UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND BIOLOGICAL SCIENCES
Divisions of Human Genetics and Infection, Inflammation and Repair

Genetic and Molecular Characterisation of A Disintegrin and Metalloprotease (ADAM) 33

by

Julie Ann Cakebread

This thesis is submitted to the University of Southampton
for the degree of Doctor of Philosophy

March 2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND BIOLOGICAL SCIENCES

DIVISIONS OF HUMAN GENETICS AND INFECTION, INFLAMMATION AND REPAIR

Doctor of Philosophy

GENETIC AND MOLECULAR CHARACTERISATION OF A DISINTEGRIN AND METALLOPROTEASE (ADAM) 33

by Julie Ann Cakebread

Asthma is a common respiratory disease that is increasingly prevalent in the UK. Polymorphism of the gene, *ADAM33*, is associated with increased susceptibility to asthma and bronchial hyperresponsiveness (BHR). *ADAM33* is expressed in the fetal lung so may regulate the development of the airways. Thus, polymorphism in the *ADAM33* gene may underlie the early differences in lung function that have been observed in children who later go on to develop asthma. This work investigates the association of *ADAM33* polymorphism with early-life lung function and characterises basal transcriptional activation of *ADAM33.*

5' nuclease allelic discrimination assays were developed and used to genotype a one month old cohort with infant lung function data (COPSAC, n=401). Haplotype trend regression, haploscore analyses and linkage disequilibrium mapping were used to assess association with lung function. Transcriptional regulation of *ADAM33* was investigated using deletion analysis of ~1500 bp of putative basal promoter sequence. Sequences were cloned into pGl3 basic vectors and Dual–Luciferase® reporter assays used to assess relative activity. Cross species and intra-species analysis was performed, using Vista, to identify conserved regions of the *ADAM33* promoter. Putative transcription binding sites were predicted using MatInspector.

ADAM33 polymorphism was not associated with lung function at four weeks of age. At aged three years, univariate analysis showed individuals homozygote for the major allele of SNPs F1 (p=0.042), Qm1 (p=0.011), STp4 (p=0.006), STp7 (p=0.038) and the minor homozygote/heterozygote group of T1(p=0.037) were associated with statistically significant increased baseline sRaw. Haplotype analysis confirmed an interaction with pre bronchodilator sRaw and ADAM33 polymorphism. Dual–Luciferase® reporter assays suggested the core promoter region lies within 359 bp (F3X) of the translation start site (ATG) and there is an inhibitory site between -637 bp and -359 bp. MatInspector predicted Sp1 sites within the F3X sequence. Over expression with Sp1 reduced *ADAM33* expression in fibroblast cell lines, and increased expression in epithelial cell lines indicating an activation mechanism involving Sp1.

This work implicates Sp1 as a key regulator of *ADAM33* transcription. Sp1 is a ubiquitous transcription factor, implicated in many developmental pathways, modification of which may alter *ADAM33* expression during lung development and maturation. Polymorphic changes were not associated with reduced lung function at one month old but were at three years. This suggests either a critical window exists between birth and age 3 years for ADAM33-environmental interactions to take place to trigger asthmatic symptoms in an already predisposed lung or the lung function tests performed on the 4 week old infant were not sensitive enough to detect an abnormality.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I extend my sincere gratitude and appreciation to all who have given their time and efforts in the making and writing of this thesis.

My special thanks go to Dr. John Holloway for his incredible patience, guidance and support and my second supervisor, Professor Stephen Holgate for the same.

Secondly, I wish to thank all who have helped me along the way, in particular Shelia Barton and Nik Maniatis for their help in statistical analysis, and members of Human Genetics labs, namely Matthew Rose-Zerilli, Dr. Ian Yang, Helen North, Lesley Hinks and Tricia Briggs. I would also like to thank Dr. Donna Davies and colleagues in the Brooke laboratory, and Dr. Derek Mann and colleagues for support in the promoter work. I must also thank Dr. Judith Holloway for IT support and critical review. Thanks also to Mrs. Ann Cakebread for proof reading this work.

I need to say a huge thank you to my husband Martin, and my children Paul, Joanna, Rachel, Becky and James for their encouragement, support and tolerance.

Lastly, I wish to thank the Medical Research Council for funding this research, and also the British Federation of Women Graduates (BFWG) and the Asthma and Allergy Inflammation (AAIR) charity for further financial support.

Abbreviations

Abbreviation	Full name
AbD	Assay by Design
AAT	Alpha-1 antitrypsin
ABI	Applied Biosystems
AoD	Assay on Demand
ADAM33	A Disintegrin and Metalloprotease 33
ADAMTS	A Disintegrin and Metalloprotease Thrombospondin
APC	Antigen-presenting cell
APP	Amyloid precursor protein
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATRN	Attractin (chr 20 p13)
ARMS	Amplification Refractory Mutation System
ASO	Allele specific oligonucleotide
ASP	Affected sib-pair
BAL	Bronchoalveolar Lavage
Baso	Basophil
BHR	Bronchial Hyperresponsiveness
BLAST	Basic Local Alignment Search Tool
$CaCl_2$	Calcium chloride
CAMR	Centre for Applied Microbiology and Research
CD	Cluster of differntiation
CEPH	Centre d'Etude du polymorphisme Humaine
CFTR	Cystic Fibrosis Transmembrane Conductance
Chr	Chromosome
cM	Centimorgan

COPD	Chronic Obstructive Pulmonary Disease
COPSAC	Copenhagen Study of Asthma in Childhood
CSGA	Collaborative Study on the Genetics of Asthma
CT	Cycle threshold
CTL	Cytotoxic T lymphocytes
CYF1PZ	Cytoplasmic fragile X mental retardation protein (FMRP) interacting proteins 2 (chr 5q 33)
DC	Dentritic cell
DEPC	Diethyl Pyrocarbonate
dHPLC	denaturing High Performance Liquid Chromatography
Der p	*Dermatophagoides pteronyssinus*
DGGE	Denaturing Gradient Gel Electrophoresis
DICE	Differential Control Elements
DNA	Deoxyribonucleic acid
DMEM	Dulbeccos Modified Eagles Medium
DMSO	Dimethyl Sulphoxide
dNTP	Deoxyribonucleotide Triphosphate
DPP10	Dipeptidyl peptidases 10 (chr 2)
ECACC	European Collection of Animal Cell Cultures
ECRHS	European Community Respiratory Health Survey
ECM	Extra cellular matrix
EDTA (disodium salt)	Ethylenediaminetetraacetic acid (disodium salt)
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-linked immunosorbant assay
EM (algorithm)	Estimation Maximisation (algorithm)
EMEM	Eagles Minimum Essential Medium
EMTU	Epithelial-mesenchymal trophic unit
Eo	Eosinophil
ERS	European Respiratory Society
FcεR	Fcε receptor
FEF	Forced expiratory flow

FEV_1	Forced Expiratory Volume in one second
FGF	Fibroblast growth factor
FRC	Functional Residual Capacity
FRET	Fluorescence Resonance Energy Transfer
FVC	Forced Vital Capacity
sGaw	Specific Airway Compliance
GM-CSF	Granulocyte Macrophage- Colony stimulating factor (chr 13)
GFRA4	GDNF family receptor alpha 4 (chr 20p13-p12)
GPRA	G protein-coupled receptor for asthma susceptibility (chr 7p14)
HBSS	Hanks Balanced Salt Solution
HB-EGF	Heparin-binding-epidermal growth factor
HCl	Hydrochloric acid
HGMP	Human Genome Mapping Project
HLA	Human Leukocyte Antigen
HLA-G	Human Leukocyte Antigen–G (chr 6p21)
HRC	Human Random Control
HRP	Horse radish peroxidase
htSNP	Haplotype tagged Single Nucleotide Polymorphism
HWE	Hardy Weinberg Equilibrium
IBD	Identical by descent
IFN-γ	Interferon-γ (chr 12q14)
Ig	Immunoglobulin
Il	Interleukin
ISAAC	International Study of Asthma and Allergy in Childhood
kD	Kilodalton
LD	Linkage disequilibrium
LNA	Locked nucleic acid
LOD	Logarithm of the Odds
NaPyr	Sodium Pyruvate
NCBI	National Centre of Bioinformatics
LAM	Lymphangioleiomyomatosis

LRTI	Lower Respiratory Tract Infections
MADGE	Microarray Diagonal Gel Electrophoresis
MALDITOF	Matrix-Assisted Laser Desorption/Ionisation Time-Of-Flight (mass spectroscopy)
MAPK	Mitogen activated protein kinase
Mast	Mast cell
Mb	Megabase
$MgCl_2$	Magnesium Chloride
MSE	Mean Squared Error
MGB	Minor groove binder
MHC	Major Histocompatibility Complex
min	Minute
mM	Millimolar
M-MLV	Maloney murine leukaemia virus
MMP	Matrix metalloproteinase
mRNA	Messenger ribonucleic acid
^{NAC} MAAS	National Asthma Campaign Manchester Asthma and Allergy Study
NEAA	Non Essential Amino Acids
NEB	New England Buffer
ng	Nanogrammes
nl	Nanolitre
nM	Nanomolar
NTC	No Template Control
OLA	Oligoligation assay
O_2	Oxygen
PBS	Phosphate Buffered Saline
PC20	Provocative concentration inducing a 20% fall from baseline FEV
PD 20	Provocative dose inducing a 20% fall from baseline FEV
PCR	Polymerase Chain Reaction
PDGF	Platelet derived growth factor
PHD	Plant Homeodomain

PHF11	PHD finger protein 11 (chr 13q14)
PMA	Phorbol Myristate Acetate
PNA	Peptide nucleic acid
psi	Pounds per square inch
Pwo DNA polymerase	*Pyrococcus woesei* DNA polymerase
QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
RAST	Radioallergosorbent test
RBM	Reticular basement membrane
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribose nucleic acid
rRNA	Ribosomal ribose nucleic acid
Rpm	Revolutions per minute
RTC	Rapid thoracoabdominal compression
RV	Residual Volume
RVRTC	Raised volume rapid thoracoabdominal compression
SAP	Shrimp alkaline phosphatase
SETDB2	SET domain, bifurcated 2 (chr 13)
sec	Second
SSCP	Single Strand Confirmation Polymorphism
SMC	Smooth muscle cell
SN	Sialoadhesin (chr 20p13)
SNP	Single Nucleotide Polymorphism
SRE	Steroid Response Element
sRaw	Specific Airways Resistance
SVMP	Snake venom metalloprotease
TACE	Tumour necrosis factor alpha converting enzyme (chr 2p25)
Taq DNA polymerase	*Thermus aquaticus* DNA polymerase
TBE buffer	Tris Borate EDTA buffer
T_c	Cytotoxic T cell
TCR	T-cell receptor

TCRS	Tuscon Children's respiratory study
TDT	Transmission Disequilibrium Test
TE	Tris EDTA
TGF	Transforming growth factor
Th	T helper cells
TIMP	Tissue inhibitor of metalloproteinase
Tm	Melting temperature
TNF-α	Tumour necrosis factor alpha (chr 6p21)
TNF-β	Tumour necrosis factor beta (chr 6p21)
TNFR	Tumour necrosis factor receptor
TRANCE	Tumour necrosis factor related activation-induced cytokine
tRNA	Transfer ribose nucleic acid
TSC	The SNP consortium
UCSC genome browser	Genome Bioinformatics Group of UC Santa Cruz.
µl	Microlitre
µM	Micromolar
UTR	Untranslated region
V_c	Vital capacity
Vt	Tidal volume
VTG	Thoracic gas volume
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Asthma

1.1.1 Incidence and mortality

Asthma is an increasingly common respiratory disease that affects both children and adults. It is estimated that as many as 300 million people of all ages and ethnicities suffer from asthma, making it an increasing burden on healthcare systems worldwide [1]. The highest asthma prevalence occurs in westernised countries, with lower prevalence occurring in developing countries [2], indeed the rate of asthma is observed to increase as communities adopt western lifestyles and become urbanised [1]. The increase in prevalence has been associated with an increase in atopic sensitisation, and is paralleled by similar increases in other allergic disorders such as eczema and rhinitis [1]. The International Study of Asthma and Allergy in Childhood (ISAAC) and European Community Respiratory Health Survey (ECRHS) studies provide up to date prevalence rates of asthma across Europe [3]. Figures from these studies suggest asthma in Western Europe has doubled over the last ten years [3], though it is unclear whether incidence is still increasing or is now declining [4, 5]. The UK has the highest prevalence of asthma in Europe (10-13% of adults) with prevalence doubling over the last twenty years. An estimated 3.4 million people (1.5 million adolescents and 1.9 million adults) experience asthma symptoms requiring treatment in the UK. According to the World Health Organisation (WHO) asthma deaths outnumbered 180,000 persons/year worldwide at the end of the 1990s [3].

1.1.2 Common symptoms and triggers

The aetiology of asthma is complex and multifactorial and involves interaction between genetic and environmental factors. No single test or physical feature can

define the presence or absence of asthma. Asthma diagnosis is made on the basis of combined information from patient history, physical examination and physiological tests [1]. It is primarily a chronic T-cell-mediated inflammatory disorder of the airways [6] in which many other cells and cellular elements play a role, in particular mast cells, eosinophils, macrophages, and epithelial cells. In susceptible individuals, the inflammation causes recurrent episodes of wheezing, chest tightness, shortness of breath and coughing particularly at night and in early morning. The episodes are associated with variable airflow obstruction, which is wholly reversible. Occasionally the recovery is spontaneous but usually treatment to reduce inflammation and relax bronchospasm is required. Symptoms are often worsened or prolonged by triggers, such as aero-allergen exposure, respiratory virus infections, exercise, cold air, smoke and pollution. Most data regarding asthma pathogenesis is in relation to atopic or allergic asthma although non-allergenic mechanisms of inflammation, such as aspirin-induced asthma, do exist [7].

1.1.3 The role of the immune system

The role of the immune system is to recognise foreign pathogens, to mount an immune response and to eliminate non-self invaders. Immune responses can be divided into humoral and cell-mediated responses. The humoral branch of the immune system involves interaction of B cells with antigen and their subsequent proliferation and differentiation into antibody-secreting plasma cells. Immunoglobulin (or antibody) binds to antigen, facilitating its subsequent elimination (via for example, phagocytosis or complement-mediated lysis). Effector cells generated in response to antigen are responsible for cell-mediated immunity. Cell-mediated immunity can only be transferred by immune T cells. Both activated T helper cells (Th) and cytotoxic T lymphocytes (CTL) serve as effector cells in cell-mediated immune reactions. Th_1 cells secrete cytokines such as IFN-γ and IL-2 and are part of an effector mechanism that directs macrophage activation. Th_2 cells secrete cytokines including IL-4, IL-13, and IL-5. These cytokines are powerful activators and differentiators of B cells, and are important in B cell growth. A third effector mechanism involves cytotoxic T cells (Tc) which are especially important in host defence against intracellular bacteria, virus-infected cells and tumour cells.

1.1.3.1 Hypersensitivity

One way in which the immune system fails is the production of an inappropriately powerful response to a harmless antigen, leading to hypersensitivity. Reactions within the humoral branch are termed immediate hypersensitivities types I, II and III, whereas reactions within the cell-mediated branch are termed delayed type hypersensitivity type IV (table 1.1). Type I hypersensitivity reactions are mediated by overproduction of IgE in response to innocuous environmental antigens such as pollen, house dust mite or animal dander. IgE is produced in response to allergen (or parasitic infection) which enters the body via mucosal surfaces. Allergen is intercepted by an antigen-presenting cell (APC), such as a dendritic cell. The allergen is processed and presented on the cell's surface to T cells. In addition, the APC secretes cytokines attracting mast cells to the immediate area. Th_2 cells secrete cytokines that induce B-cell proliferation and favour the production of an allergen specific IgE response [8]. IgE, secreted by B-cells, binds mast cells via the high affinity FcεR1 receptor on the cell surface, thus sensitising them. When allergen subsequently reaches the sensitised mast cell it cross-links surface bound IgE causing a release of pre-formed mediators such as histamine, proteases, leukotrienes and prostaglandins, leading to a response that includes inflammation, itching, coughing, lacrimation, bronchoconstriction, mucus secretion, vomiting and diarrhoea, all common symptoms of the allergic response. The resulting release of pharmacological mediators by IgE-sensitised mast cells produces an acute inflammatory reaction, also termed immediate hypersensitivity because of the rapid manifestation of inflammatory symptoms such as wheal and flare, and wheeze.

Table 1.1 Hypersensitivity and its consequences.

Hypersensitivity Type	Action	Examples of Disease Consequences
Type I	IgE binds to mast cells via Fc receptors. Allergen binds to IgE and becomes cross-linked which signals mast cells to degranulate, releasing mediators that produce allergic reactions.	Asthma, Rhinitis, Eczema, Anaphylaxis
Type II	Antibody is directed against antigen on an individual's own cells or on foreign antigen such as in transfused red blood cells. This may lead to cytotoxic action by killer cells or complement-mediated lysis	Rejection of blood transfusion, Haemolytic disease of the newborn, Myasthenia gravis
Type III	Immune complexes are deposited in the tissue. Complement is activated and polymorphs are attracted to the site of inflammation causing localised inflammation and swelling	Autoimmune diseases such as arthritis, multiple sclerosis.
Type IV	Antigen sensitised T cells release lymphokines following secondary contact with the same antigen. Cytokines induce inflammatory reactions, activating and attracting macrophages, which release inflammatory mediators.	Contact (e.g. nickel, chromate, rubber), Tuberculin, Granuloma formation

Table 1.1 The four types of hypersensitivity reactions. Asthma is a type I hypersensitivity reaction caused by overproduction of IgE resulting in degranulation of mast cells which releases a cocktail of inflammatory mediators.

In the context of the asthmatic airway, exposure to previously encountered allergen causes an early phase response (within fifteen minutes of trigger inhalation). Allergen binds to and cross-links IgE/FcεR1 on previously sensitised mast cells which are present in the airways of both asthmatics and non-asthmatics. This triggers a signalling cascade resulting in mast cell degranulation with release of histamine and other inflammatory mediators into the surrounding lung tissue. Allergen is also intercepted by antigen-presenting cells (APCs). Signals from these phagocytic cells attract more mast cells to the immediate area. The resulting localised inflammation causes smooth muscle contraction, plasma exudation, oedema and increased mucus production resulting in the characteristic wheeze and shortness of breath identified in asthma. For immediate relief of an acute attack short acting β_2-agonists and anticholinergics are used to relax smooth muscle around the airways. Their effects are short lived.

