between *HLA-DQA10301 genotypes**

The bar chart shows the mean (SE bar lines) of anti-elastin antibody titre in patients stratified by *HLA-DQA1**0301 genotype, each column illustrates the number of *HLA-DQA1*0301* alleles present. Levels are significantly higher in patients homozygous for *HLA-DQA1*0301* (p=0.027).

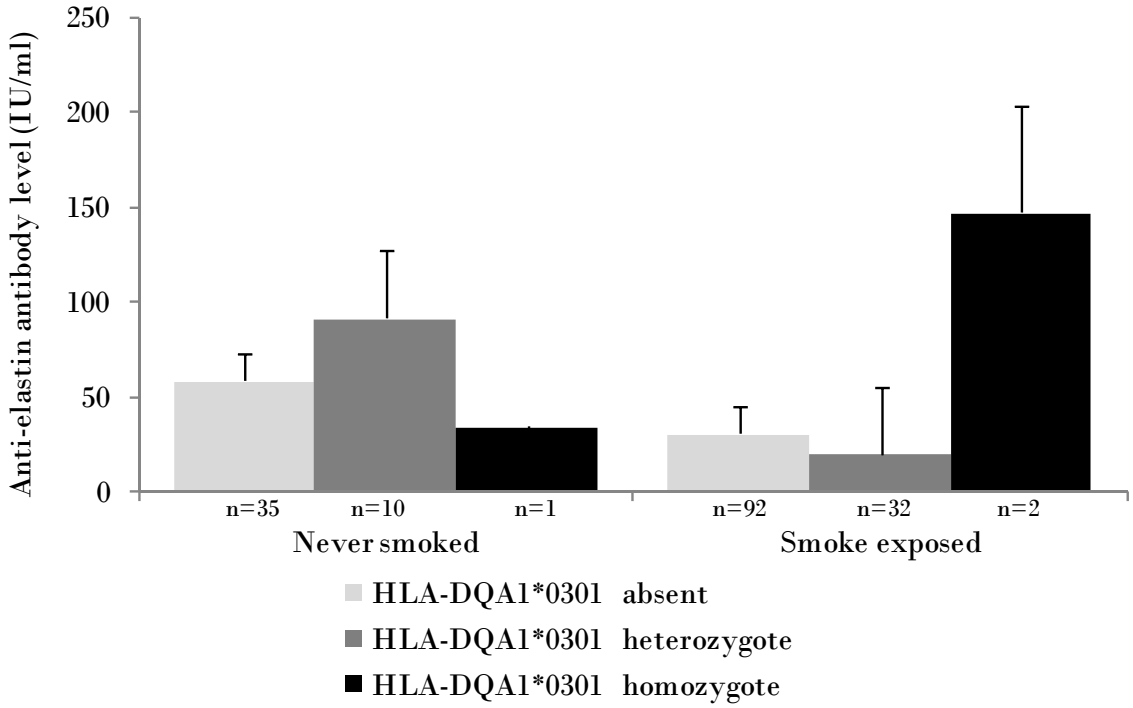


Figure 6.8: Anti-elastin antibody level stratified for both *HLA-DQA10301 genotype and smoke exposure**

The bar chart shows the mean serum antibody level with the dataset split according to whether the subject had ever smoked. In subjects who had smoked antibody levels were higher in those homozygous for the *HLA-DQA1**0301 allele (p=0.021), but in non smokers the relationship was not significant (p=0.859).

6.5 Discussion

The data show an association between the *HLA-DQA1* locus and KCO, and individual alleleic associations of *HLA-DRB1**04 and *HLA-DQA1**0301 with lower FEV1/FVC, and of the latter with lower FEV1. These 2 alleles, which exhibited LD, were the only alleles to show consistent effects across several pulmonary function parameters, although others associated with a single measure. Importantly, HLA-*DQA1**0301 remained significantly associated with FEV1/FVC and FEV1 after a conservative correction for multiple statistical testing; the validity of this correction in the HLA region will be discussed further later.

Anti-elastin antibodies showed no phenotypic associations when considering smoke exposed subjects, but did show a genotypic association, in that they were higher in those homozygous for either *HLA-DRB1**0301 or *HLA-DRB1**04. This suggests that an anti-elastin mediated autoimmune pathology, in which HLA type is a key driver of susceptibility exists in AATD, but that it may only be so in those homozygous for one, or both of *HLA-DRB1**04 and *HLA-DRB1**0301. Elastin peptides might be presented to the immune system more effectively by proteins encoded by associated alleles, leading to generation of an auto-antibody and a subsequent response directed at elastin in the lung, as illustrated in Figure 6.9. This hypothesis is consistent with previous work in usual COPD[205].

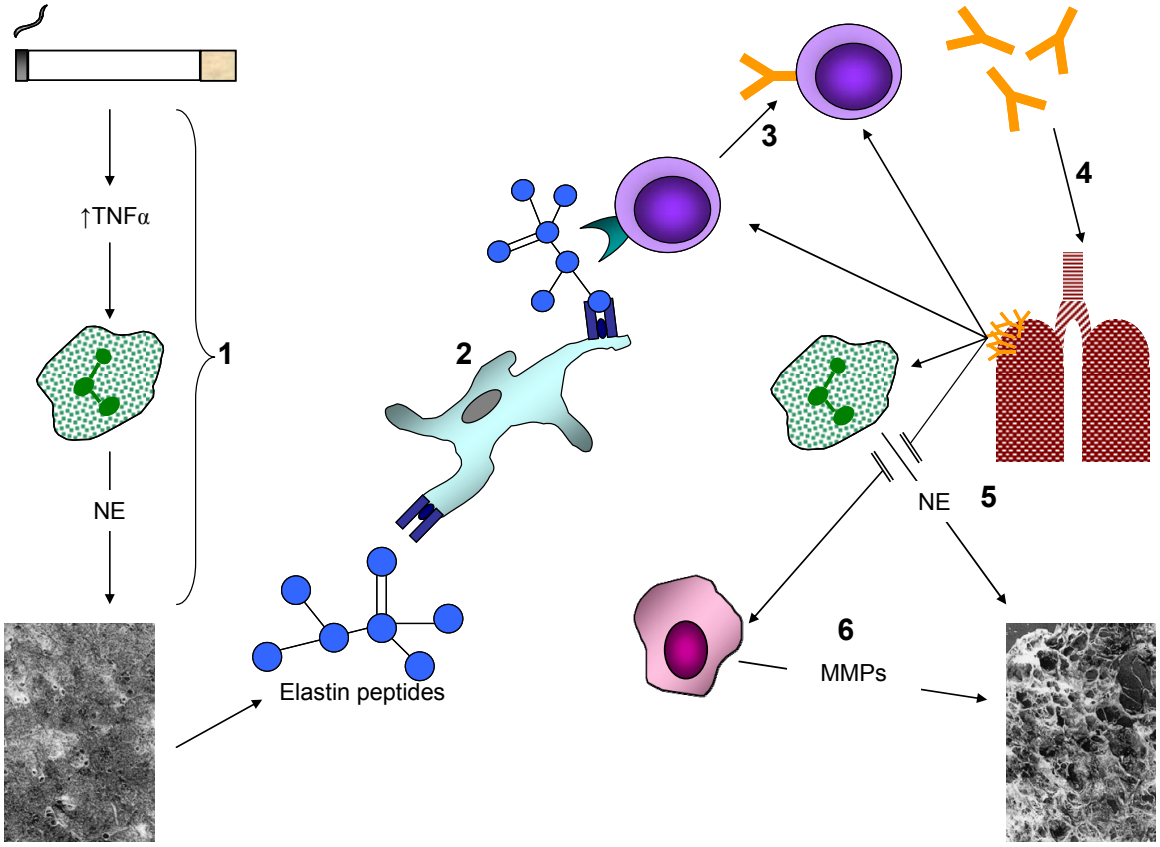


Figure 6.9: A possible role for HLA class II type and anti-elastin autoimmunity in lung disease in AATD

1 Cigarette smoke upregulates TNFα which causes neutrophil aggregation in the lung. Neutrophil elastase (NE) is released and causes elastolytic damage to normal lung, generating elastin peptides

2 Elastin peptides are presented to the immune system. At this point if *HLA-DRB1**04, or *DQA1**0301, is the HLA class II protein present the peptide could bind preferentially, such that it is presented more effectively to T cells, via the T cell receptor. Alternatively it might be more able to dock to the T cell receptor after binding. Either of these processes would amplify the subsequent immune response.

3 T-cells activate B cells, such that they express anti-elastin antibodies on their surface, and generate similar free antibodies.

4 Free antibody binds to elastin in the lung, causing an autoimmune response, such that other cells of the immune system are activated against the tissue.

5 Neutrophils release NE, which causes further elastolytic damage.

6 Macrophages release other proteases, such as matrix metalloproteases (MMPs), which also damage components of the extracellular matrix, such as collagen.

Autoimmune diseases have a complex pattern of HLA association, with the strongest evidence being for class II alleles[306]. Classical autoimmune diseases have strong predisposing effects associated with *HLA-DRB1**03 and *HLA-DRB1**04. Examples of diseases with such associations are type 1 diabetes (T1D)[314, 315] and rheumatoid arthritis (RA)[316]. Other *HLA-DRB1* alleles, such as *HLA-DRB1**1301, have been proposed to act as protective variants in RA[317]. These allelic associations may be due to an altered amino acid sequence at positions β72-74 in the DRB1 protein (the 'shared epitope')[317], an area which spans environments involved in autoantigen presentation and T-cell antigen receptor docking. This suggests that changes here could determine which antigens are presented to the immune system[308]. *HLA-DQA1* specific associations with autoimmune diseases have been reported in Graves' disease[309], and haplotype associations including this locus and *DRB1* are well known[308].

The involvement of an allele associated with autoimmunity lends support to the hypothesis that development of emphysema could depend, in part, upon an antigen presented by HLA class II. If an antigen was more able to be presented to the immune system in those with *DQA1**0301 (or *DRB1**04) the net effect could be a greater immune response and more autoimmune lung destruction, or alternatively lesser immune response and hence less tissue destruction. Various autoantigens have been suggested in emphysema [205, 298] and their binding properties with the protein products of the associated alleles are unknown. Many disease mechanisms have been proposed for HLA class II in autoimmunity, including variation in autoantigen binding and failure to select a good regulatory T cell population[306]. These results could support the former as a disease mechanism, but it is of note that lower numbers of regulatory T cells in the lungs

have been reported in emphysematous subjects relative to healthy controls[205]. Functional effects of variation in the HLA region are an area of current research in traditional autoimmune diseases where HLA associations are more well replicated[306], though are complex because of the multi-allelic nature of HLA class I and II. Such studies may be relevant to pursue in COPD if autoimmunity is accepted as an important disease mechanism.

Within HLA class II LD is observed between DR and DQ loci, resulting in classical haplotypes[313], conventionally described by the *DRB1* allele on which they are based. Any *DRB1* allele typically in LD with *DQA1**0301 might, therefore, show the same disease association. Indeed, *HLA-DRB1**04 is in LD with *HLA-DQA1**0301 here, such that their associations are in the same direction. However, the pattern of LD also makes it impossible to reliably state that the association is due entirely to *HLA-DQA1**0301 - it might be due to either, or indeed both, alleles. The other allelic associations seen with clinical features are less consistent across all measures of pulmonary function, perhaps suggesting a risk that a false positive result has occurred. Furthermore none would remain significant after applying a Bonferroni correction, unlike *HLA-DQA1**0301, so such associations would require replication in another patient group before they could be considered truly significant.

Previous work in COPD unrelated to AATD has reported HLA associations at class I and class III loci. The class I associations were reported for different sub-phenotypes of COPD - FEV1[204] and chronic bronchitis[311], and were inconsistent. The only class III associations reported are with *TNFA*, which has been considered in detail in chapter 3. LD exists between HLA classes I, II and III, such that it may be difficult to ascertain the

true associated allele. Whilst a primary HLA class I or class III effect can not be excluded, data reported in this chapter is supported by a plausible disease mechanism incorporating HLA class II molecules. *HLA-DRB1**03 has been reported to be in LD with a *TNFA* promoter region SNP (rs1800629)[312], but in the studies contained within this thesis this SNP did not associate with any clinical phenotype, making it less likely to be a confounding factor in this case.

This study is weakened by small numbers, particularly in the group with no emphysema, and by the lack of a replication group. Classically in genetic association work, large populations are needed for adequate power. However the current study demonstrated a significant association with *HLA-DQA1**0301 even after a Bonferroni correction, usually deemed overly conservative for genetic association work[68], particularly in the HLA region where it is uncertain how best to adjust for alleles in tight LD. This suggests that in this population a true effect is more likely than a chance association. Other factors that have the potential to influence the result of a genetic association study are genotyping error and population stratification. Independent verification of genotype by two researchers was used to reduce the chance of genotyping error. All subjects were racially matched, which minimises the chance of population stratification having affected the results, though this could be further confirmed by the addition of racial markers to the loci typed.

It is also of interest that antibody levels were higher in subjects who had not smoked, albeit not statistically significant. Since non-smokers generally have better lung function, as shown by the influence of smoke exposure upon all lung function measures in the regression models reported in Chapter 3 (Table 3.6) it may be that non-smokers are

influencing the correlation between FEV1/FVC and antibody level. This is supported by the observation that the correlation disappears when these subjects are excluded. It is, therefore, of interest to consider when non-smokers with AATD may present to medical services. In the UK there is no routine screening programme for AATD, and as such there is a potential ascertainment bias inherent in the patients studied, such that those with worse disease are more likely to present to medical services and undergo the tests needed to diagnose deficiency. If anti-elastin autoimmunity dependent on HLA class II type alone, rather than a combination of this and smoke exposure, is important, then non-smokers with the relevant HLA class II type could develop disease and hence present to health care services. Conversely those without the associated HLA type may be less likely to develop anti-elastin associated disease, and also less likely to develop lung disease by the more well-recognised proteolytic and oxidative mechanisms (Figures 1.4 and 1.5), thus would not present to health care services. Furthermore, smokers who had the associated HLA type could develop disease by both mechanisms and might die before diagnosis, or be too unwell to travel to our centre for assessment. All of these issues could lead to a bias in our dataset regarding the smoke exposure of individuals pre-disposed to develop anti-elastin antibodies due to their genotype. The observation that the difference in anti-elastin antibody with HLA type is not maintained in those who have never been smoke exposed suggests that other autoantigens may also be important, or may be an artefact due to low numbers in that group. Clearly this hypothesis is speculative in nature, but could be tested in a country where AATD is screened for at birth, where such a bias would not exist.

Evidence from animal models[206, 235, 300], observation of lymphocyte accumulation in the lung[11, 302], clustering of autoimmune disease in individuals with COPD[304] and demonstration of autoantibodies in emphysema[205, 298], together with our results suggest that autoimmunity may be an important disease mechanism in emphysema, which would have implications for treatment. Response to immunosuppression, such as steroids, is variable in COPD. Since autoimmune processes characteristically respond to such treatments, our results suggest that there may be identifiable groups of patients in whom immunosuppression will be valuable. Replication of this study in COPD unrelated to AATD is now indicated.

6.6 Conclusions

HLA class II shows association at the *HLA-DQA1* locus with KCO, though no individual allele could be confidently identified that might account for this. However, individual allelic analyses showed the *HLA-DQA1**0301 allele to be associated with both FEV1 and FEV1/FVC even after a conservative correction for multiple statistical testing. *HLA-DRB1**04, with which *HLA-DQA1**0301 is in tight LD, also showed a phenotypic association with FEV1. Both alleles exhibited genotypic association with anti-elastin antibody levels, suggesting that anti-elastin mediated autoimmunity may be pathophysiologically important in individuals with either of theses alleles and AATD.

Chapter 7: The role of pollution and other inhaled agents linked to lung disease

7.1 Introduction

Adverse health effects have been linked to levels of air pollution by a variety of epidemiological and toxicological studies. Pollutants may be gaseous or particulate, and primary (released directly into the atmosphere) or secondary (formed by reactions in the atmosphere). The pollutants with most evidence regarding ill effects on health are ozone, particulate matter, nitrogen dioxide (NO_2) and sulphur dioxide (SO_2)[207], though the last is somewhat controversial as toxicological studies suggest the concentrations needed to cause an effect are far higher than seen in the environment[318]. Two major studies have linked long term exposure to particulate matter with deaths from lung cancer and cardio-respiratory disease[319, 320]. Time series work also suggests similar shorter term effects [321]. Other epidemiological work assessing short term exposure effects including hospital admission rates for respiratory disease and lung function changes, has found significant effects of particulate matter[322, 323], ozone[324, 325] and sulphur dioxide (SO_2)[322, 326]. In the UK reports have been produced that focus on long term effects of all major pollutants, with the aim of estimating the overall health burden attributable to pollution, and direct public health policy as a consequence. The World Health Organisation (WHO) and American Thoracic Society (ATS) have also provided guidance on what constitutes an adverse effect of air pollution[207, 327], identifying changes in lung function as an important outcome. Both recognise that genetic factors may be important in determining

such effects[207, 327]. Subjects with AATD sustain lung damage as a result of the relatively unopposed action of neutrophil elastase, which degrades lung elastin[328]. This process leads to a greater disease burden in those exposed to inflammatory stimuli, predominantly cigarette smoke[17]. Pollutants may contribute to lung disease by inflammatory, elastase predominant pathways or by mechanisms relating to oxidative stress[44]. By studying subjects with AATD it may be possible to elucidate which pollutants contribute to disease by elastase dependent mechanisms, as their effects may be more marked in AATD.

Another important group of inhaled agents associated with lung disease[329], both asthma, interstitial disease and COPD, are those observed in the workplace. Occupational exposures may also influence lung function in AATD[330-332], and are appropriate to consider alongside pollution. It is certainly conceivable that subjects with AATD would be at greater risk of lung damage from occupational agents by the same mechanisms as pollution, and such information would be of value to them when choosing work. Occupational exposures may be ascertained by direct measurements in prospective studies or by the occupational history in retrospective study designs.

Several methodological issues must be considered when studying pollution. Retrospective study designs may consider effects of day-to-day changes in pollution on daily events, such as deaths and hospital admissions, (known as time-series studies), or may be cross sectional in order to give an indication of long term effects. Time series studies, whilst using detailed pollution data, are only able to use crude measures of health effects. Cross sectional designs are limited by the fact that pollutant levels have changed over time, and availability of data pertaining to covariates, such as smoking. Prospective

longitudinal studies may therefore be the ultimate goal, but will be expensive to conduct. The subjects recruited to ADAPT are potentially susceptible to the effects of pollution, well characterised regarding respiratory disease, and have information on covariates available. They therefore represent a group in whom a cross sectional design might be informative. Furthermore, longitudinal data on lung function is available for many, such that cumulative pollution effects might be considered, albeit in a retrospective manner.

Much of the data presented here has been published in peer reviewed journals (Appendix 7).

7.2 Aims of this chapter

This chapter describes putative long term effects of outdoor air pollution and inhaled occupational agents on lung function and of emphysema severity in AATD, as well as associations of air pollution with lung function decline. Other clinical features that predict decline will also be described.

7.3 Methods

7.3.1 Long term effects of air pollution and occupation

Spatial distribution of pollution remains relatively constant in the UK[333]. Consequently, a cross-sectional study including only patients who have lived in one area throughout their lives can give some indication of the effect of lifetime exposure. Occupation may be assessed in epidemiological work of this nature using a system that classifies lifetime exposure risk by the use of a job exposure matrix, which takes into account all jobs that a person has undertaken.

Subjects

All PiZZ subjects from the ADAPT databases who had always lived within 25km of their place of birth were selected for assessment with regard to pollution, giving a total of 304 individuals. Subjects who had moved (n=118) were studied only with regard to their demographic and clinical features, in order to ascertain if differences existed between those that had or had not moved. Residential history was ascertained by asking for place of birth, current and previous addresses.

Patients' addresses from the year of pollution data collection were mapped to OS co-ordinates using the National Statistics Postcode Directory[334]. The same database was used to obtain a census classification into urban/rural dwelling for the population. All subjects were phenotyped as described previously, and information on covariates such as smoking and occupational history noted. Lifetime risk of exposure to inhaled agents in the

workplace associated with lung disease was ascertained by use of a job exposure matrix[335]. This was created using Standard Occupational Classifications (SOC 2000) generated using the CASCOT[336] system. Job codes were divided into 3 groups, according to the likelihood and probable intensity of exposure to agents known to cause occupational lung disease. Patients who had worked at any time in a job where the exposure intensity was likely to be less than 30% of the workplace exposure limit[337] were classed as low risk, with intensities above this being deemed high risk. Those who had never worked in an exposure prone job were classed as zero risk. Clinical and demographic features were compared between those of different occupational risk groups, and the frequency of at risk professions was compared to that in the UK as a whole, obtained using the 2001 census.

Pollutants

Annual mean data for SO_2, NO_2, oxides of nitrogen (NOx), particulate matter <10μm in diameter (PM_{10}) and secondary particles was obtained from geographical information system (GIS) maps for 2006. These were used as indicators of long-term exposure to each pollutant. Annual mean ozone is not available at the level of resolution of GIS maps, hence alternative metrics from similar GIS maps were used for this pollutant: AOT40 (ozone $\mu g/m^3$.hr) and days of ozone >120$\mu g/m^3$. AOT40 is the parameter used to represent accumulated ozone dose and is the sum of the differences between the annual hourly mean ozone concentration and 40 ppb for each hour when the concentration exceeds this limit during daylight hours. Levels of primary particles were calculated by subtracting

secondary particles from PM_{10}. The maps detail pollutant levels on a 1km x 1km grid across the UK, using data from the National Atmospheric Emissions Inventory and a combination of modelling methods, including dispersion kernel approaches and weighted regression analyses to calculate effects from major industrial sources. Major roadside increments are taken into account where necessary. This methodology is described in detail elsewhere[338]. Pollutant levels were mapped to patients using the OS co-ordinates from the grid.

Statistical analysis

The relationship of pollutants to one another was checked using simple correlations, and differences across urban and rural locations ascertained (Kruskal Wallis test) in order to validate the data for the cohort against existing evidence regarding pollutant distribution. Regression models ascertaining pollution effects took into account pack years smoked, age, gender and level of occupational risk prior to addition of pollution data, using FEV1, FEV1/FVC, KCO, UZVI and LZVI as the outcome variables.

For pollutant associations from the regression models the relationship of current levels to that of previous years was checked using data from 1993, 2003 and 2005. These were assessed for variability and compared between years using the Kruskal Wallis test and Friedman's test for ozone metrics because the data violated the assumption of sphericity for repeated measures ANOVA.

7.3.2 Pollution and decline of lung function

Subjects

All PiZZ subjects with baseline and follow up data were studied (n=404). Decline of lung function for descriptive statistics of the group was calculated using linear regression over 4 data points, as described in chapter 2. The number of subjects with 4 data points for calculation of decline was 220; those with less follow up than this were assessed as well in order to reduce the effect of selection bias on association analyses of lung function decline. The years of data collection were noted, such that cumulative pollution exposure and mean exposure per year during that period could be calculated. The number of exacerbations per year over the time period was also noted.

Pollutants

Annual mean levels of SO_2, NO_2, PM_{10} and ozone (AOT40, as before) from each of the years used for calculation of lung function decline were taken from GIS maps as before, matching to patients by the use of the postcode. Secondary particle data was not available for sufficient years to calculate primary particles from PM_{10}. Similarly there were fewer years with ozone data, reducing the number of subjects this could be matched to.

Statistical analysis of pollutant data

Predictors of lung function decline were ascertained using univariate models, and those

significant added to multivariate models including a single pollutant on each occasion. Generalised estimating equations were used to ascertain predictors of decline, since these can account for repeated measures in an individual, as well as including a time function, which accounts for differing lengths of follow up. This allowed inclusion of a greater number of subjects in the initial analysis for pollution effects, thus maximising power. Analyses were then substratifed for those with and without 4 data points for their lung function. This was performed in order to see if the increased accuracy of decline measurement obtained in those with more measurements available was influencing the results in the larger dataset. Subgroup analysis by index status was also carried out, in order to exclude any influence of this on the outcome measure.

7.4 Results

7.4.1 Long term effects of air pollution

Clinical and demographic features

The characteristics of the subjects are shown in Table 7.1, with sub-stratification for level of occupational risk in Table 7.2, and the most frequent at-risk professions in Table 7.3. There was no significant difference in any demographic or clinical feature between the subjects who had never moved from their place of birth and those who had.

No clinical or demographic feature differed between occupational risk groups, though at risk professions were over-represented in the patient group compared to the UK population (9.45 v 7.48%, $p<0.0001$). Subjects had worked for a mean of 19 years and stopped work at a mean age of 48 years. There was no difference in the duration of working life or age stopped work between the occupational risk groups ($p>0.1$). Since some occupational agents are associated with interstitial lung disease this was sought using subjects' HRCT images, but was not present in any subject.

	Never moved (n=304)	Moved away from place of birth (n=118)
Subjects of male gender	58.88 (179)	63.60 (75)
Age in years	50.29 (42.59-56.22)	51.11 (40.63-58.09)
Never smoked	23.70 (72)	25.40 (30)
Pack years smoked	14.00 (1.10-25.38)	13.50 (0-23.50)
FEV1/FVC	39.60 (30.63-56.63)	36.10 (28.45-59.20)
FEV1 % predicted	34.79 (24.72-65.23)	38.04 (24.16-58.41)
KCO % predicted	66.10 (53.84-87.90)	66.67 (50.11-82.62)
UZVI	29.75 (16.03-42.60)	34.30 (17.90-47.70)
LZVI	48.30 (27.95-59.20)	49.80 (32.90-63.00)

Table 7.1: Characteristics of subjects stratified by whether they had moved their place of birth

Data are shown as median (IQR) for quantitative data, and as % (n) for frequency data. All data were non-normally distributed, except for age – this is shown as median (IQR) for consistency within the table. There were no significant differences between those that who had or had had not moved (all $p>0.05$).

	Level of occupational risk			p value
	Zero	Low	High	
% of subjects of male gender	60.93	67.65	50.67	NS
Age	50.40 (0.82)	48.81 (1.32)	49.23 (1.32)	NS
Pack years	14 (0.2-26)	13.5 (5.25-24.5)	15.75 (0-24)	NS
FEV1	36.66 (25.81-58.91)	39.87 (24.13-68.54)	34.7 (25.90-69.91)	NS
FEV1/FVC	38.2 (29-53)	41(30.6-57)	39.1 (32-60.25)	NS
KCO	68.86 (57.37-88.75)	69.87 (2.84)	66.1 (57-85.91)	NS
Bronchodilator reversibility (%)	62.16	64.71	59.46	NS

Table 7.2: Characteristics of subjects stratified for level of occupational risk to the lung

51.4% of subjects exhibited zero occupational risk, with 23.1% having low risk and 25.5% a high risk exposure history. There were no significant (NS) differences between the groups.

Occupation	Number
Labourers in process and plant operations	8
Carpenters and joiners	6
Construction trades	6
Vehicle body builders and repairers	5
Welding trades	4
Motor mechanics and auto engineers	4
Painters and decorators	4
Coal mine operatives	4
Metal working machine operatives	4
Assemblers (vehicles & metal goods)	4

Table 7.3: Common high risk professions in AATD UK resident subjects who have never moved

The 10 most common high risk professions in the subject group are shown in the table.

Pollutants

All pollutants except secondary particles differed across settlement types allocated by the census (all $p<0.0001$, shown in Table 7.4 and Figure 7.1), ozone being greater in rural locations, whilst others were greater in urban areas. The two ozone metrics correlated well, whilst NO_2 and NOx showed a curvilinear relationship (both $p<0.0001$). Significant relationships between all other pollutants were also present (all $p<0.0001$). These are illustrated in Figure 7.2.

	Urban n=241	Town n=25	Village n=26	Hamlet n=12	p value
Primary particles	14.79 (0.16)	11.70 (0.14)	11.26 (0.18)	11.12 (0.29)	<0.0001
Secondary particles	7.46 (0.08)	7.72 (0.26)	7.96 (0.22)	7.12 (0.35)	0.082
NO_2	20.29 (0.38)	11.85 (0.82)	12.93 (0.58)	9.96 (0.94)	<0.0001
SO_2	3.99 (0.16)	2.86 (0.48)	2.71 (0.13)	2.14 (0.12)	<0.0001
Ozone days >120μg/m^3	14.25 (0.22)	15.98 (0.86)	17.33 (0.61)	16.44 (1.27)	<0.0001
AOT40	11859.13 (190.24)	13602.18 (498.98)	13673.59 (735.60)	13746.80 (1056.48)	<0.0001

Table 7.4: Pollutant distribution across census classes

Pollutants are shown as mean (SE), units being μg/m^3 except CO (mg/m^3) and AOT40 (μg/m^3.hr). All differences were significant with $p<0.0001$, with all pollutants bar ozone being higher in urban areas.