Three to six hours after the immediate reaction, a second intense inflammatory reaction may occur, termed the late phase response. Mediators released during the immediate response, such as cysteinyl-leukotriennes, eotaxin, and many other chemokines, attract inflammatory cells, including eosinophils and Th_2 cells, to the site

and activate them. Eosinophils have a similar action to mast cells, their degranulation causing further localised inflammation. Th_2 cells secrete cytokines such as IL-4, IL-5 and IL-13, which continue to drive the inflammatory cascade. Treatments for long-term control of asthma include long-acting β-agonists to relax the smooth muscle of the airway, and steroids to reduce inflammation. Leukotriene modifiers are an alternative to steroids but are not as effective. Omalizumab (Xolair) is a recently developed immunomodulatory treatment that acts by binding IgE, preventing it binding to and activating mast cells [9]. Mast cells release a variety of inflammatory molecules including leukotrienes, histamine and cytokines which drive inflammation. Another immunomodulatory treatment recently developed acts on Tumour necrosis factor alpha (TNF-α) [10], a major therapeutic target in a range of chronic inflammatory disorders characterised by a Th_1 type immune response, and in which TNF-α is generated in excess. Atopic asthma is predominantly regarded as a Th_2 type disorder but as it becomes more severe and chronic, it adopts additional characteristics including corticosteroid refractoriness and involvement of neutrophils suggestive of an altered inflammatory profile towards a Th_1 type response. TNF-α levels in bronchiolar lavage (BAL), TNF-α gene expression and TNF-α immunoreactive cells were increased in subjects with severe corticosteroid-dependent asthma. An anti-TNF treatment used in the treatment of arthirits, called Etanercept, was recently trialled for use in asthmatic patients. Use of Etanercept showed an improvement in asthma symptoms, lung function, and bronchial hyperresponsiveness [10]. These new treatments hold promise for control of symptoms in therapy-resistant and severe asthma sufferers [9-11].

1.2 Lung development and remodelling

In utero the lung undergoes extensive modelling and remodelling processes that are entirely appropriate for the normal development of the lung. Indeed many of the cytokines thought to be involved in the process of remodelling are necessary for normal lung development [12]. Until recently, structural change of the airway has been assumed to result from uncontrolled inflammatory and repair processes. Increased stress, damage and shedding of the airway epithelium are reported in autopsy samples from people who have died of asthma [12-14]. In the healthy lung,

loss of epithelium induces a healing process that results in complete restoration of the columnar epithelial layer with normal proportions of goblet and ciliated cells. In an asthmatic lung, repeated injury results in a defective repair process leading to enlargement (hypertrophy) and/or abnormal multiplication (hyperplasia) of goblet cells. These phenomena are found in both asthma and chronic obstructive airways disease (COPD)) and leads to hypersecretion of mucus. In fatal or near fatal asthma the mucus has increased viscosity.

In asthma, clusters of sloughed epithelial cells (creola bodies) are observed in the sputum, especially during exacerbations. Increased numbers of epithelial cells are found in BAL fluid, and (variable) loss of surface epithelium is observed in biopsy specimens [12]. Injury to the superficial epithelium is normally accompanied by mitotic activity in the remaining cells with rapid restoration of the injured area [12]. The mitotic response in asthma is repressed giving a reduced repair response to injury [15]. In addition, susceptibility of the asthmatic airway epithelium to oxidant stress is greater than normal, which may contribute to the well established association of asthma with air pollution (e.g. nitrogen dioxide and ozone) and diets low in anti-oxidants [16].

The airflow obstruction observed in asthma is also, in part, a result of increased collagen deposition resulting from ongoing inflammation and repair processes. The extra cellular matrix (ECM) provides the lung with scaffolding to stabilise the physical structure of the tissues. Collagen is a major component of the ECM and collagen turnover (synthesis and degradation) is rapid in normal tissues. The matrix metalloproteinase (MMP) family exhibit degradative activity against a range of ECM proteins. The gene expression of MMPs is tightly regulated by cytokines and growth factors that enhance, (e.g. IL-1, epidermal growth factor (EGF), fibroblast growth factor (FGF) or platelet derived growth factor (PDGF)), or inhibit (e.g. transforming growth factor- β (TGF-β) or IL-4), their transcription. Proteinase activity in the lung is also regulated by tissue inhibitors of metalloproteinases (TIMPs).

In asthmatic airways communication between the epithelium and mesenchyme, termed the epithelial-mesenchymal trophic unit (EMTU), appears to be reactivated and is similar to that observed during development [17]. Signals from the damaged

epithelium induce proliferation of fibroblasts through the action of profibrogenic growth factors such as TNF-α and TGF-β. TNF-α is secreted predominantly by macrophages and mast cells [18]. It is a multifunctional, though mainly proinflammatory, cytokine that often acts together with IL-1. It has indirect effects on the growth of smooth muscle [19] and can stimulate collagen synthesis and fibroblast proliferation [20]. Over-production of TNF-α has been implicated in the several inflammatory diseases such as arthritis and cachexia [21]. TGF-β is an important peptide growth factor that stimulates collagen production by fibroblasts [22] and promotes differentiation of fibroblasts into myofibroblasts [23]. Myofibroblasts secrete interstitial collagens, as well as growth factors, such as endothelin 1 and vascular endothelial growth factor, which are mitogens for smooth muscle and endothelial cells [23]. Collagen deposition from the increased number of myofibroblasts leads to a thickened reticular basement membrane (RBM) in asthmatic airways [12, 24]. Collagens (types I, III and V) can entrap molecules such as growth factors. The entrapped molecules could modulate the state of differentiation, integrity and function of the overlying surface epithelium. In addition, it is proposed that entrapped and adsorbed molecules may create an osmotic potential with the resulting uptake of water increasing oedema and swelling. The thickened membrane does not prevent migration of inflammatory cells [12]. Thickening of the RBM has shown a positive correlation with airway hyperresponsiveness, the frequency of asthma attacks and the numbers of fibroblasts and myofibroblasts that lie external and adjacent to it [12]. There is an increase in smooth muscle mass (key end effector cells of bronchospasm and airway narrowing) in asthmatics, predominantly due to smooth muscle cell proliferation leading to an increase in myocyte number. Muscle fibre hypertrophy may also contribute. In addition to thickening, there is an increased sensitivity to stimulus, termed hypersensitivity, and the maximal degree of airway smooth muscle contraction usually exceeds that of 'normals' (hyperreactivity) [25]. The increased sensitivity and reactivity usually occur together; this is termed hyperresponsiveness. Figure 1.1 summarises the complex interactions involved in the pathogenesis of asthma.

Figure 1.1 The complex interactions of Asthma

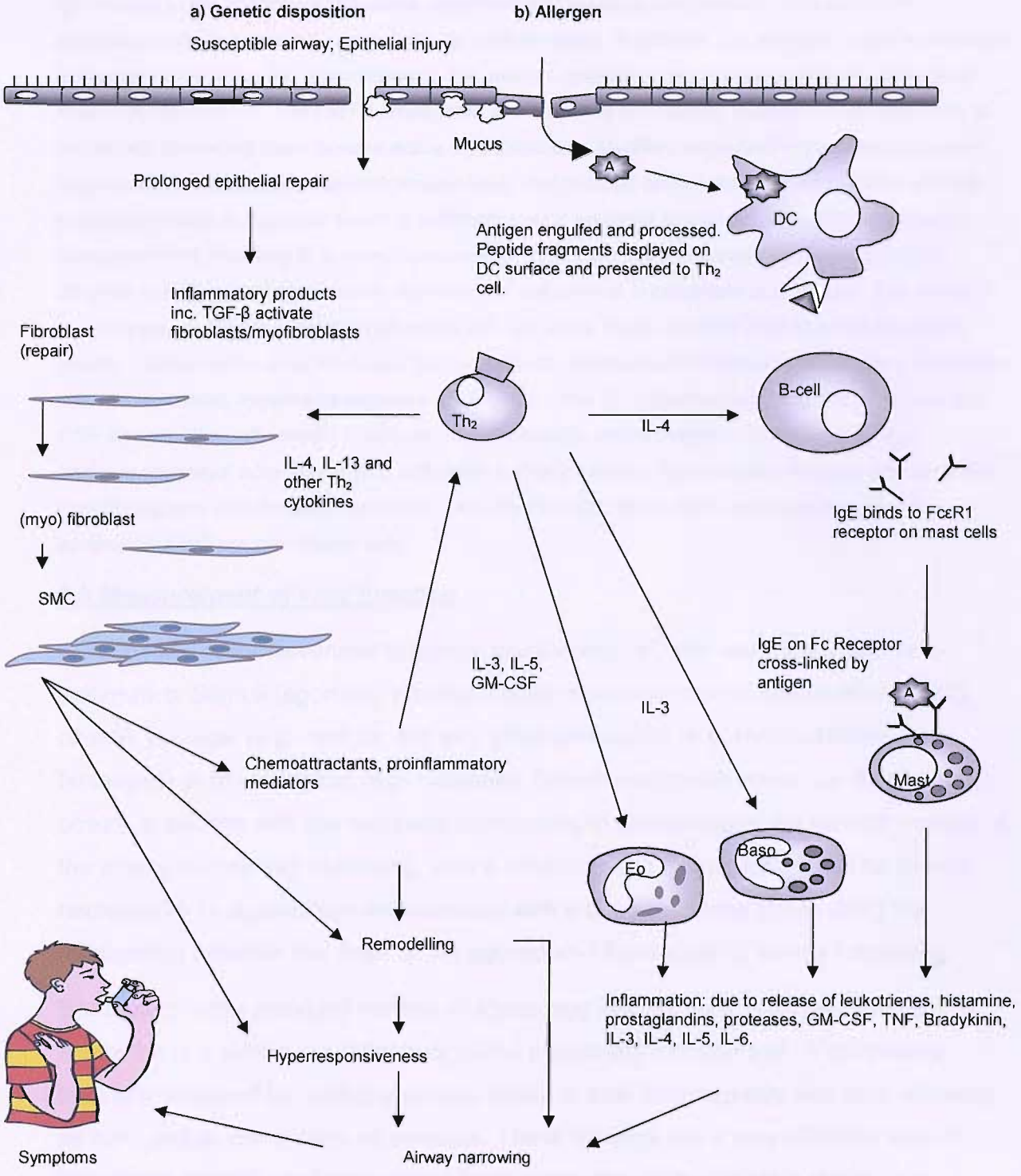


Figure 1.1 The Complex Interactions of Asthma

a) Damage to the epithelium stimulates repair and inflammatory mechanisms. The asthmatic epithelium has an increased susceptibility to oxidant stress. In addition, the epithelial repair mechanism is impaired prolonging the repair process. The altered epithelium communicates with the underlying mesenchyme (EMTU). The EMTU propagates and amplifies remodelling events through activation of fibroblasts converting them to more active myofibroblasts. Myofibroblasts proliferate in the asthmatic lung resulting in a thickened smooth muscle layer and reduced airway diameter. In addition, smooth muscle contracts to a greater extent in asthmatics, and responds to a lower dose of agonist than in non-asthmatics, resulting in hyperresponsiveness. b) Antigen is intercepted and processed by a dendritic cell. Peptide fragments displayed on the cell surface is presented to a Th_2 cell. This evokes IL-4 release, switching B cell IgG release to IgE. IgE binds FcεR1 on mast cells found in the airway vicinity. Antigen cross-links the bound IgE, and signals the mast cell to release inflammatory mediators such as histamine, trypsin and chymase. Release of other Th_2 cytokines including IL-5, IL-3 and GM-CSF attracts other inflammatory cells such as eosinophils and basophils into the area. A Th_2 'microenvironment' born from EMTU activation and inflammatory cell secretion sustains and amplifies remodelling and inflammatory processes. DC- dendritic cell; Mast- mast cell; Baso-basophil; Eo-eosinophil; SMC-smooth muscle cells.

1.3 Measurement of lung function

Non-specific responsiveness to airway insult occurs in both asthmatics and non-asthmatics. Stimuli (agonists) eliciting airway narrowing can be chemical (e.g. SO_2, ozone), physical (e.g. cold air, dry air), pharmacological (e.g. methacholine, histamine) or physiological (e.g. exercise). Specific responsiveness, i.e. to allergen, occurs in asthma with the response culminating in shortening of the smooth muscle of the airway and airway narrowing, with a resulting drop in lung function. The airway response to an agonist can be assessed with a dose-response curve using the relationship between the dose of the agonist and the degree of airway narrowing.

Spirometry is the standard method of measuring relative lung volumes. Forced expiration is a simple but extremely useful pulmonary function test. A spirometry tracing is obtained by having a person inhale to total lung capacity and then exhaling as hard and as completely as possible. These tracings are a very effective way of separating normal ventilatory states from obstructive and restrictive states. In a normal forced expiration curve, the volume that the subject can expire in one second (referred to as FEV_1) is usually about 80% (in infants this is 100%) of the total forced vital capacity (FVC), or approximately four litres out of five. In an obstructive condition

such as asthma, bronchitis or emphysema, the forced vital capacity is reduced, as well as the rate of expiratory flow. Thus, an individual with an obstructive defect might have a forced vital capacity of only 3.0 litres, and in the first second of forced expiration, exhale only 1.5 litres, giving a FEV_1/FVC of 50%.

Spirometry cannot provide information about absolute volumes of air in the lung and is difficult to use in very young or debilitated patients. Body plethysmography is used for measuring lung volumes. The subject sits in an airtight box and breathes as normal (functional residual capacity, FRC) through a breathing tube. A shutter drops across the breathing tube and the subject makes respiratory efforts against the closed shutter. Pressure and volume changes are measured and calculated during respiratory efforts, using Boyle's Law (P1V1=P2V2; the principle that at constant temperature, the volume of a gas varies inversely to the absolute pressure applied to the gas) [26].

Measurement of resistance and compliance of lung tissue is a useful measure of disease status. In an asthmatic lung thickening of the smooth muscle layer, oedema and deposition of collagenous matrix increases the 'stiffness' of the tissue thereby decreasing compliance (a measure of the pulmonary volume change over unit pressure change) and increasing resistance (the opposition of the tracheobronchial tree to airflow). Recently whole body plethysmography has been used to measure specific airway resistance (sRaw) in young children to provide insight into asthma development in early childhood [27]. Children who wheeze have higher specific airways resistance (sRaw) [28], correlating with reduced lung function. Reduced infant lung function is associated with increased BHR later in childhood, predicting increased likelihood of the development of asthma and decline in lung function later in life [29]. The degree of sensitivity of these tests to detect abnormalities in the airways varies depending on the test and the age at which it is applied.

1.3.1 Factors affecting lung function

An adult's lung function is affected by growth in lung function during childhood, eventual maximal attained level of lung function, age of onset of decline in lung function and rate of decline [30]. Wheezing during the first year of life is often a transient condition, improving with time. It appears to be related to early-life reduced

airway calibre [31]. The majority of early wheezers have normal lung function in later life. For those who develop persistent asthma symptoms evidence increasingly suggests changes in lung structure occur before the onset of asthmatic symptoms [32, 33]. Lung function continues to grow during puberty and after cessation of height growth. Those subjects with asthma symptoms have a reduced growth in lung function, and those with low FEV_1 and BHR in childhood are predicted to have poorer lung function in adulthood [34]. Environmental risk factors for asthma which could alter normal lung development and repair include active or passive smoking, exposure to allergen (dust mite, pollen, animal dander), viral respiratory infections and possibly diet and air pollution (NO_2 and ozone) [35]. Rate of decline is similar between asthmatics and non asthmatics but increased in asthmatics who smoke [36]. In asthma and COPD there is evidence to suggest decline in lung function could have a genetic component [37, 38], although the only proven genetic risk factor is severe alpha 1-antitrypsin (AAT) deficiency in COPD [39].

1.4 Genetic disease

Eighty five years ago Robert Cooke and Albert Van der Veer investigated the heritability (proportion of variance due to genetic effects) of allergy using a cohort of 621 people with allergy and 76 non-allergic people. The results showed that 48% of allergic people had a positive family history of allergy compared to 14% of the control population, suggesting heritability of these diseases. Further studies have since established that atopy and atopic diseases such as asthma, rhinitis and eczema have strong genetic components [40].

Genetic distances are generally expressed in centimorgans (cM) with one cM corresponding to 1% recombination and approximately 1 000 000 base pairs on a physical map (1Mb). Recombination of 1% means crossover occurs one in every 100 meioses. The DNA of an organism is subject to a variety of different types of heritable change or mutation. If a change occurs at greater than 1% frequency in the general population it is referred to as polymorphism. Most base substitutions involve replacement of a single base. This type of polymorphism is referred to as a single nucleotide polymorphism (SNP). Mutations are usually due to errors in DNA

replication or repair, and often have no detrimental effect. However, polymorphic variation can affect the ultimate function or control of a gene.

1.4.1 Gene Regulation

Polymorphic variation in DNA can ultimately affect all levels of regulation. There are three broad levels of gene regulation in humans. Epigenetic mechanisms such as DNA methylation operate in mammalian cells, acting as a general method to repress transcription and are involved in monoallelic gene expression. Transcriptional regulation of gene expression occurs at the level of initiation through the core promoter of a gene (section 5.1.2-5.1.3). Post transcriptional regulation of gene expression includes mechanisms operating at the level of RNA processing, transport, translation, mRNA stability, protein processing, protein targeting and protein stability.

1.4.2 Simple Mendelian disease

Simple genetic diseases are relatively rare and are the result of mutations in a single gene, which is sufficient to cause manifestation of disease. They typically display a Mendelian pattern of inheritance i.e. autosomal dominant, recessive or sex-linked. The use of genetic analysis to identify genes responsible for simple mendelian traits such as Cystic Fibrosis [41] and Huntington's Chorea [42] has become almost routine over the last twenty years, since it was recognised that genetic inheritance can be traced with naturally occurring sequence variation. Since then many genes have been mapped to specific chromosomal locations and many positionally cloned. However, most common medical conditions with a known genetic component to their aetiology have much more complex inheritance patterns.

1.4.3 Complex genetic disease

The majority of common human diseases are complex genetic diseases and include conditions such as diabetes, hypertension, heart disease, schizophrenia and asthma. A clear hereditary pattern is apparent but the disease does not follow a Mendelian pattern of inheritance. Complex genetic diseases tend to be far more prevalent than simple mendelian (monogenic) diseases, laying a huge burden on the health care system. For example, asthma occurs in 10% (and rising) of children in the UK [3], whereas one of the most common single gene disorders in Caucasians, cystic

fibrosis, affects 1:2000 live white births [43]. In complex genetic disease there may be a number of different genes that predispose individuals to disease, either more than one gene in an individual (polygenic) or different genes in different individuals (genetic heterogeneity). There may be a number of environmental factors necessary for the expression of the disease phenotype. The recent rise in asthma has occurred too quickly for it to be caused by de novo genetic mutation. It is more likely that the rise in cases is due to environmental interaction with a genetically predisposed lung. This fits with the observation that the rise in asthma in the west, far exceeds that seen in underdeveloped countries suggesting the 'western lifestyle', (under-activity, obesity, pollution, hygiene) to be a major contributory factor.

1.4.4 Identifying genes underlying complex disease

Before embarking on a genetic study of a disease several factors need to be considered including;

- assessment of heritability to establish that there is a genetic component;
- defining the phenotype of physical characteristics to be measured in a population;
- determining which markers are going to be typed in the DNA samples obtained from the population;
- how the relationship between the genetic data and phenotype measured are to be analysed;
- how the data can be used to identify the genes underlying the disease.

1.4.4.1 Heritability of atopy and asthma

Heritability is the proportion of the cause of a character that is due to genetic rather than environmental influence. Familial concordance of a disorder can be due to genetic or environmental effects. Family studies have shown that risk of developing atopic disease such as asthma and rhinitis, for example, is increased among families with atopic relatives and that children with asthmatic parents are at a far higher risk of developing the disease than those with non asthmatic parents [44-46]. Indeed in a study using a twin-family model, the incidence of asthmatic disease in twins with

affected parents was fourfold compared to the incidence in twins with unaffected parents [47]. This indicates that asthma recurs in families due to shared genes rather than just environment. Twin studies are a useful tool to estimate genetic and environmental influence on a trait or disease and help determine relative contribution of shared genes or shared environment to that phenotypic trait. Twin studies compare phenotypic similarity between monozygotic twins (100% DNA similarity) and dizygotic twins (~50% DNA similarity) and assume monozygous and dizygous twins share the same gene pool, twins are representative of the general population, self reported zygosity is correct in questionnaire based studies, and the environment for monozygote and dizygotic twins is similar [35]. The correlation between the twin pairs can be used to estimate the magnitude of genetic and environmental influence [35]. Numerous twin studies have shown there is significantly greater concordance among monozygotic (MZ) twins than dizygotic (DZ) twins with regard to asthma [48, 49] suggesting a genetic component. Using atopy rather than asthma as an end point, 50-80% concordance has been reported in twin pairs [50-52] indicating a greater concordance with atopy. However, even monozygotic twins, who have identical genomes, do not have 100% concordance with regard to the development of asthma or atopic diseases. The failure to observe 100% concordance (or even levels approaching 100%) strongly suggests that environmental as well as genetic factors influence atopic or asthma phenotype expression.

Another experimental design that can identify a genetic component to a trait is segregation analysis. Segregation analysis tests the hypothesis that the aggregation of a trait is due to the action of a major gene, but does not identify the gene or region. Segregation analysis requires family pedigrees and is used to analyse the inheritance, penetrance and heritability of a character. It identifies whether a marker co-segregates within an affected family revealing Mendelian inheritance patterns (autosomal or sex-linked and recessive or dominant), nonclassical inheritance (mitochondrial diseases, genomic imprinting), or non-Mendelian inheritance (no pattern, environmental). Segregation analysis investigating asthma has been largely inconclusive, perhaps because of variation in phenotype definition between studies or because of genetic heterogeneity (different populations with different genes acting to regulate the phenotype). Studies confirm a clear genetic component to asthma but

suggest a model where several genes may exert effects, which are then attenuated or intensified by environmental interaction. In addition different combinations of gene variants contribute to the phenotype in different families [40, 44, 46].

1.4.4.2 Defining a phenotype for asthma

To investigate the genetic basis of any disease a recognised definition should be defined. This is especially true of complex diseases, such as asthma, which is likely to be a spectrum of different phenotypes within a broad diagnosis, rather than a single specific disease with a single pathogenic mechanism. Various clinical definitions of asthma have been proposed through national and international guidelines, working groups and workshops, which incorporate symptoms, lung function, exacerbations and specific use of high dose corticosteroids [53, 54]. The phenotype of asthma is complex and involves the interaction of a number of genes. Use of an intermediate phenotype such as bronchial hyperresponsiveness (BHR) or serum IgE, that contribute to the overall broad phenotype, allows a more tightly defined group to be examined. Lung function measures, including peak flow, BHR, and reversibility of bronchoconstriction, can be measured objectively in all individuals and used to help define asthma. Increased BHR is associated with decline in lung function and increased asthma severity [55].

Allergic phenotypes can be assessed by measuring total and specific serum IgE levels, serum eosinophil counts and skin test responses to common allergens. Asthma beginning in early life is frequently associated with atopy, with more than half the cases of persistent allergic asthma beginning before the age of three and 80% before the age of six [56]. This indicates that early allergic sensitisation is a risk factor for persistent asthma [57]. Asthma, in particular severe asthma, is correlated with raised IgE levels [58, 59]. The fundamental importance of IgE in allergic asthma has been illustrated by the marked anti-inflammatory effects of the anti- IgE drug, Omalizumab [9]. Approximately 5-10% of asthmatics will progress to severe asthma [60].

Severe (or refractory) asthma is a distinct sub-group that is increasingly recognised. A variety of symptoms define this group, including widely varying peak flow (brittle asthma), severe/chronic airflow limitation (irreversible asthma), rapid progressive loss

of lung function, variable response to corticosteroids (steroid dependent and/or resistant asthma) and difficult to control or poorly controlled asthma [53]. Severe asthma is more often non-allergenic than mild/moderate asthma [61]. The irreversibility of bronchoconstriction, and steroid insensitivity seen in severe asthma is more often associated with COPD [62], a late onset disease associated with smoking. In addition severe asthma sees an increase in neutrophilia, [60, 63, 64] adding to the symptom overlap with COPD.

Phenotyping asthma in infants and children is even more problematic since there are at least three different wheeze patterns; transient (usually non atopic and associated with maternal smoking and pre existing abnormal lung function); non atopic asthma of childhood which often remits early; and atopic asthma [57].

1.4.4.3 Recruiting a study population

Having established a genetic component to the disease of interest, a study population is needed to undertake molecular genetic analysis to identify the gene(s) responsible. The type and size of the population depends on the epidemiology of the disease, the method of genetic epidemiologic analysis being used and the type of genetic markers being typed. For example, family groups are required to undertake linkage analysis, whereas an association study requires a randomly selected case-control population. If the disease has a late onset (such as Huntington's chorea) a family-based linkage study may be impractical due to absence of multigenerational families. If a disease is rare, recruiting affected cases with matched unaffected controls will be preferential to recruiting a random group, which would need to be very large to have sufficient power. Power depends on several factors including magnitude of effect, sample size (N), allele frequency and required level of statistical significance. Several websites are available offering free packages to calculate power [65-67].

Another consideration is the type and number of markers. There are 20,134 known genes in the human genome (Ensemble v36, December 2005). However, any site within the genome, intronic or exonic, that contains DNA variation (polymorphism) can be used as a marker. Microsatellites, a repetitive series of short dinucleotide or trinucleotide repeats, are commonly used for linkage analysis. The number of repeats varies between individuals allowing them to be used as markers, following

transmission of a chromosomal region from one generation to the next. Over 5000 microsatellite markers have been identified with the human genome project, spanning the genome at an average of 2cM [68]. Single nucleotide polymorphisms (SNPs) result from a single base mutation substituting one nucleotide for another. This means only 2 alleles are present at any SNP site, making them less informative than microsatellites (which may have many alleles depending on number of copies of repeat). However SNPs occur more frequently than microsatellites. The SNP consortium (TSC) has identified and validated 1.8 million SNPs to date. It has been estimated that in the world's human population about 10 million (1 per 300 bases) SNPs exist constituting 90% of the variation in the population [69, 70].

The International HapMap (haplotype mapping) Project is a large scale collaborative effort to catalogue SNP variations and determine the common patterns of DNA sequence variation in the human genome in haplotypes [69]. A haplotype is defined as the combination of alleles of different genes on the same chromosome. If a group of alleles are 'linked' in the same haplotype they are said to be in linkage disequilibrium (LD). LD is the statistical association between alleles at separate but linked loci (i.e. non independent). Alleles remain linked until a recombination event occurs, so alleles in close proximity are less likely to undergo recombination during meiosis than alleles further apart. Over time repeated recombination events break up the disequilibrium between alleles until only those tightly linked remain shared. SNPs are more suited to mapping by association and linkage disequilibrium than microsatellites [71].

1.4.4.4 Analysis of genetic data

The method chosen to analyse the data obtained is dependent on the design of the study.

1.4.4.4.1 Linkage analysis

Linkage analysis identifies chromosomal regions of interest that may harbour disease susceptibility genes. It requires family pedigrees with at least one affected individual. Genome-wide polymorphic markers (SNPs or microsatellites) are used to screen the families. One or more markers will eventually be identified that co-segregate with the

disease more often than would be expected by chance. Measure of linkage between the marker and disease allele is performed by looking at the recombination fraction and asking whether it differs from the value (0.5) that is expected from unlinked loci. The likelihood that a trait segregates with a marker is expressed as a LOD score (logarithm of the odds) i.e. likelihood that the loci are linked/likelihood that the loci are not linked. A value greater than 3.0 is considered to be significant in a whole genome screen using ~500 markers, and a value of ⁻2 excludes linkage. To study linkage using a LOD score approach a model has to include parameters such as mode of inheritance, penetrance or allele frequencies (parametric approach). If these parameters are unknown a nonparametric approach is used. A nonparametric approach involves allele-sharing methods. Affected relatives should show excess of the disease locus. Affected sib-pair (ASP) analysis is the simplest form of allele-sharing method. Both siblings are affected so disease genes are assumed to have been transmitted, thereby excluding non-penetrant individuals from the analysis. Two sibs can show identical by descent (IBD) sharing to no, one or two copies of any locus. Excess allele sharing can be measured using a χ^2 test (section 1.4.4.4.3).

1.4.4.4.2 Association analysis

Association analysis uses SNPs to test for statistic association with a disease phenotype. It requires a much smaller number of samples than linkage studies to detect a contributing gene effect. Association studies are a form of case-control study where the frequency of a particular allele at a locus is compared between affected and unaffected individuals. The strength of association is measured by odds ratio. A positive association between an allele and a trait suggests the allele in question is the disease-causing mutation, the allele is in linkage disequilibrium (LD) with the disease gene or the association is a result of population stratification. Confounding by stratification occurs when individuals are selected from two genetically different populations in different proportions in cases and controls; thus, the cases and controls are not matched for their genetic background. This may cause spurious associations, or it may mask true associations. A way to test for stratification is by genotyping approximately ten and twenty SNP or microsatellite markers across the genome and assessing differences between case and control populations. These

markers must not be related to each other or to the trait under investigation. The transmission disequilibrium test (TDT) [72, 73] is not affected by population stratification but requires family-based cohorts. It can be used to test for linkage and association. Alleles of heterozygote parents are divided into transmitted and non-transmitted alleles, and the preferential transmission of an allele to the affected child is tested.

1.4.4.4.3 Hardy Weinberg equilibrium

Gene frequencies and genotype ratios in a randomly breeding population remain constant from generation to generation. This is a basic principle of population genetics identified by Mendel when crossing monohybrids and is known as the law of Hardy Weinberg equilibrium. When two individuals who are heterozygote for a trait (i.e. Aa) are mated the probability of the offspring being homozygous for the dominant trait (AA) is 25%, the probability of heterozygosity (Aa) 50% and homozygosity for the recessive trait (aa) 25%. This is because meiosis separates the two alleles of each heterozygous parent so that 50% of the gametes will carry one allele and 50% the other. Mating mixes the gametes as illustrated in table 1.2

Table 1.2 Probability of heterozygous random matings

	A(p)	**a(q)**
A(p)	AA (pxp)	Aa (pxq)
a(q)	aA (qxP)	Aa (qxq)

Table 1.2 if the 'A' allele, is represented by p and 'a' allele by q the relationship may be expressed algebraically. It can be seen that the fraction of offspring being homozygote for the dominant allele is pxp (p^2), the fraction of offspring heterozygous for the trait is twice pxq (2pq) and the fraction of offspring homozygous recessive is qxq (q^2).

So: $p+q=1$, therefore $p^2+2pq+q^2=1$

Using an example from the data from the ECACC population (chapter 3) SNP Stp5, expected frequencies for a population in Hardy Weinberg equilibrium would be calculated as follows:

Using 191 samples (+1 negative control, not included in calculations)

49 of the population were homozygote GG, 98 were heterozygote GA and 44 were homozygote AA

So the number of alleles represented is:

G: (49 x 2) + 98 = 196; and A: (44 x 2) + 98 = 186

The total number of alleles is 191 x 2 = 382

So observed frequency of G = 196/382 = 0.51 (51%) and A = 186/382 = 0.49 (49%).

To observe what frequencies we would expect to see in a population (n = 191) if p = 0.51 and q = 0.49, when Hardy Weinberg equilibrium is obeyed:

Let G be represented by 'p' and A by 'q':

$p^2 + 2pq + q^2 = 1$

$0.51^2 + 2(0.51 \times 0.49) + 0.49^2 = 1$

0.2601 + 0.4998 + 0.2401 = 1

So: p (GG) (26/100) x 191 = 49.6 (50)

pq (GA) (49.98/100) x 191 = 95.46 (95)

q (AA) (24/100) x 191 = 45.84 (46)

Visual comparison of observed and expected suggests HWE is obeyed, but the ratios are not identical. The Chi-square (χ^2) test can be used to test that observed ratios are statistically close enough to the expected hypothesis driven ratios (The hypothesis here is that the two populations are the same). χ^2 can be calculated using this formula: $\chi^2 = \sum (O-E)^2 / E$

(where $\sum$ = sum of, O= observed and E = expected). The calculation may be performed using a table (table 1.3)

Table 1.3 Example of a χ^2 calculation

Class	O	E	$(O-E)^2$	$(O-E)^2 / E$
GG	49	50	1	0.02
GA	98	95	9	0.094
AA	44	46	4	0.086
			Total ($\sum$)	0.2

Table 1.3 The genotypes observed are compared to expected genotypes in a similar population. A χ^2 distribution table is used to find a probability (p) value for the sum (here 0.2). It is general scientific convention that if the probability of observing a deviation from expected is less than 5% (that expected by chance) the hypothesis is rejected (i.e. the observed and expected samples are not similar) So a p-value less than (p=0.05) suggests it is probable the observed population is significantly different.

In this case 0.2 (with 2 degrees of freedom) gives a value p = 0.9 so we can accept the hypothesis that observed is similar to expected. Most deviations from HWE indicate a poorly performing genotyping assay; however HWE calculations assume random mating of populations so several other scenarios may affect equilibrium. For example, if the rate of mutation of one allele (i.e. A) exceeds that of the other allele (a), the frequency of the alleles in the offspring will not remain in Hardy Weinberg equilibrium. Evolution depends on mutation through creation of new alleles, which when shuffled with the rest of the gene pool provides material for natural selection. The rate of mutation is so low however, that mutation by itself plays a minor evolutionary role. Many species are made of local populations whose members tend to breed within their immediate group. This is also true of some human populations. Each local population develops a gene pool distinct from that of other local populations. When one distinct population breeds with that of another, new genes may be introduced or existing gene frequencies may be altered, thus affecting Hardy Weinberg equilibrium. This is known as gene migration. Interbreeding is often limited to members of local populations. If the population is small then chance alone may eliminate alleles that are out of proportion in that population. The frequency of the allele may then drift to higher or lower proportions, ultimately with one allele being in 100% of the population, or equally as likely, disappearing. This is called genetic drift and may produce evolutionary change, but it is not a change that is adaptive to circumstances.

One of the cornerstones of Hardy Weinberg equilibrium is that mating must be random. If one or other of the 'mates' is choosy in his or her selection the gene frequencies may become altered. Darwin called this sexual selection. Humans tend to be attracted to similar genotypes (e.g. size, ethnicity, age) so seldom mate randomly. This is called assortative mating. Marriage between close relatives is a special case of assortive mating. The closer the kinship the more alleles are shared. Isolated

(founder) populations such as the Hutterites, a religious isolated population of European descent, in the USA, demonstrate assortive mating. They offer a more homogenous population so are useful for the study of genetic diseases [74]. If individuals with certain genes are better able to pass on their genes to offspring who mature to adulthood, then the frequency of the alleles within those genes will increase when compared to an individual who is less able to produce offspring. This is natural selection.

1.4.4.4.4 Multiple testing of hypotheses

When conducting association studies large numbers of allelic markers may be included. Even without a true affect 5% of results will show significance at the p=0.05 level, and 1% at p=0.01. This is termed type I error. Raw p-values should be corrected to account for the number of statistical tests applied using a statistical method such as the Bonferroni correction (p-value divided by the number of tests to give an adjusted p-value required for significance) [75]. Conversely a negative association may be seen even when there is a true effect. This is termed type II error and may be due to an under-powered study (section 1.4.4.3).

1.4.4.5 Approaches to gene identification

Two main strategies have been used to identify susceptibility genes in complex disease: positional cloning (reversed genetics) which scans the whole genome using a panel of polymorphic markers to find linkage in a chromosomal region, and the candidate gene approach that selects genes appearing relevant to the pathophysiology of the disease. Figure 1.2 highlights the pathways of these methods.