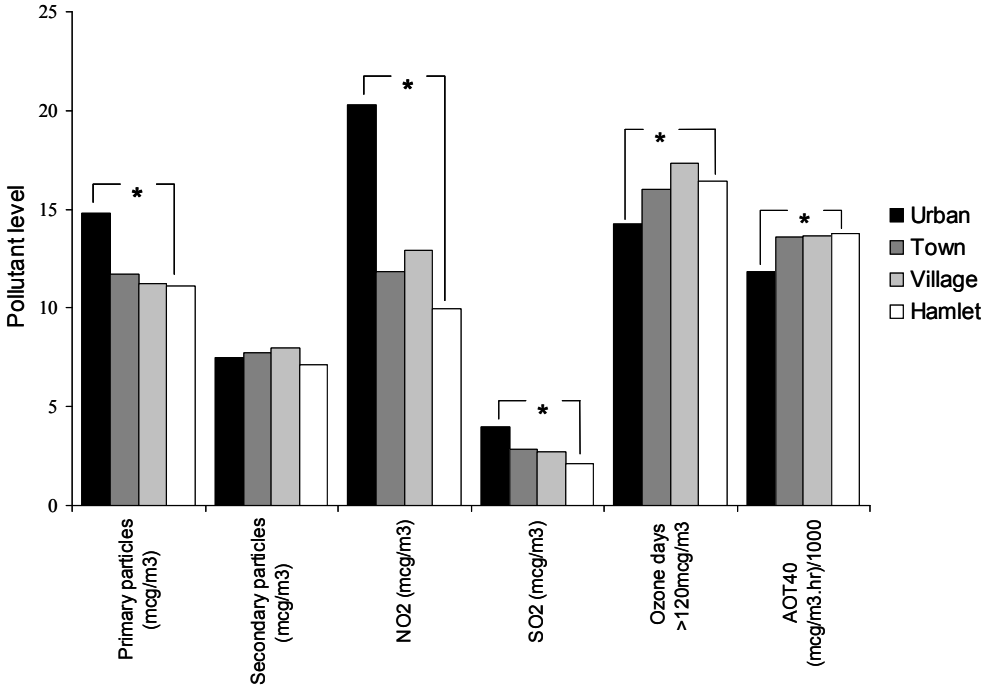


Figure 7.1: Pollution levels between residential areas in the UK

The graph shows the annual mean level from 2006 for each pollutant with the units shown on x axis labels. Primary particles, NO_2 and SO_2 were higher in urban areas, whilst ozone was higher in rural areas. Significant differences are marked with an asterisk (* = $p<0.0001$).

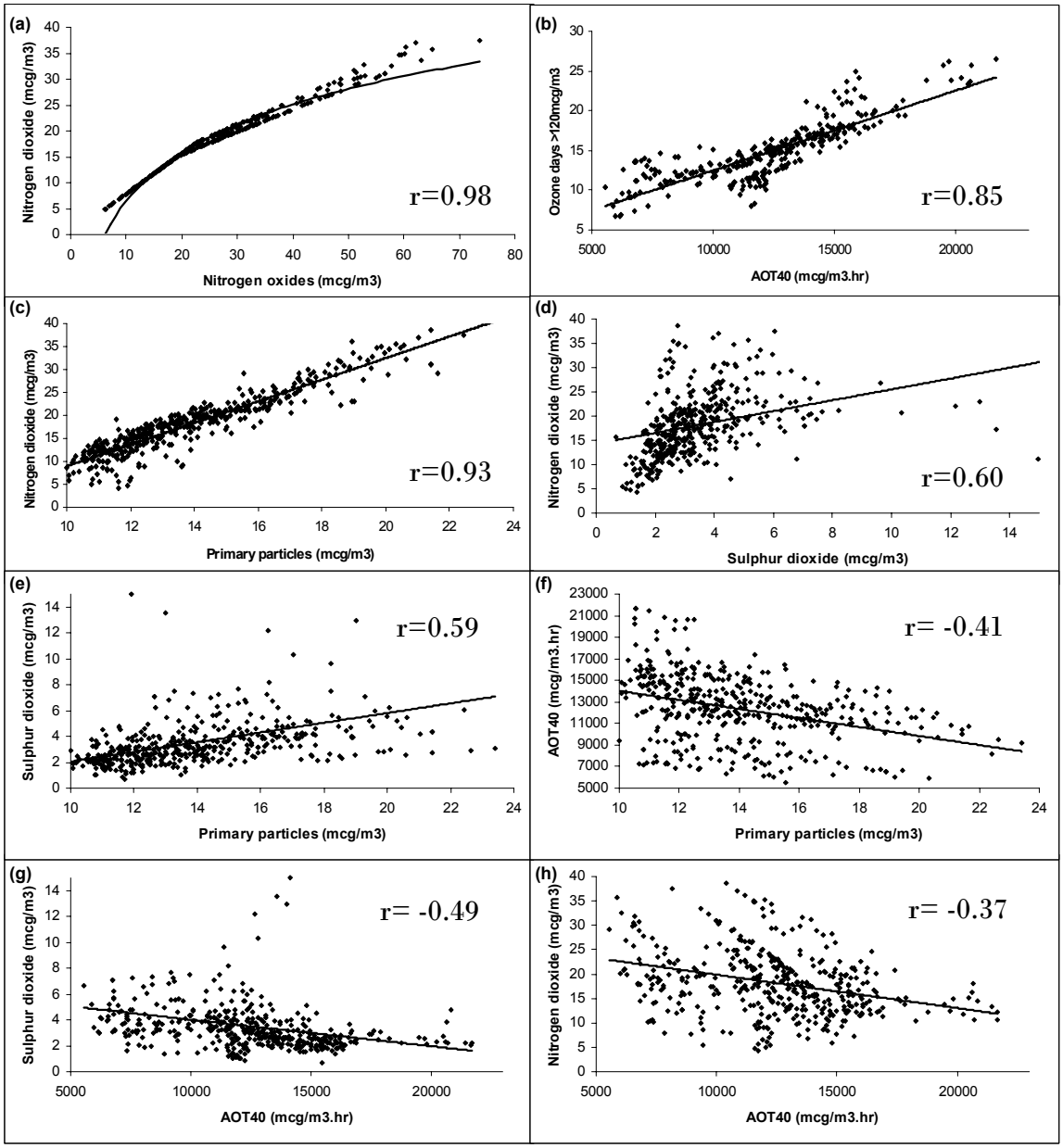


Figure 7.2: Intra-pollutant and inter pollutant correlations in the cross sectional study
(a) and (b) show correlations between measurements pertaining to the same pollutant, being for nitrogen oxides and their derivative, nitrogen dioxide, and ozone measures respectively. (c), (d) and (e) show the positive correlations observed between nitrogen dioxide, primary particles and sulphur dioxide. (f), (g) and (h) show the negative correlations between these pollutants and ozone. The correlation coefficients are shown on the graphs; all relationships are significant with $p<0.0001$.

Regression modelling

Each regression model was first constructed without pollution data, in order to ascertain associations with demographic features. Only those that were statistically significant were included in subsequent multiple regression models that included pollutants. Increasing age and cigarette smoke exposure were associated with lower KCO, FEV_1 and FEV_1/FVC and worse CT densitometry. The r^2 values suggest that smoking accounts for 6% of variability in KCO and FEV1/FVC, 16% of FEV1 variability, 2% of UZVI variability and 12.7% of LZVI variability. Male subjects had lower FEV1/FVC and more severe emphysema as measured by densitometry. Occupational risk was not a significant predictor of any clinical outcome.

Single pollutant models were used in the regressions, because of multi-collinearity between pollutants. AOT40 was used to represent ozone exposure. Although primary particles, NO_2 and SO_2 contributed significantly to regression models the B values suggested that lung function and UZVI were better in the presence of high levels. Conversely the values for ozone showed that KCO was lower and both FEV_1/FVC and UZVI higher with increasing ozone levels, the urban-rural difference in ozone (Table 7.4) being associated with a 2% decrease in KCO. Addition of ozone to the model accounted for only 1% of KCO variability. The regression results are summarised in Table 7.5.

Relationship of current to previous years' pollution levels

Ozone exhibited significant variability over time ($p<0.0001$, Table 7.6). Despite this the cumulative ozone exposure, calculated from the sum of means over 35ppb for all 3

years (the only metric available for 1993, 2003 and 2005), correlated well with 2006 ozone values. AOT40 showed a closer relationship than ozone days >120μg/m^3 (r=0.68 v 0.44, both p<0.0001, Figure7.3). The sum of annual mean values for other pollutants also showed good linear relationships with the 2006 values (all p<0.0001, Figure 7.3).

Feature	KCO		FEV1		FEV1/FVC		UZVI		LZVI	
	B	p	B	p	B	p	B	p	B	p
Primary particles	1.06	0.02	0.01	0.05	0.01	0.03	-1.10	0.01	-1.45	0.02
NO_2	0.51	0.01	0.004	0.06	0.003	0.052	-0.41	0.02	-0.45	0.06
AOT40	-0.001	0.05	-5.90×10^{-06}	0.19	-6.25×10^{-06}	0.03	0.001	0.02	0.001	0.03
SO_2	0.94	0.11	0.002	0.74	0.006	0.16	-1.14	0.01	-1.23	0.01
Age	-0.48	<0.01	-0.003	0.03	-0.003	<0.01	0.41	<0.01	0.68	<0.01
Pack years	-0.41	<0.01	-0.01	<0.01	-0.004	<0.01	0.19	0.05	0.48	<0.01
Gender	-0.73	0.31	-0.03	0.08	-0.06	<0.01	7.30	<0.01	4.57	0.04
Occupation	-0.95	0.54	-0.002	0.90	0.01	0.38	-1.51	0.28	-0.32	0.84

Table 7.5: Results of pollutants in regression models in the cross-sectional study

The table shows the regression coefficients and significance of each variable. Models for FEV1 and FEV1/FVC used log values whilst those for KCO and CT densitometry used untransformed data. All pollutants were measured in $\mu g/m^3$ except AOT40 ($\mu g/m^3.hr$). Results for covariates are from models prior to addition of pollutants. Gender was coded male=1, female=0.

	Sum of ozone>35ppb				
Year	All	Urban	Town	Village	Hamlet
1993	3085.38 (57.79)	2920.63 (62.52)	4049.00 (195.02)	3238.22 (120.62)	4055.43 (239.90)
2003	3417.98 (51.80)	3291.91 (54.87)	4186.04 (182.21)	3647.42 (188.15)	3852.49 (243.55)
2005	2185.28 (42.29)	2061.80 (44.45)	2460.95 (99.81)	2596.42 (146.51)	2185.28 (249.24)

Table 7.6: Ozone variability over time

Levels showed significant variability over time ($p<0.0001$) but were lower in more recent years ($p<0.0001$ in all locations).

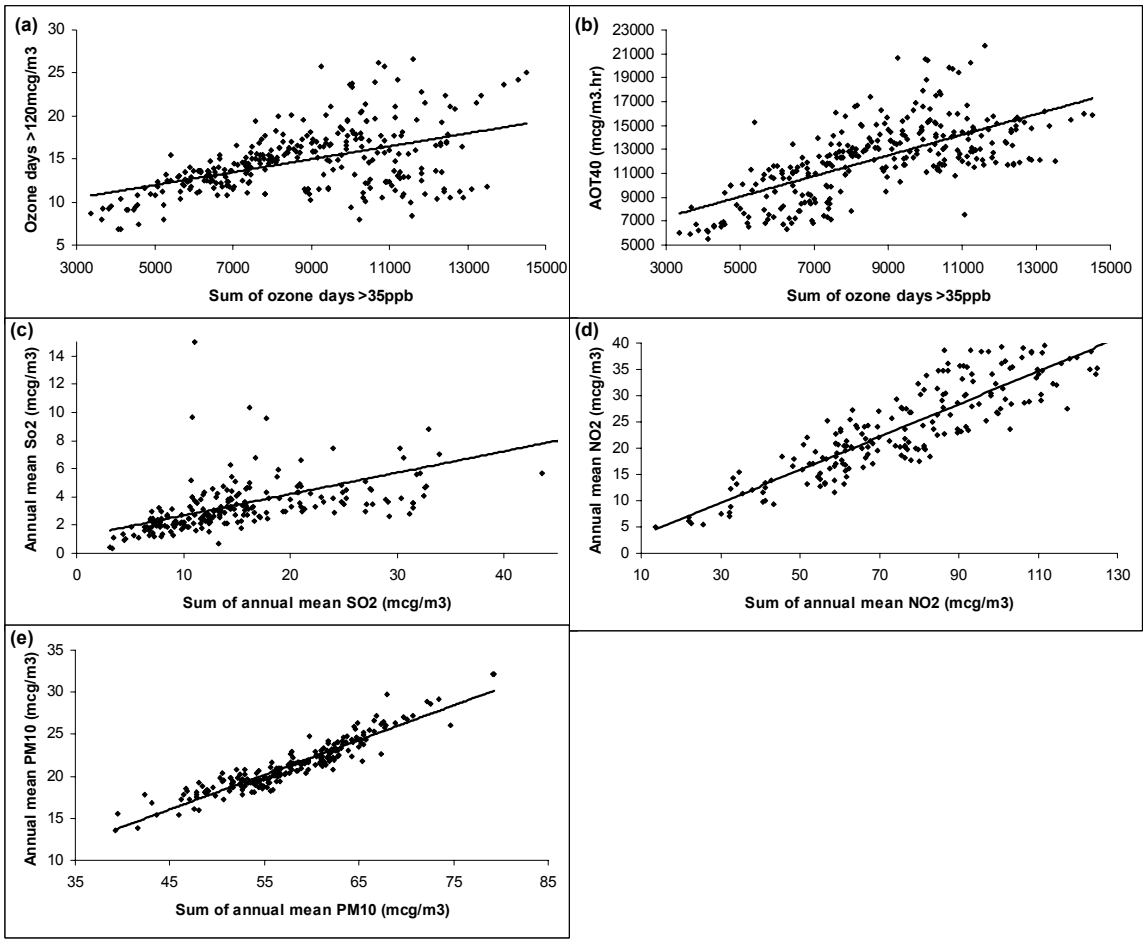


Figure 7.3: Correlations of 2006 pollutant metrics to cumulative metrics

The figure shows correlations between (a) 2006 AOT40 and (b) 2006 ozone days >120μg/m^3 with cumulative ozone >35ppb over 1993, 2003 and 2005. Both are significant at $p<0.0001$, the r values being 0.68 and 0.44 respectively. (c), (d) and (e) show correlations between 2006 levels of SO_2, NO_2 and PM_{10} with a similar cumulative measure, the r values being 0.51, 0.88 and 0.94 respectively. All are significant ($p<0.0001$).

7.4.2 Pollution and decline of lung function

Clinical determinants of lung function decline

The characteristics of the subjects included in this part of the study are shown in Table 7.7. All features shown in the table were compared between the whole group and the subset with most lung function measures. The only difference between them was in the distribution of occupational risk, where those followed up longer were generally working in higher risk professions ($p<0.01$). All other comparisons were non-significant ($p>0.05$). This indicates that if a subset of only those with at least 4 data points to calculate decline were studied there may be some selection bias. Some of the observed risk occupations are shown in Table 7.8.

Age, gender, smoke exposure, current smoking, exacerbation frequency, occupational risk and baseline lung function were examined as predictors of decline. More rapid decline of FEV_1 was associated with male gender (Mean (SE) -38.6 (8.7) v -23.4 (9.7) ml/yr, $p=0.02$) and higher baseline FEV1 ($r=0.-22$, $p=0.003$, Figure 7.4). Rapid decline of KCO was associated with higher baseline KCO ($r=-0.20$, $p=0.02$, Figure 7.5). Occupational risk did not predict FEV_1 or KCO decline ($p=0.50$ and 0.64 respectively). Correlation coefficients for comparisons of continuous variables are shown in Table 7.9.

	Study subjects n=399	Subgroup with at least 4 lung function data points n=218
Subjects of male gender %	61.2 (244)	62.84 (137)
Mean age in years	50.12 (0.17)	51.09 (0.70)
Current smokers %	8.7 (35)	7.80 (17)
Pack years	16.89 (0.24)	13.50 (1.25-24.88)
Index cases %	74.1 (296)	79.6 (174)
Currently working %	67.3 (268)	70.8 (154)
Level of occupational risk in working people %	High = 10.5 (28) Low = 20.4 (55) Zero = 69.1 (185)	High = 18.2 (28) Low = 32.5 (50) Zero = 49.4 (76)
FEV1 %predicted	51.28 (0.50)	52.04 (36.00-78.93)
KCO %predicted	73.31 (0.45)	68.47 (55.82-84.91)
Decline of FEV1 in ml/yr	-32.66 (2.06)	-34.84 (3.12)
Decline of KCO in [mmol/min/kPa/l]/yr	-0.03 (0.004)	-0.02 (0.007)
Exacerbations per year	1.00 (0.33-1.67)	1.00 (0.33-1.67)

Table 7.7: Characteristics of subjects assessed for decline of lung function

Data for the whole group and for the subset with 4 data points for lung function are shown in the 2 columns, and is expressed as mean (SEM) for normally distributed quantitative measures and median (IQR) for non-normally distributed quantitative measures. Categorical measures are shown as a percentage of the group, with the raw numerical data in brackets. The only statistically significant difference between the two groups was in the distribution of occupational risk ($p<0.01$).

Risk level	Occupation	Number of subjects
High	Metal working machine operatives	4
	Builder	4
	Metal welder	3
	Motor vehicle assembly worker	2
	Collier	2
	Wood machinist	2
	Glass making	2
	Printer	2
	Service engineer	2
	Motor mechanic	2
	Farmer	1
Low	HGV driver	6
	Bus driver	4
	Van driver	3
	Glazier	3
	Heating engineer	3
	Electrician	3
	Maintenance engineer	2
	Docker	2

Table 7.8: Risk professions observed in the longitudinal model

The table shows all high risk professions observed in those currently working within the group included in the longitudinal study, and all low risk professions occupied by more than one subject. All subjects shown here were part of the subgroup who had at least 4 data points for lung function.

	FEV_1	KCO	Age	Pack years	Exacerbation frequency
FEV_1 decline	**-0.22**	0.15	0.03	-0.03	-0.05
KCO decline	0.04	**-0.20**	0.07	-0.02	-0.03

Table 7.9: Correlation coefficients from univariate models

Significant correlations are highlighted in the table by the use of bold type, and were present for FEV_1 decline with baseline FEV1 ($p=0.003$), and between baseline KCO and KCO decline ($p=0.02$).

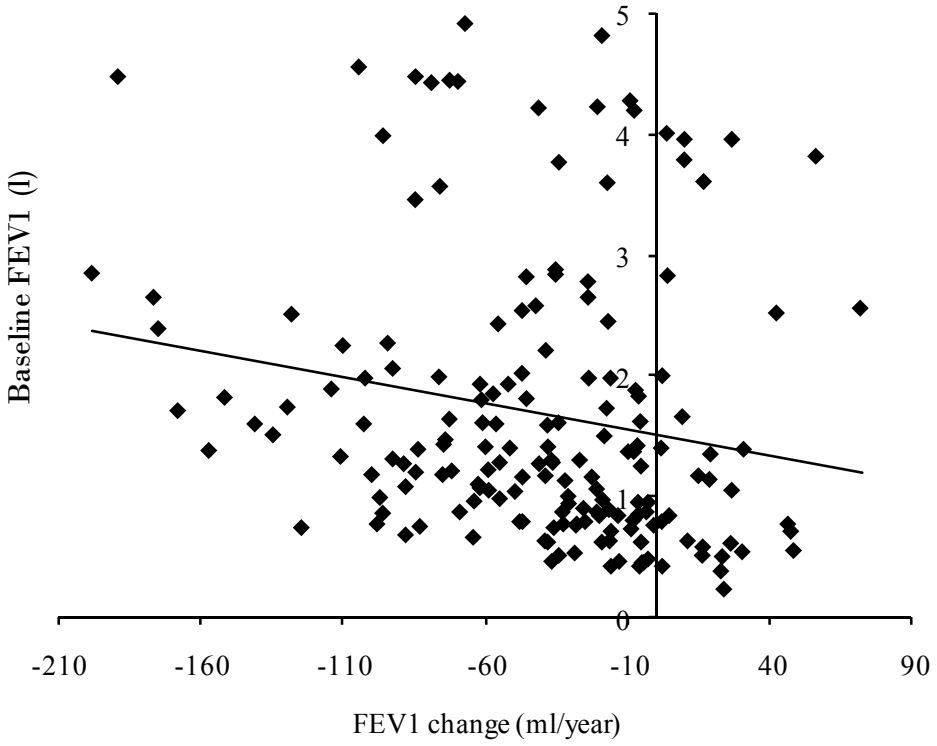


Figure 7.4: FEV1 decline relates to baseline FEV1

The graph shows the correlation between FEV1 decline and baseline FEV1 - higher baseline FEV1 is associated with rapid FEV1 decline (r=-0.22, p=0.003).

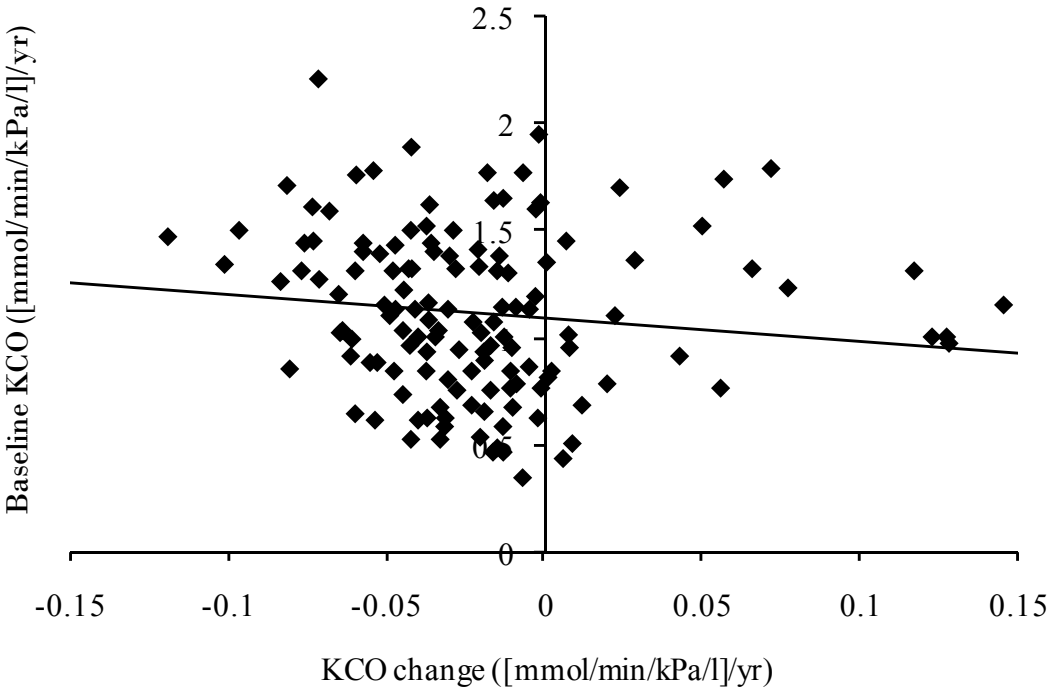


Figure 7.5: KCO decline relates to baseline KCO

The graph shows the correlation between KCO decline and baseline KCO – those with higher baseline KCO tend to have faster subsequent decline (r=-0.20, p=0.02).

Pollution and lung function decline

Full data on SO_2, NO_2 and PM_{10} was available, such that exposure could be calculated for all subjects; cumulative and mean data is shown in Table 7.10. All measured pollutants showed significant variability over time (all $p<0.0001$), the trend being to lower levels of SO_2 and NO_2, but higher levels of particles and ozone in the most recent year. Annual AOT40 data has been collected since 2003, so exposure to this metric could only be calculated for those subjects whose period of follow up was after this date, such that only 42 subjects could be studied. Over half of the subjects had decline calculated using 4 data points over 3 years, however this was not the case for all, hence the cumulative exposure for the decline period could not be compared between subjects without the use of generalised estimating equations, which take both this and the repeated measurement of lung function into account.

In the multivariate analyses performed using generalised estimating equations male gender and baseline FEV1 remained predictors of FEV_1 decline ($p=0.02$ and 0.01 respectively). There was no association with any of the pollutants studied (all $p>0.02$). Baseline KCO associated with more rapid KCO decline ($p=0.01$), as did PM_{10}, NO_2, SO_2 and ozone ($p=0.03$, <0.001, <0.001 and 0.02 respectively). The results from the generalised estimating equations are summarised in Table 7.11.

	SO_2	NO_2	PM_{10}	AOT40
1997	7.81 (0.34)	34.06 (0.71)	18.25 (0.17)	
1998	6.26 (0.25)	31.01 (0.63)	16.92 (0.13)	
1999	5.01 (0.21)	29.71 (0.62)	16.31 (0.11)	
2000	4.35 (0.19)	27.89 (0.61)	15.91 (0.12)	
2001	3.75 (0.11)	26.07 (0.60)	15.51 (0.13)	
2002	3.82 (0.16)	20.87 (0.51)	14.98 (0.14)	
2003	4.13 (0.16)	23.81 (0.55)	17.56 (0.14)	9887.97 (219.51)
2004	3.47 (0.15)	18.25 (0.50)	20.43 (0.22)	3236.36 (78.84)
2005	3.38 (0.13)	18.88 (0.51)	22.04 (0.18)	4335.35 (103.81)
Cumulative exposure over decline period	10.99 (8.08-16.32)	78.36 (2.28)	52.21 (46.22-59.52)	19553.09 (815.81)
Mean exposure per annum	3.48 (2.60-4.87)	24.13 (0.64)	16.82 (15.19-18.66)	6517.69 (271.94)

Table 7.10: Pollution exposures in decline cohort

Data are shown as mean (SEM) or median (IQR) dependent on its' distribution. All pollutants are measured in $\mu g/m^3$ except AOT40 ($\mu g/m^3.hr$)

		B	p value
FEV_1 decline	Baseline FEV1	0.02	0.01
	Gender	-0.17	0.02
KCO decline	PM_{10}	-0.001	0.03
	SO2	-0.006	<0.001
	NO2	-0.002	<0.001
	Ozone	-0.00002	0.02
	Baseline KCO	0.02	0.01

Table 7.11: Multivariate models of lung function decline

The table shows significant predictors of decline from the generalised estimating equations. The B value is the regression co-efficient from the equation. A negative value indicates an association with faster decline, a positive one with slower decline. Since males were coded as 1 and females 0, this means that male gender was associated with faster decline, whilst higher baseline lung function values and all pollutants were associated with faster KCO decline.

7.5 Discussion

UK pollution data have been collected for over 60 years, though the range of pollutants measured, and the number of measuring stations, has increased with time[339]. This (together with sophisticated statistical modelling methods that can take account of meteorological effects, dispersion from large sources and roads in addition to background pollution levels) has allowed the generation of GIS maps, which resolve pollution levels on a 1km by 1km grid across the UK. The modelling data have been validated by those responsible for map generation and published elsewhere[340, 341]. The validation data show good correlations between modelled and measured data at either networked or verification sites or both for NO_2, PM_{10}[341] and ozone[340], with r^2 values for the correlations lying between 0.63 and 0.95. SO2 showed less robust correlations outside network sites[341] probably because the relative contribution of point sources (such as coal burning power stations) is reducing, which makes the importance of background levels greater than previously. Less is known about these background sources of SO_2 than for the other pollutants considered in this chapter, which may have influenced the accuracy of this GIS map.

Extensive longitudinal data exists for some, but not all, pollutants, and for recently added datasets (such as $PM_{2.5}$) relatively few data exist. However, levels of most pollutants in the UK have decreased gradually with time[339] and their spatial distribution remains relatively constant[333]. The longitudinal data analysis herein concurs, showing a good relationship between single and cumulative years. This was the main justification for the cross-sectional study, but has the potential for error in exposure

assessment. Whilst the longitudinal study of decline uses more accurate exposure than the cross-sectional work, and could therefore give a threshold level for an adverse effect, there was insufficient data on ozone to study this pollutant well. The two studies must therefore be viewed as complementary.

Cross sectional model

The clinical data on covariates and clinical attributes concurs with previous studies that increasing age[149] and cigarette smoke exposure[17] influence phenotype. The pollution data also concurs with known UK data[207] (in terms of correlations and spatial distribution), giving a degree of validation to the measurements used in the pollution analyses.

There was no significant effect of occupation despite a high prevalence of at risk professions. This is a relatively small dataset to confidently exclude an effect of occupational exposure, and it is possible that ascertainment bias influenced results, since higher risk professions for the studied dataset were more common than in the general UK population. An arbitrary workplace exposure limit of 30% (used in most studies of occupational risk) was set and lifetime risk graded based on any prior or current level of exposure using a matrix. Whilst misclassification is possible with matrices[342] it is a method which could be used more widely to correct for workplace exposure in epidemiological studies. More detailed quantification of risk would require interviews and/or contemporaneous data on risk agents in each work environment. For large scale studies, as needed to assess pollution effects in the general population, this level of risk

measurement would be prohibitively complex and expensive.