1.4.4.5.1 Positional Cloning by genome-wide screens

The positional cloning approach has successfully identified genes in simple mendelian disorders such the Cystic Fibrosis Transmembrane Conductance Regulator gene (CFTR) for cystic fibrosis and is now a useful approach for complex disorders such as asthma. Many genome-wide screens for atopy and atopic susceptibility disorders have been completed (table 1.4). A large number of genetic loci have been located, implicating a plethora of genes including IL-10 (chromosome 1), IL-3, IL-4, IL-5, IL-9, IL-13, GM-CSF, β_2 adrenergic receptor, CD14, (the 5q 31-33

cytokine cluster), HLA and TNF (chromosome 6), FcεRI-γ chain (chromosome 11), TCR α/δ and the IL-4 receptor (chromosome 16) [76, 77].

Figure 1.2 Approaches to gene identification

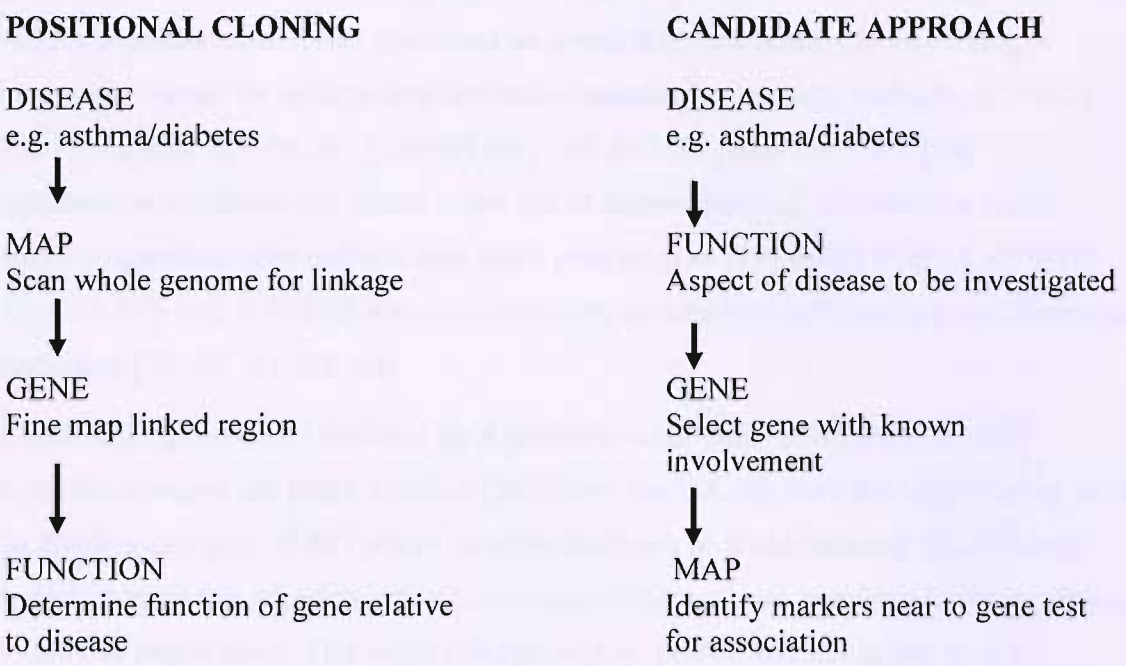


Figure 1.2 Two pathways used to identify genes involved in disease.

The positional cloning approach screens the whole genome for preferentially transmitted polymorphic markers in disease-affected individuals. Linkage analysis identifies regions of linkage. This region is then screened with a higher density of marker to reduce the size of region. Genes in the region are sequenced to identify polymorphic markers (SNPs). Association studies identify which SNPs are associated with the disease phenotype, identifying a candidate disease gene. Functional studies investigate the role of identified SNPs in the disease process. The candidate gene approach considers a disease and selects genes known to be involved in the disease pathway or are proposed as a candidates by linkage analysis. Association analysis, usually case-control design, compares frequencies of polymorphic alleles in and around the gene in affected and non affected individuals. Analysis identifies association of that marker with a given trait.

Table 1.4 summarises genetic loci that contribute to asthma and/or atopy susceptibility using a linkage-based approach. Three genome scans for asthma have been undertaken in genetically isolated populations [74, 78] identifying multiple regions of the genome confirming the heterogeneity of asthma and allergy. Recently six asthma genes have been identified as a result of positional cloning using a genome-wide scan for allergy and asthma disease phenotypes, namely *ADAM33* [79], *PHF11* [80], *DPP10* [81], *GPRA* [82] , *HLA*-G [83] and *CYF1PZ* [84] (chromosome positions are listed in the list of abbreviations). *ADAM33* is most strongly associated with asthma and BHR phenotypes [79] whilst *PHF11*, *DPP10*, *GPRA*, *HLA*-G and *CYF1PZ* are predominantly associated with allergic and immune phenotypes [78, 80, 81, 83, 84].

The *ADAM33* gene was identified by a genome-wide scan conducted on 460 Caucasian affected sib pairs families (362 from the UK, 98 from the USA) using multi-point linkage analysis of 401 micro satellite markers at 9 cM spacing. Each family included at least two siblings with doctor-diagnosed asthma and who were receiving anti-asthma medication. The initial linkage signal, performed using the broad phenotype 'asthma', resulted in two peaks at 20p13 (LOD score 2.94 in both cases). At both locations the probability that two affected siblings shared two alleles identical by descent (IBD) was 31% compared to the null expectation of 25%. More homogenous phenotypes (asthma plus BHR, asthma plus IgE, asthma plus specific IgE) were used to further characterise the linkage signal. Despite a reduced sample size (218 nuclear families), the asthma plus BHR phenotype resulted in an increase in evidence for linkage over the second peak (LOD score 3.93 with 35% excess allele sharing). Asthma plus total IgE (274 families) and asthma plus specific IgE (288 families) also supported the area under the second peak. The proportion of IBD suggested the asthma plus BHR phenotype contributed most strongly to the linkage signal. Gene discovery efforts then focused on a 1-LOD drop support interval (90% confidence interval) around the linkage signal for the asthma plus BHR phenotype. This region spanned 4.28 centimorgans (cM). A 2.5 megabase (Mb) physical map was developed across the region and BAC clones sequenced to facilitate gene identification. Forty genes were identified and were prioritised for association studies based on potential function, tissue expression and location with respect to the peak

LOD score for BHR. Association studies were then performed using case control study design. Unrelated affected offspring, showing evidence for linkage, were selected as cases, whilst controls were selected from subjects with no asthma symptoms ever, and with no family history of asthma. Selecting the case-controls in this way increases the potential differences between the groups, so intensifying any association signal. Analyses of 135 SNPs across 23 genes spanning the 90% confidence interval revealed *ADAM33* as having the strongest association signal in the linkage region. LD studies of the five genes flanking *ADAM33* showed that LD extended up to 130 kb to the left and 40 kb to the right of *ADAM33*. Within this 185 kb LD interval, twenty four SNPs, in three genes (*ADAM33*, GDNF family receptor alpha 4 (*GFRA4*) and sialic acid binding Ig-like lectin 1, (or sialoadhesin, *SIGLEC1, Sn*)), were significant in the association analysis. Fourteen of the twenty four associated SNPs were in *ADAM33*. This positive association of *ADAM33* polymorphism in the case-control study was confirmed by analysis using a family based test of association (TDT) thus excluding confounding as a factor.

A second asthma gene, Plant Homeodomain Finger 11 (*PHF11*), on Chromosome 13q14, was identified as accounting for a proportion of variation in total serum IgE levels and to be associated with severe asthma [80]. A genome-wide scan identified a region of linkage on chromosome 13q14 to serum IgE levels, with confirmation of this linkage in another cohort [85]. The third asthma gene identified using positional cloning was Dipeptidyl Peptidase 10 (*DPP10*), located on chromosome 2q14 [81]. An interval between 2q14 and 2q32 have been linked to asthma and related phenotypes by four separate reports [86-89] (Table 1.4). The fourth gene is G-protein coupled receptor for asthma susceptibility (*GPRA*), found on chromosome 7p14, and associated with asthma and high total serum IgE levels in two Finnish populations and a Canadian population. Findings concerning this gene have not been replicated or refuted so far [82, 90] The fifth gene, Human Leukocyte Antigen-G (*HLA-G*), located on chromosome 6p21, was associated with asthma and BHR. Linkage of asthma and related phenotypes in this region have been reported in seven separate genome scans making it the most replicated region of the genome [83]. Finally, the sixth gene CYF1PZ (chr 5q33) was associated, by means of a family association test, with childhood atopic asthma and may be important in differentiation of T cells [84].

This work concentrates on *ADAM33*, a novel gene implicated in lung remodelling and contributing to BHR, so *PHF11*, *DPP10*, *GPRA*, *HLA-G* and *CYF1PZ* will not be discussed further. *ADAM33's* role in the pathophysiology of asthma still needs to be elucidated, but ultimately this may lead to development of new and more effective therapies.

Table 1.4 Summary of regions highlighted in genome screens for Atopy and Allergic Disease Phenotypes

*Chr and Position	Phenotype	Population	Author/Year
1p21-1p22	Asthma (strict/loose) and Atopy[a]	US Hutterites	Ober et al 1998 [74]
1p3132	Asthma[b]	France	Dizier et al 2000 [91]
1p	Asthma[c]	US Hispanics	Xu et al 2000 [92]
1q21	Atopic dermatitis	UK	Cookson et al 2001 [93]
1q25	Atopy and asthma[d]	China	Xu et al 2001 [94]
2pter	Asthma, BHR, IgE & RAST	Germany	Wjst et al 1999 [89]
2p25	Atopy and asthma[d]	China	Xu et al 2001 [94]
2q14	Atopic asthma	Australia and UK	[h]Allen et al 2003 [81]
2q21-q23	Der p specific IgE	US Caucasians	CSGA 1998 [87]
2q33	Asthma	US Hispanics	CSGA 1997 [95]
2q	Atopy & IgE	Holland	Koppelman et al 2002 [88]
3p24.2-22	Asthma symptoms	US Hutterites	Ober et al 1998 [74]
3q21	Atopic dermatitis	Europe	Lee et al 2000 [96]
4p15.3	Atopy and asthma[d]	China	Xu et al 2001 [94]
4q24-27	Allergic Rhinitis	Denmark	Haagerup et al 2001 [97]
4q35	Asthma[e]	Australia	Daniels et al 1996 [86]
4q35	Mite sensitive atopic asthma[f]	Japan	Yokouchi et al 2000 [98]
5p15	Asthma	US African americans	CSGA 1997 [95]
5q23-31	Asthma	US Caucasians	CSGA 1997 [95]
5q31-33	Asthma, BHR & symptoms	US Hutterites	Ober et al 1998 [74]
5q31[g]	Asthma[c]	US Total	Xu et al 2001 [92]
5q31	IgE	Holland	Xu et al 2000 [99]
5q31-33	Mite sensitive atopic asthma[f]	Japan	Yokouchi et al 2000 [98]
5q33	Mite sensitive atopic asthma[f]	Japan	[h]Noguchi et al 2005 [84]
6p21	Asthma[c]	US Caucasions	CSGA 1997 [95] & Xu et al 2001 [92]
6p21	Atopy[a]	US Hutterites	Ober et al 2000 [100]
6p21	Asthma	US Caucasian and Hutterites	[h]Nicolae et al 2005 [83]
6p21.3	Asthma, IgE, RAST& eo	Germany	Wjst et al 1999 [89]
6p22-p21.3	Mite sensitive atopic asthma	Japan	Yokouchi et al 2000 [98]
6p	Atopy & IgE	Holland	Koppelman et al 2002 [88]
6q16.3 - 6q25.2	Asthma[e]	Australia	Daniels et al 1996 [86]

Table 1.4 Summary of Genome scans for Atopy and Allergic Disease Phenotypes (continued)

*Chr and Position	Phenotype	Population	Author/Year
7p21-7p15	Asthma[e]	Australia	Daniels et al 1996 [86]
7p14-15	Asthma, IgE	Finland	Laitinen 2001 [78]
7p14	Asthma, IgE	Finland	[h]Laitinen 2004 [82]
7q	Total and specific IgE	Holland	Xu et al 2000 [99]
7q	Atopy & IgE	Holland	Koppelman et al 2002 [88]
8p23[g]	Asthma	US Total	Xu et al 2001 [92]
8p23-21	Der p specific IgE	US African American	CSGA1998 [87]
9p	Asthma, IgE & RAST	Germany	Wjst et al 1999 [89]
10p25	Atopy and asthma[d]	China	Xu et al 2001 [94]
11p13	Asthma[b]	France	Dizier et al 2000 [91]
11p15	Asthma	US Caucasian	CSGA 1997 [95]
11q13	Asthma[e]	UK & Australia	Daniels et al 1996 [86]
11q13	Asthma[b]	France	Dizier et al 2000 [91]
11q21	Asthma[c]	US African Americans	Xu et al 2001 [92]
11q	Atopy	Holland	Koppelman et al 2002 [88]
12q	Total IgE	Holland	Xu et al 2000 [99]
12q13	Asthma	Germany	Wjst et al 1999 [89]
12q 15-24.1	Asthma , BHR & symptoms	US Hutterites	Ober et al 1998 [74]
12q14-24.2	Asthma, BHR & IgE	US Caucasians and Hispanics	Xu et al 2001 [92]
12q21-q23	Mite sensitive atopic asthma, BHR	Japan	Yokouchi et al 2000 [98]
12q24	Asthma[b]	France	Dizier et al 2000 [91]
13q11	Mite sensitive atopic asthma[f]	Japan	Yokouchi et al 2000 [98]
13q12-13	IgE	Holland	Koppelman et al 2002 [88]
13q14.1-13q14.3	Asthma[e]	Australia & UK	Daniels et al 1996 [86]
13q14.1-q14.3	Mite sensitive atopic asthma, BHR	Japan	Yokouchi et al 2000 [98]
13q14	Asthma, IgE	UK & Australia	[h]Zhang et al [80]
13q21.3	Asthma	Caucasian	CSGA 1997 [95]
13q31	Asthma[b]	France	Dizier et al 2000 [91]
14q11-13	Asthma	US Caucasian	CSGA 1997 [95]
14q32	IgE and BHR	US Total	CSGA 1997 [95] and Xu et al 2001 [92]
15q13[g]	Asthma[c]	US	Xu et al [92] 2001

Table 1.4 Summary of Genome scans for Atopy and Allergic Disease Phenotypes (continued)

*Chr and Position	Phenotype	Population	Author/Year
16p12	Atopy[a]	US Hutterites	Ober et al 1998 [74]
16p12	Atopy and asthma[d]	China	Xu et al 2001 [92]
16q24.1	Asthma[e]	Australia & UK	Daniels et al 1996 [86]
17p11.1–q11.2	Asthma	US Afro-American	CSGA 1997 [95]
17q12-21	Asthma[b]	France	Dizier et al 2000 [91]
17q25	Atopic dermatitis	UK	Cookson et al 2001 [93]
17q25	Eosinophils	Holland	Koppelman et al 2002 [88]
19q13	Asthma	US Hutterites	Ober et al 1998 [74]
19q13	Asthma[b]	France	Dizier et al 2000 [91]
19q13	Asthma	Caucasian	CSGA 1997 [95]
19q13.1	Atopy and asthma[d]	China	Xu et al 2001 [94]
20p12	Atopic dermatitis & asthma	UK	Cookson et al 2001 [93]
20p13	Asthma and BHR	UK and US	[h]Van Eerdewegh et al 2002 [79]
21q21	Asthma	US Hutterites	Ober et al 1998 [74]
21q21	Asthma	US Hispanics	CSGA 1997 [95]
22q11.2	Atopy and asthma[d]	China	Xu et al 2001 [94]
22q	Atopy	Holland	Koppelman et al 2002 [88]

*Table 1.4 The results of the genome-wide screens for susceptibility genes reflect the genetic and environmental heterogeneity seen in allergic disorders. Multiple regions of the genome have been linked to varying phenotypes, with differences between cohorts recruited from different populations: this illustrates the difficulty of identifying susceptibility genes for complex genetic diseases. a Five intermediate phenotypes associated with asthma, b Asthma and four intermediate phenotypes, c Asthma intermediate phenotype, d Intermediate phenotypes (quantitative trait loci) including BHR, forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC), total IgE, eosinophils, skin-prick test wheal size, and specific IgE, e Four intermediate phenotypes associated with asthma, f House dust mite sensitive, g Significant only after conditioning for another locus (see Xu et al, 2001 [92]), h Identified by positional cloning. *Chr Chromosome BHR, Bronchial Hyperresponsiveness; CSGA, Collaborative Study on the Genetics of Asthma; RAST, Radioallergosorbant test.*

1.5 Genotyping methods

Many studies have been performed looking at polymorphic changes, with the human genome project generating huge amounts of data on polymorphism throughout the genome. Once a disease-associated gene has been located and screened for polymorphism it is necessary to identify which polymorphism(s) is important in disease aetiology. Thus there is a need for affordable, reliable, reproducible and high-throughput methodologies to screen populations for polymorphisms which may be involved in disease processes. Identification of a single base pair change within the whole genome is a difficult task. This was rectified in 1985 when Kary Mullis developed the polymerase chain reaction (PCR) [101], allowing easy copying and amplification of a specific region of DNA, thus permitting an expansion of SNP genotyping methods. For convenience the genotyping methods presented here will be divided into endonuclease cleavage, conformation analysis, hybridisation techniques (incorporating Fluorescence Resonance Energy Transfer (FRET) and allele specific PCR), and primer extension [102]. Further information of the technologies available can be found in reviews by Syvanen [102] and Kristensen *et al [103]*.