The cross sectional study results illustrate the difficulties of assessing pollutant effects on disease. Firstly, the most significant associations are seen with KCO and CT densitometry, neither of which are widely measured in epidemiological studies, nor in routine clinical assessment of COPD[8]. Although CT densitometry seems most sensitive at detecting disease effects, it is perhaps more realistic to emphasise the need to assess KCO in subjects with COPD, especially as it tends to reflect upper zone emphysema (common in usual COPD) better than FEV_1[18]. Secondly, since all measured pollutants correlate closely, study results can be difficult to interpret. However whilst single pollutant models, such as ours, may be thought limited, multi-pollutant models do not necessarily improve either the direction or degree of associations between ozone exposure and health end points (H.Walton, Health Protction Agency, UK; personal communication). Different time windows of air pollution exposure, or differences in the impact of absolute levels compared to changes in exposure levels, were not possible to assess, though such analyses may be more relevant to a fluctuating condition, such as asthma, rather than a relatively fixed condition, such as COPD. Finally, as with any long term study, it is possible that the data represents a survivor bias.

Many measured pollutants have been associated with respiratory disease in epidemiological or toxicological studies. Particulate matter contributes to lung disease predominantly via pro-inflammatory effects, though oxidative stress also plays a role [44, 343]. Epidemiological associations of NO_2 with admissions and mortality[344] are unlikely to be due to the gas itself, rather to a correlated particle metric such as particle numbers[345]. SO_2 has been associated with bronchoconstriction in asthmatics[346] and

mortality[347]. Thus at first glance our regression results for these pollutants do not appear to make sense, as higher levels are associated with better lung function.

There are, however, a number of interpretations of these results. Firstly, that the apparent effects of high levels of NO_2, SO_2 and particles could be attributed to concurrently lower levels of ozone, consistent with its patho-physiological effects. This interpretation assumes that current levels are proportional to total exposure, which is supported by the data (Figure 7.3). An alternative interpretation could be that this model is insufficiently representative of long term exposure to other pollutants to detect their effects. Finally indoor exposures may be more important in our cohort, since the UK population in general spends most of its' time indoors. Indoor exposures comprise both indoor sources (such as from fuel combustion) and pollutants diffusing in from outdoors. The extent to which the latter occurs depends on the penetration coefficient, ventilation and decay rates[348]. Indoor exposure could therefore only be included if measured prospectively for each individual. However, for ozone there are no relevant indoor sources, and indoor concentrations are likely to be reflective of outdoor concentrations, so the findings related to this pollutant are unlikely to be affected by daily habits pertaining to indoor exposure.

Ozone is known to induce airway inflammation, impair host defence to bacterial insults, decrease macrophage activity, impair mucociliary clearance and enhance bronchial hyperresponsiveness[349]. Any, or all, of these effects could contribute to the associations with lung function observed in our cohort, and it is notable that most are pathological processes accepted to be important in the pathogenesis of COPD. AAT levels in the lung rise within an hour of exposure to ozone [47], and provide over 80% of the

lung's anti-elastase defences in models of ozone induced inflammation [350]. This suggests a rationale for enhanced susceptibility to ozone effects in AATD, and lends support to elastase driven pathways being of most importance in the effects of this pollutant. It is conceivable that in AATD this is so pronounced that any effects from other pollutants are overwhelmed. Indeed, the results reported here suggest relative protection from lung disease in the presence of low ozone levels, manifesting as a protective effect of high levels of other pollutants because of the direction in which they correlate.

Ozone is a secondary pollutant: the level varies as it depends less on emissions and more on meteorology and atmospheric effects[339]. Since ozone is scavenged by NOx any intervention controlling NOx, although beneficial in reducing NO_2, will also tend to increase ozone. This, together with a rise in background ozone due to air travel across the Atlantic, and predicted changes in the UK climate (likely to make it warmer and sunnier) underlie projected increases in ozone over the next 15 years[351, 352]. This makes it more critical to investigate health effects of ozone, and highlights the importance of the observation described here.

Longitudinal model

This model assesses factors influencing the decline in FEV_1 and KCO with particular reference to ambient air pollution. It may be considered more accurate than the cross-sectional model, because it could account for cumulative exposures reliably, by specifying the exposure window as the period over which lung function decline was calculated for each individual.

In order to interpret the data consideration of the methods used to calculate decline is important. FEV1 has a variability of about 100ml between repeated measures[353] which exceeds the average annual decline of FEV_1 in AATD. For this reason a simple calculation of decline based on two annual measures would not be sufficiently sensitive to determine the real change. However the standard error of the measure is inversely proportional to the number of years over which it is measured, such that over 3 years the error of the slope reduces to 33.3ml/yr, which is less than the decline described in this group of patients elsewhere[175]. In the current study decline was roughly equivalent to this, however the addition of two more data points to this slope reduces the error further (to approximately 11.1ml/yr), so that the calculated decline should be reliable for any given individual where 4 data points are used. This subgroup is by definition smaller in number than the whole group, thus there is loss of power, and selection bias could have been an issue. Gas transfer shows a greater degree of variability in its measurement than FEV_1[175] but is a more sensitive measure of decline in AATD than FEV1[354]. The same methodology was chosen for gas transfer as for FEV_1, in order to minimise error whilst maximising the number of subjects studied.

Comparison of the subgroup with at least 4 data ponts for lung function and the whole group did not show any significant differences in their pattern of lung disease, in that lung function measures were similar. However those in the subgroup that had at least 4 data points from which to calculate decline tended to work in higher risk professions than the whole group. This indicates some selection bias present – it may be that those working in high risk professions are more motivated to continue follow up, perhaps because they are used to regular assessments as part of occupational health follow up in their workplace, or

because of greater concern regarding their respiratory health. Conversely it might be that some of the higher risk professions lend themselves to being self-employed, such that subjects can choose more easily when to take time off, thus can attend the appointments offered. Since subjects travel from across the UK to the registry it will often involve a whole day off, which those employed by a company, or in the public sector, might find more difficult to do. Such conclusions can only be speculative in nature, but could be followed up with questionnaires administered to subjects within the registry, either by post or by telephone, to enquire about access issues and occupational health involvement at work. To maximise power the analyses were carried out in the larger group before moving to this subgroup. Pollution and clinical predictors of decline in the subgroup with at least 4 lung function measurements were no different from the whole in terms of direction or significance, suggesting that selection bias would not have influenced conclusions. In the subgroups defined by index status the results from the multivariate analyses were similar once again. An exception to this is ozone, where few measures were available from which to calculate exposure, such that the whole group could not be split and used in generalised estimating equations because the minimum number for the equations to work in SPSS was not reached in non index subjects.

FEV_1 decline was associated with higher baseline FEV1, consistent with the fact that decline tends to be fastest in those with moderately impaired, rather than severely impaired or normal FEV1[238]. Male subjects tended to have smoked more, which may account for the gender association seen, as smoking is usually associated with faster decline in AATD[152]. We did not find any associations between ambient pollution and FEV_1 decline, which may be because other measures are more reflective of disease in

AATD, or because the study was underpowered. A previous population based study of FEV_1 decline found an influence of PM_{10}, with higher levels being associated with faster decline. This was determined from the use of two spirometric measures 10 years apart, such that the methodology would be accurate to within 10ml/year during the follow up period [355]. However, differences in FEV_1 decline attributable to PM_{10} in the multivariate models were less than 10ml, so it is possible the results represent natural variation.

KCO decline was predicted by baseline KCO which is perhaps intuitive, and in line with the association of FEV1 decline with its baseline value. All of the outdoor air pollutants studied were associated with KCO decline after adjusting for covariates. In AATD the predominant pathology is parenchymal disease[17], such that changes in KCO may reflect disease progression better than FEV1. Thus our results could concur with previous work [355] that particulates contribute to lung function decline, the difference in the parameter associated being reflective of the disease susceptibility and pathophysiology of the subjects studied. It is also suggestive of other pollutant effects – in the cross-sectional direct linear relationships between all outdoor air pollutants were shown (Figure 7.2) such that we would expect both NO_2 and SO_2 to be high in the presence of high PM_{10}, but ozone to be lower[356]. PM_{10} is associated with airway inflammation[44], acute changes in lung function[357] and mortality[319, 320] whilst ozone [6,7] and SO_2 [4,8] are associated with hospital admissions for respiratory disease and lung function. In the cross sectional study of pollution effects, association was seen between ozone and markers consistent with parenchymal lung disease[356]. However, the cross sectional study of pollution and occupation contained herein was smaller because it considered only subjects who had

never moved from their place of birth, in order to provide a model of lifelong pollution exposure. Also, although linear relationships were present between current and past pollution exposures to validate the model, this could be considered flawed, since it could not account for true cumulative pollution exposure, unlike the subsequent longitudinal model. The results for KCO decline support the long term exposure model in concluding that ozone is important in determining lung function in AATD, but also suggests that any pollutant associated with inducing lung inflammation could have an influence. This is more consistent with pathophysiology than the cross sectional model.

As for the cross sectional work, the study is limited somewhat by patient numbers. It is also limited by the lack of available longitudinal data pertaining to ozone, which may be of particular importance in AATD, since AAT provides approximately 80% of the lung's protection from ozone induced inflammation [350]. A final point to note is that the analyses have not been corrected for multiple statistical testing. This is because both FEV1 and KCO decline correlate with one another, as do the pollutants studied. A Bonferroni correction would not be appropriate under these circumstances.

In summary these data show a significant association between exposure to outdoor air pollution of all types and disease progression in AATD, results that concur with population based work, and highlight the importance of measurements other than FEV_1 in the follow of patients at risk of COPD.

7.6 Conclusions

Ozone and other outdoor air pollutants may be of importance in determining disease severity and progression in AATD, but further prospective studies are needed to clarify the associations described in this chapter.

Chapter 8: Discussion

8.1 Summary of results

This thesis concerns genetic and environmental influences on clinical phenotype in AATD. As such the results can be split into those describing observed clinical features, those describing genetic associations and those describing environmental factors. The interpretation of the results in the context of current knowledge of COPD is considered using this format.

8.1.1 Clinical phenotypes

All of the studies reported in this thesis have relied on detailed clinical phenotyping of subjects with AATD. The results concur with previous reports[17] that indicate a range of phenotypes occur in people with AATD, with basal emphysema being the most commonly observed. In all of the groups of subjects studied here, the majority had impaired FEV1 and KCO, which usually reflects emphysema (Tables 3.1, 3.2, 6.1, 7.1 and 7.7). This was confirmed on CT in most subjects, with a higher lower zone than upper zone voxel index being the most common pattern, showing that the disease was more severe in the basal areas of the lung (Tables 3.1, 3.2 and 6.1). This is in contrast to usual COPD where emphysema is usually worse in the upper zone [358].

Chronic bronchitis occurred in around 40% of subjects (Tables 3.1, 3.2 and Figure 3.3); a significantly lower proportion than reported in a recent study describing usual COPD sub-phenotypes, where it was seen in 72% of subjects with spirometrically defined COPD

[359]. This may be because of differences in reporting between the subjects with usual COPD in New Zealand [359] and those with AATD in the UK. Alternatively, subjects with AATD could exhibit different airways pathology to COPD, a concept worthy of further research. In general airway inflammation is thought to be more severe in AATD[201], which intuitively would suggest a higher risk of chronic bronchitis. However, it is possible that this is not reflected solely by sputum production and that other measures of airways disease may be helpful in future. The radiological assessment of airways disease has improved markedly over recent years[360] and now includes more subtle markers, such as airway wall thickness, a feature whose clinical significance is now being studied. However, since airway wall thickness relates to FEV1[361] and smoke exposure[362], and airway remodelling is an accepted part of pathophysiology[11], it is likely that such measurements also reflect part of the disease process.

Bronchiectasis was the predominant lung pathology in 24% of subjects (Figure 3.3), and age was a significant predictor of its presence (Table 3.6). A possible explanation for this is that bronchiectasis may develop because of repeated infections, and hence the effect will accumulate with time. Detailed scoring systems to describe bronchiectasis on CT in AATD were published subsequent to the clinical data collection part of the studies reported here[22]. When such systems are used to analyse the CT, bronchiectasis is present (though not predominant) in a much larger proportion, and tends to co-localise with emphysema[22], which might support the hypothesis of bronchiectasis developing in diseased, infection prone areas. In usual COPD, bronchiectasis has been reported in 30 to 50% of subjects, and may be associated with longer periods of critical illness[363]. Widely accepted diagnostic algorithms for reporting of bronchiectasis on CT are needed before

reliable comparisons of incidence between AATD and COPD can be undertaken. Few longitudinal studies of airways disease on CT scan have been performed, but these will be central in determining if subtle airway wall changes predict an infection prone phenotype or precede bronchiectasis. In asthma a recent longitudinal study has reported gradual accumulation of airway wall thickening, and bronchiectasis[364], and in COPD changes in airway wall thickness have been shown to relate to FEV decline[365]. Neither study reported data specifically relating airway wall thickness to bronchiectasis. Indeed, in cystic fibrosis the two pathologies are distinct, and progress separately [366]. Further research is needed to clarify these issues in COPD.

Lung function decline was studied in a subset (n=110) of the unrelated case-control group who participated in the genetic studies reported here, and a larger group (many of whom had no stored blood for DNA extraction) who participated in the pollution study. In the latter work there were 220 subjects who had at least 4 data points over which to calculate decline, and a further 181 with 2 or 3 data points. In general, those participating in the genetic studies were recruited earlier in the evolution of ADAPT, though not exclusively so. Decline of FEV1 was 42.00ml/yr on average in the genetic study group, 34.84ml/yr in the 4 data point pollution group, and 32.66ml/yr in the whole pollution group. The difference between the latter two was non-significant. It may be that decline was faster in the genetic group because they had smoked more heavily than the majority of registered subjects (Tables 3.1 and 5.2), or their earlier recruitment date and the more limited range of treatments available previously. This level of change over time is, however, better than that reported in previous studies of AATD[367] and of usual COPD[368], being much closer to that of a healthy individual. This may reflect good

management of subjects' lung disease which has effectively stabilised decline or inclusion of more healthy, non-index subjects in the analyses. Alternatively it may be that the measures reported here are more accurate than others reported previously, because of the number of data points available, and the strict quality control measures employed in the lung function laboratory in Birmingham. Gender was a significant predictor of FEV1 decline in both the genetic and pollution groups (Tables 3.6 and 7.5), even after adjustment for smoking. This suggests gender specific differences in the pathophysiology of COPD, possibly a consequence of gender differences in immune response [369]. Oestrogen is also recognised as a protective factor in acute lung injury, sepsis and myocardial infarction, where effects are thought to be mediated by differences in vasomotor tone, resulting in relative protection from hypoxic damage in females[370]. Oestrogen metabolites may also interact with components of the oxidant-antioxidant system in some tissues [371]. All of these mechanisms are potentially relevant to COPD, but further research is needed to clarify why gender differences occur.

Decline of KCO was very similar in the genetic group and the 2 pollution groups (0.03, 0.03 and 0.02 [mmol/min/kPa/l]/yr respectively, Tables 5.1 and 7.7). This has recently been reported to be a more sensitive measure than FEV1 to measure decline in AATD[159]. Since some of the subjects in this study were from the UK, and indeed participated in the studies herein prior to the augmentation therapy described by the authors, it would be inappropriate to make comparisons between the groups as they are not truly independent. No studies of this feature in usual COPD have been published. Hence the data here are novel and may help to inform future work in the area.

Bacterial colonisation of the airway was the final clinical phenotype assessed in the

studies reported here. It occurred in the stable state in 70% of subjects who produced sputum, with mixed normal flora alone in 20% and potential pathogens in 50% (Table 4.10). Colonised subjects did not differ from those who were not (Table 4.11), though all exhibited significant lung function impairment and emphysema on CT. The incidence of potential pathogens in the stable state has been studied widely in usual COPD, with varying results. Incidence ranges from 33%[372] to 70.9%[373] in recent studies. The bacterial species' observed also show significant variation, though *Haemophilus influenza* is often the most commonly seen [276, 373, 374], as in the AATD group reported here. *Pseudomonas spp* were uncommon in this AATD group, but have been highly prevalent in some COPD cohorts [373, 375], where selection may have influenced the results. The AATD group studied here are, however, similar to usual COPD in their incidence and type of bacterial colonisation.

8.1.2 Genetic associations

When the studies contained in this thesis were designed, published genetic association studies in COPD were generally relatively small, variable in their level of phenotyping and poorly replicated. Some exceptions to this rule were seen, notably studies of the Boston Early Onset COPD cohort and the NETT cohort [72, 87, 88, 108, 376, 377]. Although the groups studied for genetic association in this thesis were also relatively small, particularly the familial group, the detailed level of phenotyping was, at the time, unique. The aim of the studies in general was to dissect specific sub-phenotype associations, in order to lend support to the hypothesis that COPD is a heterogenous group of conditions, and

determine which aspects of pathophysiology were most important in each sub-phenotype. Case-control with familial replication was chosen as the study design, since no other AATD cohort with this level of phenotyping, and stored DNA, was available for study. It was, therefore, necessary to accept, at the outset, that there were weaknesses including probable under powering of both the test (Tables 3.5, 4.3, 4.5 and 4.8) and replication groups to detect associations confirming the level of OR seen in complex disease, and lack of a replication group for some phenotypes. In an attempt to minimise the effect of these issues on the studies, relatively well replicated loci (*TNFA, SFTPB, GC, TGFB, MMP1*), and those with prior evidence in AATD (*MMP1, 3* and *12*) were chosen for the candidate gene work. This was combined with a focus on those phenotypes most likely to relate to each protein product, from current knowledge of pathophysiology. Issues of power and replication have also been discussed in sections 3.5 and 4.5.

Since the start of my work progress in COPD genetics has been rapid, and genome-wide association studies (GWAS) are now being performed. These have the potential to discover associations in areas not previously considered important in COPD, implying pathophysiological roles for their protein products, and thus extending our understanding of the condition more than candidate gene work. Disadvantages of GWAS include the large numbers needed for adequate power (probably at least 3000 cases and controls[217]), the difficulty of interpreting associations in intergenic areas and high costs. Problems with reproducibility of results may also occur if different genotyping platforms are used, such that different SNPs in a given area are typed, whose LD is either unknown, or incomplete. The expansion in GWAS in all diseases has occurred in part because of technological advances in genotyping, with consequent improvements in both feasibility and cost. Older

genotyping methods, such as traditional gel based techniques (like the HLA typing performed for the studies in chapter 6), are time consuming and require a lot of DNA. Fluorescence based techniques, such as TaqMan®, are more reliable at low DNA concentrations and lend themselves better to high throughput. Advances in statistical programming for analysis of large datasets have also aided progress towards GWAS. Nevertheless, determining statistical significance remains a hotly debated subject and it is still unclear if correction for multiple statistical testing, particularly in areas of high LD, is appropriate (an issue discussed in section 6.5) even when methods exist to do so[378]. In parallel with these changes in genetics it has become widely accepted in respiratory medicine that COPD constitutes many sub-phenotypes [359], that quantitative CT is important in defining them [361] and that genetic associations may not be meaningful without full clinical data pertaining to delineate parenchymal and airways disease. Large well characterised groups of subjects with usual COPD are now being collected, which will have adequate power for GWAS, such as ECLIPSE (Evaluation of COPD longitudinally to identify surrogate end points) [379] and COPDGene®[380].

One GWAS based on the Bergen cohort, with replication in the Boston Early Onset COPD, International COPD Genetics Network (ICGN) family cohort[90] and NETT cohorts has now been published, in which 3 loci have achieved genome-wide significance [73]. From this it is possible to be reasonably confident of association with COPD for 2 SNPs (rs8034191 and rs1051730, both $p<1x10^{-9}$) within the α-nicotinic acetylcholine receptor (*CHRNA3/5*) and 2 SNPs within hedgehog interacting protein (*HHIP*, rs1851851 and rs13118928, both $p<2x10^{-7}$). The risk alleles within *CHRNA3/5* conferred a mean OR of 1.32 for disease, whilst the 2 SNPs within *HHIP* exhibited a protective effect with a

mean OR of 0.72[73]. The findings for the *HHIP* loci are supported by a GWAS examining lung function in the general population [381]. The functional significance of these associations is not yet clear – the associations with *CHRNA3/5* may represent nicotine addiction effects (though the authors attempted to control for this in their analyses) or a role for non-neuronal nicotonic signaling in the lung[382]. This relates to immune function via lymphocytes[383], which are known to be present in the lungs of subjects with COPD[11]. *HHIP* encodes a signalling molecule (HIP1), which is present in most mammalian tissues[384] and interacts with other hedgehog proteins to control morphogenesis[385]. Its precise role in the lung is not yet known. This GWAS has, therefore, revealed at least one putative new pathophysiological process in COPD, and will require follow up with functional work. A second GWAS has been reported solely in abstract form, and found association in the ECLIPSE cohort and in the Bergen COPD cohort for several SNPs with a quantitive measurement of emphysema on CT scan [79]. The most associated SNP was rs615098 in *MMP3*, the combined p value from meta-analysis of the 2 groups (a controversial method) being 1.92×10^{-7}. The 2 genes flanking *MMP3* (*MMP1* and *MMP12*) have been the subject of candidate gene studies in COPD, so this is a less novel result.

A candidate gene study, using the same test cohort as the above GWAS, with replication in the Bergen COPD cohort, achieved statistical significance with the overly conservative Bonferroni correction[90]. Meta-analyses have also been published on many candidate genes [228, 386, 387]; both these strategies are accepted as proof of association. Loci from these non-GWAS with good evidence of association lie within *SERPINE2*[90], *TNFA*[228], *TGFB*[386], *GSTP*s[386], *IL1B*[386], and *EPHX1*[387]. The potential role

within COPD pathophysiology for most of these candidates and the function of associated SNPs has been considered in Chapter 1 (section 1.1.4).

It is in the context of this expanding evidence, and changing levels of confidence in their results, that the genetic associations reported here must be seen. One SNP in *TNFA* (rs361525) conferred an OR of 1.99 for chronic bronchitis (section 3.2.3), and an elevated level of its protein product (TNFα) was seen in airway secretions of those with the associated allele (Figure 3.11). Although it was not possible to assess for replication, this functional effect, consistent with known pathophysiology, together with supportive evidence in a recent meta-analysis (albeit in Asians)[386], suggests it is a true effect. It is not yet clear why a compartment specific genotypic effect is seen - the level of TNFα was not elevated in the plasma of subjects with the associated allele (Figure 3.10). Local differences in gene expression are a possible explanation, but further research will be needed. Minor allele homozygotes for rs2118177, in *SFTPB*, had an OR of 0.28 of COPD in the case-control dataset and the same genotype associated with higher FEV1 in the familial replication dataset (section 4.4.1). Since the direction of effect was replicated it is likely to be a true association, but the gene is not close to any of the GWAS peaks seen so far[73] and has not been studied in a meta-analysis, hence has little other evidence supporting this at present. It is possible the existing GWAS was underpowered to detect the effect, or that it is specific to AATD. Three SNPs in *TGFB* showed a suggestion of association with lung function, but this did not replicate in the familial group. However, this locus now has meta-analysis data to support a role for it in COPD, so perhaps warrants further analysis in a better powered replication cohort in AATD, to ascertain if the effects are truly absent in this population. The association seen between

bronchiectasis, airway bacterial colonisation and SNPs within GC in the case-control dataset could not be tested in the family cohort, and has to be viewed with caution, despite a biologically plausible role for its protein product in airway host defence. It should however prompt research into the role of the protein in the lung, which will be discussed in section 8.2.2. The association of the *MMP* cluster detailed in chapter 5 represented follow up of a prior linkage study in AATD, and although this is not true replication it is suggestive of an effect. It is also of interest that an area on chromosome 11, not vastly distant from this (approximately 3Mbp) appeared protective in one of the study populations in the recent GWAS [73], and that a SNP within *MMP3* was associated in a smaller GWAS looking at emphysema on CT[79] Further work on this region is indicated, given the relatively strong candidate gene work in the area (section 1.1.4), and the surprising lack of study in recent meta-analyses.

The association of an HLA type commonly seen in autoimmunity with lung function (Table 6.4), which remained significant after a Bonferroni correction for the locus is a novel finding. Genotype specific differences in anti-elastin antibody level (Figures 6.7 and 6.8) suggest a possible local gene-environment interaction, such that genetically predisposed individuals perpetuate the lung damage initiated by smoke because of autoimmune activity directed against lung elastin. In this case the environmental factor is the circulating elastin peptide, which interacts only with the associated HLA protein. In individuals with the risk genotype (*HLA-DQA**0301 or *HLA-DRB1**04) a high level of antibody results, and there is then subsequent perpetuation of elastin breakdown (Figure 6.9). Although some confidence in the association is gained from the Bonferroni correction, and putative functional effect, it is by no means certain. Methods for

confirming the association are considered in section 8.2.3. Further support for genetic susceptibility to autoimmune processes in COPD is taken from the report of *CTLA4* associations [388] in the ICGN and Bergen cohorts. These polymorphisms are associated with a number of traditional autoimmune diseases, including diabetes and Grave's disease [389].

8.1.3 Environmental factors

The most important environmental factor to influence type or severity of lung disease in AATD is cigarette smoking. This is well known[17] and is confirmed by the regressions analyses pertaining to both lung function and CT diagnosed emphysema presented in chapter 3 (Table 3.6). A role for other inhaled agents with the potential to cause lung inflammation was assessed in chapter 7. The results showed a consistent deleterious effect of ozone on lung function, whether a long term, cross-sectional model, or a medium term, longitudinal model was used (Tables 7.5 and 7.9). The cross-sectional model had several weaknesses, discussed in section 7.5. It is possible that the lack of effect of other outdoor air pollutants is due in part to this, since further possible harmful effects of PM_{10}, NO_2 and SO_2 were seen in the longitudinal model. This is the first cohort of AATD subjects to undergo study with regard to their risks from outdoor air pollution, and it is not yet clear whether the associations will be specific to AATD, or seen in usual COPD as well. Intuitively, since the pathophysiology of the respiratory problem is much the same in the 2 conditions, it is likely an effect will be seen. However, given that AAT provides the

majority of the lung's protection from inflammation [350], it may be that effects in usual COPD will be even more subtle and very large cohorts will be needed to have adequate power to detect them. Relatively speaking, cigarette smoke accounted for 6 times the amount of variability in KCO of ozone in the study reported here (section 7.4.1), so it remains more important to advise against smoking that advising on an ideal residence site for subjects with AATD.

Occupational agents may cause a range of lung diseases and were assessed in the studies reported here only by questionnaire based information about their job, or past work. This graded the possible risk of their work to the lung, but is not without its problems, as discussed in section 7.5. No significant associations between respiratory phenotype and occupation were seen in either long term (Table 7.2) or medium term models of this risk (section 7.4.2). It is, however, likely that the study was underpowered, as even common risk professions were observed in less than 10 individuals (Tables 7.3 and 7.8). Prospective studies of occupational risk in subjects with AATD are unlikely to be funded or carried out, since accurate exposure measurement would be prohibitively expensive, adequate numbers difficult to recruit and the potential benefits of such results to the population small. It will probably be of greater value to pursue such investigations in subjects with COPD, or in the general population.

8.2 Future directions

8.2.1 Replication dataset

As with all genetic studies a statistically significant result requires further independent validation. Therefore, the studies within this thesis will require replication not only in AATD, but also in COPD unrelated to AATD, particularly for those which were neither significant at genome wide level, nor replicated in the family dataset. A protocol has been designed to recruit such a dataset in usual COPD and ethical approval obtained (South Birmingham LREC, reference 07/H1207/231). The patient information leaflet and consent form are included in Appendix 1. Direct, reliable comparisons between usual COPD and the UK AATD cohorts will be facilitated by use of the same data collection techniques, and an identical database to that used throughout this project. Data collection for this cohort is now underway.

Replication in AATD is also necessary. The ADAPT programme continues to recruit subjects so this will be possible in the next 5 years. In addition, collaboration with the American group headed by Professor EK Silverman, who is also studying genetic modifiers in AATD, at Harvard University (Boston, USA) has also been arranged. This has an added benefit with regard to the HLA region, which has already been typed in their patient group, albeit by a different method. It will now be critical to cross validate the two methods in order to compare the results between the two patient groups. It should also be noted that differences between the 2 groups due to geographical variation could be present, and this will need to be taken into account in any replication work in the USA.

Ultimately a genome-wide association study could be performed in the above cohorts,

as the genotyping technology and statistical methods to interpret the results are now available. Commercially available chips capable of genotyping hundreds of thousands of SNPs for each individual have been available for some time, but are now falling in price, making such a study more viable. Many companies now provide software for integrated results formatting and analysis. It would be critical to have at least 2000 well characterised subjects for adequate power[217], for comparison to control subjects. An application could be made to the 1958 birth cohort[390] for DNA geographically and ethnically matched to the subject group, or it may be possible to request access to genotypes of controls already typed in other studies. The concept of using shared control subjects has been demonstrated in a large Wellcome Trust sponsored study in the UK[217]. A replication dataset in usual COPD would be the most sensible choice for replication. Genome wide studies have been performed in COPD [73, 79, 391], so it may be possible to approach those managing these collections for their genotype data, or to perform new genotyping where necessary. Since Professor Silverman has been closely involved with these collections, and is currently collecting even larger COPD cohorts in several populations worldwide a comprehensive replication study should be possible via this existing collaboration.

8.2.2 The role of protein products of associated genes in COPD

Since the majority of the studies contained in this thesis concerned candidate genes and their protein products, their role in disease pathogenesis has been studied in some detail,

most notably for TNFα. However, neither SFTPB nor GC has been studied widely in any COPD phenotype.

SFTPB

Levels of SFTPB are low in animal models of emphysema[252], so it would be useful to establish if this is the case in human subjects. Since the protein is likely to be of importance in the alveoli, rather than larger airways, the most appropriate samples in which to measure it would be broncho-alveolar lavage (BAL) fluid. Alternatively it might be possible to study the protein in tissue cultures and histological specimens obtained during lobectomies performed on emphysematous subjects for suspected malignant processes, or resections performed for lung volume reduction surgery. Tissue samples could also be used to extract RNA to analyse gene expression and measure product formation and compare to samples from subjects without emphysema. Ethical approval for collection of such tissue samples from subjects with and without emphysema, which could contribute to the above work, has recently been obtained (Midlands Lung Tissue Consortium, North West REC reference 07/MRE08/42; consent form located in Appendix 1) and local systems to facilitate sample collection are currently being put in place. It is also of note that another surfactant protein (SFTPD) has shown promise as a biomarker in COPD when measured in the circulation[392, 393]. Similar studies could be performed for SFTPB in stored samples from the AATD subject group, which could also be compared to associated SNP genotype to give an indication of the functional role of the variant.

GC

GC has a role in macrophage activation and neutrophil chemotaxis[267, 268], and whilst genetic variation has been shown to influence the former, effects on neutrophil chemotaxis have not been proven. New techniques in the measurement of neutrophil migration in COPD[394] mean that studies of chemotactic patterns can now be more sensitive. Fresh blood samples from patients of known *GC* genotype could be compared for the direction and speed of neutrophil migration towards a known chemoattractant, such as IL8. It will also be important to study this protein further in the airway, given the genetic association with bronchiectasis. The sol phase of sputum would be appropriate for this, and such samples are available for some of the AATD subjects already genotyped. Alternatively bronchial biopsies, lung tissue samples taken at resection or tissue cultures could be used for studies of gene and protein expression. Such a study should select groups with and without AATD, with and without bronchiectasis and other COPD sub-phenotypes in order to ascertain precisely the sub-phenotype in which the protein plays a role.

Vitamin D

Circulating levels of vitamin D relate to disease severity in asthma[395] and the vitamin D endocrine system is known to be important in animal models of the disease[396]. Genetic association with vitamin D binding protein suggests that a similar axis may be important in COPD. It would be necessary to compare vitamin D levels between sub-phenotypes in the AATD group, using plasma or serum samples (of which

many are available) and to relate the data to associated genotypes. This analysis, in combination with study of its binding protein's role might indicate which aspect of the vitamin D axis would be important to target in drug design in the future for therapies of COPD. Comparisons of exacerbation frequency or disease progression between subjects receiving either vitamin D or placebo would also be considered if studies of vitamin and GC levels suggest a biological role for vitamin deficiency in protection from infection.

8.2.3 Mechanism of disease in anti-elastin autoimmunity

The association of HLA and anti-elastin antibody levels with features of COPD suggests that anti-elastin autoimmunity may be of importance. However, this remains theoretical until the study has been replicated, and further functional work has been performed. At the most basic level the functional relevance of HLA type would need to be studied by assessing the binding properties of anti-elastin antibodies with the associated HLA proteins. Since the level of resolution of the technique used in the studies reported here is relatively low, it would be insufficient for studies of protein structure. A higher resolution typing method, such as the Dynal kit, would need to be used in a larger dataset to ascertain the associated genotypes and therefore associated protein products. Once this step is completed the relation of anti-elastin antibodies to these proteins could be studied further.

An alternative approach to obtain evidence of this putative mechanism would be to correlate anti-elastin antibody levels to elastin breakdown products. A desmosine cross-link is unique to elastin and may be used as a marker of its degradation[397]. Desmosine

and elastin peptides are elevated in the plasma, urine and sputum of COPD patients[398], whilst urinary levels positively correlate with the annual rate of decline in FEV1 in smokers[399]. These studies show the importance of tissue destruction in COPD, but not the reason by which it occurs. If antibody driven elastin breakdown is the key disease mechanism then antibody and desmosine levels would be directly related. The HLA association of antibody levels would suggest that this mechanism will only be of importance in those with associated HLA types, so an HLA typed group should be studied. Desmosine may be measured in a number of ways[397-401], the most sensitive is likely to be mass spectrometry[398].

Finally, antibody-elastin complexes could be sought in lung tissue, such as that obtained by the Midlands Lung Tissue Consortium. This may be possible by immunohistochemistry using formalin fixed tissue slices, or by immunofluorescence using frozen sections. Both techniques will be explored when the tissue becomes available.

8.2.4 Gene-environment interactions

If the AAT level is important in susceptibility to the effects of ozone in the lung, as suggested in toxicological models [47], then this could indicate a gene-environment interaction. This could be confirmed by studying pollution associations in subjects with different AAT levels, for instance studying PiSZ and PiMZ subjects together with PiMM controls. This should be performed using the same methodology as the studies contained in Chapter 7 in order to facilitate comparison of the results with those reported here.

It would also be possible to study other COPD associated genes with regard to gene-

environment interaction. The results here suggest that ozone and particulate matter are the most relevant air pollutants to consider in this respect. However, very large datasets will be needed to have adequate power, particularly if large numbers of SNPs are to be studied in the same subject group. Thus far genome-wide studies of gene-environment interaction have been limited by the exploratory nature of the statistical analyses needed.

The definition of interaction also needs clarification. Currently the main type of interaction considered is at the statistical level, such that interaction is deemed present if the effect of a genetic and environmental influence together deviated from the additive model. However, this does not consider interaction at the biological level (as suggested by the ozone results in AATD) a combination of statistical and biological interaction, or gene-gene interaction. Newer statistical methods employing neural network technology[402] may be able to overcome these uncertainties, assuming a sufficiently large cohort could be recruited to have adequate power. Such techniques might also be employed in pharmacogenetic studies[403], which are perhaps the ultimate gene-environment interaction in modern medicine, potentially allowing targeted treatment for every individual.

8.3 Conclusion

COPD is a complex group of diseases with many components contributing to the pathophysiology and consequent clinical manifestations. There are multiple genetic and environmental influences acting upon clinical phenotype, which are likely to be additive, or synergistic. In AATD genes shown to influence usual COPD show such additive effects, but they segregate to clearly defined clinical phenotypes. This gives some indication of disease mechanisms, and in the future may help to guide individual therapy.

Appendix 1: Consent forms and ethics

CONSENT FORM FOR PARTICIPATION IN RESEARCH 3359

Assessment for Alpha$_1$ Antitrypsin Deficiency

What is the study about?

You or a member of your family has been identified as having alpha$_1$ antitrypsin deficiency. This is an inherited condition that is believed to increase the risk of development of lung health problems. However very little detailed information has been collected on the way this deficiency affects patients and some studies have suggested that lung disease may run in families even without the deficiency. It is likely however that the deficiency highlights the tendency to develop these diseases and when present will make them worse.

We wish to learn as much as possible about the deficiency and its relationship to lung disease and for this reason invite you to participate in our alpha$_1$ antitrypsin deficiency assessment programme.

What will I have to do?

In broad terms you will undergo all the routine questioning, examination and tests that we normally undertake when assessing somebody who presents with lung disease. However we hope to do this more carefully and in more detail than is routinely carried out by your own doctor or specialist.

We will ask many questions about your past and present symptoms, health and wellbeing. In addition you will be examined thoroughly to determine the presence of signs related to lung disease. After this screening programme you will be asked to perform some lung function testing which assesses how your lungs work and their ability to take oxygen in and out of your body. We may also perform a specialised CT scan of your lungs (if you have not had one) which is a very sensitive technique of detecting damage that has occurred. Finally we will ask you to provide one blood sample and if you have a cough productive of sputum we will arrange for you to collect this over several hours on one day before coming to see us.

Once all these tests have been performed we will be able to determine whether you have lung disease related to alpha$_1$ antitrypsin deficiency. This will be explained to you and any modifications in your treatment that are indicated will be communicated to both yourself and your own doctor.

It is our general clinical routine to follow patients with established lung disease on a long-term basis. Patients are usually seen once every 4-6 months to assess their wellbeing and follow any progress in the condition. If you have alpha$_1$ antitrypsin deficiency we would

wish to see you once a year to assess your symptoms, clinical signs and repeat the extensive lung function tests and the CT scan to confirm the extent or any progression of your disease will be carried out every 2 years. After the first year the lung function may be repeated less frequently (2 or 3 yearly) depending on whether these are changing or are stable.

NOTE The CT scan exposes you to a small degree of radiation – about the equivalent of 6 months background radiation in the UK. Although this dose is safe (it is the same as a single x-ray of the abdomen), it is important that you inform us if you are likely to be pregnant as we will not carry out the test in these circumstances.

What are the benefits?

The major purpose of the study is to find out as much as possible about the lung disease associated with alpha$_1$ antitrypsin deficiency. This will provide the background information that enables us to design studies to assess the role of alpha$_1$ antitrypsin replacement therapy in both the short and long term. The investigations that we undertake will allow us to advise upon the degree of lung disease that you have and simple measures that you can undertake with your current treatment in order to try and stabilize the lung disease. In addition the breathing tests that we will perform will help us to optimize your current treatment in order to improve your breathlessness where possible.

What are the risks?

All the investigations that are taking place are entirely routine, used in the assessment of patients with lung disease. As such they are repeated on many occasions in the same patient without any adverse effects. The only minor problem that is likely to occur is a slight degree of bruising in some patients when they have their blood taken.

What are the alternatives?

There are currently no alternatives to finding the information that is required other than the assessment programme outlined above.

What happens if I do not wish to take part?

If you do not wish to participate in the assessment programme or the short term follow up programme this will be fully understood. Your own general practitioner, or the consultant chest physician who normally looks after you, will be informed of the diagnosis and provided with advice on how to assess and manage your follow up along the lines outlined above. If for any reason they or you require further advice from us in the future we will be only too pleased to see you. It is important to emphasize that your overall management by your doctor will not be affected by your decision.

What happens to the information?

The important information that we collect over the next year or two will be written up as a report and submitted to medical journals so that other doctors may read about the problem. The research information obtained from the samples that we have collected will be the basis for future studies on the role of alpha$_1$ antitrypsin replacement therapy which we hope to start within the next 12-18 months. Neither your name or any details relating

to you personally will be released to any other person outside the research programme.

Who is taking part?
All subjects that we identify with alpha$_1$ antitrypsin deficiency and in some cases members of their family will be asked to take part. At present we know of over 800 such patients as yourself and it is likely that there are several thousand similar people in the West Midlands alone. It is hoped over the next year or two to identify and assess all people with the deficiency in the Midlands.

What if something goes wrong?
The question largely relates to clinical trials and at present that is not part of the assessment programme. If you develop new symptoms for any reason during the assessment day this will be in the presence of a doctor who will take any steps that are necessary to help you. In between visits to the assessment centre your own doctor will be largely responsible for your care but (depending on where you live) we may collaborate with your doctor and help by seeing you if you become unwell.

What happens at the end of a study?
At the end of the study we will have learnt a lot about alpha$_1$ antitrypsin deficiency. If you remain part of the programme for assessment we will be having regular meetings for all patients to attend to discuss how the assessment programme is going and how our understanding is developing. If for any reason you are unable to attend we will be keeping in touch with you by letter and newsletter to inform you of the progress of the programme.

What if I have more questions or do not understand something?
The doctors, co-ordinator and nurse involved in the alpha$_1$ antitrypsin assessment programme will happily answer your questions on any occasion when you visit. If questions arise between visits you will be able to contact the centre and either speak to somebody at that time or arrange to do so if for any reason if it is inconvenient.

What happens now if I decide to take part?
If you decide to take part in the programme now we will arrange an appointment in the not too distant future for you to come to the assessment centre for the investigations outlined above. This will be arranged to suit everybody's convenience and all the assessment will be completed where possible on a single visit.

What happens if I change my mind during the study?
If you change your mind during the study, it is important that you notify the assessment centre. This will enable any investigations that have been organized or visits to be cancelled. Your decision will be passed on to both your own doctor and where appropriate your own specialist in order that they can arrange for appropriate appointments to monitor your progress. Providing you are agreeable we would like to contact your own doctor or specialist from time to time in order to find out how you are progressing. However if in the future you once again decide to join the programme, we would be only too pleased to see you.

Bayer Lung Resource Centre, (Antitrypsin Deficiency Assessment and Programme for Treatment), 1st Floor, Nuffield House, Queen Elizabeth Hospital, Birmingham. B15 2TH

Mrs. Rebecca Bray, Co-ordinator, Registry Office Tel: 0121 697 8257

Please Initial

I ..... voluntarily agree to participate in the alpha$_1$ antitrypsin deficiency assessment programme.

Yes	No

I have been given a full explanation of the programme and read the patient information sheet and have had all my questions answered and agreed to cooperate where possible with this programme.

Yes	No

I understand that that if I suffer from any unexpected problems that it may be important to contact both my own doctor and the staff at the alpha-1-antitrypsin resource centre.

Yes	No

I understand fully that I am free to withdraw from the programme at any time without giving a reason and that this will not adversely affect my future management.

Signed (Patient)……..……..…
Date ……..

Signed (Witness)…….
Date……...

CONSENT FORM FOR PARTICIPATION IN RESEARCH 3359a

Alpha-1-Antitrypsin Deficiency Genetic Variation Study

What is the study about?

You or a member of your family has been identified as having alpha$_1$ antitrypsin deficiency. This is an inherited condition that is believed to increase the risk of development of lung and liver health problems. However very little detailed information has been collected on the way this deficiency affects patients and some studies have suggested that lung and liver disease may run in families even without the deficiency. It is likely however that the deficiency highlights the tendency to develop these diseases and when present will make them worse.

However, our studies have shown that even in families with alpha-1-antitrypsin deficiency the tendency to develop these diseases also varies. This suggests that there may be some other genetic influences of importance even in alpha-1-antitrypsin deficiency.

We wish to learn as much as possible about the deficiency and its relationship to lung and liver disease and for this reason invite you to participate in our alpha$_1$ antitrypsin deficiency assessment programme. In particular, we wish to study your blood for any other genetic influence that may determine whether lung or liver disease develops. This may be of importance both to alpha-1-antitrypsin deficiency and to emphysema or liver disease in patients who don't have deficiency. Our recent study funded by the European Union has identified two areas of genes that relate to the likelihood that lung or liver disease will develop. In order to determine the reasons we now need to repeat the study, focusing on these genes to understand their role, and hence how to prevent or treat the lung and liver disease.

What will I have to do?

As part of the genetic study we store part of a sample of your blood, or of a mouthwash, from which we can extract your DNA. It does not matter which sample we use – if you are attending the centre as a patient or with your relative we can take a blood sample, or if it is more convenient we can send you a mouthwash kit through the post. No information about the genetic research study will be released to you (unless you specifically request this) or your family or family doctor. The sample collected will be coded once it is processed and separated from any information that identifies it as yours. It will have a code but no one will be able to tell directly that it has come from you. There will be no reference in your medical file that you have taken part in this gene variation part of the study. However as medical science progresses new ideas and new genes related to lung or liver disease are likely to become known. For this reason we will store the sample for such future studies. It is important to emphasise that the whole study will be confidential and specific procedures have been put in place to separate all confidential information that could help identify you from the results of genetic DNA analysis.

What are the benefits?
The major purpose of the study is to find out as much as possible about the lung or liver disease associated with alpha$_1$ antitrypsin deficiency including the association with any other unrecognised genetic factors.

What are the risks?
The only minor problem that is likely to occur is a slight degree of local bruising in some patients if they have their blood taken.

What are the alternatives?
There are currently no alternatives to finding the information that is required other than the genetic programme outlined above.

What happens if I do not wish to take part?
If you do not wish to participate in the genetic study this will be fully understood. It will not affect your care, or that of any relative of yours who may be participating in the alpha$_1$ antitrypsin deficiency programme.

What happens to the information?
The important information that we collect over the next few years will be written up as a report and submitted to medical journals so that other doctors may read about the problem. The research information obtained from the samples that we have collected will help us understand the development of emphysema and liver disease. The data may provide clues to help develop new forms of treatment. Neither your name nor any details relating to you personally will be released to any other person outside the research programme (including insurance companies).

Who is taking part?
All subjects that we identify with alpha$_1$ antitrypsin deficiency will be asked to take part. At present we know of over 1700 patients with the deficiency and it is likely that there are several thousand similar people in the West Midlands alone. It is hoped over the next year or two to identify and assess as many people with the deficiency as possible and obtain DNA samples from most of them for these studies. Wherever possible we hope to assess patients' family members too, so that we can be more certain of other genetic influences on the development of emphysema and liver disease.

What if something goes wrong?
The question largely relates to clinical trials and at present that is not part of the DNA study.

What happens at the end of a study?
At the end of the study we will have learnt a lot about alpha$_1$ antitrypsin deficiency. There are regular meetings for all patients and their relatives to attend to discuss how the assessment programme is going and how our understanding is developing. You may wish to attend these, or we can keep in touch with you by letter and newsletter.

What if I have more questions or do not understand something?

The doctors, co-ordinator and nurse involved in the alpha$_1$ antitrypsin genetic programme will happily answer your questions on any occasion when you visit, or over the phone. If questions arise you will be able to contact the centre and either speak to somebody at that time or arrange to do so.

What happens now if I decide to take part?

If you decide to take part in the programme now we ask that you contact us using the details below. We can then send you a mouthwash kit, containing instructions for use. You should follow the instructions and return to us. Alternatively if you are attending the centre soon with your relative and you would prefer to give a sample of blood or mouthwash whilst you are here, we can arrange this.

What happens if I change my mind during the study?

If you change your mind during the study, it is important that you notify the assessment centre. The sample of DNA that we have stored will then be destroyed.

Bayer Lung Resource Centre, (Antitrypsin Deficiency Assessment and Programme for Treatment), 1st Floor, Nuffield House, Queen Elizabeth Hospital, Birmingham. B15 2TH

Mrs. Rebecca Bray, Co-ordinator,
Registry Office Tel: 0121 697 8257

I …….. voluntarily agree to participate in the alpha$_1$ antitrypsin genetic study.	**Yes**	
I have been given full information including reading the information sheet and had all my questions answered.	**Yes**	
I understand that I can change my mind at any time and the sample will be destroyed.	**Yes**	
I am happy for my name to be held on the genetic database but know that this will not be disclosed to anyone outside the research study group.	**Yes**	
I understand the information will not be filed in my general notes.	**Yes**	

Signed (Patient) ……..……
Name (Print) ……
Date ……

Signed (Witness) ……
Name (Print) ……
Date ……

PATIENT INFORMATION LEAFLET AND CONSENT FORM

Study code: RRK3404

Title: A study of clinical phenotypes, disease progression and genetic susceptibility in patients with chronic obstructive pulmonary disease and bronchiectasis

You have been asked to participate in a clinical study for research purposes. Before you decide to take part it is important that you understand why the study is being performed, what it involves, and the possible risks and benefits for you. Take your time reading the following information and discuss it with others if you wish.

Introduction to the research & invitation to take part

This study is being carried out by the research department in the Lung Investigation Unit at the Queen Elizabeth Hospital, which is part of University Hospitals Birmingham NHS Foundation Trust, and is led by Professor RA Stockley. Several hospitals in the West Midlands are working closely with Professor Stockley in order to help with the study. The department has a special interest in understanding and treating chronic lung disease, in particular chronic obstructive pulmonary disease, or COPD for short, and bronchiectasis. You have been chosen because your hospital doctor has identified you as suffering from COPD or bronchiectasis and believes that you are suitable to take part in the study.

What is the research about?

The main purpose of the study is to find patterns of symptoms, breathing tests, blood and sputum tests that identify particular groups of patients. We also hope to establish the importance of environmental influences, such as chest infections, in the development of COPD and bronchiectasisand how they progress over time. Other research suggests there may be genetic influences in COPD, which may affect particular groups of patients, or affect how their breathing tests change over time. We plan to study this in more detail to find out which genes are important.

What will I have to do?

Please read this section carefully as it details the procedures that are specific to research and not part of usual clinical care

If you agree to take part you will be seen by a research doctor at the time of your usual out-patient clinic appointment. They will explain more about the study and what it would mean for you. The study will take place over 5 years, and will involve seeing a research doctor after one of your out-patient appointments each year for at least 3 years. There will be no extra visits to the hospital other than those for your usual out-patient

appointments. If you would rather participate for only one visit, rather than several, please inform the research doctor.

If you have brought a sample of sputum with you, you will be asked if some could be used for the research. If you have not brought any sputum with you, and usually cough it up every day, you will be given a container in which you may bring a sample on your next visit.

A blood sample will be collected from you. This can be done at the same time as any other blood tests your hospital doctor has asked for. We will collect 4 small tubes of blood for the study, which is the equivalent of about 2 tablespoons.

We will process the blood and sputum samples so that they can be stored safely in secure freezers at the University of Birmingham. For this study they will be used to examine markers relevant to COPD and bronchectasis. They may also be valuable for future research in other areas, but will not be used for other studies without prior ethical approval.

We will also extract your DNA from your blood, which will be stored in the same way. Your DNA will be used to look for genes that may influence the development of COPD. Your DNA will not be used for other studies.

As well as collecting these samples we will also collect information about you by talking to you and by looking at your medical notes. This will include your date of birth, gender, height, weight, date you were diagnosed and whether you smoke now, or have done so in the past. We will also ask some questions about symptoms that people with COPD and/or bronchiectasis often experience, and ask you to fill in a questionnaire about the way your lung disease affects you. We will also look in your medical notes to see what medicines you take, the results of any breathing tests you have had done recently, and scans (if you have had one). All of this information will be recorded in written form only.

All of the above procedures should take no more than half an hour on the first occasion, and be a little shorter on later visits. This will be in addition to seeing your usual hospital doctor.

What are the benefits?
You will have no personal benefit from the study. Your participation and donation of samples may benefit patients with COPD and bronchiecatsis in the future by helping us to understand the disease better, and helping us to develop new treatments.

What are the risks?
We do not expect any harm to come to you as a result of providing samples or talking to the researchers. Sometimes blood tests can be uncomfortable, or leave bruising, but this will be temporary.

What are the alternatives?
There are many ways of looking at blood and sputum samples, but relating them to breathing tests, symptoms and genes has not been done together before. It is therefore a new area of research, and there are no similar studies being done in the UK at present.

What if I do not want to take part?
The study is entirely voluntary and if you do not wish to participate it will not affect your future care.

What happens to the information?
If you decide to take part you will need to allow access to your medical records. They may be looked at by the research team, by the hospital research and development department and by regulatory authorities who check that the study is being carried out properly. By signing this form you are giving permission for this to be done.

The information collected will be stored on a secure computer, but your name will not. This is known as linked anonymised data, meaning that only Professor Stockley will be able to link any of your information to your name. He will have sole access to a written record of your information, stored in a secure facility at University Hospitals Birmingham. All the data collected, samples you provide, and their results, including any information about your genes, will be coded with a number. The results of tests on your samples and about your genes will not be available to anyone outside of the research team and our collaborators. The link to your name will be destroyed after 15 years. Once the data is collected it will be the property of the research department.

The results of the study may be published in a medical journal, but your identity will not be revealed. The results may be used in statistical tests, research and development of new treatments, diagnostic tests and medical aids.

Who else is taking part?
At least 500 other patients with COPD and with bronchiectsis will be asked to take part.

What if something goes wrong?
Since the study involves only simple tests that could form part of your routine care, we do not expect any harm to come to you. Whatever part of the study you choose or decide not to take part in will not affect your future care. If you are harmed by taking part in this research project there are no special compensation arrangements.

What happens at the end of the study?
Throughout the study, and when it ends, your hospital doctor and general practitioner will continue to treat your chest problems.

What happens if I have more questions?
If you do not understand something in this leaflet, or have further questions you may ask the researcher now, or your hospital doctor.

What happens now if I decide to take part?
If you decide to take part you will be asked to read, sign and date the Written Consent Form attached to this sheet. By signing it you acknowledge that you have understood the aims of the research, and what you are being asked to do.

In summary

- The research aims to find out patterns of symptoms, breathing tests, blood and sputum tests and scans in patients with COPD and bronchiectasis
- If you take part you will be asked to provide blood and sputum samples, and answer questions about yourself and your chest problems at one out-patient appointment per year for 3 years, or as many as you can manage
- All information will be stored in such a way that you cannot be identified by anyone without Professor Stockley's permission

Will my General Practitioner (GP) be informed?
If you give your permission, your GP will be told about your participation in the study.

What happens if I change my mind during the study?
You are free to withdraw your participation at any time, and it will not affect your future care. If you withdraw your consent after your samples have been analysed it will be the responsibility of your hospital doctor to ensure that the samples are destroyed if you so wish.

Who can I contact about the study?
In the first instance any concerns or questions should be addressed to either your GP or hospital doctor. If you have further concerns you can contact

Professor Robert Stockley Telephone: 0121 472 1311
Thank-you for reading this information leaflet

I…… (Name in BLOCK CAPITALS)

Initials

Have read the attached information concerning my participation in this study and have had the opportunity to discuss it and ask questions. All my questions have been answered in a satisfactory way. ☐

I voluntarily consent to take part in this study. ☐

I know that at any time, and without giving a reason, withdraw my participation in the study and that my future care and management will not be affected. ☐

I understand that I will have a copy of this Patient Information Leaflet and Written Consent to keep. ☐

I hereby give my permission to authorise personnel from the research team to examine my medical records. ☐

I hereby give permission for samples taken as part of this study to be stored for future use by the research team. ☐

I understand that samples taken as part of this study will not be used in other studies without prior ethical approval. ☐

I hereby give permission for my GP and hospital consultant to be informed about my participation in this research study. ☐

……..
…….
Patient's signature Date

……..
Name in BLOCK CAPITALS

Responsible investigator
I have explained the nature and purpose of this study for the person named above

……
…….
Responsible investigator/representative signature Date

……
Name in BLOCK CAPITALS

STUDY INFORMATION SHEET AND CONSENT FORM

Study Title: Molecular and Functional Mechanisms of Human Lung Disease

Midlands Lung Tissue Consortium

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Introduction

Lung disease causes pain, discomfort and can prevent sufferers from carrying out everyday activities. Whilst available treatments including steroids and other drugs may relieve symptoms, none provide a cure. Operations may help some people. More research is needed to find new treatments that can cure lung disease.

What is the purpose of the study?

In order to find the causes of lung disease such as COPD and lung cancer and to find new ways of treating these diseases we have to do more research. It is ideal to do the research on tissue from human lungs because we are investigating a human disease. The lung research teams at Glenfield Hospital, Leicester Birmingham Heartlands Hospital, Walsgrave Hospital, Coventry and pharmaceutical companies (including Astrazeneca) have joined together in order to collaborate on studies into lung disease using human lung tissue. A numbers of diseases will be studied and the lung tissue will be used in a number of different laboratory studies. Pharmaceutical companies (including AstraZeneca) will use the lung tissue in research to further the understanding of lung disease and development of new treatments for lung disease. These experiments will be done on lung tissue that has been removed from patients as part of their medical treatment, which would otherwise be destroyed. In some cases we may also wish to take a blood sample to compare the findings in the lung tissue and blood.

Why have I been chosen?

You have been chosen because your doctor has said that you may need to have some of your lung removed to treat your disease.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will

be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
Your surgeon has recommended having an operation that should help you with your current condition. Some samples of lung tissue removed as l part of your operation would normally be destroyed. If you are willing to take part in our research, your surgeon will provide the collaborating hospitals and pharmaceutical partners including AstraZeneca with some of the lung tissue that has been removed as a normal part of your operation. The surgeon will **not** remove extra lung tissue for this research. We will also record some information about your recent medical history, medicines taken and reason for the operation from your medical records. In terms of your operation, stay in hospital and subsequent follow up there will be no difference to what will happen to you whether you take part in the study or not except that in a few cases we may wish to take an extra blood sample of approximately 30mL (about three tablespoons full) before your operation.

What happens if I don't want to take part?
Nothing, you simply don't sign this form. This will not affect your medical care or your legal rights in any way.