1.5.1 Endonuclease cleavage

Restriction fragment length polymorphism (RFLP) is an example of endonuclease cleavage and was one of the first methods used to examine SNPs [104]. It can be used if a SNP alters the recognition sequence for an endonuclease by elimination of an existing site or creation of a new recognition site. PCR is used to amplify target DNA and then a restriction endonuclease, specific to the SNP site, is used to identify presence or absence of the SNP. Resolution of the products by gel electrophoresis separates the molecules based on their size. The development of microarray diagonal gel electophoresis (MADGE) [105] enables a high-throughput, low-cost method of electrophoretic fragment analysis. MADGE reduces the track lengths and incorporates high-density wells in a format compatible with industrial 96-well plates. Gel preparation is simpler and quicker than that of acrylamide, and resolution higher than that of agarose gels [106]. Data can be easily analysed using software such as Phoretix®.

1.5.2 Conformation analysis

A large number of polymorphic typing strategies exploit the change in conformation or secondary structure resulting from polymorphism of the DNA sequence of a DNA fragment. Single stranded DNA has a tendency to fold up and form complex structures stabilised by weak intramolecular bonds. The electrophoretic mobility of these structures on non-denaturing gels depends on the chain length and conformation, which is dictated by DNA sequence. One method of conformation analysis, Cleavase fragment length polymorphism analysis, uses engineered, structure specific endonucleases, such as Cleavase I, to recognise and cut stem loops that form imperfect hybrids following thermal denaturing and renaturing. The series of fragments can be resolved by gel electrophoresis [107]. A similar method is single strand conformation polymorphism (SSCP). This involves the denaturing of PCR amplified fragments with the subsequent formation of sequence specific secondary and tertiary structures of the single strands during non-denaturing gel electrophoresis. The electrophoretic mobility depends on the three dimensional shape of the single stranded molecules with changes in secondary and tertiary structure identifying polymorphic changes in the DNA. Originally fragments were labelled with ^{32}P, but this has been largely replaced by fluorescent labelling [103].

Heteroduplex analysis emanates from SSCP. Most mutations occur in heterozygous form. Heteroduplexes can be formed by heating the heterozygous test-PCR product to denature it and then cooling slowly, forming DNA duplexes. Homoduplexes (same sequence on both strands) or heteroduplexes (single base pair mismatch on one strand) have different phoretic mobility on a non-denaturing gel [103, 108]. This method can be used for fragments up to 200 bp long and can detect insertions, deletions and some SNPs [109]. High performance liquid chromatography (dHPLC) and denaturing gradient gel electrophoresis (DGGE) exploit the fact that heteroduplexes have abnormal denaturing profiles. Heteroduplexes are formed as before and in both cases the mobility of a fragment changes markedly when it denatures. These methods require tailoring (i.e. by PCR amplification) to the particular DNA sequence under test and are best suited to routine analysis of a given fragment in many samples [110].

1.5.3 Hybridisation and allele specific PCR

Hybridisation techniques include oligonucleotide ligation, 5' nuclease assay (Taqman®) [111-113] and Molecular Beacons [114, 115]. Oligonucleotide ligation uses pairs of oligonucleotide probes that anneal to their target sequence adjacent to each other and have an allele specific 3' or 5' nucleotide at the probe junction. When the probes are perfectly matched to their target sequence they will be joined by a ligase, whereas a mismatch will not [116]. Modification of the probes by adding a biotin label to one and digoxygenin to another allows use of an Enzyme-linked immunosorbant assay (ELISA) based method to read the reactions. ELISA methods are carried out in microtitre plates and use colorimetric detection [102, 117].

Taqman® [111-113] was originally used for quantitative PCR analysis but has been adapted for allelic discrimination. Taqman® is based on the principle of energy transfer in which fluorescence is detected as a result of change in physical distance between a fluorophore moiety and a quencher molecule on hybridisation of the allele specific oligonucleotide (ASO) probe to a perfectly matched target sequence. Molecular Beacon [114, 115] also uses FRET technology. The probe forms a stem and loop structure which keeps the quencher and fluorophore close, thus preventing fluorescence. Complete hybridisation to a complementary sequence unfolds the probe and fluorescence occurs. Allele-specific PCR incorporates the use of two forward primers with the 3' ends complementary to one or other nucleotide of the SNP, with a common reverse primer to selectively amplify the SNP alleles [118]. Monitoring of products can be attained by incorporating a fluorescent label (e.g. molecular beacons [119]) or by size separation on a high-throughput system such as microplate array diagonal gel electrophoresis (MADGE) [105, 120]. Hybridisation and ASO assays require reaction conditions and primers designed for each specific PCR, making them best suited to analysis of a limited number of SNPs in large sample collections.

1.5.4 Primer extension

Primer extension chemistry is a simple robust method for analysing multiple SNPs in the same tube, minimising effort for assay design and optimisation. The method employs a PCR-derived target DNA containing the SNP together with 2 universally

tagged allele-specific primers whose 3' ends define the alleles. A thermophilic DNA polymerase is used for label incorporation into extended products. Because the two tagged allele-specific primers overlap the SNP site in the target DNA, only the correctly hybridized primer(s) will be extended to generate a labeled product(s). A non-complementary primer will not be extended or labeled due to the 3' mismatched base. Pyrosequencing is an adapted primer extension method that uses enzyme-mediated luminometric detection of pyrophosphate, released on incorporation of deoxynucleotide triphosphates [121, 122]. The genotype of a SNP is deduced by sequential addition and degradation of the nucleotides using apyrase in a dedicated instrument that operates a 96-well or 384-well microtitre plate format [123]. Pyrosequencing allows short 30-50 bp sequences that flank a SNP to be determined. A limitation of the method is that sequential identification of the bases prevents genotyping of several SNPs per reaction. Mass spectrometry (MALDITOF, matrix-associated laser desorption time-of-flight mass spectrometry) can also be used to detect primer extension products. Mass spectrometry is particularly useful, as primers of different lengths can be used in combination with mixtures of deoxy and dideoxynucleoside triphosphates to yield allele-specific primer extension products with non-overlapping mass distributions allowing multiplexing [102].

1.6 ADAMs

1.6.1 Family and their functions

The ADAM family of transmembrane proteins belong to the metzincin subgroup of the zinc protease superfamily. The metzincin subgroup is further divided into serralysins, astracins, and the closely related matrixins (Matrix metalloproteases (MMPs)), adamalysins (A Disintegrin And Metalloprotease (ADAMs) and A Disintegrin And Metalloprotease with Thrombospondin motif (ADAMTS) [124]. The matrixins and adamalysins are the principle agents responsible for extracellular matrix degradation and remodelling, and play important roles in development, wound healing, and the pathology of diseases such as arthritis and cancer [125]. Mechanisms of metalloprotease action underlying these events include the proteolytic cleavage of growth factors so that they can become available to cells not in direct physical contact, degradation of the ECM so that founder cells can move across tissues for

repair, and regulated receptor cleavage to terminate migratory signalling. Most of these processes require a delicate balance between the functions of MMPs, ADAMs and natural tissue inhibitors of metalloproteases (TIMPs) [125].

1.6.2 The ADAM family

Thirty four ADAM genes have been identified to date, with nineteen of these found in humans. Up to date registries of ADAM family members in different species can be found at http://www.uta.fi/~loiika/ADAMs/HADAMs.htm. Members of the ADAM family are found in vertebrates *Caenorhabditis Elegans*, *Drosophila melanogaster* and *Xenopus laevis*. ADAMs (and ADAMTSs) are similar to matrixins regarding their metalloprotease domains but contain a unique integrin receptor-binding disintegrin domain [21]. In mammals many ADAMs are expressed exclusively or predominantly in the testis and associated structures (e.g. *ADAM*s 2, 7, 18, 20, 21, 29 and 30), whilst others have a broader somatic distribution (e.g. *ADAMs* 8, 9, 10, 11, 12, 15, 17, 19, 22, 23, 28 and 33). Metalloprotease-disintegrins have been implicated in diverse functions including sperm-egg binding and fusion, myoblast fusion, protein ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins, and other extracellular protein domains. Some ADAM functions can be attributed to an individual domain (e.g. ectodomain shedding to the metalloproteinase domain [126] and cell adhesion to the disintegrin domain [127]), but it is likely that the other domains play important regulatory roles to these functions by conferring specificity and selectivity. The domain structure of the ADAMs consist of a signal sequence, a prodomain, a catalytic (metalloprotease) domain, a disintegrin domain, a cysteine-rich domain, an EGF-like domain, a transmembrane domain and a cytoplasmic tail (Figure 1.3).

Figure 1.3 The exonic structure, domain organisation and SNP location of human ADAM33 in relation to function (adapted from Holloway et al (2004)

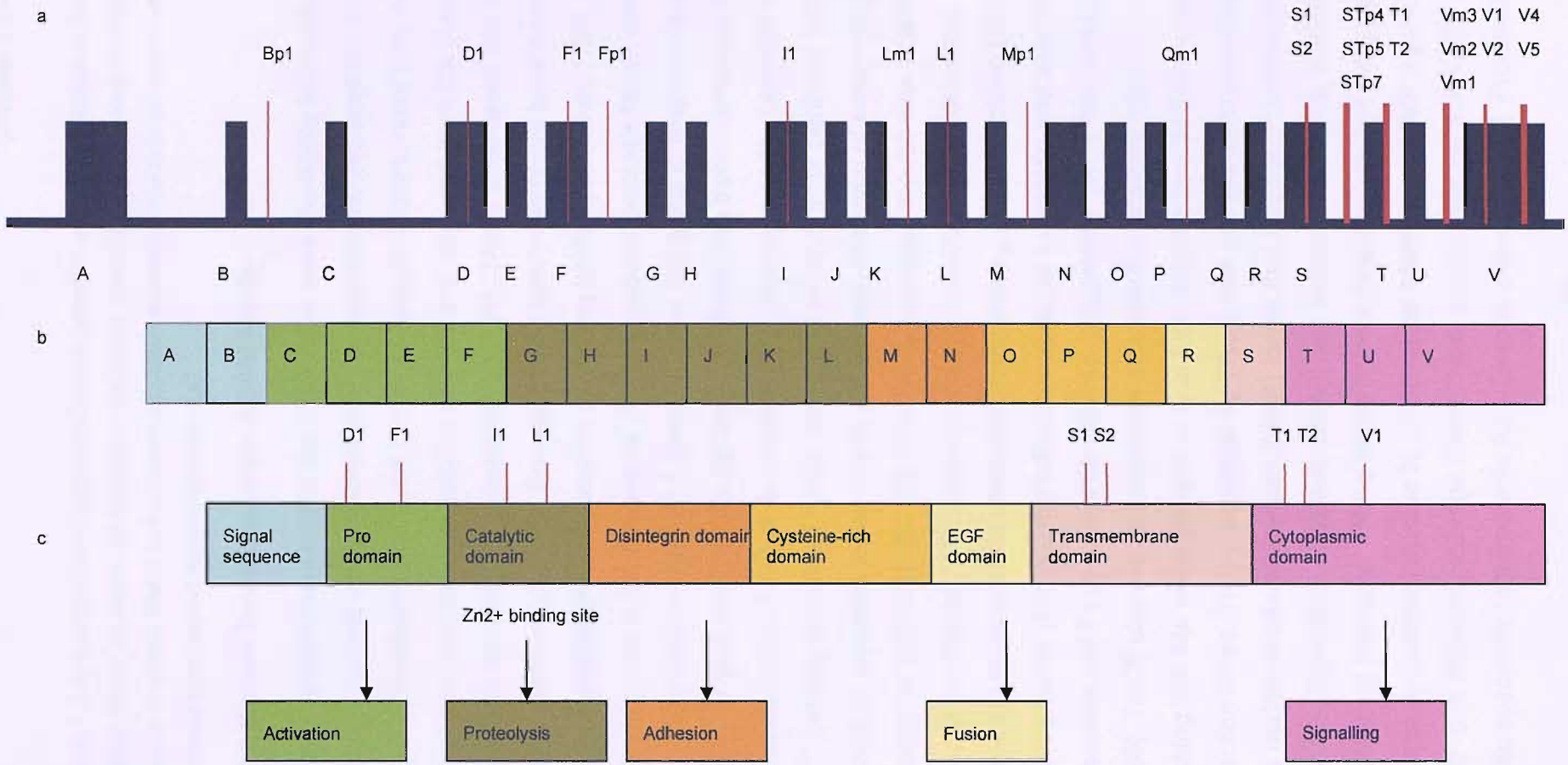


Figure 1.3 a) illustrates the intronic and exonic structure of ADAM33 including the position of the most studied SNPs. b) shows the position of the exons in relation to the protein domains c) shows the position of exonic SNPs and allocates a putative function to the domain based on experimental evidence from related ADAMs (section 1.6).

1.6.3 Known functions of the ADAM family

During development and in the adult, cells have the ability to modify their surface to regulate various kinds of functions. For example the extracellular domain of over 40 plasma membrane anchored cytokines, growth factors, receptors, adhesion molecules and enzymes can be cleaved and released from the plasma membrane by proteases (sheddases or secretases) [128]. These sheddases are themselves transmembrane proteins. Several ADAM family members have been shown to have active metalloprotease domains displaying proteolytic activity [129]. One of the best-studied cases is the release of TNF-α by tumour necrosis factor-alpha-converting enzyme (TACE or ADAM17) from various cell types in response to injury or infection. ADAM17 was identified for its ability to cleave the transmembrane form of tumour necrosis factor (TNF) at its physiological processing site. Targeted mutation of the Zn^{2+} binding domain of TACE resulted in embryonic or perinatal lethality in mice [130] providing evidence for its role in ectodomain shedding of essential growth factors like ligands for EGFR [130-132]. The TACE knockout mouse was unable to shed TNF-α. Loss of functional TACE caused embryonic lethality in most mutant fetuses as a result of failure to shed embryonic transforming growth factor-α (TGF-α). In addition TNF receptor (TNFR), the adhesion molecule L-selectin and amyloid protein precursor (APP) could not be cleaved, suggesting TACE has multiple substrates [130]. TACE cleaves its substrates at a specific cleavage site [133]. Substrates lacking this site were cleaved only at high enzyme concentrations and prolonged reaction times. TNF-α cleavage by TACE was compared to cleavage by members of the matrix metalloproteinase (MMP) family that had been previously implied in TNF-α release. Specificity constants were much reduced for MMPs. For example, MMP 7 processed precursor TNF-α at the correct cleavage site but did so with a 30-fold lower specificity constant relative to TACE whilst MMP 1 and MMP 9 processed precursor TNF-α at different cleavage sites. These results indicate that TACE is unique in terms of its specificity requirements for substrate cleavage [133]. Processing of TNF-α can

still occur in TACE-deficient mice, but residual sheddase activity is inhibited by the metalloprotease inhibitor IC-3, suggesting that other family members may process TNF-α [134].

Sheddase activity by ADAMs 9, 10, 12 and 17 has been observed on epidermal growth factor receptor (EGFR) ligands [133, 135]. ADAM9 also cleaves insulin growth factor binding protein 3 and 5 as well as fibronectin [127, 136]. ADAM10 has been shown to be a control point in the relay between G-protein coupled receptors (GPCR) and the epidermal growth factor receptor (EGFR) by acting as a sheddase for heparin-binding epidermal growth factor (HB-EGF). EGFR transactivation occurs not only in the same cell where both ADAM10 and EGFR reside, but also affects the neighbouring cells that sense released HB-EGF [137]. ADAM12 (meltrin α) is a catalytically active metalloprotease-disintegrin [138] but other functions include binding of cell surface proteins such as syndecans via the cysteine-rich domain affecting cell adhesion and signalling events [139]. ADAM12 and 17 are both up-regulated in smooth muscle lesions of lymphangioleiomyomatosis (LAM), a rare lung disease affecting young women [140] implying a role in adhesion, cell migration and possibly myogenic proliferation. ADAM23 does not have an active metalloprotease domain, suggesting it may be exclusively involved in the cell adhesion process [141]. The roles of fertilin (ADAM2) and cyritestin (ADAM3) during fertilisation further illustrate the importance of ADAM proteins in cell-cell fusion and adhesion [142-144].

ADAM proteins are anchored in the trans-Golgi network or plasma membrane [145], and several secreted splice variants have been identified [146]. For example, ADAM12 has two splice forms, one (L) produces a membrane-bound protein, the other (S) a secreted form [147]. The spliced secreted form of ADAM12 (ADAM12-S) provokes myogenesis in a nude mouse tumour model [147] as well as displaying protein cleavage activity [148]. Alternate splicing of ADAM28 yields isoforms with different subcellular localisation patterns and tissue expression. Murine ADAM28 may have three forms, two larger ones predicted to encode membrane-anchored proteins and expressed in the epididymis and lung, and a smaller one predicted to encode a secreted protein with testis specific expression [149]. Human ADAM28 has two isoforms. The secreted form is preferentially expressed in spleen and the membrane

bound form is lymph node specific [150]. ADAM9 and ADAM10 also have alternatively spliced variants with both secreted and membrane-associated forms [151, 152] as do ADAM11 [153] and ADAM33 [154]. Signalling roles are displayed by ADAM10, a key regulator of Notch signalling pathways in Drosophila [155]. The cytoplasmic tail of ADAM12 interacts with Src family proteins [156] and the cytoplasmic tail of ADAM19 binds to ArgBP1 [156, 157]. Altered ADAM activity caused by hypermethylation of the promoter region has been observed in high grade breast cancers [158], pancreatic cancers [159] and gastric cancers [160] again confirming important roles in cell adhesion and migration. Many ADAMs have no ascribed function to date.

1.6.4 ADAM33 protein structure and function

The selective expression of ADAM33 in mesenchymal cells of asthmatic airways strongly suggests that alterations in its activity might underlie abnormalities in airway smooth muscle cells and fibroblasts. Its function to date is unknown but the absence of the MP domain in many isoforms suggests that some of the functions in smooth muscle and mesenchymal cells may be linked to other domains [154, 161]. Expression studies of ADAM33, characterisation of its isoforms and known activities of relatives suggests roles including cell adhesion, fusion, migration and cell signalling [139-142, 155-157, 162] (section 1.6.3). Epigenetic regulation of the gene via methylation is likely [158-160]. Human ADAM33 (hADAM33) protein is 78% identical to murine ADAM33 (mADAM33) with particularly high homology in the extracellular region containing the metalloprotease, disintegrin, cysteine-rich and EGF-like domains [163]. It is also closely related to hADAMs 8, 12, 15, 19 and Xenopus ADAM13 [79]. ADAM33 is 812 amino acids long. The protein has eight domains in total. These are signal sequence, pro-domain, catalytic (metalloprotease) domain, disintegrin-like domain, cysteine-rich, EGF-like domain, transmembrane domain (anchoring the molecule to membranes at the cell surface or Golgi), and a cytoplasmic domain with intracellular signalling sequences [129].