What rights do I have to the results of the research?
Any information derived directly or indirectly from this research by the collaborating hospitals or by pharmaceutical companies such AstraZeneca, as well as any patents, diagnostic tests, drugs, or biological products developed directly or indirectly as a result, are the sole property of the company (or their successors, licensees, and assigns) and may be used for commercial purposes. You have no right to this property or to any share of the profits that may be earned directly or indirectly as a result of this research. However, in signing this form and donating a blood sample, you do not give up any rights that you would otherwise have as a participant in research.

What do I have to do?
There is nothing extra to do as a result of being part of this study

What are the possible disadvantages and risks of taking part?
There are no disadvantages or risks to taking part in the study over and above the normal risks associated with this surgery, which you require as part of your care. If you are asked to donate an additional blood sample there may be some discomfort of the needle being inserted into a vein in your arm and the possibility of bruising developing afterwards around the area that the needle was inserted. This should disappear in a few days

What are the possible benefits of taking part in the study?
There are no direct benefits. Taking part in this study means that you may possibly help suffers of lung disease in the future, as information about the changes that occur in the lung may be used to develop new treatments.

What if something goes wrong?
We do not think there is any significant risk of any harm occurring as a result of participating in this study. However if you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms would be available to you.'

Will my taking part in this study be kept confidential?
All information resulting from you taking part in the study will be stored and analysed in a computer and will be treated confidentially. You will be identified in the computer by a number and only your doctor will be able to identify the number as belonging to you. The study records will not be made available in any form to anyone other than authorised representatives of the health authorities and our pharmaceutical partners such as AstraZeneca. In all instances, your confidentiality will be maintained, in accordance with the Data Protection Act or as local laws permit.

The pharmaceutical partners such as AstraZeneca and Regulatory authorities may wish to check that this research has been done properly, they may have access to your files and know your identity, but they are under a duty of confidentiality not to disclose details to others.

What will happen to the samples that I have donated?
The samples will be processed by the research team and used in a range of experiments into the causes of lung disease. Samples may be transported to our pharmaceutical partners e.g. AstraZeneca or other hospitals in the collaborating group to do further experiments including tests to develop new drugs. Those samples that are not fully used up in experiments may be stored by the research team or by pharmaceutical partners including AstraZeneca, for use in future experiments, for up to 20 years.

Who is organising and funding the research?
The research is a collaboration between the lung research teams at the hospitals in Leicester, Coventry, Birmingham and pharmaceutical partners including AstraZeneca. The pharmaceutical partners including AstraZeneca will be contributing to costs of the research at the collaborating hospitals. The income obtained from the pharmaceutical partners will only be used to support the work carried out as part of this project.

Can I withdraw my consent?
You may withdraw your consent to the use of your data and samples at any time. If you withdraw your permission consent before your donated tissue and data are used, we will not use the data and the samples will be destroyed. If you withdraw your consent after your tissue sample has been sent for analysis we will ensure that your sample(s) are destroyed. However, if analysis has already been performed neither AstraZeneca nor ourselves are obliged to destroy results of this research.

Who has reviewed the study?
The study has been reviewed by the research teams within the consortium and by the members of the Department of Respiratory Medicine, Allergy and Thoracic Surgery. Individual research projects where we use the lung tissue have been reviewed by a variety of charities and funding organisations

If you have any further questions about this study please do discuss them with:

Dr David Thickett, Honorary Consultant in Intensive Care, Critical Care Research Department, 2nd Floor Lincoln House, Birmingham Heartlands Hospital, Bordesley Green East, Birmingham – B9 5SS. Tel. 0121 424 0662. Fax: 0121 424 2971

CONSENT FORM

Initials

1. I confirm that I have read and understood the patient information form on the above project and have been given a copy to keep. I have had the opportunity to ask questions about the project and understand why the research is being done and any foreseeable risks involved. ☐

2. I agree to give a sample of tissue for research in the above project. I understand how the sample will be collected and that giving the sample is voluntary. I am free to withdraw my approval for use of the sample at any time without giving any reason and without my medical treatment or legal rights being affected. ☐

3. I give permission for my medical records to be looked at and information taken from them to be treated in strict confidence by responsible people from *(Insert name of relevant hospital)* and AstraZeneca. ☐

4. I understand that my doctor will be informed if any of the results of the tests done as part of the research are important for my health. ☐

5. I understand that I will not benefit financially if this research leads to a new treatment or medical test. ☐

6. I do know where to contact *(David Thickett)*, if I need further information. ☐

7. Do you agree to take part in this study? **YES/NO**

Signed: …… Date: ……..

Name (Block capitals) ……..

I, (Name of investigator, block letters) ……
have explained the nature and purpose of the study to …….
and believe that he/she understands what the study involves.

Signed: …….. Date: ……..

Appendix 2: Reference equations

Pulmonary function tests

Reference equations that generate predicted lung function values are available from the Association for Respiratory Technology and Physiology in the UK (www.artp.org.uk). These were generated using regression models from a cohort of subjects aged 18-60, and include height, age and gender components. The equations are as follows.

	Unit	Regression equation	
		Male	Female
FEV1	l	4.30H - 0.029A - 2.49	3.95H - 0.025A - 2.60
FVC	l	5.76H - 0.026A - 4.34	4.43H - 0.026A - 2.89
TLC	l	7.99H - 7.08	6.60H - 5.79
RV	l	1.31H + 0.22A -1.23	1.81H + 0.016A - 2.00
TLCO	$mmol.min^{-1}.kPa^{-1}$	11.11H - 0.066A - 6.03	8.18H - 0.049A - 2.74
KCO	$mmol.min^{-1}.kPa^{-1}.l^{-1}$	predTLCO/predTLC	

CT densitometry

This is not an area that has been pubished upon widely. Values at -912HU from healthy subjects who have never smoked have been published[214], and are taken as the threshold for normality for the purposes of this thesis. These are as follows:

Upper zone voxel index	Lower zone voxel index
5	13

Appendix 3: Solutions

REAGENT A
10ml Triton X
110g sucrose
1ml Tris HCl pH 7.4
1.02ml 4.9M $MgCl_2$
Make up to 1l with distilled water

REAGENT B
400ml 1M Tris HCl
120ml 0.5M EDTA
30ml 5M NaCl
40ml 25% SDS
Make up to 1l with distilled water

PBS
8.0g NaCl
0.2g KCl
1.44g Na_2HPO_4
0.24g K_2HPO_4
Add 1l of distilled water, followed by acid or alkali as needed to achieve a pH of 7.2-7.4

TE
10ml Tris HCl
0.1ml 0.5M EDTA
Make up to 1l with distilled water

10 x TBE
108g Tris buffer
55g Boric acid
40ml 0.5M EDTA pH8.0
Make up to 1l with distilled water

Loading dye
0.25g Bromophenol blue
40g sucrose
Make up to 100ml with distilled water

TDMH
16.7ml MillQ water
6.75ml 25mM $MgCl_2$
8.45ml 10 x buffer
600µl dNTPs

Appendix 4: Statistical methods

Power

Both false positive and false negative associations are more likely to occur in small, underpowered datasets. In an ideal world for complex disease genetics the dataset should have at least 80% power to detect an allele OR of 1.5, but this may not be achievable if the allele frequency is low. The power of the studies in this thesis to detect a disease effect, as an OR, was predominantly calculated after genotyping to aid interpretation of the results, since the dataset used was the largest possible in the UK.

Power was calculated using a programme designed and provided to my supervisor by Heather Cordell (University of Cambridge, UK).

Hardy Weinberg equilibrium

Hardy-Weinberg law states that allele and genotype frequency in a large population does not change formone generation to the next provided there is random mating with regard to a trait (i.e. mating does not depend upon the trait). Factors that can change this include mutation, population migration, genetic drift and selection. In general in genetic association work the Hardy-Weinberg equibrium is checked to aid detection of genotyping error, and as such it may be of more importance to check the control cohort rather than cases, since by definition cases have been selected. However this is controversial, thus any cohorts out of Hardy-Weinberg equilibrium in this thesis were excluded from further analysis, regardless of their case/control status.

Hardy Weinberg equibrium was checked using the χ^2 test, where the expected genotype frequencies are as follows:

	Observed	Expected
Minor allele homozygote	x	$\frac{x + (y/2)}{(x + y + z)(x + y/2)}$
Heterozygote	y	$\frac{2(x + y/2)}{(x + y + z)(x + y/2)}$
Major allele homozygote	z	$\frac{z + (y/2)}{(x + y + z)(z + y/2)}$

Appendix 5: Abbreviations

For all abbreviations if the text is in italics it refers to a gene, and if in plain text a protein of the same name.

°C	Temperature in degrees celcius
6MWT	6 minute walk test
AAT	Alpha 1 antitrypsin
AATD	Alpha 1 antitrypsin deficiency
AB	Antibody
ABI	Applied Biosystems
ADAPT	Antitrypsin deficiency Assessment and Programme for Treatment
Ala	Alanine
Arg	Arginine
AOT40	Sum of the differences between the hourly mean ozone concentration and 40 ppb for each hour when the concentration exceeds this limit during daylight hours
ATS	American Thoracic Society
BAL	Bronchoalveolar lavage
BMI	Body mass index
bp	Base pair
CHRNA	Alpha nictonic acid receptor
COMEAP	Committee on Medical Effects of Air Pollutants
COPD	Chronic obstructive pulmonary disease
DLCO	Diffusing capacity for carbon monoxide
DNA	Deoxyribonucleic acid
dTNP	Deoxynucleotide triphosphates
ECLIPSE	Evaluation of COPD longitudinally to predict surrogate end points
EDTA	Ethylene diamine tetracetic cid
ELISA	Enzyme linked immmunoadsorbent assay
EPHX	Micorsomal epoxide hydrolase
FEF	Forced expiratory flow
FEV1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
GC	Vitamin D binding protein
GIS	Geographical information system
Glu	Glutaminc acid
Gly	Glysine
GROα	Growth related oncogene alpha
GST	Glutathione S transferase
GWAS	Genome wide association study
His	Histidine
HHIP	Hedgehog interacting protein
HLA	Human Leukocyte Antigen

HMOX	Heme oxygenase
HRCT	High resolution computed tomography
HU	Hounsfield Units
IBD	Identity by descent
IC	Intracellular
ICGN	International COPD genetics network
IL	Interleukin
Ile	Isoleucine
IQR	Interquartile range
KCO	Diffusing capacity for carbon monoxide corrected for alveolar volume
LD	Linkage disequilibrium
Leu	Leucine
LHS	Lung Health Study
LOD	Logarithm of the odds
Lsy	Lysine
LTB4	Leukotriene B4
LZDE	Lower zone dominant emphysema
LZVI	Lower zone voxel index
MAF	Minor allele frequency
MMP	Matrix metalloprotease
NCBI	National Centre for Biotechnology
NE	Neutrophil elastase
NETT	National Emphysema Treatment Trial
NFκβ	Nuclear factor kappa beta
NHLBI	National Heart Lung and Blood Institute
NO_2	nitrogen dioxide
NOS3	Nitrogen oxide synthase 3
NOx	oxides of nitrogen
OR	Odds ratio
p	Level of significance (p value)
PAR	Proteinase activated receptor
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDE4	Phosphodiesterase 4
Phe	Phenylalanine
PLAU	Plasminogen activator urokinase
PM_{10}	Particulate matter <10μm in diameter
ppb	Parts per billion
Pro	Proline
RANKL	Receptor activator for nuclear factor kappa beta
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rs	Reference SNP - refers to the number allocated to the SNP in NCBI databases
RV	Residual volume
SE	Standard error

Ser	Serine
SERPINA3	Alpha 1 antichymotrypsin
SFTPB	Surfactant protein B
SNP	Single nucleotide polymorphism
SO_2	Sulphur dioxide
SOD3	Superoxide dismutase 3
spp	Species
T1D	Type 1 diabetes
TBE	Tris borate EDTA buffer
TDMH	Taq polymerase reaction buffer
TE	Tris-EDTA buffer
TGF	Transforming growth factor
Thr	Threonine
TLC	Total lung capacity
Tm	Annealing temperature
TNF	Tumour necrosis factor
Tyr	Tyrosine
UV	Ultraviolet
UZDE	Upper zone dominant emphysma
UZVI	Upper zone voxel index
Val	Valine
WHO	World health organisation
$\chi 2$	Chi squared

Appendix 6: HLA typing sheets

Phototyping method

When using this typing sheet if a band is seen in a given lane other than the control bands this corresponds to the allele listed, though an additional check of the PCR product size to ensure that it corresponds to the expected size should also be carried out.

Locus	Alleles amplified	Position in 96 well plate	Apply photo from gel
DRB1	0101/2/4	1A	
	0103	1B	
	1501-5	1C	
	1601-6	1D	
	03011/12/2/3/4/5, 1107	1E	
	0301/4	1F	
	0302/5, 1302/3, 1109/20, 1402/3/9/13/19	1G	
	0401-22, 1410, 1122	1H	
	0701	2A	
	0801-11, 1415	2B	
	0901	2C	
	1001	2D	
	1101-21	2E	
	1201/2/3	2F	
	1301/2/4/8/15-17/19/20/22, 1416, 1102/3/11/14/16/20/21	2G	
	1301/2/4/8/9/16/20, 1116/20, 1416	2H	
	1301/2/10/15/16, 1116/20, 1419/20	3A	
	1305/6/10, 1109, 1419/21	3B	
	1303/10, 1419/21	3C	
	1303/4/12/21, 1413	3D	
	1305/7/11/14/18/21, 1101/4/6/9/10/(11)/12/15	3E	
	1301/2/5-9/10/11/14-16/18/19/20/22, 1402/3/6/9/12/14/17/19-21, 0301-5	3F	
	1401/3-5/7/8/10-12/14/15/18, 1318, 0809, 1117	3G	
	1401/7/16	3H	
	1402-3/6/9/12/13/17, 1318, (1418/19/21), (0301/2/3/5)	4A	
DQB1	0305	4H	

	02	5A
	04	5B
	05	5C
	0601-3	5D
	0603-9	5E
	0301/4	5F
	0302	5G
	03032	5H
DQA1	0101/4	6A
	0102/3	6B
	0201	6C
	03011/12	6D
	0302	6E
	0401	6F
	0501/2	6G
	0601	6H

Dynal DRB1 typing sheets

The first page is used to indicate positive lanes, and the subsequent pages for interpretation.

These pages are better viewed or printed at A3 size, and can be accessed at: http://products.invitrogen.com/ivgn/en/US/adirect/invitrogen?cmd=catProductDetail&entryPoint=adirect&messageType=catProductDetail&showAddButton=true&productID=4515010&CID=Search-45150-10

The pages can be found in the Manuals and Documents section, under lot specific data.

The lot specific worksheet for lot number 010 is that used for the studies herein.

PEL-FREEZ
A DYNAL BIOTECH BRAND

Gel Documentation Form
DRB96 SSP UniTray®

Dynal Biotech
Telephone: 800.558.4511
Fax:414.357.4518
www.dynalbiotech.com

Kit Name: DRB96 SSP Lot#: 010 Batch#: _____ Exp. Date:_____ Tested by:_____

Institution:_____ Name:_____ Sample I.D.:_____ Date:_____

Gel Picture **Tray Identification Number:** _____

Well location	**A1**	**B1**	**C1**	**D1**	**E1**	**F1**	**G1**	**H1**		**A2**	**B2**	**C2**	**D2**	**E2**	**F2**	**G2**	**H2**	**A3**	**B3**	**C3**	**D3**	**E3**	**F3**	**G3**	**H3**
Lane number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**M**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Size (bp)	210	85	120	210	85	110	230	105		205	230	180	175	150	200	220	150	210	150	100	200	200	120	200	130
	200	75	105	200		100	150			175	160	155	115			200	85	200	145		130		105		100
			100	130			110					150	70			110		195					85		
Well location	**A4**	**B4**	**C4**	**D4**	**E4**	**F4**	**G4**	**H4**		**A5**	**B5**	**C5**	**D5**	**E5**	**F5**	**G5**	**H5**	**A6**	**B6**	**C6**	**D6**	**E6**	**F6**	**G6**	**H6**
Lane number	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**M**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
Size (bp)	225	105	265	105	160	180	200	145		220	170	195	90	165	230	165	160	255	120	120	115	160	235	235	170
		70	190		100	115	155			215	150	180	75	150	95		70	180	95		110	130	215	225	165
							115					135	65	130/105	85			175	65			120	200	170	75
Well location	**A7**	**B7**	**C7**	**D7**	**E7**	**F7**	**G7**	**H7**		**A8**	**B8**	**C8**	**D8**	**E8**	**F8**	**G8**	**H8**	**A9**	**B9**	**C9**	**D9**	**E9**	**F9**	**G9**	**H9**
Lane number	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**M**	**57**	**58**	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**
Size (bp)	105	225	225	270	265	220	155	130		185	235	205	250	165	215	165	185	85	160	175	170	225	195	240	215
	90	175	220	180	255	210	150				230	195	195	105	170	150	130	75	130	125	70	185	185	205	
			200	125	200	200	145				200	95	130			95									
			175			150	140				110		95												
Well location	**A10**	**B10**	**C10**	**D10**	**E10**	**F10**	**G10**	**H10**		**A11**	**B11**	**C11**	**D11**	**E11**	**F11**	**G11**	**H11**	**A12**	**B12**	**C12**	**D12**	**E12**	**F12**	**G12**	**H12**
Lane number	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**	**M**	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**	**92**	**93**	**94**	**95**	**96**
Size (bp)	200	175	175	175	200	225	145	100		215	150	225	215	185	85	145	260	200	205	165	215	185	240	215	
			105	105		185	70	90		210	145		170			105	175	115	110	155	130	140	220	210	
				75		145				140			160			80								85	
						105																			

Positive Lanes Failed Lanes Contamination (lane 96):

Test 1 _____ _____ Yes ☐ No ☐

Size of internal control PCR fragment: app. 796 bp
Base pair sizes are approximate.
For detailed information, please see the lot specific worksheet or primer mix specificity table.

PEL-FREEZ
A DYNAL BIOTECH BRAND

DRB96 SSP UNITRAY© WORKSHEET

Code no.: 45150-10 45151-10

Dynal Biotech
Telephone: 800.558.4511
Fax: 414.357.4518
www.dynalbiotech.com

Institution
Sample I.D.
Name
Ethnic Origin
Donor/Patient

Purpose of Test
DNA Extraction Method
DNA Conc. (ng/µl)
Tested by
Reviewed by

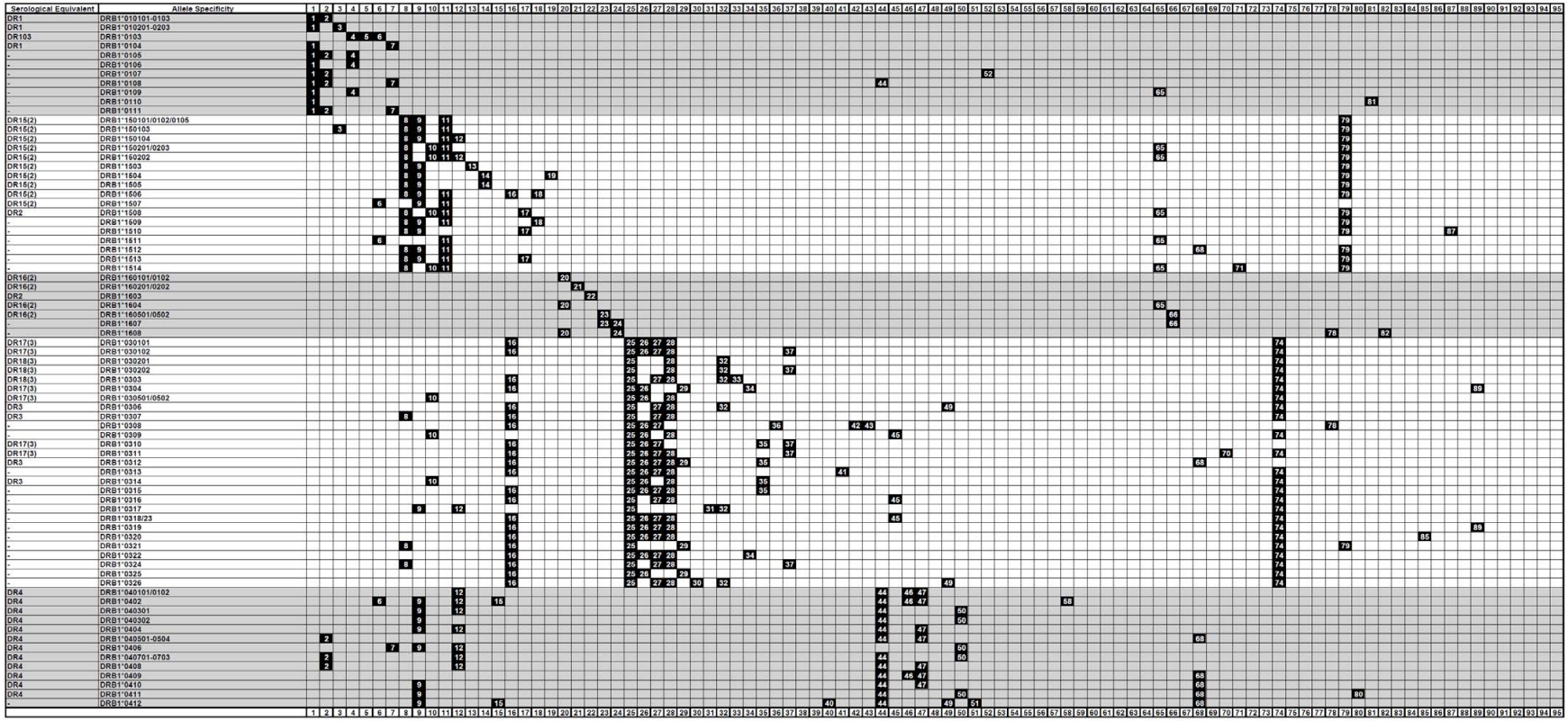
Taq Lot#
Kit Lot# 010 Batch#
Expiration Date
Test Date
Review Date

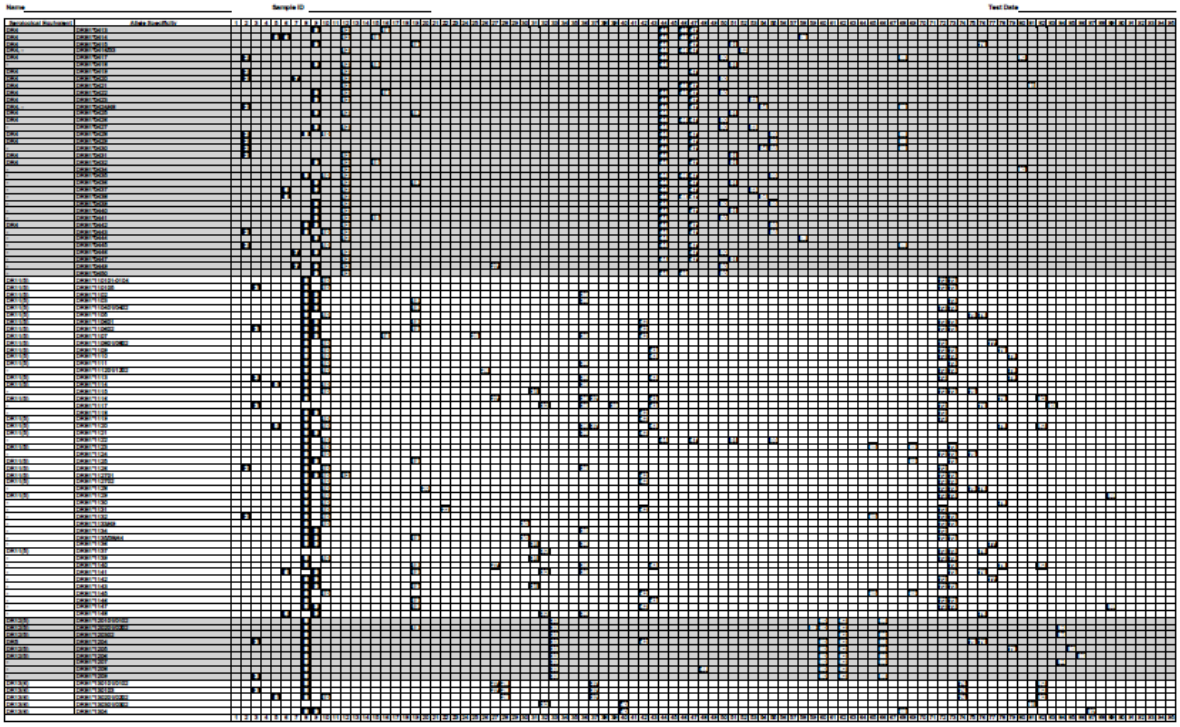
Low Resolution Typing Result

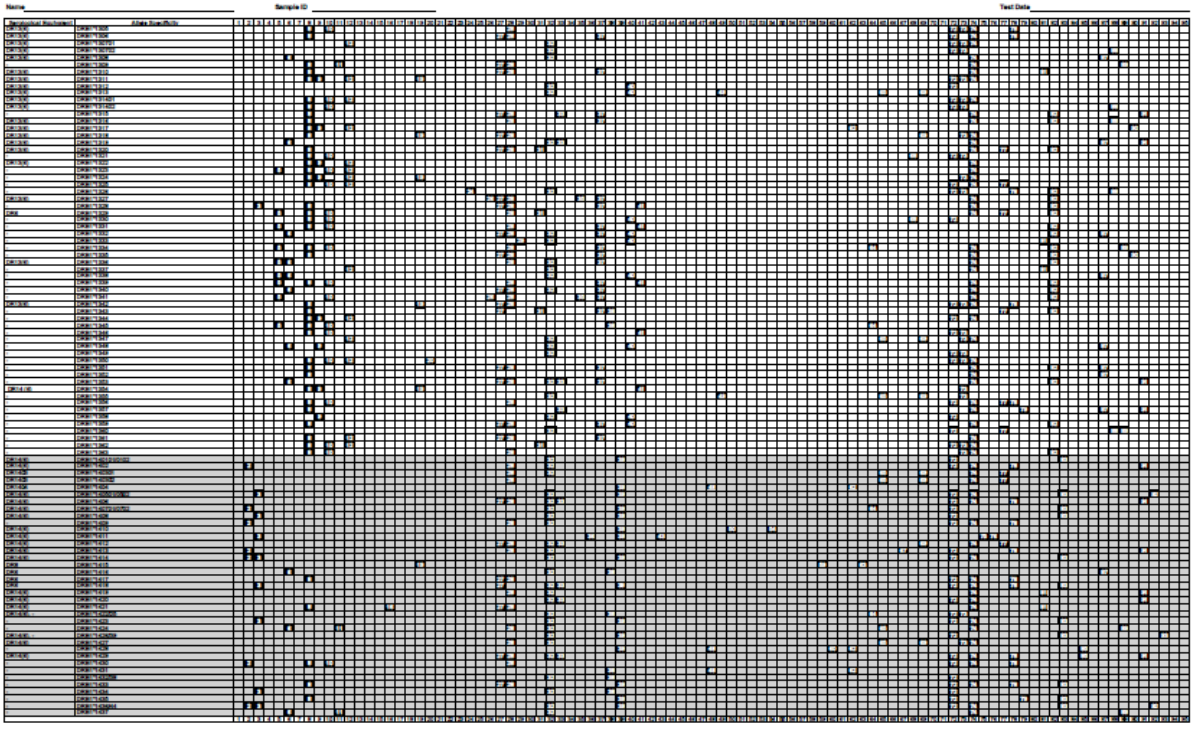
DRB1*
DRB1*
DRB3/4/5*
DRB3/4/5*

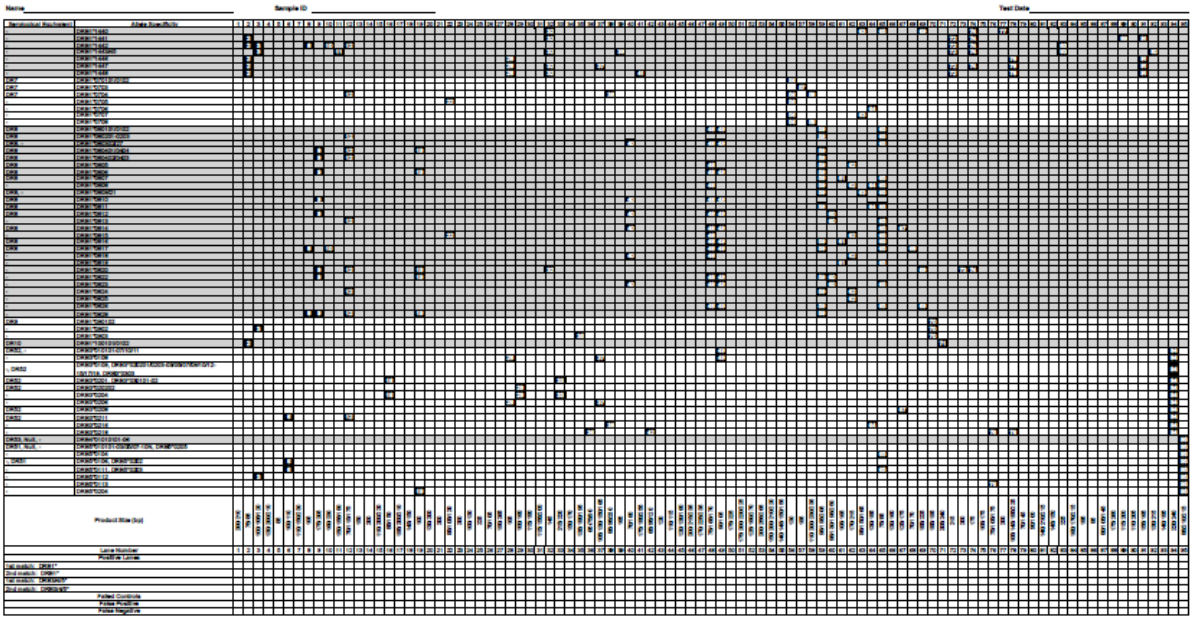
High Resolution Typing Result

1st Allele Assignment DRB1*
2nd Allele Assignment DRB1*
DRB3/4/5*
DRB3/4/5*









Dynal DQB1 typing sheet

The table shows the specificity of each well in the batch number EK0128, being that used for the studies herein.

Lane	PCR product size (bp)	Amplified *DQB1* alleles
A	220	050101-0504
B	215	060101-0622, 0624-0627
C	205	020101-0204
D	130, 150	020101-0202, 0204, 030201, 030202, 030501, 030503, 0307, 0308, 0311, 0318
E	170, 165	030101, 030102, 0304, 0309, 0310, 0312-0314
F	125	030201-030203, 030501-0308, 0311, 0312, 0315, 0317, 0318
G	125	0203, 030101, 030102, 030302, 030303, 0306, 0309, 0310, 0312, 0313, 0315-0317
H	210	0401, 0402

Appendix 7: Publications resulting from this thesis

In order to avoid repetition within the following list for data that has been presented in both abstract and paper form only the paper is listed.

Original papers

Chapter 3

1. AM Wood *et al*, The TNFα gene relates to clinical phenotype in AATD, Resp Res (2008) 9: 52-67

Chapter 4

1. AM Wood *et al*, Phenotypic differences in alpha 1 antitrypsin deficient sibling pairs may relate to genetic variation, Journal of COPD (2008) 5: 353-9

Chapter 5

1. CJ McAloon, AM Wood *et al*, Matrix metalloprotease polymorphisms are associated with gas transfer in AATD, Therapeutic advances in respiratory disease (2009) 3: 23-30

Chapter 7

1. AM Wood et al, Outdoor air pollution associates with pulmonary function in AATD, ERJ (2009) 34: 346-53

Abstracts

Chapter 3

1. AM Wood *et al*, Sputum TNFα levels are associated with *TNFA* genotype, AJRCCM (2009) 179: A2942

Chapter 4

1. AM Wood *et al*, Variation in the transforming growth factor β gene is associated with small airways disease in AATD, AJRCCM (2008) 178: A200
2. AM Wood *et al*, Variation in the Vitamin D binding protein gene is associated with the development of bronchiectasis in AATD, AJRCCM (2008) 178: A776
3. AM Wood *et al*, Bacterial colonisation in AATD, ERJ (2008) 32: E4496

Chapter 6

1. AM Wood *et al*, Anti-elastin autoimmunity in AATD, AJRCCM (2009) 179: A1009

Chapter 7

1. AM Wood *et al*, Occupational exposure in AATD, ERJ (2008) 32: P3825
2. AM Wood *et al*, Air pollution & decline of lung function in AATD, Thorax (2008) 63 (Suppl VII): A41

Reviews, editorials and book chapters

1. AM Wood & RA Stockley, The genetics of COPD , Resp Res (2006) 7: 130-167
2. AM Wood & RA Stockley, AATD: from gene to treatment, Respiration (2007) 74: 481-92
3. AM Wood & RA Stockley, Unifying the genetics, comorbidities and management of COPD, Therapeutic advances in respiratory disease (2008) 2: 113-117
4. SL Tan & AM Wood, Chronic obstructive pulmonary disease and co-morbidities: a review and consideration of pathophysiology, Panminerva Medica (2009) In press

References

1. Mortality Statistics: review of the Registrar General on deaths by cause, sex and age in England and Wales, 2005. Office for National Statistics, (2006). DH2 number 32
2. Chronic obstructive pulmonary disease. National clinical guideline on management of chronic obstructive pulmonary disease in adults in primary and secondary care. Thorax. (2004) 59 Suppl 1:1-232.
3. WHO, NHLBI. Global Initiative for Obstructive Lung Disease (GOLD), (2003)
4. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. (2005) 26(5):948-68.
5. ATS. Cigarette smoking and health. . Am J Respir Crit Care Med. (1996) 153(2):861-5.
6. Burrows B, Knudson R, Cline M, Lebowitz M. Quantitative relationship between cigarette smoking and ventilatory function. Am Rev Resp Dis. (1977) 115:195-205.
7. Coultas DB, Hanis CL, Howard CA, Skipper BJ, Samet JM. Heritability of ventilatory function in smoking and nonsmoking New Mexico Hispanics. Am Rev Respir Dis. (1991) 144(4):770-5.
8. Management of chronic obstructive pulmonary disease in adults in primary and secondary care [http://www.nice.org.uk/pdf/CG012_nicceguideline.pdf]. 2004 [cited 2006].
9. ATS. Chronic bronchitis, asthma and pulmonary emphysema: a statement by the committee on diagnostic standards for non-tuberculous respiratory diseases. Am Rev Resp Dis. (1962) 85:762-8.
10. Snider G, Kleineman J, Thurlbeck W, Bengali Z. The definition of emphysema: a report of the National Heart, Lung and Blood Institute, Division of Lung Diseases Workshop. Am Rev Resp Dis. (1985) 132:182-5.
11. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The nature of small-airway obstruction in chronic obstructive pulmonary disease. N Engl J Med. (2004) 350(26):2645-53.
12. Curkendall SM, DeLuise C, Jones JK, Lanes S, Stang MR, Goehring E, Jr., et al. Cardiovascular disease in patients with chronic obstructive pulmonary disease, Saskatchewan Canada cardiovascular disease in COPD patients. Annals of epidemiology. (2006) 16(1):63-70.
13. Schmidt MI, Duncan BB, Sharrett AR, Lindberg G, Savage PJ, Offenbacher S, et al. Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities study): a cohort study. Lancet. (1999) 353(9165):1649-52.
14. Sin DD, Man JP, Man SF. The risk of osteoporosis in Caucasian men and women with obstructive airways disease. Am J Med. (2003) 114(1):10-4.
15. Definition and classification of chronic bronchitis for clinical and epidemiological purposes. A report to the Medical Research Council by their Committee on the Aetiology of Chronic Bronchitis. Lancet. (1965) 1(7389):775-9.
16. Webb WR. High-resolution computed tomography of obstructive lung disease. Radiologic clinics of North America. (1994) 32(4):745-57.
17. Needham M, Stockley RA. Alpha 1-antitrypsin deficiency. 3: Clinical manifestations and natural history. Thorax. (2004) 59(5):441-5.
18. Parr DG, Stoel BC, Stolk J, Stockley RA. Pattern of emphysema distribution in

alpha1-antitrypsin deficiency influences lung function impairment. Am J Respir Crit Care Med. (2004) 170(11):1172-8.
19. Demeo DL, Hersh CP, Hoffman EA, Litonjua AA, Lazarus R, Sparrow D, et al. Genetic determinants of emphysema distribution in the national emphysema treatment trial. Am J Respir Crit Care Med. (2007) 176(1):42-8.
20. Ito I, Nagai S, Handa T, Muro S, Hirai T, Tsukino M, et al. Matrix metalloproteinase-9 promoter polymorphism associated with upper lung dominant emphysema. Am J Respir Crit Care Med. (2005) 172(11):1378-82.
21. Naidich DP, McCauley DI, Khouri NF, Stitik FP, Siegelman SS. Computed tomography of bronchiectasis. Journal of computer assisted tomography. (1982) 6(3):437-44.
22. Parr DG, Guest PG, Reynolds JH, Dowson LJ, Stockley RA. Prevalence and impact of bronchiectasis in alpha1-antitrypsin deficiency. Am J Respir Crit Care Med. (2007) 176(12):1215-21.
23. Rennard SI, Fogarty C, Kelsen S, Long W, Ramsdell J, Allison J, et al. The safety and efficacy of infliximab in moderate to severe chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2007) 175(9):926-34.
24. Wechsler ME, Lehman E, Lazarus SC, Lemanske RF, Jr., Boushey HA, Deykin A, et al. beta-Adrenergic receptor polymorphisms and response to salmeterol. Am J Respir Crit Care Med. (2006) 173(5):519-26.
25. Israel E, Drazen JM, Liggett SB, Boushey HA, Cherniack RM, Chinchilli VM, et al. The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma. Am J Respir Crit Care Med. (2000) 162(1):75-80.
26. Seitz M, Wirthmuller U, Moller B, Villiger PM. The -308 tumour necrosis factor-alpha gene polymorphism predicts therapeutic response to TNFalpha-blockers in rheumatoid arthritis and spondyloarthritis patients. Rheumatology (Oxford, England). (2007) 46(1):93-6.
27. Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. (2004) 23(6):932-46.
28. Rennard SI, Schachter N, Strek M, Rickard K, Amit O. Cilomilast for COPD: results of a 6-month, placebo-controlled study of a potent, selective inhibitor of phosphodiesterase 4. Chest. (2006) 129(1):56-66.
29. Compton CH, Gubb J, Nieman R, Edelson J, Amit O, Bakst A, et al. Cilomilast, a selective phosphodiesterase-4 inhibitor for treatment of patients with chronic obstructive pulmonary disease: a randomised, dose-ranging study. Lancet. (2001) 358(9278):265-70.
30. Rabe KF, Bateman ED, O'Donnell D, Witte S, Bredenbroker D, Bethke TD. Roflumilast--an oral anti-inflammatory treatment for chronic obstructive pulmonary disease: a randomised controlled trial. Lancet. (2005) 366(9485):563-71.
31. Gompertz S, Stockley RA. A randomized, placebo-controlled trial of a leukotriene synthesis inhibitor in patients with COPD. Chest. (2002) 122(1):289-94.
32. Mahler DA, Huang S, Tabrizi M, Bell GM. Efficacy and safety of a monoclonal antibody recognizing interleukin-8 in COPD: a pilot study. Chest. (2004) 126(3):926-34.
33. Decramer M, Rutten-van Molken M, Dekhuijzen PN, Troosters T, van Herwaarden C, Pellegrino R, et al. Effects of N-acetylcysteine on outcomes in chronic obstructive pulmonary disease (Bronchitis Randomized on NAC Cost-Utility Study, BRONCUS): a randomised placebo-controlled trial. Lancet. (2005) 365(9470):1552-60.

34. Bruce C, Thomas PS. The effect of marimastat, a metalloprotease inhibitor, on allergen-induced asthmatic hyper-reactivity. Toxicology and applied pharmacology. (2005) 205(2):126-32.
35. Mao JT, Goldin JG, Dermand J, Ibrahim G, Brown MS, Emerick A, et al. A pilot study of all-trans-retinoic acid for the treatment of human emphysema. Am J Respir Crit Care Med. (2002) 165(5):718-23.
36. Celik P, Sakar A, Havlucu Y, Yuksel H, Turkdogan P, Yorgancioglu A. Short-term effects of montelukast in stable patients with moderate to severe COPD. Respir Med. (2005) 99(4):444-50.
37. Rubinstein I, Kumar B, Schriever C. Long-term montelukast therapy in moderate to severe COPD--a preliminary observation. Respir Med. (2004) 98(2):134-8.
38. Stockley RA. Neutrophils and protease/antiprotease imbalance. Am J Respir Crit Care Med. (1999) 160(5 Pt 2):S49-52.
39. Shapiro SD, Senior RM. Matrix metalloproteinases. Matrix degradation and more. Am J Respir Cell Mol Biol. (1999) 20(6):1100-2.
40. Eriksson S. Studies in alpha 1 antitrypsin deficiency. Acta MedScand. (1965) 177 (Suppl.):432.
41. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J. (2003) 22(4):672-88.
42. Dery O, Corvera CU, Steinhoff M, Bunnett NW. Proteinase-activated receptors: novel mechanisms of signaling by serine proteases. Am J Physiol. (1998) 274(6 Pt 1):C1429-52.
43. Hunninghake GW, Crystal RG. Cigarette smoking and lung destruction. Accumulation of neutrophils in the lungs of cigarette smokers. Am Rev Respir Dis. (1983) 128(5):833-8.
44. Li XY, Gilmour PS, Donaldson K, MacNee W. Free radical activity and pro-inflammatory effects of particulate air pollution (PM10) in vivo and in vitro. Thorax. (1996) 51(12):1216-22.
45. Young RP, Hopkins R, Black PN, Eddy C, Wu L, Gamble GD, et al. Functional variants of antioxidant genes in smokers with COPD and in those with normal lung function. Thorax. (2006) 61(5):394-9.
46. MacNee W. Oxidants/antioxidants and COPD. Chest. (2000) 117(5 Suppl 1):303S-17S.
47. Devlin RB, McDonnell WF, Becker S, Madden MC, McGee MP, Perez R, et al. Time-dependent changes of inflammatory mediators in the lungs of humans exposed to 0.4 ppm ozone for 2 hr: a comparison of mediators found in bronchoalveolar lavage fluid 1 and 18 hr after exposure. Toxicology and applied pharmacology. (1996) 138(1):176-85.
48. Zheng T, Zhu Z, Wang Z, Homer RJ, Ma B, Riese RJ, Jr., et al. Inducible targeting of IL-13 to the adult lung causes matrix metalloproteinase- and cathepsin-dependent emphysema. J Clin Invest. (2000) 106(9):1081-93.
49. Churg A, Wang RD, Tai H, Wang X, Xie C, Wright JL. Tumor necrosis factor-alpha drives 70% of cigarette smoke-induced emphysema in the mouse. Am J Respir Crit Care Med. (2004) 170(5):492-8.
50. Parr DG, White AJ, Bayley DL, Guest PJ, Stockley RA. Inflammation in sputum relates to progression of disease in subjects with COPD: a prospective descriptive study. Respir Res. (2006) 7:136.

51. Gan WQ, Man SF, Senthilselvan A, Sin DD. Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. Thorax. (2004) 59(7):574-80.
52. Vernooy JH, Kucukaycan M, Jacobs JA, Chavannes NH, Buurman WA, Dentener MA, et al. Local and systemic inflammation in patients with chronic obstructive pulmonary disease: soluble tumor necrosis factor receptors are increased in sputum. Am J Respir Crit Care Med. (2002) 166(9):1218-24.
53. Sapey E, Ahmad A, Bayley D, Newbold P, Snell N, Rugman P, et al. Imbalances between interleukin-1 and tumor necrosis factor agonists and antagonists in stable COPD. J Clin Immunol. (2009) 29(4):508-16.
54. Wouters EF. Local and systemic inflammation in chronic obstructive pulmonary disease. Proc Am Thorac Soc. (2005) 2(1):26-33.
55. Redline S, Tishler PV, Rosner B, Lewitter FI, Vandenburgh M, Weiss ST, et al. Genotypic and phenotypic similarities in pulmonary function among family members of adult monozygotic and dizygotic twins. Am J Epidemiol. (1989) 129(4):827-36.
56. Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. Am J Epidemiol. (1984) 120(4):518-30.
57. Kueppers F, Miller RD, Gordon H, Hepper NG, Offord K. Familial prevalence of chronic obstructive pulmonary disease in a matched pair study. Am J Med. (1977) 63(3):336-42.
58. Silverman EK, Province MA, Rao DC, Pierce JA, Campbell EJ. A family study of the variability of pulmonary function in alpha 1-antitrypsin deficiency. Quantitative phenotypes. Am Rev Respir Dis. (1990) 142(5):1015-21.
59. Patel BD, Coxson HO, Pillai SG, Agusti AG, Calverley PM, Donner CF, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2008) 178(5):500-5.
60. Fletcher C, Peto R. The natural history of chronic airflow obstruction. Br Med J. (1977) 1(6077):1645-8.
61. Strachan T, Read A. Human Molecular Genetics. Oxford: Garland Science Publishers; 2003.
62. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science. (1998) 280(5366):1077-82.
63. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature. (2004) 429(6990):446-52.
64. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. (2004) 305(5683):525-8.
65. Cappuzzo F, Hirsch FR, Rossi E, Bartolini S, Ceresoli GL, Bemis L, et al. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. J Natl Cancer Inst. (2005) 97(9):643-55.
66. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. (1955) 7(3):277-318.
67. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. (1995) 11(3):241-7.
68. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic

associations with complex outcomes. Lancet. (2003) 361(9360):865-72.
69. 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.
70. The International HapMap Consortium. A Haplotype Map of the Human Genome. Nature. (2005) 437:1299-320.
71. Wood AM, Needham M, Simmonds MJ, Newby PR, Gough SC, Stockley RA. Phenotypic differences in alpha 1 antitrypsin-deficient sibling pairs may relate to genetic variation. COPD. (2008) 5(6):353-9.
72. Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. Am J Respir Cell Mol Biol. (2005) 33(1):71-8.
73. 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 Genet. (2009) 5(3):e1000421.
74. Zhang B, Ye S, Herrmann SM, Eriksson P, de Maat M, Evans A, et al. Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. Circulation. (1999) 99(14):1788-94.
75. Zhou M, Huang SG, Wan HY, Li B, Deng WW, Li M. Genetic polymorphism in matrix metalloproteinase-9 and the susceptibility to chronic obstructive pulmonary disease in Han population of south China. Chin Med J (Engl). (2004) 117(10):1481-4.
76. Minematsu N, Nakamura H, Tateno H, Nakajima T, Yamaguchi K. Genetic polymorphism in matrix metalloproteinase-9 and pulmonary emphysema. Biochem Biophys Res Commun. (2001) 289(1):116-9.
77. Joos L, He JQ, Shepherdson MB, Connett JE, Anthonisen NR, Pare PD, et al. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. Hum Mol Genet. (2002) 11(5):569-76.
78. Hersh CP, Demeo DL, Lazarus R, Celedon JC, Raby BA, Benditt JO, et al. Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2006) 173(9):977-84.
79. Kong X, Coxson HO, Lomas DA, Silverman EK, Pillai SG. Genome-wide association of high resolution CT phenotypes in two independent COPD populations. American Thoracic Society; San Diego: American Journal of Respiratory and Critical Care Medicine; 2009. p. A2444.
80. Hirano K, Sakamoto T, Uchida Y, Morishima Y, Masuyama K, Ishii Y, et al. Tissue inhibitor of metalloproteinases-2 gene polymorphisms in chronic obstructive pulmonary disease. Eur Respir J. (2001) 18(5):748-52.
81. Poller W, Faber JP, Scholz S, Weidinger S, Bartholome K, Olek K, et al. Mis-sense mutation of alpha 1-antichymotrypsin gene associated with chronic lung disease. Lancet. (1992) 339(8808):1538.
82. Poller W, Faber JP, Weidinger S, Tief K, Scholz S, Fischer M, et al. A leucine-to-proline substitution causes a defective alpha 1-antichymotrypsin allele associated with familial obstructive lung disease. Genomics. (1993) 17(3):740-3.
83. Ishii T, Matsuse T, Teramoto S, Matsui H, Hosoi T, Fukuchi Y, et al. Association between alpha-1-antichymotrypsin polymorphism and susceptibility to chronic obstructive pulmonary disease. Eur J Clin Invest. (2000) 30(6):543-8.

84. Benetazzo MG, Gile LS, Bombieri C, Malerba G, Massobrio M, Pignatti PF, et al. alpha 1-antitrypsin TAQ I polymorphism and alpha 1-antichymotrypsin mutations in patients with obstructive pulmonary disease. Respir Med. (1999) 93(9):648-54.
85. Scott RW, Bergman BL, Bajpai A, Hersh RT, Rodriguez H, Jones BN, et al. Protease nexin. Properties and a modified purification procedure. J Biol Chem. (1985) 260(11):7029-34.
86. Baker JB, Low DA, Simmer RL, Cunningham DD. Protease-nexin: a cellular component that links thrombin and plasminogen activator and mediates their binding to cells. Cell. (1980) 21(1):37-45.
87. 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. Hum Mol Genet. (2002) 11(6):623-32.
88. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet. (2002) 70(5):1229-39.
89. Demeo DL, Mariani TJ, Lange C, Srisuma S, Litonjua AA, Celedon JC, et al. The SERPINE2 Gene Is Associated with Chronic Obstructive Pulmonary Disease. Am J Hum Genet. (2006) 78(2):253-64.
90. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, Anderson WH, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med. (2007) 176(2):167-73.
91. Chappell S, Daly L, Morgan K, Baranes TG, Roca J, Rabinovich R, et al. The SERPINE2 gene and chronic obstructive pulmonary disease. Am J Hum Genet. (2006) 79(1):184-6; author reply 6-7.
92. Ashitani J, Mukae H, Arimura Y, Matsukura S. Elevated plasma procoagulant and fibrinolytic markers in patients with chronic obstructive pulmonary disease. Intern Med. (2002) 41(3):181-5.
93. Wang IM, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, et al. Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. Am J Respir Crit Care Med. (2008) 177(4):402-11.
94. He JQ, Ruan J, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function in smokers. Am J Respir Crit Care Med. (2002) 166(3):323-8.
95. Gilliland FD, Gauderman WJ, Vora H, Rappaport E, Dubeau L. Effects of glutathione-S-transferase M1, T1, and P1 on childhood lung function growth. Am J Respir Crit Care Med. (2002) 166(5):710-6.
96. Ishii T, Matsuse T, Teramoto S, Matsui H, Miyao M, Hosoi T, et al. Glutathione S-transferase P1 (GSTP1) polymorphism in patients with chronic obstructive pulmonary disease. Thorax. (1999) 54(8):693-6.
97. He JQ, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Glutathione S-transferase variants and their interaction with smoking on lung function. Am J Respir Crit Care Med. (2004) 170(4):388-94.
98. Harrison DJ, Cantlay AM, Rae F, Lamb D, Smith CA. Frequency of glutathione S-transferase M1 deletion in smokers with emphysema and lung cancer. Hum Exp Toxicol. (1997) 16(7):356-60.

99. Baranova H, Perriot J, Albuisson E, Ivaschenko T, Baranov VS, Hemery B, et al. Peculiarities of the GSTM1 0/0 genotype in French heavy smokers with various types of chronic bronchitis. Hum Genet. (1997) 99(6):822-6.
100. Smith CA, Harrison DJ. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. Lancet. (1997) 350(9078):630-3.
101. Yoshikawa M, Hiyama K, Ishioka S, Maeda H, Maeda A, Yamakido M. Microsomal epoxide hydrolase genotypes and chronic obstructive pulmonary disease in Japanese. Int J Mol Med. (2000) 5(1):49-53.
102. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Pare PD. Susceptibility genes for rapid decline of lung function in the lung health study. Am J Respir Crit Care Med. (2001) 163(2):469-73.
103. Guenegou A, Leynaert B, Benessiano J, Pin I, Demoly P, Neukirch F, et al. Association of lung function decline with the heme oxygenase-1 gene promoter microsatellite polymorphism in a general population sample. Results from the European Community Respiratory Health Survey (ECRHS), France. J Med Genet. (2006) 43(8):e43.
104. Yamada N, Yamaya M, Okinaga S, Nakayama K, Sekizawa K, Shibahara S, et al. Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. Am J Hum Genet. (2000) 66(1):187-95.
105. Sakao S, Tatsumi K, Igari H, Shino Y, Shirasawa H, Kuriyama T. Association of tumor necrosis factor alpha gene promoter polymorphism with the presence of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2001) 163(2):420-2.
106. Sakao S, Tatsumi K, Igari H, Watanabe R, Shino Y, Shirasawa H, et al. Association of tumor necrosis factor-alpha gene promoter polymorphism with low attenuation areas on high-resolution CT in patients with COPD. Chest. (2002) 122(2):416-20.
107. Huang SL, Su CH, Chang SC. Tumor necrosis factor-alpha gene polymorphism in chronic bronchitis. Am J Respir Crit Care Med. (1997) 156(5):1436-9.
108. 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). Hum Mol Genet. (2004) 13(15):1649-56.
109. Ishii T, Keicho N, Teramoto S, Azuma A, Kudoh S, Fukuchi Y, et al. Association of Gc-globulin variation with susceptibility to COPD and diffuse panbronchiolitis. Eur Respir J. (2001) 18(5):753-7.
110. Horne SL, Cockcroft DW, Dosman JA. Possible protective effect against chronic obstructive airways disease by the GC2 allele. Hum Hered. (1990) 40(3):173-6.
111. Schellenberg D, Pare PD, Weir TD, Spinelli JJ, Walker BA, Sandford AJ. Vitamin D binding protein variants and the risk of COPD. Am J Respir Crit Care Med. (1998) 157(3 Pt 1):957-61.
112. van der Pouw Kraan TC, Kucukaycan M, Bakker AM, Baggen JM, van der Zee JS, Dentener MA, et al. Chronic obstructive pulmonary disease is associated with the -1055 IL-13 promoter polymorphism. Genes Immun. (2002) 3(7):436-9.
113. Foreman MG, Demeo DL, Hersh CP, Carey VJ, Fan VS, Reilly JJ, et al. Polymorphic variation in surfactant protein B is associated with COPD exacerbations. Eur Respir J. (2008).
114. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al.