The N terminus of all ADAMs contains a signal sequence that directs ADAMs into the secretory pathway and a pro-domain that functions in ADAM maturation, regulating the activity of the catalytic (metalloprotease) domain via a cysteine switch

mechanism. The cysteine switch mechanism involves formation of an intramolecular complex between the single free cysteine residue in the pro-peptide domain and the essential zinc atom in the catalytic domain [164]. Expression of the full length ADAM33 in mammalian cell lines has shown that the metalloprotease domain of ADAM33 is functional [165] but the biological targets of the metalloprotease activity remains unknown. Zhou *et al* examined the capacity of ADAM33 to cleave a wide range of substrates (39 in total) and only 4 were cleaved, namely stem cell factor (SCF, c-kit), β-amyloid precursor protein (β APP), TNF-related activation-induced cytokine (TRANCE) and insulin B. The kinetics suggested these may not be the natural substrates [166]. Other studies of ADAM33 show very low expression level of transcripts with a functional pro- and catalytic domain in airway fibroblasts (<2%) [167]. Full-length clones of ADAM33 have been identified in lung mesenchymal cells (expressing high levels of ADAM33 transcript), but the great majority of transcripts encode predicted proteins that lack this activity [154]. This suggests that the ADAM33 domains downstream of the catalytic unit have distinct functions independent of the pro-/catalytic domain and that the contribution of ADAM33 to BHR may have little to do with its ability to cleave protein substrates. Alternatively, the potency of the transcript containing the metalloprotease domain may be such that only minute concentrations are required for activity.

The disintegrin domain is named for its presence in the snake venom metalloproteases (SVMPs), in which it is involved in the binding of platelet integrin receptors. This prevents the association of platelets with natural ligands such as fibrinogen and results in a block in platelet aggregation at the wound site, and breakdown in basement membrane components leading to haemorrhage. It has been shown that the isolated disintegrin domain of ADAM33 supports $\alpha_9\beta_1$integrin-dependent cell adhesion of leukocytes [168] supporting a hypothesis of cell adhesion activity. The cysteine-rich and EGF domains are poorly understood, but have been linked to cell adhesion and membrane fusion events respectively [139, 147]. One isoform of ADAM33 does not have any means of cellular attachment (ADAM33-S). This isoform could display similar actions to its close relative ADAM12-S which induces myogenesis when expressed in rhabdomyosarcoma cells [147]. It could be that ADAM33-S has a role in the increase in smooth muscle observed in the

asthmatic lung. Indeed, if the domains downstream of the catalytic domain were found to influence the migration of smooth muscle cells, this might provide an important functional link between ADAM33 and BHR. Downstream of the transmembrane domain, anchoring the protein is the cytoplasmic tail. ADAM33 has a relatively short cytoplasmic tail when compared to other ADAMs [169]. It is proline-rich and contains an SH3 binding site [169], a casein kinase F/II phosphorylation site and a MAPK (mitogen activated protein kinase) consensus sequence all of which may be important in ADAM33 intra-cellular localisation, protein trafficking or function [169]. The 3' portion of the gene encoding the transmembrane and cytoplasmic domains and 3' UTR is one of the key regions linked to asthma and BHR [79].

1.6.5 Processing of ADAM33

Processing and activation of ADAM33 occurs in the trans Golgi network [165] Localisation experiments on ADAM33 revealed most protein was localised in the endoplasmic reticulum and proximal Golgi with a small amount of protein converted into mature ADAM33 at the cell surface [165]. A similar mechanism is observed in ADAM12 whose cytoplasmic tail contains a retention signal, keeping processed ADAM12 in the endoplasmic reticulum (ER) until required [145, 170].

1.6.6 Expression of ADAM33

In situ hybridisation using antisense ADAM33 probes and quantitative RT- PCR has shown ADAM33 to be preferentially expressed in tissues of mesenchymal origin such as fibroblasts, myofibroblasts and smooth muscle with an absence of expression in non-mesenchymal haematopoietic cells (T cells, inflammatory leukocytes), epithelial and endothelial cells [79, 171]. ADAM 33 is expressed in smooth muscle bundles and around embryonic bronchi, strongly suggesting an important role in smooth muscle development and function [161]. A role in development is possible as the expression in undifferentiated mesenchymal cells is higher than that of α-SMA positive cells and development studies in mice show ADAM33 expression increases during lung embryogenesis [172]. ADAM33 protein identified in bronchoalveolar lavage fluid (BAL) [173] is thought to derive from alveolar macrophages [174] or activated fibroblasts [175]. Its absence from inflammatory cells and presence in mesenchymal

cells [79, 171] makes it a strong contender as the 'remodelling' gene irrespective of allergic status. It could therefore be a true asthma, rather than atopy, gene.

1.6.7 ADAM33 splice variants

Northern blot analysis has identified two transcripts of the ADAM33 gene of 5.0 and 3.5 kb, but only the latter has been found in cytoplasmic RNA [79]. A transcript of 7-8 kb has also been noted in tissues expressing ADAM33 at higher levels including uterus, small intestine and heart [171]. The 3.5 kb clone contains the entire open reading frame comprising 22 exons with the canonical sequence (CT/AG) present at each splice junction with no polymorphic changes in the immediate vicinity [79]. Of the multiple splice variants identified, three use splice junctions that are predicted by the full-length clone. These clones encode predicted proteins that are in-frame with the wild-type translation and possess both a secretion signal and a trans-membrane anchor. The Alpha (α) and Beta (β) isoforms were cloned in conjunction with the identification of the mouse and human ADAM33 gene [163]. The skipping of exon Q gives rise to the β-form of ADAM33, which is present in approximately 25% of transcripts in airway mesenchymal cells. This variant has also been detected in testes, brain and placenta and its biological relevance remains unknown. Since exon Q spans both the cysteine-rich and EGF domains, its absence could lead to the disruption of function of both domains. Partial removal of exon R (EGF domain) results in the deletion of 37 bases, resulting in a frame-shift in the resulting transcript predicting a protein that does not appear to have any means of cellular attachment (ADAM33-S). In keeping with other soluble ADAM/ADAM-TS proteins, the level of mRNA expression of this transcript is very low [154].

In situ hybridisation studies by Laporte et al [176] compared expression of ADAM33 in asthmatics and non-asthmatics. Results showed that in asthmatic biopsies, expression of both α and β isoforms of *ADAM33* mRNA was increased when compared to normal tissue, with localisation to the smooth muscle [176]. However overall expression of *ADAM33* and relative abundance of isoforms was not significantly different between asthmatics and non-asthmatics, suggesting regulation of *ADAM33* isoform levels is unlikely to account for its role in asthma [161]. Analysis of the ADAM33 transcript reveals a lack of 3' UTR AU-rich elements. These are often

involved in protein/mRNA binding [171] that can influence mRNA stability/degradation, [177], control nuclear export and cellular localisation and also influence translational efficiency [178]. Umland et al have shown that the 3'UTR of *ADAM33* mRNA is necessary for processing of the mature peptide (zymogen), specifically for removal of the prodomain [171]. Domains downstream of the catalytic domain were also implicated as influencing the efficiency of pro-domain removal. In these studies, post-translational processing, but not expression of *ADAM33*, was affected by the 3'UTR region, indicating that it is able to affect sub-cellular localisation rather than translational efficiency or mRNA degradation, since the process of pro-domain removal from ADAMs occurs in the Golgi apparatus [165, 171].

1.6.8 Genetic variation in *ADAM33*

ADAM33 is a highly polymorphic gene with over 100 SNPs, though not all have been experimentally verified. The original study by Van Eerdewegh *et al* [79] found association between *ADAM33* SNPs and the asthma/BHR phenotype. This was corroborated by the existence of a syntenic region on mouse chromosome 2 at 74 cM that has been linked to BHR [179] and overlying an ortholog of *ADAM33* that exhibits ca. 70% homology with its human counterpart [163, 180]. Replication of the association between *ADAM33* SNPs and asthma has been seen in adult studies of African American, Hispanics, and US Caucasian outbred populations as well as Dutch [181], two German populations [182], and a Korean population [183]. In addition, a study by Simpson et al showed an association between *ADAM33* SNPs and reduced lung function in children aged three and five [184]. Two further studies in Dutch populations identified association with *ADAM33* and accelerated decline in lung function [37, 38]. Although significant associations were observed with asthma susceptibility in all these populations, no individual SNP was consistently significant in all groups. However, a meta-analysis by Blakey et al suggested Fp1 and STp7 variants were significantly associated with asthma. Given the incidence of asthma in the population and the frequency of these polymorphisms in the general population the additional risk imparted by these variations could account for 50,000 excess asthma cases in the UK alone [185]. Haplotype analysis revealed susceptibility haplotypes which again were not consistent between studies [38, 79, 181-183]. A

further study found association with *ADAM33* SNPs and allergic rhinitis in a Japanese population [186]. Three studies have failed to replicate the original findings [185, 187, 188]. *ADAM33* SNPs associated with asthma and allergy are summarised in table 1.5

Table 1.5 Summary of the ADAM33 single nucleotide polymorphisms (SNPs) associated with asthma phenotypes and allergy

Study	SNPs [a]	Associated phenotype
Van Eerdewegh et al [79] Case-control		
UK (n=362 sib pairs)	Fp1, Qm1, S1, S2, STp4, Vm1, V4	Asthma
US (n=98 sib pairs)	I1, Lm1, Mp1, T1, T2, Tp1	BHR
Combined (US + UK)	Qm1, S1, STp4, STp7, Vm1, V4	
Howard et al [181] Case-control		
African Americans (n=160) [b]	S2, STp4, V4	Asthma
White (n=219)	S2, T1, T2	BHR
Hispanic (n=112)	S2, STp4, T1, T2, Vm1	
Dutch (N=153)	S1, S2, STp7, V4	
Werner et al [182]		
German families (n=732)	Fp1, STp4, STp5	Asthma
Case-control (n=590)	STp7	BHR
Simpson et al [184]		
Low lung function: aged 3 (n=302) [c]	Fp1	Reduced lung function in childhood
Low lung function: aged 5 (n=504)	Fp1, Mp1,S1, STp5, T1, T2, V4	
Lee et al [183]		
Korean population Case (n=326)-control (n=151)	T1	BHR
Cheng et al [186]		
Japanese population Case (n=95)–control (n= 95)	Fp1, Lm1, S2, T1, T2, Tp1	Allergic rhinitis
Jongepier et al [37]		
Dutch population: decline in lung function in asthmatics (n=200)	S2, T1, T2, V4	Decline in lung function
Van Diemen [38]		
Dutch population: decline in lung function	Fp1, Qm1, S1,S2, T2	Decline in lung function
in general population (n=1390)		

Table 1.5 No single SNP was found to be consistently associated between populations. [a]SNP designations are derived from exon labeling (see fig 1.5) [b]Number of cases matched by similar number of controls [c]Children born of atopic and asthmatic parents.

1.7 Aim of research and hypothesis

For most patients with asthma, the disease begins prior to six years of age [189, 190]. There is evidence that processes involved in the pathogenesis of asthma begins in early childhood [191, 192], or even *in utero* [193-195] and that abnormalities in childhood lung function are associated with respiratory disease in later life [196, 197]. While asthma is a disorder of the conducting airways characterised by Th_2 inflammation, a second set of mechanisms, termed remodelling, is being increasingly recognised as fundamental to disease chronicity and severity. Asthma is a complex disease resulting from interactions between environmental and multiple small-effect genetic factors. The identification of *ADAM33* as a susceptibility gene primarily associated with BHR, together with evidence presented in the preceding sections suggests *ADAM33* may play a role in lung morphogenesis, cell adhesion, and cell migration, ultimately affecting remodelling. *ADAM33* SNPs have been associated with BHR suggesting polymorphic change is instrumental to change in *ADAM33* function.

The first part of this thesis seeks to investigate the relationship between *ADAM33* polymorphism and early life lung function. The hypothesis is that polymorphisms of *ADAM33* are associated with reduced measures of lung function in early life before the development of clinical disease. In other words, that the polymorphism affects fetal airway development thereby increasing the risk of subsequent BHR and asthma (chapters 3 and 4).

The second part of this work will concentrate on regulation of *ADAM33* at the transcriptional level (chapter 5). Regulatory sequences, called promoters and enhancers, are necessary for controlling where and when the gene is transcribed. Transcription is the primary means by which gene expression is controlled to produce different proteins in different cell types or in response to different stimuli. This process is controlled by transcription factors that activate or repress expression of a specific gene. The second part of this work seeks to identify the core promoter region of *ADAM33* and identify which transcription factors are important for *ADAM33* transcription. I will also investigate how *ADAM33* transcription is up or down regulated in response to inflammatory and other changes that occur in asthma.

Chapter 2

General Materials and Methods

2.1 Materials

Stock solutions were stored at room temperature unless otherwise stated. All chemicals and reagents were of molecular biology or biotechnology grade.

Table 2.1 Reagents and media

Reagent	Supplier
Absolute Ethanol	Hayman, England
Agar granules solidifying agent	Fisher Scientific, Leicestershire, UK
Agarose	Eurogentec, Hampshire, UK
Ammonium persulphate (APS)	Sigma-Aldrich, Dorset, UK
Ampicillin	Sigma-Aldrich, Dorset, UK
Big dye terminators version 3 (stored at -20°C)	ABI Cheshire, UK
Calcium chloride	Sigma-Aldrich, Dorset, UK
Chloroform	Sigma-Aldrich, Dorset, UK
Collagen IV (stored at 4°C)	Invitrogen, Paisley, UK
Deoxyribonucleotide triphosphate (dNTP) 100mM (stored at -20°C)	Promega, Southampton, UK
Diethyl pyrocarbonate (DEPC)	Ambion, Cambridgeshire, UK
Dimethyl Sulphoxide (DMSO)	Sigma-Aldrich, Dorset, UK
Disodium Ethylene Diamine Tetra-acetate, pH 8.0 (EDTA) (Disodium salt: dehydrate)	Sigma-Aldrich, Dorset, UK
Dulbeccos Modified Eagles Medium (DMEM) without L-glutamine (stored at 4°C)	Cambrex, Nottingham, UK

Table 2.1 Reagents and media (continued)

Reagent	Supplier
Eagles Minimum Essential Medium (EMEM) without L-glutamine (stored at 4°C)	Cambrex, Nottingham, UK
Fetal Calf Serum (stored at -20°C)	Invitrogen, Paisley, UK
Gene page plus 5% 6 M urea (stored at 4°C)	Amresco, Ohio,USA
Glycerol	Sigma-Aldrich, Dorset, UK
Half sequencing buffer (2.5x) (stored at -20°C)	ABI Cheshire, UK
Hanks Balanced Salt Solution (HBSS) (stored at 4°C)	Cambrex, Nottingham, UK
Hydrochloric acid (12 M)	Sigma-Aldrich, Dorset, UK
Isopropanol	Sigma-Aldrich, Dorset, UK
Kanamycin	Sigma-Aldrich, Dorset, UK
L-glutamine (stored at -20°C)	Invitrogen, Paisley, UK
Luria broth (10 g tryptone, 5 g yeast extract, 5 g sodium chloride in 1 litre)	Fisher Scientific, Leicestershire, UK
Magnesium Chloride, 25 mM or 50 mM (stored at -20°C)	Promega, Southampton, UK
N,N,N',N'-tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Dorset, UK
Non Essential Amino Acids (NEAA) (stored at 4°C)	Invitrogen, Paisley, UK
Penicillin / streptomycin (stored at -20°C)	Invitrogen, Paisley, UK
Phosphate buffered saline (PBS) tablets	Oxoid, Basingstoke, Hampshire, UK
*Quantitative PCR mix (2x) (stored at -20°C)	Eurogentec, Hampshire, UK
Roswell Park Memorial Institute (RPMI-1640) without L-glutamine (stored at 4°C)	Cambrex, Nottingham, UK
Smart ladder molecular marker (stored at 4°C)	Eurogentec, Hampshire, UK
SOC medium	Invitrogen, Paisley, UK
Sodium pyruvate (NaPyr) (stored at 4°C)	Invitrogen, Paisley, UK
*Taqman® universal PCR (2x) mastermix (stored at 4°C)	Applied biosystems (ABI) Cheshire, UK
Tris (pH 8.8)	Anachem, Bedfordshire, UK
Tris- Borate- EDTA (TBE) buffer (10x)	National Diagnostics, Yorkshire, UK
TRIzol reagent	Invitrogen, Paisley, UK
Trypan Blue	Sigma-Aldrich, Dorset, UK
Trypsin	Invitrogen, Paisley, UK
Vitrogen 100	Celltrix Laboratories, USA

*Table 2.1 *Comparison experiments show these products to be comparable.*

Table 2.2 Enzymes and enzyme buffers

Enzyme	Supplier
DNase (1U/µl) and DNase buffer	Ambion, Cambridgeshire, UK
Exonuclease I (20000U/ml)	New England Biolabs, Herts, UK
Maloney murine leukaemia virus (M-MLV) reverse transcriptase (100U/µl) and transcription buffer	Ambion, Cambridgeshire, UK
New England Buffer 2 (NEB2)	New England Biolabs, Herts, UK
New England Buffer 4 (NEB4)	New England Biolabs, Herts, UK
Pwo (*Pyrococcus woesei*) DNA polymerase (5u/µl)	Roche Diagnostics, East Sussex, UK
Restriction endonuclease Bgl II (10000U/ml)	New England Biolabs, Herts, UK
Restriction endonuclease Hph I (5000U.ml)	New England Biolabs, Herts, UK
Restriction endonuclease Kpn I (10000U/ml)	New England Biolabs, Herts, UK
Shrimp alkaline phosphatase (SAP) (1U/µl) and SAP buffer	Promega, Southampton, UK
T4 DNA Ligase (400,000 units/ml) and ligation buffer	New England Biolabs, Herts, UK
Taq (*Thermus aquaticus*) DNA polymerase (5U/µl) and 10X buffer without magnesium	Promega, Southampton, UK

Table 2.2 All enzymes and their buffers were stored at stored at -20°C.