dbSNP: the NCBI database of genetic variation. Nucleic acids research. (2001) 29(1):308-11.
115. Mannervik B. The isoenzymes of glutathione transferase. Adv Enzymol Relat Areas Mol Biol. (1985) 57:357-417.
116. Sundberg K, Johansson AS, Stenberg G, Widersten M, Seidel A, Mannervik B, et al. Differences in the catalytic efficiencies of allelic variants of glutathione transferase P1-1 towards carcinogenic diol epoxides of polycyclic aromatic hydrocarbons. Carcinogenesis. (1998) 19(3):433-6.
117. Ryberg D, Skaug V, Hewer A, Phillips DH, Harries LW, Wolf CR, et al. Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. Carcinogenesis. (1997) 18(7):1285-9.
118. Bartsch H, Petruzzelli S, De Flora S, Hietanen E, Camus AM, Castegnaro M, et al. Carcinogen metabolism in human lung tissues and the effect of tobacco smoking: results from a case--control multicenter study on lung cancer patients. Environ Health Perspect. (1992) 98:119-24.
119. Tingle MD, Pirmohamed M, Templeton E, Wilson AS, Madden S, Kitteringham NR, et al. An investigation of the formation of cytotoxic, genotoxic, protein-reactive and stable metabolites from naphthalene by human liver microsomes. Biochem Pharmacol. (1993) 46(9):1529-38.
120. Hassett C, Aicher L, Sidhu JS, Omiecinski CJ. Human microsomal epoxide hydrolase: genetic polymorphism and functional expression in vitro of amino acid variants. Hum Mol Genet. (1994) 3(3):421-8.
121. Hosagrahara VP, Rettie AE, Hassett C, Omiecinski CJ. Functional analysis of human microsomal epoxide hydrolase genetic variants. Chem Biol Interact. (2004) 150(2):149-59.
122. Raaka S, Hassett C, Omiencinski CJ. Human microsomal epoxide hydrolase: 5'-flanking region genetic polymorphisms. Carcinogenesis. (1998) 19(3):387-93.
123. Oury TD, Chang LY, Marklund SL, Day BJ, Crapo JD. Immunocytochemical localization of extracellular superoxide dismutase in human lung. Lab Invest. (1994) 70(6):889-98.
124. Bowler RP, Nicks M, Tran K, Tanner G, Chang LY, Young SK, et al. Extracellular superoxide dismutase attenuates lipopolysaccharide-induced neutrophilic inflammation. Am J Respir Cell Mol Biol. (2004) 31(4):432-9.
125. Juul K, Tybjaerg-Hansen A, Marklund S, Lange P, Nordestgaard BG. Genetically increased antioxidative protection and decreased chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2006) 173(8):858-64.
126. Tenhunen R, Marver HS, Schmid R. The enzymatic conversion of heme to bilirubin by microsomal heme oxygenase. Proc Natl Acad Sci U S A. (1968) 61(2):748-55.
127. Maestrelli P, El Messlemani AH, De Fina O, Nowicki Y, Saetta M, Mapp C, et al. Increased expression of heme oxygenase (HO)-1 in alveolar spaces and HO-2 in alveolar walls of smokers. Am J Respir Crit Care Med. (2001) 164(8 Pt 1):1508-13.
128. Okinaga S, Takahashi K, Takeda K, Yoshizawa M, Fujita H, Sasaki H, et al. Regulation of human heme oxygenase-1 gene expression under thermal stress. Blood. (1996) 87(12):5074-84.
129. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional

activation. Proc Natl Acad Sci U S A. (1997) 94(7):3195-9.
130. Whitsett JA, Weaver TE. Hydrophobic surfactant proteins in lung function and disease. N Engl J Med. (2002) 347(26):2141-8.
131. Pastva AM, Wright JR, Williams KL. Immunomodulatory roles of surfactant proteins A and D: implications in lung disease. Proc Am Thorac Soc. (2007) 4(3):252-7.
132. Seifart C, Plagens A, Brodje D, Muller B, von Wichert P, Floros J. Surfactant protein B intron 4 variation in German patients with COPD and acute respiratory failure. Dis Markers. (2002) 18(3):129-36.
133. Guo X, Lin HM, Lin Z, Montano M, Sansores R, Wang G, et al. Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. Eur Respir J. (2001) 18(3):482-90.
134. van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Nolte IM, Boezen HM. Decorin and TGF-beta1 polymorphisms and development of COPD in a general population. Respir Res. (2006) 7:89.
135. Wu L, Chau J, Young RP, Pokorny V, Mills GD, Hopkins R, et al. Transforming growth factor-beta1 genotype and susceptibility to chronic obstructive pulmonary disease. Thorax. (2004) 59(2):126-9.
136. Ito I, Nagai S, Hoshino Y, Muro S, Hirai T, Tsukino M, et al. Risk and severity of COPD is associated with the group-specific component of serum globulin 1F allele. Chest. (2004) 125(1):63-70.
137. van der Pouw Kraan TC, van Veen A, Boeije LC, van Tuyl SA, de Groot ER, Stapel SO, et al. An IL-13 promoter polymorphism associated with increased risk of allergic asthma. Genes Immun. (1999) 1(1):61-5.
138. Sadeghnejad A, Meyers DA, Bottai M, Sterling DA, Bleecker ER, Ohar JA. IL13 promoter polymorphism 1112C/T modulates the adverse effect of tobacco smoking on lung function. Am J Respir Crit Care Med. (2007) 176(8):748-52.
139. Zhu Z, Homer RJ, Wang Z, Chen Q, Geba GP, Wang J, et al. Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. J Clin Invest. (1999) 103(6):779-88.
140. Laurell CB, Eriksson S. The electrophoretic alpha 1 globulin pattern of serum in alpha 1 antitrypsin deficiency. Scand J Clin Lab Invest. (1963) 15:132-40.
141. Koj A, Regoeczi E, Toews CJ, Leveille R, Gauldie J. Synthesis of antithrombin III and alpha-1-antitrypsin by the perfused rat liver. Biochim Biophys Acta. (1978) 539(4):496-504.
142. Mornex JF, Chytil-Weir A, Martinet Y, Courtney M, LeCocq JP, Crystal RG. Expression of the alpha-1-antitrypsin gene in mononuclear phagocytes of normal and alpha-1-antitrypsin-deficient individuals. J Clin Invest. (1986) 77(6):1952-61.
143. Venembre P, Boutten A, Seta N, Dehoux MS, Crestani B, Aubier M, et al. Secretion of alpha 1-antitrypsin by alveolar epithelial cells. FEBS Lett. (1994) 346(2-3):171-4.
144. Cichy J, Potempa J, Travis J. Biosynthesis of alpha1-proteinase inhibitor by human lung-derived epithelial cells. J Biol Chem. (1997) 272(13):8250-5.
145. Lomas DA, Evans DL, Finch JT, Carrell RW. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. Nature. (1992) 357(6379):605-7.
146. Wilczynska M, Fa M, Karolin J, Ohlsson PI, Johansson LB, Ny T. Structural

insights into serpin-protease complexes reveal the inhibitory mechanism of serpins. Nat Struct Biol. (1997) 4(5):354-7.
147. Elliott PR, Lomas DA, Carrell RW, Abrahams JP. Inhibitory conformation of the reactive loop of alpha 1-antitrypsin. Nat Struct Biol. (1996) 3(8):676-81.
148. Stoller JK, Sandhaus RA, Turino G, Dickson R, Rodgers K, Strange C. Delay in diagnosis of alpha1-antitrypsin deficiency: a continuing problem. Chest. (2005) 128(4):1989-94.
149. Holme J, Stockley JA, Stockley RA, editors. When should we start monitoring alpha 1 antitrypsin deficient subjects? British Thoracic Society; 2007; London.
150. Eden E, Hammel J, Rouhani FN, Brantly ML, Barker AF, Buist AS, et al. Asthma features in severe alpha1-antitrypsin deficiency: experience of the National Heart, Lung, and Blood Institute Registry. Chest. (2003) 123(3):765-71.
151. Piitulainen E, Sveger T. Effect of environmental and clinical factors on lung function and respiratory symptoms in adolescents with alpha1-antitrypsin deficiency. Acta Paediatr. (1998) 87(11):1120-4.
152. Piitulainen E, Eriksson S. Decline in FEV1 related to smoking status in individuals with severe alpha1-antitrypsin deficiency (PiZZ). Eur Respir J. (1999) 13(2):247-51.
153. Dawkins PA, Dawkins CL, Stockley JA, Needham M, Stockley RA, editors. Associations with Annual Decline in Lung Function in Alpha-1-Antitrypsin Deficiency. American Thoracic Society Conference; 2006; San Diego.
154. Shaker SB, Stavngaard T, Stolk J, Stoel B, Dirksen A. Alpha1-antitrypsin deficiency. 7: Computed tomographic imaging in alpha1-antitrypsin deficiency. Thorax. (2004) 59(11):986-91.
155. Abusriwil H, Stockley RA. Alpha-1-antitrypsin replacement therapy: current status. Current opinion in pulmonary medicine. (2006) 12(2):125-31.
156. Parfrey H, Dafforn TR, Belorgey D, Lomas DA, Mahadeva R. Inhibiting polymerization: new therapeutic strategies for Z alpha1-antitrypsin-related emphysema. Am J Respir Cell Mol Biol. (2004) 31(2):133-9.
157. Devlin GL, Parfrey H, Tew DJ, Lomas DA, Bottomley SP. Prevention of polymerization of M and Z alpha1-Antitrypsin (alpha1-AT) with trimethylamine N-oxide. Implications for the treatment of alpha1-at deficiency. Am J Respir Cell Mol Biol. (2001) 24(6):727-32.
158. Lu Y, Choi YK, Campbell-Thompson M, Li C, Tang Q, Crawford JM, et al. Therapeutic level of functional human alpha 1 antitrypsin (hAAT) secreted from murine muscle transduced by adeno-associated virus (rAAV1) vector. J Gene Med. (2006) 8(6):730-5.
159. Dirksen A, Piitulainen E, Parr DG, Deng C, Wencker M, Shaker SB, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in alpha1-antitrypsin deficiency. Eur Respir J. (2009) 33(6):1345-53.
160. Cox DW, Markovic VD, Teshima IE. Genes for immunoglobulin heavy chains and for alpha 1-antitrypsin are localized to specific regions of chromosome 14q. Nature. (1982) 297(5865):428-30.
161. Cox DW, Woo SL, Mansfield T. DNA restriction fragments associated with alpha 1-antitrypsin indicate a single origin for deficiency allele PI Z. Nature. (1985) 316(6023):79-81.
162. Schroeder WT, Miller MF, Woo SL, Saunders GF. Chromosomal localization of the

human alpha 1-antitrypsin gene (PI) to 14q31-32. Am J Hum Genet. (1985) 37(5):868-72.
163. Lai EC, Kao FT, Law ML, Woo SL. Assignment of the alpha 1-antitrypsin gene and a sequence-related gene to human chromosome 14 by molecular hybridization. Am J Hum Genet. (1983) 35(3):385-92.
164. Haq I, Chappell S, Johnson SR, Daly L, Morgan K, Baranes TG, et al., editors. Genetic variants of matrix metalloproteases 1, 9 and 12 in COPD. American Thoracic Society; 2009; San Diego.
165. Fagerhol MK, Laurell CB. The Pi system-inherited variants of serum alpha 1-antitrypsin. Prog Med Genet. (1970) 7:96-111.
166. Fagerhol MK, Laurell CB. The polymorphism of "prealbumins" and alpha-1-antitrypsin in human sera. Clin Chim Acta. (1967) 16(2):199-203.
167. DeMeo DL, Silverman EK. Alpha1-antitrypsin deficiency. 2: genetic aspects of alpha(1)-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. Thorax. (2004) 59(3):259-64.
168. Cook L, Burdon JG, Brenton S, Knight KR, Janus ED. Kinetic characterisation of alpha-1-antitrypsin F as an inhibitor of human neutrophil elastase. Pathology. (1996) 28(3):242-7.
169. Curiel DT, Vogelmeier C, Hubbard RC, Stier LE, Crystal RG. Molecular basis of alpha 1-antitrypsin deficiency and emphysema associated with the alpha 1-antitrypsin Mmineral springs allele. Mol Cell Biol. (1990) 10(1):47-56.
170. Scott CF, Carrell RW, Glaser CB, Kueppers F, Lewis JH, Colman RW. Alpha-1-antitrypsin-Pittsburgh. A potent inhibitor of human plasma factor XIa, kallikrein, and factor XIIf. J Clin Invest. (1986) 77(2):631-4.
171. Morgan K, Chappell S, Guetta-Baranes T, Morley S, Kalsheker N. The alpha-1-antitrypsin gene promoter in human A549 lung derived cells, and a novel transcription initiation site. Int J Biochem Cell Biol. (2009) 41(5):1157-64.
172. Kalsheker N, Morley S, Morgan K. Gene regulation of the serine proteinase inhibitors alpha1-antitrypsin and alpha1-antichymotrypsin. Biochem Soc Trans. (2002) 30(2):93-8.
173. Chappell S, Hadzic N, Stockley R, Guetta-Baranes T, Morgan K, Kalsheker N. A polymorphism of the alpha1-antitrypsin gene represents a risk factor for liver disease. Hepatology. (2008) 47(1):127-32.
174. 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. Eur Respir J. (2006) 27(1):77-84.
175. American Thoracic Society/European Respiratory Society Statement: Standards for the Diagnosis and Management of Individuals with Alpha-1 Antitrypsin Deficiency. Am J Respir Crit Care Med. (2003) 168(7):818-900.
176. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax. (2004) 59(10):843-9.
177. Wu Y, Whitman I, Molmenti E, Moore K, Hippenmeyer P, Perlmutter DH. A lag in intracellular degradation of mutant alpha 1-antitrypsin correlates with the liver disease phenotype in homozygous PiZZ alpha 1-antitrypsin deficiency. Proc Natl Acad Sci U S A. (1994) 91(19):9014-8.
178. Matsunaga E, Shiokawa S, Nakamura H, Maruyama T, Tsuda K, Fukumaki Y.

Molecular analysis of the gene of the alpha 1-antitrypsin deficiency variant, Mnichinan. Am J Hum Genet. (1990) 46(3):602-12.
179. Seyama K, Nukiwa T, Souma S, Shimizu K, Kira S. Alpha 1-antitrypsin-deficient variant Siiyama (Ser53[TCC] to Phe53[TTC]) is prevalent in Japan. Status of alpha 1-antitrypsin deficiency in Japan. Am J Respir Crit Care Med. (1995) 152(6 Pt 1):2119-26.
180. Seyama K, Nukiwa T, Takabe K, Takahashi H, Miyake K, Kira S. Siiyama (serine 53 (TCC) to phenylalanine 53 (TTC)). A new alpha 1-antitrypsin-deficient variant with mutation on a predicted conserved residue of the serpin backbone. J Biol Chem. (1991) 266(19):12627-32.
181. Lomas DA, Finch JT, Seyama K, Nukiwa T, Carrell RW. Alpha 1-antitrypsin Siiyama (Ser53-->Phe). Further evidence for intracellular loop-sheet polymerization. J Biol Chem. (1993) 268(21):15333-5.
182. Elliott PR, Stein PE, Bilton D, Carrell RW, Lomas DA. Structural explanation for the deficiency of S alpha 1-antitrypsin. Nat Struct Biol. (1996) 3(11):910-1.
183. Mahadeva R, Chang WS, Dafforn TR, Oakley DJ, Foreman RC, Calvin J, et al. Heteropolymerization of S, I, and Z alpha1-antitrypsin and liver cirrhosis. J Clin Invest. (1999) 103(7):999-1006.
184. Luisetti M, Seersholm N. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax. (2004) 59(2):164-9.
185. Turino GM, Barker AF, Brantly ML, Cohen AB, Connelly RP, Crystal RG, et al. Clinical features of individuals with PI*SZ phenotype of alpha 1-antitrypsin deficiency. alpha 1-Antitrypsin Deficiency Registry Study Group. Am J Respir Crit Care Med. (1996) 154(6 Pt 1):1718-25.
186. Nukiwa T, Takahashi H, Brantly M, Courtney M, Crystal RG. alpha 1-Antitrypsin nullGranite Falls, a nonexpressing alpha 1-antitrypsin gene associated with a frameshift to stop mutation in a coding exon. J Biol Chem. (1987) 262(25):11999-2004.
187. Sifers RN, Brashears-Macatee S, Kidd VJ, Muensch H, Woo SL. A frameshift mutation results in a truncated alpha 1-antitrypsin that is retained within the rough endoplasmic reticulum. J Biol Chem. (1988) 263(15):7330-5.
188. Fraizer GC, Siewertsen M, Harrold TR, Cox DW. Deletion/frameshift mutation in the alpha 1-antitrypsin null allele, PI*QObolton. Hum Genet. (1989) 83(4):377-82.
189. Curiel D, Brantly M, Curiel E, Stier L, Crystal RG. Alpha 1-antitrypsin deficiency caused by the alpha 1-antitrypsin Nullmattawa gene. An insertion mutation rendering the alpha 1-antitrypsin gene incapable of producing alpha 1-antitrypsin. J Clin Invest. (1989) 83(4):1144-52.
190. Garver RI, Jr., Mornex JF, Nukiwa T, Brantly M, Courtney M, LeCocq JP, et al. Alpha 1-antitrypsin deficiency and emphysema caused by homozygous inheritance of non-expressing alpha 1-antitrypsin genes. N Engl J Med. (1986) 314(12):762-6.
191. Takahashi H, Crystal RG. Alpha 1-antitrypsin Null(isola di procida): an alpha 1-antitrypsin deficiency allele caused by deletion of all alpha 1-antitrypsin coding exons. Am J Hum Genet. (1990) 47(3):403-13.
192. Fregonese L, Stolk J, Frants RR, Veldhuisen B. Alpha-1 antitrypsin Null mutations and severity of emphysema. Respir Med. (2008) 102(6):876-84.
193. Dahl M, Hersh CP, Ly NP, Berkey CS, Silverman EK, Nordestgaard BG. The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J. (2005) 26(1):67-76.

194. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC. Variability of pulmonary function in alpha-1-antitrypsin deficiency: residual family resemblance beyond the effect of the Pi locus. Hum Hered. (1990) 40(6):340-55.
195. Rodriguez F, de la Roza C, Jardi R, Schaper M, Vidal R, Miravitlles M. Glutathione S-transferase P1 and lung function in patients with alpha1-antitrypsin deficiency and COPD. Chest. (2005) 127(5):1537-43.
196. Moncada S, Palmer RM, Higgs EA. Nitric oxide: physiology, pathophysiology, and pharmacology. Pharmacological reviews. (1991) 43(2):109-42.
197. Novoradovsky A, Brantly ML, Waclawiw MA, Chaudhary PP, Ihara H, Qi L, et al. Endothelial nitric oxide synthase as a potential susceptibility gene in the pathogenesis of emphysema in alpha1-antitrypsin deficiency. American journal of respiratory cell and molecular biology. (1999) 20(3):441-7.
198. Ali-Osman F, Akande O, Antoun G, Mao JX, Buolamwini J. Molecular cloning, characterization, and expression in Escherichia coli of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. J Biol Chem. (1997) 272(15):10004-12.
199. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, et al. IL10 Polymorphisms are Associated with Airflow Obstruction in Severe alpha 1-antitrypsin Deficiency. Am J Respir Cell Mol Biol. (2007).
200. Churg A, Dai J, Tai H, Xie C, Wright JL. Tumor necrosis factor-alpha is central to acute cigarette smoke-induced inflammation and connective tissue breakdown. Am J Respir Crit Care Med. (2002) 166(6):849-54.
201. Hill AT, Bayley DL, Campbell EJ, Hill SL, Stockley RA. Airways inflammation in chronic bronchitis: the effects of smoking and alpha1-antitrypsin deficiency. Eur Respir J. (2000) 15(5):886-90.
202. Churg A, Wang X, Wang RD, Meixner SC, Pryzdial EL, Wright JL. {alpha}-1-Antitrypsin Suppresses TNF{alpha} and MMP-12 Production by Cigarette Smoke-Stimulated Macrophages. Am J Respir Cell Mol Biol. (2007) 37:144-51.
203. Stolk J, editor. Identification of genetic modifiers of emphysema in phenotypically discordant PiZZ alpha 1 antitrypsin deficient patients. ERS; 2007; Stockholm.
204. Kauffmann F, Kleisbauer JP, Cambon-De-Mouzon A, Mercier P, Constans J, Blanc M, et al. Genetic markers in chronic air-flow limitation. A genetic epidemiologic study. Am Rev Respir Dis. (1983) 127(3):263-9.
205. Lee SH, Goswami S, Grudo A, Song LZ, Bandi V, Goodnight-White S, et al. Antielastin autoimmunity in tobacco smoking-induced emphysema. Nature medicine. (2007) 13(5):567-9.
206. Taraseviciene-Stewart L, Scerbavicius R, Choe KH, Moore M, Sullivan A, Nicolls MR, et al. An animal model of autoimmune emphysema. Am J Respir Crit Care Med. (2005) 171(7):734-42.
207. WHO. Air Quality Guidelines: Global Update 2005. Copenhagen: World Health Organisation; 2006.
208. Anthonisen NR, Manfreda J, Warren CP, Hershfield ES, Harding GK, Nelson NA. Antibiotic therapy in exacerbations of chronic obstructive pulmonary disease. Ann Intern Med. (1987) 106(2):196-204.
209. Dowson LJ, Guest PJ, Hill SL, Holder RL, Stockley RA. High-resolution computed tomography scanning in alpha1-antitrypsin deficiency: relationship to lung

function and health status. Eur Respir J. (2001) 17(6):1097-104.
210. Meneely GR, Kaltreider NL. The volume of the lung determined by helium dilution; description of the method and comparison with other procedures. J Clin Invest. (1949) 28(1):129-39.
211. Blakemore WS, Forster RE, Morton JW, Ogilvie CM. A standardized breath holding technique for the clinical measurement of the diffusing capacity of the lung for carbon monoxide. J Clin Invest. (1957) 36(1 Part 1):1-17.
212. Parr DG. Quantitative computed tomography in chronic obstructive pulmonary disease: University of Birmingham; 2005.
213. Naidich DP. High-resolution computed tomography of cystic lung disease. Seminars in roentgenology. (1991) 26(2):151-74.
214. Soejima K, Yamaguchi K, Kohda E, Takeshita K, Ito Y, Mastubara H, et al. Longitudinal follow-up study of smoking-induced lung density changes by high-resolution computed tomography. Am J Respir Crit Care Med. (2000) 161(4 Pt 1):1264-73.
215. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, et al. Ensembl 2007. Nucleic acids research. (2007) 35(Database issue):D610-7.
216. Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, et al. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue antigens. (1995) 46(5):355-67.
217. WTCCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. (2007) 447(7145):661-78.
218. Simmonds MJ, Heward JM, Howson JM, Foxall H, Nithiyananthan R, Franklyn JA, et al. A systematic approach to the assessment of known TNF-alpha polymorphisms in Graves' disease. Genes Immun. (2004) 5(4):267-73.
219. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics (Oxford, England). (2005) 21(2):263-5.
220. Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. Behavior genetics. (1972) 2(1):3-19.
221. Statistical analysis for genetic epidemiology
National Centre for Research Resources; 2008.
222. Nedwin GE, Naylor SL, Sakaguchi AY, Smith D, Jarrett-Nedwin J, Pennica D, et al. Human lymphotoxin and tumor necrosis factor genes: structure, homology and chromosomal localization. Nucleic acids research. (1985) 13(17):6361-73.
223. Spriggs DR, Deutsch S, Kufe DW. Genomic structure, induction, and production of TNF-alpha. Immunology series. (1992) 56:3-34.
224. Higham MA, Pride NB, Alikhan A, Morrell NW. Tumour necrosis factor-alpha gene promoter polymorphism in chronic obstructive pulmonary disease. Eur Respir J. (2000) 15(2):281-4.
225. Ferrarotti I, Zorzetto M, Beccaria M, Gile LS, Porta R, Ambrosino N, et al. Tumour necrosis factor family genes in a phenotype of COPD associated with emphysema. Eur Respir J. (2003) 21(3):444-9.
226. Wilson AG, de Vries N, Pociot F, di Giovine FS, van der Putte LB, Duff GW. An allelic polymorphism within the human tumor necrosis factor alpha promoter region is strongly associated with HLA A1, B8, and DR3 alleles. J Exp Med. (1993) 177(2):557-60.
227. The International HapMap Project. Nature. (2003) 426(6968):789-96.

228. Gingo MR, Silveira LJ, Miller YE, Friedlander AL, Cosgrove GP, Chan ED, et al. Tumour necrosis factor gene polymorphisms are associated with COPD. Eur Respir J. (2008) 31(5):1005-12.
229. Howard TD, Koppelman GH, Xu J, Zheng SL, Postma DS, Meyers DA, et al. Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma. Am J Hum Genet. (2002) 70(1):230-6.
230. Keatings VM, Collins PD, Scott DM, Barnes PJ. Differences in interleukin-8 and tumor necrosis factor-alpha in induced sputum from patients with chronic obstructive pulmonary disease or asthma. Am J Respir Crit Care Med. (1996) 153(2):530-4.
231. Mueller R, Chanez P, Campbell AM, Bousquet J, Heusser C, Bullock GR. Different cytokine patterns in bronchial biopsies in asthma and chronic bronchitis. Respir Med. (1996) 90(2):79-85.
232. Di Francia M, Barbier D, Mege JL, Orehek J. Tumor necrosis factor-alpha levels and weight loss in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (1994) 150(5 Pt 1):1453-5.
233. Wright JL, Farmer SG, Churg A. Synthetic serine elastase inhibitor reduces cigarette smoke-induced emphysema in guinea pigs. Am J Respir Crit Care Med. (2002) 166(7):954-60.
234. Hill AT, Campbell EJ, Bayley DL, Hill SL, Stockley RA. Evidence for excessive bronchial inflammation during an acute exacerbation of chronic obstructive pulmonary disease in patients with alpha(1)-antitrypsin deficiency (PiZ). Am J Respir Crit Care Med. (1999) 160(6):1968-75.
235. Sethi S, Maloney J, Grove L, Wrona C, Berenson CS. Airway inflammation and bronchial bacterial colonization in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2006) 173(9):991-8.
236. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, et al. IL10 polymorphisms are associated with airflow obstruction in severe alpha1-antitrypsin deficiency. Am J Respir Cell Mol Biol. (2008) 38(1):114-20.
237. Dowson LJ, Guest PJ, Stockley RA. The relationship of chronic sputum expectoration to physiologic, radiologic, and health status characteristics in alpha(1)-antitrypsin deficiency (PiZ). Chest. (2002) 122(4):1247-55.
238. Dawkins PA, Dawkins CL, Wood AM, Nightingale PG, Stockley JA, Stockley RA. Rate of progression of lung function impairment in alpha1-antitrypsin deficiency. Eur Respir J. (2009) 33(6):1338-44.
239. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. (2002) 57(10):847-52.
240. Mucolytic agents for chronic bronchitis or chronic obstructive pulmonary disease [database on the Internet]. Cochrane database of systematic reviews. 2007 [cited 04.08.2009]. Available from: http://www.cochrane.org/reviews/en/ab001287.html.
241. Sevenoaks MJ, Stockley RA. Chronic Obstructive Pulmonary Disease, inflammation and co-morbidity--a common inflammatory phenotype? Respir Res. (2006) 7:70.
242. Nightingale JA, Rogers DF, Barnes PJ. Effect of repeated sputum induction on cell counts in normal volunteers. Thorax. (1998) 53(2):87-90.
243. Bayley JP, de Rooij H, van den Elsen PJ, Huizinga TW, Verweij CL. Functional

analysis of linker-scan mutants spanning the -376, -308, -244, and -238 polymorphic sites of the TNF-alpha promoter. Cytokine. (2001) 14(6):316-23.
244. Rahman P, Siannis F, Butt C, Farewell V, Peddle L, Pellett F, et al. TNFalpha polymorphisms and risk of psoriatic arthritis. Annals of the rheumatic diseases. (2006) 65(7):919-23.
245. Kaluza W, Reuss E, Grossmann S, Hug R, Schopf RE, Galle PR, et al. Different transcriptional activity and in vitro TNF-alpha production in psoriasis patients carrying the TNF-alpha 238A promoter polymorphism. The Journal of investigative dermatology. (2000) 114(6):1180-3.
246. Parmar JS, Mahadeva R, Reed BJ, Farahi N, Cadwallader KA, Keogan MT, et al. Polymers of alpha(1)-antitrypsin are chemotactic for human neutrophils: a new paradigm for the pathogenesis of emphysema. Am J Respir Cell Mol Biol. (2002) 26(6):723-30.
247. Pilot-Matias TJ, Kister SE, Fox JL, Kropp K, Glasser SW, Whitsett JA. Structure and organization of the gene encoding human pulmonary surfactant proteolipid SP-B. DNA (Mary Ann Liebert, Inc. (1989) 8(2):75-86.
248. O'Reilly MA, Weaver TE, Pilot-Matias TJ, Sarin VK, Gazdar AF, Whitsett JA. In vitro translation, post-translational processing and secretion of pulmonary surfactant protein B precursors. Biochim Biophys Acta. (1989) 1011(2-3):140-8.
249. Rova M, Haataja R, Marttila R, Ollikainen V, Tammela O, Hallman M. Data mining and multiparameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia. Hum Mol Genet. (2004) 13(11):1095-104.
250. Miles PR, Bowman L, Rao KM, Baatz JE, Huffman L. Pulmonary surfactant inhibits LPS-induced nitric oxide production by alveolar macrophages. Am J Physiol. (1999) 276(1 Pt 1):L186-96.
251. Ikegami M, Whitsett JA, Martis PC, Weaver TE. Reversibility of lung inflammation caused by SP-B deficiency. American journal of physiology. (2005) 289(6):L962-70.
252. Han X, Ruan Y, Zhang X, Wang H, Chu Y. [Changes of surfactant A and B in alveolar type II cells in hamster with elastase-induced emphysema]. Zhonghua bing li xue za zhi Chinese journal of pathology. (2000) 29(1):43-5.
253. Ingenito EP, Tsai LW, Majumdar A, Suki B. On the role of surface tension in the pathophysiology of emphysema. Am J Respir Crit Care Med. (2005) 171(4):300-4.
254. Gorrini M, Lupi A, Iadarola P, Dos Santos C, Rognoni P, Dalzoppo D, et al. SP-A binds alpha1-antitrypsin in vitro and reduces the association rate constant for neutrophil elastase. Respir Res. (2005) 6:146.
255. Derynck R, Rhee L, Chen EY, Van Tilburg A. Intron-exon structure of the human transforming growth factor-beta precursor gene. Nucleic acids research. (1987) 15(7):3188-9.
256. Derynck R, Jarrett JA, Chen EY, Eaton DH, Bell JR, Assoian RK, et al. Human transforming growth factor-beta complementary DNA sequence and expression in normal and transformed cells. Nature. (1985) 316(6030):701-5.
257. Morris DG, Huang X, Kaminski N, Wang Y, Shapiro SD, Dolganov G, et al. Loss of integrin alpha(v)beta6-mediated TGF-beta activation causes Mmp12-dependent emphysema. Nature. (2003) 422(6928):169-73.
258. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med. (2000) 342(18):1350-8.