Table 2.3 Taqman® probes for 5' nuclease allelic discrimination assays

Assay ID	VIC Sequence MGB-NFQ	6-Fam Sequence MGB-NFQ
Bp1	AGGCTAGGTgTCCCCT	AGGCTAGGTaTCCCCT
D1	TGGGCCGTaGTGGG	GGGCCGTgGTGG
F1	AAAGGAaCCTGTGGCC	AAAGGAgCCTGTGGC
FP1	AGAAGAGACgGGAATT	AGAGACaGGAATTCA
I1	CACAGcCCCCGGC	CCCACAGtCCCCG
Lm1	AGTCCcCCAACCC	CAGTCCtCCAACCC
L1	AGAAGgCGCGCAGC	CGGAAGAAGaCGCG
MP1	AAGCTCCcCGAAAC	AAGCTCCaCGAAAC
Qm1	CCCGACcCCTCCT	CCGACtCCTCCTT
S1	CTCAGCCgTCCTGCT	TCCTCAGCaTCCTG
S2	TCCCAGGgGCCGGC	CCCAGGcGCCGGC
STp4	TCTGTCACCCaCTTGA	TCTGTCACCCcCTTGA
STP5	AGCACCgCAGCTG	AGCACCaCAGCTGA
STp7	CCATCCCATCcCC	CCATCCCATCtCC
T1	CCCCAtGGAGTTGG	CCCCAcGGAGTTG
T2	CTGGACAGtCCTGG	CACTGGACAGcCCT
TP2	AGCAGGAGaCCCT	AGCAGGAGcCCCT
Vm3	AGCCCAcCCTCACT	AGCCCAtCCTCACTC
Vm1	CCCCTAAgGACT	TCCCCTAAtGACT
Vm2	TGGGCTCcCTGCCA	TGGGCTCtCTGCCAC
V1	CAGCAGATCAtGTCC	AGCAGATCAaGTCC
V2	AGCTTcGACCCACC	TGGAGCTTtGACCCAC
V4	CTCCCCTGcAGCCT	CCCTGgAGCCT
V5	TGGCCTCTGCAAaCA	TCTGCAAgCAAACAT

Table 2.3 All probes were supplied by Applied Biosystems and contained a fluorescent marker (Vic or 6-Fam) at the 5' end. The 3' end has attached a minor groove binder (MGB) and non fluorescent quencher. Position of the polymorphic base is indicated by lower case letters.

Table 2.4 Primers for Taqman® 5' nuclease allelic discrimination assays

Assay ID	Forward primer sequence	Reverse primer sequence
Bp1	GCCAAGATGGTACAGAAGAAAGAGT	GTGAGTGTTTCCTGCCTTTTGG
D1	CTGGCCCCAGGATACATAGAAA	GAAGCATCTCACCGTGTGGTT
F1	CTCAACCCACGAGATCTTTCG	GGTCATGCCCGCTTTGTT
FP1	AATGCTGTATCTATAGCCCTCCAAA	GGAACTTGTTTCTTCAGACTTCAATAAA
I1	CGTCACGCAGGACGCC	ACCCACGTGAGCAGCTGC
Lm1	GCGTCTGGCGGCTGTAG	AACACGCGCGGAAACG
L1	TTTCCGCGCGTGTTCAG	GCATTGGAGAGGCAAGCG
MP1	TGAGGGCATGGAAGGTTCAG	GCACCAATTAACTAAGGCCAACAG
Qm1	CCCAGTGTGGACCTAGAATGGT	GGCCTGCGGGATTCAAAC
S1	CCGCAGACCATGACACCTT	CCTGGGAGTCGGTAGCAACA
S2	CGCAGACCATGACACCTTCCT	GCACGCAGGGTCCCTTCT
STp4	GCCACTTCCTCTGCACAAATC	CCAGTGGTATGGAGTGAAAAGATG
STP5	AGCACATCTTTTCACTCCATACCA	AGGTCACAGAGAACTGGGTTAAGG
STp7	AAGGAACATCACAGGAAATGACAA	CCTGCTGCCCTTGATGATTC
T1	CCTGGGCGGCGTTCA	CCCCTGGTGCCTCACTCA
*T2	TGCCAACCTCCTGGACTCTT	CCCCTGGTGCCTCACTCA
TP2	AGTCACCTTGAGTCCCCTATGC	GGAGGGTAAATGTTGTGAATATGGT
Vm3	CTGTGGCTTTTCCTGGATCACT	GGCCCCAGGCTGCAA
Vm2	TCCTGGATCACTGGTCCTCACT	ATGAGGAGGATATGTTGTCCCCTAA
Vm1	TGGATCACTGGTCCTCACTGAGT	CTGCTGAGAATGAGGAGGATATGTT
V1	GGGACAACATATCCTCCTCATTCT	AGAGGCAGGATCTTGGCATCT
V2	GCAGGCAGCTTGGAAGTTTCT	CTCCAGCCCTCAGGAACTTCTA
V4	GGCCCTATGGTTCGACTGAGT	AGGAAGGTCCCCAAAATTATGTTT
V5	TATGGTTCGACTGAGTCCACACTC	GGAGAAGACAGGGTGGGAAGA

*Table 2.4 *Primers supplied by Sigma-Genosys, Cambridgeshire, UK. All others supplied by ABI, Cheshire, UK.*

Table 2.5 Amplification and Sequencing Primers for sequence verification of 5' nuclease allelic discrimination assays using ABI377 BDT sequencer

Assay ID	Forward primer sequence	Reverse primer sequence
Bp1	GAGGCCCCCAGAGACCCA	*GCTGGAATGCCACTTGCTTTAT
F1	*GCCTGGGATGGGGAAAGAGC	*CCAATCCCAGCTCCCAGAAGG
FP1	*GCCTGGGATGGGGAAAGAGC	*CCAATCCCAGCTCCCAGAAGG
Lm1	*TGTTGTGTGACTTTGCGCG	GAGCCGACTTAACCTGGCC
L1	*TGTTGTGTGACTTTGCGCG	GAGCCGACTTAACCTGGCC
MP1	*GGAAGCGGGCGAGGAGTGTGA	CTGCGCAGGTGACGGGTGG
Qm1	*GTGCCAGTGGACTCTACCGTTCA	TGGAGCACAGTGGCAGTTATGG
S1	**CCTAATCCTACGTGACCCTCCTC	GCAGGAGTAGGCTCAGGAAGC
STp4	*CCCTTGACTTCGCCCTGTT	GGCAGCTATGTCTGCTCTCCTT
STP5	*CCCTTGACTTCGCCCTGTT	GGCAGCTATGTCTGCTCTCCTT
STp7	**GCAGAGGAGCAAGGTGGC	GGCAGCTATGTCTGCTCTCCTT
T1	*CACTCAGGATCACTGTGCCAA	GGCTGCTGGGCTCATGA
T2	*CACTCAGGATCACTGTGCCAA	GGCTGCTGGGCTCATGA
TP2	*CACTCAGGATCACTGTGCCAA	GGCTGCTGGGCTCATGA
Vm3	*GCAGACCAGGATGCTGGAA	GGAGAGGCTGTCAGATTCTGAG
Vm1	*GCAGACCAGGATGCTGGAA	GGAGAGGCTGTCAGATTCTGAG
V1	*GCAGACCAGGATGCTGGAA	GGAGAGGCTGTCAGATTCTGAG
V2	*GCAGACCAGGATGCTGGAA	GGAGAGGCTGTCAGATTCTGAG
V4	*AGTGGAGGCTGGGTAGTGCT	CGGTGTCTTGCTGTGTTGC
V5	*AGTGGAGGCTGGGTAGTGCT	CGGTGTCTTGCTGTGTTGC

Table 2.5 Primers were supplied by MWG Biotech, London, U.K.

** denotes primer used for sequencing reaction*

***separate primer, detailed below, used for sequencing following amplification:*

S1:CAGGCTGAAAGTATGCCAGTG; STp7:GCAGAGGAGCAAGGTGGC

Table 2.6 DNA collections

DNA collection	Stock Concentration	Supplier
Copenhagen study on asthma in childhood (COPSAC)	10 ng/µl	Copenhagen University Hospital, Copenhagen, Denmark
European community collection of animal cell cultures (ECACC) Human random control panel 1 (HRC-1) and HRC-2,	HRC1 4 ng/µl and 8 ng/µl HRC2 8 ng/µl	Centre for Applied Microbiology and Research (CAMR), Salisbury, Hampshire, UK.

Table 2.7 Primers designed for deletion constructs of the ADAM33 promoter region

Primer name	Forward primer sequence	Reverse primer sequence	Amplicon size bp
ADAM331650	TATGCTTTCTGTACTTTTTGATTT	TGGCCAGAGCAGCAGCAGTAGT	1619
F1	gcggtaccTTCGAGACCAGCCTGACCAACAT	cgaagatctGCTGTGAGCTCCTCGGCCTCTAG	1346
F2	gcggtaccGGTGGCTCATGCCTGTAATCC	cgaagatctGCTGTGAGCTCCTCGGCCTCTAG	1087
F3	gcggtacCTGCTGCATCGCCTTTGCC	cgaagatctGCTGTGAGCTCCTCGGCCTCTAG	637
F3X	gcggtaccGTTCTCTCCTCCTCTCAGGAGTAGAGGC	cgaagatctGCTGTGAGCTCCTCGGCCTCTAG	359
R2		cgaagatctGCTGTGAGCTCCTCGGCCTCTAG	

Table 2.7 Primers were supplied by MWG Biotech, London, U.K.

Table 2.8 Sequencing primers for deletion constructs

Primer name	Primer sequence
pGL3 RV3	TAGCAAAATAGGCTGTCCC
pGL3 GL2	CTTTATGTTTTTGGCGTCTTCCA
Pro C	GGTGGATCACGAGATCAGGAGATTG

Table 2.8 Sequencing primers for verification of deletion constructs. pGL3 RV3 and GL2 are sequencing primers for the pGL3 plasmid (Promega). Pro C was designed to anneal within the F2 sequence to allow verification of the complete sequence. All primers were supplied by MWG, Biotech, London, UK.

Table 2.9 Plasmids

Plasmid	Supplier
Green fluorescent protein pEGFP-N1	Clontech, Takara-bio inc, California, USA
pGL3 basic	Promega, Southampton, UK
Renilla luciferase pRL	Promega,Southampton, UK
Sp1	Gift from Dr. J. Mann and Dr I Clarke, University of Southampton, UK
XL1-blue	Stratagene, California, USA

Table 2.10 Eukaryotic cell lines

Cell line	Supplier
H292 (passage 38-46) epithelial cell line	Prof. D. Davies, University of Southampton, UK
MCF7 (passage 2-4 from frozen stocks) epithelial cell line	Prof. D. Davies, University of Southampton, UK
MRC5 (passage 16-22) fibroblasts cell line	European Community of Animal Cell Cultures (ECACC) Salisbury, Wiltshire, UK.
WI38 (passage 16-18) fibroblasts cell line	European Community of Animal Cell Cultures (ECACC) Salisbury, Wiltshire, UK.

Details of cells are found in section 2.2.2.

Table 2.11 Cell treatments

Cytokine	Stock concentration	Supplier
Dexamethasone	25 mM	Sigma-Aldrich, Dorset, UK
IFN-γ	1 mg/ml	Peprotech, New Jersey, USA
IL-13	10 µg/ml	Peprotech, New Jersey, USA
Phorbol myristate acetate (PMA)	5 mM	Merck Biosciences (Calbiochem), Nottingham, UK

Table 2.11 Treatments were reconstituted as stated in section 2.2.6.1.

Table 2.12 Probes and primers for Quantitative PCR

Template target	Forward primer	Reverse primer	Probe (5'Fam 3'Tamra)
18S Ribosome	18S (Eurogentech)	18S (Eurogentech)	18S (Eurogentech)
3' UTR (750)	GGCCTCTGCAAACAAACATA ATT	GGGCTCAGGAACCACCT AGG	CTTCCTGTTTCTTCCCACCCTG TCTTCTCT

Table 2.12 Probes and primers used for quantitative PCR (qPCR). Probe primer mixes were supplied premixed as a gift from Dr.R. Powell, University of Southampton. Each reaction contained 1 µl probe/primer mix (ADAM33 or 18S rRNA probe/primer mix) containing 250 nM fluorogenic probe and 900 nM each of forward and reverse primers. 18S rRNA primer and probe mix was supplied by Eurogentec.

Table 2.13 Laboratory Kits

Name of Kit	Supplier
Promega Dual-Luciferase® reporter system	Promega, Southampton, England
Qiagen Effectene® transfection reagent	Qiagen Ltd, West Sussex, UK
Qiagen Gel extraction kit	Qiagen Ltd, West Sussex, UK
Qiagen HiSpeed® Plasmid purification kit	Qiagen Ltd, West Sussex, UK
Qiagen Mini prep kit	Qiagen Ltd, West Sussex, UK
Qiagen PCR purification kit	Qiagen Ltd, West Sussex, UK

Table 2.14 Plasticware

Product	Supplier
Bijou tube (10 ml)	Fisher Scientific, Leicestershire, UK
Cell culture plates (6-well)	Fisher Scientific, Leicestershire, UK
Cell scraper	Fisher Scientific, Leicestershire, UK
Cryotubes (nunc)	Fisher Scientific, Leicestershire, UK
Disposable cuvettes	Promega, Southampton, UK
Eppendorf tube (1.5 ml)	Greiner, UK
Fisherbrand disposable plastic cuvets	Fisher Scientific, Leicestershire, UK
Micro centrifuge tube (0.5 ml)	Alpha labs, Hampshire, UK
Millipore 0.22 µM filter and syringe	Fisher Scientific, Leicestershire, UK
Nunc easy flask, cell culture flasks, 25 cm^3, 75 cm^3 and 175 cm^3	Fisher Scientific, Leicestershire, UK
O ring tubes (2 ml)	Alpha labs, Hampshire, UK
Optical plates (384-well)	ABI Cheshire, UK
PCR tube (0.2 ml)	Alpha labs, Hampshire, UK
Pipette tips (P10 filter)	Biohit (Laser) Northampton, UK
Pipette tips (p10)	Alpha labs, Hampshire, UK
Pipette tips (p100 filter)	Biohit (Laser), Northampton, UK
Pipette tips (P100)	Alpha labs, Hampshire, UK
qPCR plate for icycler (96-well)	AB-gene, Surrey, UK
Thermofast 96-well skirted 0.2 ml PCR plate.	AB-gene , Surrey, UK

Table 2.15 Apparatus

Equipment	Company
ABI377 sequencer	ABI, Cheshire, UK
ABI 7900HT absolute quantification and allelic discrimination instrument	ABI, Cheshire, UK
Beckman DU-7000 DNA quantitator	Beckman Coulter, Buckinghamshire, UK
Biofuge, Fresco bench centrifuge	Heraeus Instruments, Essex, UK
Bio-rad i-cycler real time PCR detection system	Bio-rad, Hertfordshire, UK
DNA Gel tank: Horizon 58	Life Technologies, GIBCO BRL, Maryland, USA
Falcon 6/300 Centrifuge	Sanyo Gallenkamp, Loughborough, UK
FluorImager 595	Molecular Dynamics
HERA safe, Class 2 cell culture hood	Heraeus Instruments, Essex, UK
Heraeus cell culture incubator	Heraeus Instruments, Essex, UK
Innova 4200 incubator shaker	New Brunswick Scientific, Edison, NJ, USA
Jenway laboratory colorimeter	Jenway, Essex, UK
Labcaire PCR workstation (model number: PCR6)	Labcaire, Clevedon, N.Somerset, UK
Leica dm1rb light microscope	Leica Microscopes, Wetzler, Germany
Multi channel pipettes, Biohit	Alpha labs, Hampshire, UK
PTC-225 DNA Engine Tetrad	MJ Research, Massachusetts, USA
Quadra 96 SV liquid handling machine	Tomtec, Receptor Technologies, Oxon, UK
RNA Gel tank: Model No: MGV-202T	CBS Scientific Co, California, USA.
Single channel Pipette, Finnpipette	Fisher Scientific, Leicestershire, UK
Single channel Pipette, Gilson	Anachem Bioscience, Bedfordshire, UK
TD 20/20 Turner single sample luminometer	Turner Biosystems, California, USA
Thermo sealer	AB-gene, Surrey, UK
UVP dual intensity transilluminator	Fisher Scientific, Leicestershire, UK
Zeiss Axiovert 50M	Advance Optical Microscope Facility (AOMF), Toronto, Canada

Table 2.16 Software packages

Program	Version	Supplier
Excel	-	Microsoft
Haploscore	-	Available from D.J Schaid [198]
HTR	-	Available from D.V. Zaykin [199]
Lasergene	6.0	DNAStar, Madison, Wi, USA
LDPlotter		Available through hhtp://innateimmunity.net//IIPGA2/Bioinformatics
SDS	2.0	ABI, Cheshire, UK
SNPHAP	-	Available through www.litbio.org
SPSS	11.0	Microsoft

Table 2.17 Miscellaneous

Miscellaneous	Supplier
AB-gene clear seal strong	AB-gene, Surrey, UK
ABI loading dye	ABI, Cheshire, UK
Adhesive foil PCR plate covers	AB-gene, Surrey, UK
Alconox powdered detergent	SPI supplies, USA
Bromophenol blue	Sigma-Aldrich, Dorset, UK
Diethyl pyrocarbonate (DEPC)	Ambion, Cambridgeshire, UK
Ethidium bromide	Sigma-Aldrich, Dorset, UK
Formamide	Sigma-Aldrich, Dorset, UK
Glycerol	Sigma-Aldrich, Dorset, UK
Haemocytometer	VWR international, Poole, UK
Optical adhesive cover applicator	ABI, Cheshire, UK
Optical adhesive covers 384-well plates	ABI, Cheshire, UK
Random hexamer primers for reverse transcription	Eurogentec, Southampton, UK
RNase wipes	Ambion, Cambridgeshire, UK
Trypan blue	Sigma-Aldrich, Dorset, UK
Virkon	Anachem Bioscience, Bedfordshire, UK
Xylene cyanol	Sigma-Aldrich, Dorset, UK

2.2 Methods

2.2.1 5' Taqman® 5' Allelic Discrimination Assays

The first part of this thesis aims to investigate the association of *ADAM33* SNPs with lung function in a cohort of infants. The Taqman® 5' Allelic Discrimination Assay was the chosen method for genotyping because it demonstrates high specificity and reproducibility and uses minimal DNA concentrations whilst permitting high-throughput (sections 1.5.3 and 3.2.1.2). The development and verification of the assays are described in chapter three, and in chapter four the assays are used to genotype the COPSAC cohort.