259. Mucida D, Park Y, Kim G, Turovskaya O, Scott I, Kronenberg M, et al. Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. Science. (2007) 317(5835):256-60.
260. Wedzicha JA, Seemungal TA, MacCallum PK, Paul EA, Donaldson GC, Bhowmik A, et al. Acute exacerbations of chronic obstructive pulmonary disease are accompanied by elevations of plasma fibrinogen and serum IL-6 levels. Thrombosis and haemostasis. (2000) 84(2):210-5.
261. Roth MD, Connett JE, D'Armiento JM, Foronjy RF, Friedman PJ, Goldin JG, et al. Feasibility of retinoids for the treatment of emphysema study. Chest. (2006) 130(5):1334-45.
262. Grainger DJ, Heathcote K, Chiano M, Snieder H, Kemp PR, Metcalfe JC, et al. Genetic control of the circulating concentration of transforming growth factor type beta1. Hum Mol Genet. (1999) 8(1):93-7.
263. Suthanthiran M, Li B, Song JO, Ding R, Sharma VK, Schwartz JE, et al. Transforming growth factor-beta 1 hyperexpression in African-American hypertensives: A novel mediator of hypertension and/or target organ damage. Proc Natl Acad Sci U S A. (2000) 97(7):3479-84.
264. Takizawa H, Tanaka M, Takami K, Ohtoshi T, Ito K, Satoh M, et al. Increased expression of transforming growth factor-beta1 in small airway epithelium from tobacco smokers and patients with chronic obstructive pulmonary disease (COPD). Am J Respir Crit Care Med. (2001) 163(6):1476-83.
265. de Boer WI, van Schadewijk A, Sont JK, Sharma HS, Stolk J, Hiemstra PS, et al. Transforming growth factor beta1 and recruitment of macrophages and mast cells in airways in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (1998) 158(6):1951-7.
266. Witke WF, Gibbs PE, Zielinski R, Yang F, Bowman BH, Dugaiczyk A. Complete structure of the human Gc gene: differences and similarities between members of the albumin gene family. Genomics. (1993) 16(3):751-4.
267. Yamamoto N, Homma S. Vitamin D3 binding protein (group-specific component) is a precursor for the macrophage-activating signal factor from lysophosphatidylcholine-treated lymphocytes. Proc Natl Acad Sci U S A. (1991) 88(19):8539-43.
268. Kew RR, Webster RO. Gc-globulin (vitamin D-binding protein) enhances the neutrophil chemotactic activity of C5a and C5a des Arg. J Clin Invest. (1988) 82(1):364-9.
269. DiMartino SJ, Shah AB, Trujillo G, Kew RR. Elastase controls the binding of the vitamin D-binding protein (Gc-globulin) to neutrophils: a potential role in the regulation of C5a co-chemotactic activity. J Immunol. (2001) 166(4):2688-94.
270. Viau M, Constans J, Debray H, Montreuil J. Isolation and characterization of the O-glycan chain of the human vitamin-D binding protein. Biochem Biophys Res Commun. (1983) 117(1):324-31.
271. Poletti V, Casoni G, Chilosi M, Zompatori M. Diffuse panbronchiolitis. Eur Respir J. (2006) 28(4):862-71.
272. Jobe AH. Pulmonary surfactant therapy. N Engl J Med. (1993) 328(12):861-8.
273. Su ZG, Wen FQ, Feng YL, Xiao M, Wu XL. Transforming growth factor-beta1 gene polymorphisms associated with chronic obstructive pulmonary disease in Chinese population. Acta pharmacologica Sinica. (2005) 26(6):714-20.
274. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, Crystal RG.

Modification of gene expression of the small airway epithelium in response to cigarette smoking. Journal of molecular medicine (Berlin, Germany). (2007) 85(1):39-53.
275. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol. (2004) 31(6):601-10.
276. Patel IS, Seemungal TA, Wilks M, Lloyd-Owen SJ, Donaldson GC, Wedzicha JA. Relationship between bacterial colonisation and the frequency, character, and severity of COPD exacerbations. Thorax. (2002) 57(9):759-64.
277. Newby PR, Roberts-Davies EL, Brand OJ, Heward JM, Franklyn JA, Gough SC, et al. Tag SNP screening of the PDCD1 gene for association with Graves' disease. Clin Endocrinol (Oxf). (2007) 67(1):125-8.
278. Brand OJ, Barrett JC, Simmonds MJ, Newby PR, McCabe CJ, Bruce CK, et al. Association of the thyroid stimulating hormone receptor gene (TSHR) with Graves' disease. Hum Mol Genet. (2009) 18(9):1704-13.
279. Finlay GA, Russell KJ, McMahon KJ, D'Arcy E M, Masterson JB, FitzGerald MX, et al. Elevated levels of matrix metalloproteinases in bronchoalveolar lavage fluid of emphysematous patients. Thorax. (1997) 52(6):502-6.
280. Ohnishi K, Takagi M, Kurokawa Y, Satomi S, Konttinen YT. Matrix metalloproteinase-mediated extracellular matrix protein degradation in human pulmonary emphysema. Lab Invest. (1998) 78(9):1077-87.
281. Howard EW, Bullen EC, Banda MJ. Preferential inhibition of 72- and 92-kDa gelatinases by tissue inhibitor of metalloproteinases-2. J Biol Chem. (1991) 266(20):13070-5.
282. Pendas AM SI, Alvarez MV, Pritchard M, Lopez-Otin C. Fine physical mapping of the human matrix metalloproteinase genes clustered 11q22.3. Genomics. (1996) 37:266-82.
283. Brinckerhoff CE, Ruby PL, Austin SD, Fini ME, White HD. Molecular cloning of human synovial cell collagenase and selection of a single gene from genomic DNA. J Clin Invest. (1987) 79(2):542-6.
284. Saus J, Quinones S, Otani Y, Nagase H, Harris ED, Jr., Kurkinen M. The complete primary structure of human matrix metalloproteinase-3. Identity with stromelysin. J Biol Chem. (1988) 263(14):6742-5.
285. Sellers A, Murphy G. Collagenolytic enzymes and their naturally occurring inhibitors. International review of connective tissue research. (1981) 9:151-90.
286. Shapiro SD, Kobayashi DK, Ley TJ. Cloning and characterization of a unique elastolytic metalloproteinase produced by human alveolar macrophages. J Biol Chem. (1993) 268(32):23824-9.
287. Rutter JL, Mitchell TI, Buttice G, Meyers J, Gusella JF, Ozelius LJ, et al. A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. Cancer Res. (1998) 58(23):5321-5.
288. D'Armiento J DS, Okada Y, Berg RA, Chada K. Collagenase expression in the lungs of transgenic mice causes pulmonary emphysema. Cell. (1992) 71:955-61.
289. Hautamaki RD KD, Senior RM, Shapiro SD. Requirment for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science.* (1997) 277:2002-4.
290. Chen XH, Rodriguez S, Hawe E, Talmud PJ, Miller GJ, Underhill P, et al. Evidence of admixture from haplotyping in an epidemiological study of UK Caucasian males: implications for association analyses. Hum Hered. (2004) 57(3):142-55.

291. Nerusu KC, Warner RL, Bhagavathula N, McClintock SD, Johnson KJ, Varani J. Matrix metalloproteinase-3 (stromelysin-1) in acute inflammatory tissue injury. Experimental and molecular pathology. (2007) 83(2):169-76.
292. Tsukahara S, Shinozaki M, Ikari K, Mochizuki T, Inoue E, Tomatsu T, et al. Effect of matrix metalloproteinase-3 functional SNP on serum matrix metalloproteinase-3 level and outcome measures in Japanese RA patients. Rheumatology (Oxford, England). (2008) 47(1):41-4.
293. Betsuyaku T, Nishimura M, Takeyabu K, Tanino M, Venge P, Xu S, et al. Neutrophil granule proteins in bronchoalveolar lavage fluid from subjects with subclinical emphysema. Am J Respir Crit Care Med. (1999) 159(6):1985-91.
294. Culpitt SV, Rogers DF, Traves SL, Barnes PJ, Donnelly LE. Sputum matrix metalloproteases: comparison between chronic obstructive pulmonary disease and asthma. Respir Med. (2005) 99(6):703-10.
295. Belvisi MG, Bottomley KM. The role of matrix metalloproteinases (MMPs) in the pathophysiology of chronic obstructive pulmonary disease (COPD): a therapeutic role for inhibitors of MMPs? Inflamm Res. (2003) 52(3):95-100.
296. Agusti A, MacNee W, Donaldson K, Cosio M. Hypothesis: does COPD have an autoimmune component? Thorax. (2003) 58(10):832-4.
297. Rose NR, Bona C. Defining criteria for autoimmune diseases (Witebsky's postulates revisited). Immunology today. (1993) 14(9):426-30.
298. Feghali-Bostwick CA, Gadgil AS, Otterbein LE, Pilewski JM, Stoner MW, Csizmadia E, et al. Autoantibodies in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2008) 177(2):156-63.
299. Taraseviciene-Stewart L, Kraskausciene V, Burns N, Voelkel NF. Anti-endothelial cell antibodies in COPD. Eur Respir J. (2007) 30 (ERS Supplement):55s.
300. Goswami S, Barranco W, Lee SH, Grudo A, Corry DB, Kheradmand F. An autoimmune basis in the disease pathogenesis of COPD/Emphysema. Journal of Immunology. (2007) 178:130.41.
301. Grumelli S, Corry DB, Song LZ, Song L, Green L, Huh J, et al. An immune basis for lung parenchymal destruction in chronic obstructive pulmonary disease and emphysema. PLoS medicine. (2004) 1(1):e8.
302. Sullivan AK, Simonian PL, Falta MT, Mitchell JD, Cosgrove GP, Brown KK, et al. Oligoclonal CD4+ T cells in the lungs of patients with severe emphysema. Am J Respir Crit Care Med. (2005) 172(5):590-6.
303. Motz GT, Eppert BL, Sun G, Wesselkamper SC, Linke MJ, Deka R, et al. Persistence of lung CD8 T cell oligoclonal expansions upon smoking cessation in a mouse model of cigarette smoke-induced emphysema. J Immunol. (2008) 181(11):8036-43.
304. Birring SS, Brightling CE, Bradding P, Entwisle JJ, Vara DD, Grigg J, et al. Clinical, radiologic, and induced sputum features of chronic obstructive pulmonary disease in nonsmokers: a descriptive study. Am J Respir Crit Care Med. (2002) 166(8):1078-83.
305. Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, et al. The DNA sequence and analysis of human chromosome 6. Nature. (2003) 425(6960):805-11.
306. Simmonds MJ, Gough SC. The HLA region and autoimmune disease: associations and mechanisms of action. Current genomics. (2008) In press.
307. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS

genetics. (2008) 4(4):e1000024.
308. Simmonds MJ, Gough SC. The HLA region and autoimmune disease: associations and mechanisms of action. Current genomics. (2007) 8(7):453-65.
309. Simmonds MJ, Howson JM, Heward JM, Cordell HJ, Foxall H, Carr-Smith J, et al. Regression mapping of association between the human leukocyte antigen region and Graves disease. Am J Hum Genet. (2005) 76(1):157-63.
310. Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. (2007) 450(7171):887-92.
311. Anagnostopoulou U, Toumbis M, Konstantopoulos K, Kotsovoulou-Fouskaki V, Zervas J. HLA-A and -B antigens in chronic bronchitis. Journal of clinical epidemiology. (1993) 46(12):1413-6.
312. Noble JA, Valdes AM, Lane JA, Green AE, Erlich HA. Linkage disequilibrium with predisposing DR3 haplotypes accounts for apparent effects of tumor necrosis factor and lymphotoxin-alpha polymorphisms on type 1 diabetes susceptibility. Human immunology. (2006) 67(12):999-1004.
313. IMGT/HLA database release 2.20 [database on the Internet]2008. Available from: http://www.ebi.ac.uk/imgt/hla/.
314. Nerup J, Platz P, Andersen OO, Christy M, Lyngsoe J, Poulsen JE, et al. HL-A antigens and diabetes mellitus. Lancet. (1974) 2(7885):864-6.
315. Singal DP, Blajchman MA. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes. (1973) 22(6):429-32.
316. Milicic A, Lee D, Brown MA, Darke C, Wordsworth BP. HLA-DR/DQ haplotype in rheumatoid arthritis: novel allelic associations in UK Caucasians. The Journal of rheumatology. (2002) 29(9):1821-6.
317. de Vries N, Tijssen H, van Riel PL, van de Putte LB. Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67-74 of the HLA-DRB1 molecule. Arthritis and rheumatism. (2002) 46(4):921-8.
318. Tunnicliffe WS, Harrison RM, Kelly FJ, Dunster C, Ayres JG. The effect of sulphurous air pollutant exposures on symptoms, lung function, exhaled nitric oxide, and nasal epithelial lining fluid antioxidant concentrations in normal and asthmatic adults. Occupational and environmental medicine. (2003) 60(11):e15.
319. Dockery DW, Pope CA, 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med. (1993) 329(24):1753-9.
320. Pope CA, 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama. (2002) 287(9):1132-41.
321. Samet JM, Dominici F, Curriero FC, Coursac I, Zeger SL. Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994. N Engl J Med. (2000) 343(24):1742-9.
322. Schwartz J. Short term fluctuations in air pollution and hospital admissions of the elderly for respiratory disease. Thorax. (1995) 50(5):531-8.
323. Atkinson RW, Anderson HR, Sunyer J, Ayres J, Baccini M, Vonk JM, et al. Acute effects of particulate air pollution on respiratory admissions: results from APHEA 2

project. Air Pollution and Health: a European Approach. Am J Respir Crit Care Med. (2001) 164(10 Pt 1):1860-6.
324. Schouten JP, Vonk JM, de Graaf A. Short term effects of air pollution on emergency hospital admissions for respiratory disease: results of the APHEA project in two major cities in The Netherlands, 1977-89. Journal of epidemiology and community health. (1996) 50 Suppl 1:s22-9.
325. Bell ML, Dominici F, Samet JM. A meta-analysis of time-series studies of ozone and mortality with comparison to the national morbidity, mortality, and air pollution study. Epidemiology (Cambridge, Mass. (2005) 16(4):436-45.
326. Sunyer J, Spix C, Quenel P, Ponce-de-Leon A, Ponka A, Barumandzadeh T, et al. Urban air pollution and emergency admissions for asthma in four European cities: the APHEA Project. Thorax. (1997) 52(9):760-5.
327. American Thoracic Society. What constitutes an adverse health effect of air pollution? Official statement of the American Thoracic Society. Am J Respir Crit Care Med. (2000) 161(2 Pt 1):665-73.
328. Stockley RA, Luisetti M. Alpha-1-antirypsin deficiency: the biochemical basis of pulmonary emphysema. Pavia: IRCCS; 1999.
329. Crapo JD, Glassroth JL, Karlinsky JB, King TE, editors. Baum's textbook of pulmonary diseases. 7th edition ed: Lippincott, Williams and Wilkins; 2003.
330. Piitulainen E, Tornling G, Eriksson S. Environmental correlates of impaired lung function in non-smokers with severe alpha 1-antitrypsin deficiency (PiZZ). Thorax. (1998) 53(11):939-43.
331. Senn O, Russi EW, Imboden M, Probst-Hensch NM. alpha1-Antitrypsin deficiency and lung disease: risk modification by occupational and environmental inhalants. Eur Respir J. (2005) 26(5):909-17.
332. Piitulainen E, Tornling G, Eriksson S. Effect of age and occupational exposure to airway irritants on lung function in non-smoking individuals with alpha 1-antitrypsin deficiency (PiZZ). Thorax. (1997) 52(3):244-8.
333. Elliott P, Shaddick G, Wakefield JC, de Hoogh C, Briggs DJ. Long-term associations of outdoor air pollution with mortality in Great Britain. Thorax. (2007) 62(12):1088-94.
334. ONS. National Statistics Postcode Directory. 2006 [cited 2007]; Available from: http://www.statistics.gov.uk/geography/nspd.asp.
335. Semple SE, Dick F, Cherrie JW. Exposure assessment for a population-based case-control study combining a job-exposure matrix with interview data. Scandinavian journal of work, environment & health. (2004) 30(3):241-8.
336. CASCOT: computer assisted structured coding tool [database on the Internet]. Warwick Institute for Employment Research. [cited 15.1.2008]. Available from: http://www2.warwick.ac.uk/fac/soc/ier/publications/software/cascot/.
337. List of Approved Workplace Exposure Limits. Health and Safety Executive; 2007; Available from: http://www.hse.gov.uk/coshh/table1.pdf
338. Kent AJ, Grice S, Stedman JR, Bush TJ, Vincent KJ, Abbott J, et al. UK Air Quality modelling for annual reporting 2005 on ambient air quality assessment under Council Directives 96/62/EC, 1999/30/EC and 2000/69/EC AEA Energy and Environment to the Department for Environment, Food and Rural Affairs, (2005). ED48208
339. Bower J, Lampert J, Broughton G, Stedman JR, Kent AJ, Targa J, et al. Air

pollution in the UK: 2005. DEFRA, (2005)
340. Stedman JR, Kent AJ. An analysis of the spatial patterns of human health related ozone metrics across the UK in 1995, 2003 and 2005. Atmospheric Environment. (2008) 42:1702-16.
341. Kent AJ, Grice S, Stedman JR, Bush TJ, Vincent KJ, Abbott J, et al. UK Air Quality modelling for annual reporting 2005 on ambient air quality assessment under Council Directives 96/62/EC, 1999/30/EC and 2000/69/EC. AEA Energy and Environment to the Department for Environment, Food and Rural Affairs, (2005). ED48208
342. Kromhout H, Symanski E, Rappaport SM. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. The Annals of occupational hygiene. (1993) 37(3):253-70.
343. Schlesinger RB, Kunzli N, Hidy GM, Gotschi T, Jerrett M. The health relevance of ambient particulate matter characteristics: coherence of toxicological and epidemiological inferences. Inhalation toxicology. (2006) 18(2):95-125.
344. Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. Lancet. (2002) 360(9341):1203-9.
345. Ayres JG. Long-term exposures to air pollution. Thorax. (2007) 62(12):1037-8.
346. Sheppard D, Wong WS, Uehara CF, Nadel JA, Boushey HA. Lower threshold and greater bronchomotor responsiveness of asthmatic subjects to sulfur dioxide. Am Rev Respir Dis. (1980) 122(6):873-8.
347. Katsouyanni K, Touloumi G, Spix C, Schwartz J, Balducci F, Medina S, et al. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. Air Pollution and Health: a European Approach. BMJ (Clinical research ed. (1997) 314(7095):1658-63.
348. Wallace L. Indoor particles: a review. Journal of the Air & Waste Management Association (1995). (1996) 46(2):98-126.
349. Hollingsworth JW, Kleeberger SR, Foster WM. Ozone and pulmonary innate immunity. Proc Am Thorac Soc. (2007) 4(3):240-6.
350. Nadziejko C, Finkelstein I, Balmes JR. Contribution of secretory leukocyte proteinase inhibitor to the antiprotease defense system of the peripheral lung: effect of ozone-induced acute inflammation. Am J Respir Crit Care Med. (1995) 152(5 Pt 1):1592-8.
351. Health effects of climate change in the UK 2008. In: DoH/HPA, editor.2008.
352. Ground level ozone in the 21st century: future trends, impacts and policy implications. The Royal Society, (2008). RS1276
353. Enright PL, Connett JE, Kanner RE, Johnson LR, Lee WW. Spirometry in the Lung Health Study: II. Determinants of short-term intraindividual variability. Am J Respir Crit Care Med. (1995) 151(2 Pt 1):406-11.
354. Dirksen A, Piitulainen E, Parr DG, Deng C, Wencker M, Shaker SB, et al. Exploring the role of CT densitometry: a randomised study of augmentation therapy in AATD. European Respiratory Journal. (2009) In press:online only doi: 10.1183/09031936.0159408.
355. Downs SH, Schindler C, Liu LJ, Keidel D, Bayer-Oglesby L, Brutsche MH, et al. Reduced exposure to PM10 and attenuated age-related decline in lung function. N Engl J Med. (2007) 357(23):2338-47.
356. Wood AM, Harrison RM, Semple S, Ayres JG, Stockley RA. Outdoor air pollution

is associated with disease severity in alpha-1-antitrypsin deficiency. Eur Respir J. (2009).
357. Pope CA, 3rd, Kanner RE. Acute effects of PM10 pollution on pulmonary function of smokers with mild to moderate chronic obstructive pulmonary disease. Am Rev Respir Dis. (1993) 147(6 Pt 1):1336-40.
358. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. N Engl J Med. (2003) 348(21):2059-73.
359. Marsh SE, Travers J, Weatherall M, Williams MV, Aldington S, Shirtcliffe PM, et al. Proportional classifications of COPD phenotypes. Thorax. (2008) 63(9):761-7.
360. Coxson HO. Quantitative computed tomography assessment of airway wall dimensions: current status and potential applications for phenotyping chronic obstructive pulmonary disease. Proc Am Thorac Soc. (2008) 5(9):940-5.
361. Kim WJ, Silverman EK, Hoffman E, Criner GJ, Mosenifar Z, Sciurba FC, et al. CT Metrics of Airway Disease and Emphysema in Severe COPD. Chest. (2009) Online ahead of print; doi 10.1378/chest.08-2858.
362. Grydeland TB, Dirksen A, Coxson HO, Pillai SG, Sharma S, Eide GE, et al. Quantitative CT: Emphysema and airway wall thickness by gender, age and smoking. Eur Respir J. (2009) Online ahead of print: doi 10.1183/09031936.00167908.
363. Gursel G. Does coexistence with bronchiectasis influence intensive care unit outcome in patients with chronic obstructive pulmonary disease? Heart Lung. (2006) 35(1):58-65.
364. Kurt E, Ozkan R, Orman A, Calisir C, Metintas M. Irreversiblity of remodeled features on high-resolution computerized tomography scans of asthmatic patients on conventional therapy: a 6-year longitudinal study. J Asthma. (2009) 46(3):300-7.
365. Ohara T, Hirai T, Sato S, Terada K, Kinose D, Haruna A, et al. Longitudinal study of airway dimensions in chronic obstructive pulmonary disease using computed tomography. Respirology. (2008) 13(3):372-8.
366. de Jong PA, Nakano Y, Hop WC, Long FR, Coxson HO, Pare PD, et al. Changes in airway dimensions on computed tomography scans of children with cystic fibrosis. Am J Respir Crit Care Med. (2005) 172(2):218-24.
367. Seersholm N, Kok-Jensen A, Dirksen A. Decline in FEV1 among patients with severe hereditary alpha 1-antitrypsin deficiency type PiZ. Am J Respir Crit Care Med. (1995) 152(6 Pt 1):1922-5.
368. Celli BR, Thomas NE, Anderson JA, Ferguson GT, Jenkins C, Jones PW, et al. Effect of Pharmacotherapy on Rate of Decline of Lung Function in COPD: Results from the TORCH Study. Am J Respir Crit Care Med. (2008).
369. Han MK, Postma D, Mannino DM, Giardino ND, Buist S, Curtis JL, et al. Gender and chronic obstructive pulmonary disease: why it matters. Am J Respir Crit Care Med. (2007) 176(12):1179-84.
370. Lahm T, Crisostomo PR, Markel TA, Wang M, Weil BR, Novotny NM, et al. The effects of estrogen on pulmonary artery vasoreactivity and hypoxic pulmonary vasoconstriction: potential new clinical implications for an old hormone. Crit Care Med. (2008) 36(7):2174-83.
371. Gao N, Rahmani M, Shi X, Dent P, Grant S. Synergistic antileukemic interactions between 2-medroxyestradiol (2-ME) and histone deacetylase inhibitors involve Akt down-regulation and oxidative stress. Blood. (2006) 107(1):241-9.

372. Miravitlles M, Marin A, Monso E, Vila S, de la Roza C, Hervas R, et al. Efficacy of moxifloxacin in the treatment of bronchial colonisation in COPD. Eur Respir J. (2009) Online ahead of print; doi 10.1183/09031936.00195608.
373. Marin A, Monso E, Garcia M, Sauleda J, Noguera A, Pons J, et al. Variability and effects of bronchial colonisation in patients with moderate COPD. Eur Respir J. (2009) Online ahead of print; doi 10.1183/09031936.00126808.
374. Sethi S, Sethi R, Eschberger K, Lobbins P, Cai X, Grant BJ, et al. Airway bacterial concentrations and exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2007) 176(4):356-61.
375. Martinez-Solano L, Macia MD, Fajardo A, Oliver A, Martinez JL. Chronic Pseudomonas aeruginosa infection in chronic obstructive pulmonary disease. Clin Infect Dis. (2008) 47(12):1526-33.
376. 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. Am J Respir Crit Care Med. (2004) 170(12):1294-301.
377. 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. Hum Mol Genet. (2003) 12(10):1199-210.
378. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet. (2004) 74(4):765-9.
379. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). Eur Respir J. (2008) 31(4):869-73.
380. Crapo J, Silverman EK. [06.08.2009]; Available from: www.copdgene.org.
381. 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 Genet. (2009) 5(3):e1000429.
382. Wessler I, Kirkpatrick CJ, Racke K. Non-neuronal acetylcholine, a locally acting molecule, widely distributed in biological systems: expression and function in humans. Pharmacol Ther. (1998) 77(1):59-79.
383. Kawashima K, Fujii T. The lymphocytic cholinergic system and its contribution to the regulation of immune activity. Life Sci. (2003) 74(6):675-96.
384. Bak M, Hansen C, Friis Henriksen K, Tommerup N. The human hedgehog-interacting protein gene: structure and chromosome mapping to 4q31.21-->q31.3. Cytogenet Cell Genet. (2001) 92(3-4):300-3.
385. Gao B, Hu J, Stricker S, Cheung M, Ma G, Law KF, et al. A mutation in Ihh that causes digit abnormalities alters its signalling capacity and range. Nature. (2009) 458(7242):1196-200.
386. Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on Suspected COPD Genes - A Summary of 20 Years' Research. Am J Respir Crit Care Med. (2009) Online ahead of print; doi 10.1164/rccm.200905-0722OC.
387. Hu G, Shi Z, Hu J, Zou G, Peng G, Ran P. Association between polymorphisms of microsomal epoxide hydrolase and COPD: results from meta-analyses. Respirology. (2008) 13(6):837-50.

388. Zhu G, Agusti A, Gulsvik A, Bakke P, Coxson H, Lomas DA, et al. CTLA4 gene polymorphisms are associated with chronic bronchitis. Eur Respir J. (2009) Online ahead of print; doi 10.1183/09031936.00141808.
389. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. Nature. (2003) 423(6939):506-11.
390. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). International journal of epidemiology. (2006) 35(1):34-41.
391. Pillai SG, Shianna K, Ge D, Kong X, Zhu G, Gulsvik A, et al. Genome-Wide Association Study of COPD in a case control population from Norway. AJRCCM. (2008) 177:A776.
392. Sin DD, Leung R, Gan WQ, Man SP. Circulating surfactant protein D as a potential lung-specific biomarker of health outcomes in COPD: a pilot study. BMC pulmonary medicine. (2007) 7:13.
393. Sin DD, Man SF, Marciniuk DD, Ford G, FitzGerald M, Wong E, et al. The effects of fluticasone with or without salmeterol on systemic biomarkers of inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. (2008) 177(11):1207-14.
394. Sapey E, Ahmad A, Bayley DL, Insall RH, Stockley RA. Chemotactic direction and speed of neutrophils from patients with COPD. American Journal of Respiratory and Critical Care Medicine. (2008) 177:A605.
395. Litonjua AA, Hollis BW, Schuemann BS, Celedon JC, Fuhlbrigge JL, Raby BA, et al. Low serum vitamin D levels are assoicated with greater risks for severe exacerbations in childhood asthmatics. American Journal of Respiratory and Critical Care Medicine. (2008) 177:A993.
396. Wittke A, Weaver V, Mahon BD, August A, Cantorna MT. Vitamin D receptor-deficient mice fail to develop experimental allergic asthma. J Immunol. (2004) 173(5):3432-6.
397. Harel S JA, Yu SY, Hurewitz A, Bergofsky EH. Desmosine radioimmunassay for measurng elastin degradation *in vitro*. Rev Respir Dis. (1980) 122:769-73.
398. Ma S, Lin YY, Turino GM. Measurements of desmosine and isodesmosine by mass spectrometry in COPD. Chest. (2007) 131(5):1363-71.
399. Gottlieb DJ SP, Sparrow D, Gale ME, Weiss ST, Snider GL, O'Conner GT. Urinary desmosine excretion in smokers with and without rapid decline of lung function: the normal aging study. Am J Resp Crit Care Med. (1996) 154:1290-5.
400. Viglio S, Iadarola P, Lupi A, Trisolini R, Tinelli C, Balbi B, et al. MEKC of desmosine and isodesmosine in urine of chronic destructive lung disease patients. Eur Respir J. (2000) 15(6):1039-45.
401. Gottlieb DJ, Stone PJ, Sparrow D, Gale ME, Weiss ST, Snider GL, et al. Urinary desmosine excretion in smokers with and without rapid decline of lung function: the Normative Aging Study. Am J Respir Crit Care Med. (1996) 154(5):1290-5.
402. Motsinger AA, Lee SL, Mellick G, Ritchie MD. GPNN: power studies and applications of a neural network method for detecting gene-gene interactions in studies of human disease. BMC bioinformatics. (2006) 7:39.
403. Motsinger AA, Ritchie MD, Reif DM. Novel methods for detecting epistasis in pharmacogenomics studies. Pharmacogenomics. (2007) 8(9):1229-41.