2.2.1.1 Construction of submission files for Assay by Design (ABI)

The principles behind selection of single nucleotide polymorphisms (SNPs) for assay design are discussed in section 3.2.1.1. Sequence AF244688, containing 23,574 base pairs spanning the *ADAM33* gene, was identified using National Centre of Bioinformatics (NCBI) http://www.ncbi.nlm.nih.gov/entrez/nucleotide, and BLASTn (Basic Local Alignment Search Tool at http://www.ncbi.nlm.nih.gov/BLAST) to check the known sequence identified by Van Eerdewegh *et al* against the current whole human genome. The confirmed sequence was imported into Lasergene software (DNAStar, USA) and annotated with known single nucleotide polymorphisms (SNPs) in the *ADAM33* gene. This was used to identify sequences flanking the SNPs. Full details of how to write a submission file are found on the ABI website at; http://home.appliedbiosystems.com/support. Briefly, a target sequence of approximately 600 bp total length, with the target SNP towards the centre of the sequence, was imported into a text document. The submission file contained a header line, (containing a maximum of 255 characters), with a greater than (>) symbol, the submitters name, contact telephone number and the assay part number. The part number specified the number of reactions purchased (table 2.18).

Table 2.18 SNP Human Assay Part Numbers

Scale	Number of 5µl reactions- 384-well	Part Number
V	scale 1,000	4331349
S	scale 3,000	4332072
A	scale 12,000	4332073

Table 2.18 Details for ordering of Taqman® 5' nuclease assays, taken from Assays-by-Designs Service Submission Guidelines Part Number 4334672 Rev.A 03/2002.

The second line contained the record name (comprising of 'seq_unique number'), the sequence (5' to 3' orientation) containing the target SNP (which is enclosed in square brackets e.g. for a G to A polymorphism [G/A]), followed by a unique assay identity coordinate providing the name of the target site (e.g.1=fp1 for *ADAM33* F+1 polymorphism). Any sequence containing multiple SNPs was re-submitted for each individual SNP, together with a unique sequence number, identity coordinate, and with the relevant SNP in square brackets. All other known SNPs within the sequence were recoded as 'N'. Only one assay was manufactured from each submitted sequence.

A software package called Assay by Design File Builder, developed by ABI, was downloaded from the website:

http:/ www.appliedbiosystems.com/support/software/assaybydesign/installs.cfm and the prepared sequences imported into the ABI file builder file along with other information including shipping and invoicing details. The orders were sent via the ABI website. The ABI assay by design manufacture process has two levels of quality control (QC). The first level of quality control (QC1) confirms suitability of the submitted sequence for assay manufacture, within the criteria written into the computer program. The submitted file is loaded into a computer program at the ABI site, which examines the sequence and predicts whether a probe/primer set can be manufactured successfully. The process is fully automated, so sequences will only be accepted for manufacture if they conform to the assay conditions stated in the program. There is no room for optimisation. Notification of sequence rejection was received by e-mail within a week of submission. On passing QC1, the assay was manufactured, and tested by ABI on twenty high-quality DNA samples. Fluorescent

change was measured and if the change was above 0.5 units the assay passed quality control 2 (QC2) and was shipped to the University of Southampton. The process was complete within three weeks from date of submission.

2.2.1.2 Custom design of assays

Following a notification of failure at QC1 or QC2, the Primer Select program (Lasergene) was used to design a custom probe/primer set. Probes were designed before the primers, and contained 30-80% GC content, with no G's on the 5' end. Runs of identical nucleotide were avoided as far as possible. The Tm of the probes were between 68-70°C and were 10°C greater than the primer Tm. Primers were placed as close to the probe as possible without overlapping, with a GC of 30-80% again avoiding identical nucleotide runs. The five nucleotides at the 3' end contained no more than two G's or C's. Orders sent to ABI were manufactured and dispatched within three weeks. Custom design sets were not pre-tested by ABI before dispatch.

2.2.1.3 Preparing DNA for the assay

ECACC DNA samples were originally in two 96-well plates. The DNA was pipetted into 384-well plates at 8 ng/well. The first fifteen assays used DNA plated out using the Quadra liquid handling robot. A minimum volume of 20 µl was needed in the 96-well loading plate so that the Quadra could execute the mixing program adequately. Eight 384-well plates were prepared using the Quadra robot. The volumes of the loading plate were too low to continue using this method, so further plates were prepared manually using the Biohit eight-channel multipipette. For the COPSAC DNA, twenty plates of 10 ng/well were prepared using the Quadra liquid handling machine. In both cases, DNA was dispensed into 384-well plates, lightly covered with white tissue and air-dried overnight at room temperature. The following day foil PCR covers were applied to the plates, which were then frozen at -20°C until required. Samples could be stored for up to a month once dried. Before use, plates were defrosted at room temperature for twenty minutes, the plates were spun-down and the foil lid removed ready for immediate use.

2.2.1.4 Storage and care of assays

2.2.1.4.1 Assay by design pre-mixed assays

Each probe/primer set arrived ready mixed (probe, primer, acetonitrile based buffer) in 40x concentration, frozen and packaged in dry ice. The assays cannot withstand repeated freeze/thaw cycles and over-exposure to light causes probe degradation. For these reasons, on receipt the assays were thawed, mixed gently then briefly vortexed and spun before aliquoting into volumes suitable for use.

The aliquot volume was based on the following calculation:

36 µM each primer and 8 µM each probe is contained in 40x mix

36 µM/40X = 0.9 µM = 900 nM (1x) primer

8 µM/40X = 0.2 µM = 200 nM (1x) probe

So for a 5 µl reaction containing 1x mix (900 nM primer and 200 nM probe);

5 µl/40X = 0.125 µl mix per 5 µl reaction. A negative control (NTC) omitting DNA was included for each plate.

The first assays were used to genotype 192 DNA samples (including NTC) so the assays were dispensed into 25 µl aliquots. (0.125 µl x 200 reactions = 25 µl). Light exposure was minimised by wrapping the tubes in aluminium foil and covering with a foil-covered plastic box when on the bench. After aliquoting, the assays were labelled, wrapped in aluminium foil and stored at –20°C.

Over time, the volume of the stored aliquots appeared to reduce. ABI had dispatched the assays in glass tubes with a rubber seal. Aliquots had been stored in 'O'-ring tubes (Alpha labs). On discussion with ABI it was thought the reduction in volume was due to acetonitrile evaporation (the buffering solution used in the assays). To minimise further evaporation the lids of all the aliquots were sealed with parafilm. Lost volume was replaced on thawing before use using distilled water which did not affect the performance of the assays.

2.2.1.4.2 Custom design assays

Primers were supplied by Sigma-Genosys, Cambridgeshire, UK. They were shipped in a lyophilised state and were reconstituted using the required amount of TE (10 mM Tris pH 7.5-8 1 mM EDTA) to give a 100 µM stock. This was diluted to 10 µM using sterile distilled water. In each 5 µl reaction 0.45 µl of each primer (900 nM) and 0.2 µl (200 nM) of each probe was added to 2.5 µl mastermix and the volume made up with water. A negative control, omitting DNA was included.

2.2.1.4.3 Mastermix

Ready-mixed mastermixes (2x) were used. These provide all the components needed for polymerase chain reaction (PCR). They contain deoxyribonucleotides (dNTP), and a hot start DNA Polymerase. This enzyme catalyzes the polymerisation of dNTPs alongside the DNA template. Hot start polymerase is modified so that it is inactive at room temperature, improving the specificity of PCR by preventing the extension of non-specifically bound primers during reaction set-up and the first heating cycle. The 5'-3' exonuclease activity of the enzyme degrades matched probe, separating fluorophores, and permitting continuation of the PCR reaction. Other components include a passive internal control dye (Rox), Magnesium Chloride and buffer. The mastermix (final conc 1x) was added to probe and primer with sufficient sterile water to make up to a 5 µl volume. Applied Biosystems mastermix was compared to Eurogentec's mastemix. Performance was similar on both wet and dry DNA (data not shown). With comparable results, price became the only deciding factor. For consistency, ABI mastermix was used for ECACC DNA genotyping (Chapter 3) since work had already started. Eurogentec mastermix was used for the COPSAC study (Chapter 4). Storage differed with ABI mastermix requiring storage in the fridge at 2-8°C and Eurogentec mastermix requiring storage at -20°C.

2.2.1.5 Setting up the 5' nuclease assay

All new assays were run on the ABI 7900HT instrument to monitor the change in fluorescence throughout the reaction, thereby observing the probe activity. Once the assay was established (i.e. genotype frequencies were in HWE (section 1.4.4.4.3) and representative samples were sequence verified (section 2.2.1.6)), we used ptc-

225 DNA engine tetrad machines for PCR and the ABI 7900HT machine for endpoint analysis. Heat-sealed lids were used on the tetrad to prevent evaporation.

2.2.1.5.1 Assay by Design

Using dried DNA, for each 5 µl reaction 2.5 µl mastermix (2x) was added to 0.125 µl (40x) probe/primer mix to give a final concentration of 900 nM primer and 200 nM probe. This was made up to volume with 2.375 µl water. This was flick-mixed, and 5 µl pipetted into each well of the 384-well plate using an eight-channel pipette. An optical lid was applied to the 384-well plate before spinning up to 1000rpm.

2.2.1.5.2 Custom Design Assays

Dried DNA was used. For each 5 µl reaction, 2.5 µl mastermix (2x) was added to 0.45 µl each primer (900 nM final) and 0.2 µl each probe (200 nM final). This was made up to 5 µl using 1.2 µl water and dispensed as before.

2.2.1.5.3 Cycle program

The reactions were placed on a ptc-225 DNA engine tetrad or ABI 7900HT and amplified using the following cycling conditions: 50°C for two minutes, 95°C for ten minutes, 95°C for fifteen seconds (denature) and 60°C for one minute (anneal and extension) over 40 cycles.

2.2.1.6 Verification of Genotyping

The frequencies observed using the genotyping assays were compared with expected frequencies to confirm the observed frequencies did not deviate from Hardy Weinberg equilibrium (section 1.4.4.4.3). Representative samples of each assay were verified using Sanger sequencing.

2.2.1.6.1 PCR amplification

For each SNP, samples representing all genotypes were sequenced. Primers were designed to anneal over 100 bp away from the SNP of interest using primer express® software. Sequencing primers were eighteen to twenty four bases long, with a GC content of 40 to 60%, and a melting temperature (Tm) between 52°C and 72°C. Primers with long runs of a single base were avoided, especially runs of three or more G's or C's. In addition, primers forming secondary structures or primer dimers were

avoided. Amplification primers were also used as sequencing primers, with the exception of S1 and STp7 (table 2.5). The first step was to amplify the genomic region. A 20 µl PCR reaction was set up using reagents detailed in table 2.19 and sequencing primers listed in table 2.5.

Table 2.19 PCR protocol for amplification of genomic region for SNP verification

Reagent	Final Concentration	Amount
Promega PCR buffer (10x)	1x	2 µl
dNTP (10 mM)	200 µM	0.4 µl
Forward primer (10 µM)	500 nM	1 µl
Reverse primer (10 µM)	500 nM	1 µl
Magnesium Chloride (50 mM)	see table 2.20	variable
PromegaTaq polymerase (5U/µl)	1U	0.4 µl
Water	Up to 20 µl	variable
ECACC	10 ng/ µl	1 µl

Magnesium chloride concentrations and annealing temperatures for each reaction are detailed in table 2.20. The reactions were placed on a ptc-225 DNA engine tetrad and amplified using the following cycling conditions: 95°C for two minutes, 95°C denature for twenty seconds, 57-64.5 °C anneal for thirty seconds, 72°C extension for twenty seconds over thirty five cycles followed by a final extension of 72°C for twenty minutes. The PCR products underwent electrophoresis as detailed in section 2.2.1.6.2

Table 2.20 Reaction conditions for sequence verification of Taqman® 5' nuclease assays

Assay ID	$MgCl_2$	Anneal Temp°C	Amplicon bp
Bp1	1.5 mM	64.5	478
F1	1.5 mM	62.5	407
FP1	1.5 mM	62.5	407
Lm1	1.0 mM	59	335
L1	1.0 mM	59	335
MP1	1.5 mM	62.5	428
Qm1	1.5 mM	57	406
S1	1.5 mM	61	416
STp4	2.5 mM	58	754
STP5	2.5 mM	58	754
STp7	1.5 mM	58	754
T1	1.5 mM	62	346
T2	1.5 mM	62	346
TP2	1.5 mM	62	346
Vm3	2.0 mM	62	798
Vm1	2.0 mM	62	798
V1	2.0 mM	62	798
V2	2.0 mM	62	798
V4	1.5 mM	63.9	513
V5	1.5 mM	63.9	513

2.2.1.6.2 Electrophoresis of PCR products

PCR products were electrophoresed through 2% agarose gels. To make the gels 2 g agarose was added to 100 ml 1x TBE and heated until dissolved. Once cool enough the gel was poured into a gel former and left to set. PCR product (3 µl) was added to 1 µl agarose loading dye (6x) and loaded into each well. A molecular weight and size marker (Smartladder (Eurogentec)) was also loaded. This underwent electrophoresis

at 100 V for twenty to thirty minutes depending on the amplicon size. The gel was stained using Ethidium Bromide (EtBr; final conc. 0.5 µg/ml) for fifteen minutes and visualised using a fluorimager (model 595, Molecular Dynamics). DNA/RNA agarose loading dye (6x) was prepared using 0.25% bromophenol blue, 0.25% xylene cyanol and 30% glycerol in water.

2.2.1.6.3 Purification of PCR products using SAP/Exo

Amplified PCR product was purified using shrimp alkaline phosphatase (SAP) and exonuclease I (Exo I). SAP catalyses the dephosphorylation of 5' phosphates from DNA whilst Exo I degrades (3' to 5') excess single-stranded primer oligonucleotide from the reaction mixture containing double-stranded extension products. SAP/Exo purification was performed by mixing 1 µl SAP, 0.1 µl ExoI and 0.9 µl SAP buffer which was added to 5 µl of the PCR product. This was flick-mixed and briefly spun down and incubated at 37°C for thirty minutes followed by a SAP deactivation step of 80°C for fifteen minutes.

2.2.1.6.4 Sequencing reaction using Big Dye Terminators version 3

Sequencing reactions were set up to verify Taqman® assays and cloned sequences. Sequencing primers are listed in tables 2.5 and 2.8 respectively. All sequencing reagents were removed from the freezer and defrosted on ice. Exposure to light was minimised by using a foil shield. For each reaction 4 µl (½x) sequencing buffer, 4 µl Big Dye Terminators and 1 µl primer (forward or reverse at 3.2 pmol/µl) were added to 1 µl DNA template and made up to a 20 µl volume with water. The reactions were placed on a ptc-225 DNA engine tetrad and amplified using the following cycling conditions: 96°C for ten minutes, 50°C for five seconds and 60 °C for four minutes over 25 cycles finishing at 4°C.

Ethanol precipitation of the products removed unincorporated dye. To each 20 µl reaction, a mastermix containing 3 µl Sodium acetate, 62.5 µl absolute ethanol and 14.5 µl water was added. This was left to stand on the bench top for fifteen minutes before spinning for twenty five minutes at 13000rpm. The ethanol mixture was pipetted off and 250 µl 70% ethanol added to wash the pellet. This was spun again for five minutes at 13000rpm and the excess ethanol removed. The pellet was then dried

on a hot block at 95°C for one minute. Once dried, the product was either frozen at -20°C or reconstituted straight away using ABI loading buffer/formamide mix. To make the loading buffer, 1 µl loading buffer was added to 5 µl formamide (for each sample). The sample (1 µl) was loaded onto the prepared gel (section 2.2.1.6.5).

2.2.1.6.5 DNA sequencing

Sequencing reactions were run on an Applied Biosystems slab gel electrophoresis unit (DNA sequencer ABI 377). To prepare the gel, the sequencing plates were first washed with 1% alconox detergent, rinsed thoroughly with hot distilled water and left to dry. The plates and spacers were then assembled onto the ABI377 frame according to the instructions.

To make the gel, a solution containing 30 ml 5% 6M Urea gene page plus, 30 µl TEMED and 300 µl 10% APS was prepared and mixed by inversion. This was immediately poured between the sequencing plates. An inverted 36-well comb was inserted and the gel left to set for two hours at room temperature. Once set, the comb was removed, the area was cleaned of excess gel, and the comb reinserted to form wells. The gel cassette was loaded onto the machine and a pre-run was performed to warm the gel and run any contaminants out of the gel. Wells were flushed with TBE prior to loading. Sequencing was analysed using ABI377 analysis software programs Sequencing Analysis and Sequence Navigator.

2.2.1.7 Analysis of 5' nuclease assays

2.2.1.7.1 SNPHAP, Haplotype tagging SNPs and Linkage Disequilibrium plots

DNA from ECACC, a random Caucasian population, was genotyped as a sample population to verify the reliability of the 5' nuclease assays. Following amplification of DNA using 5'nuclease assays, an end-point reading was taken using an ABI 7900HT instrument. Genotype calls were assigned and the data was then exported to Notepad and then into Excel. Here the samples were matched with their individual identification numbers, producing an Excel file for each SNP. This was imported into SPSS in gene order. Genotype calls were transformed into alleles (allele 1 and allele 2). Samples with more than three missing calls were removed from the data set before further analysis. The SPSS file was saved as a tab-delimited text file. All

headings were then removed from the file before uploading to SNPHAP (www.litbio.org), a haplotype analysis program [200]. SNPHAP uses the E.M algorithm to identify common (20-SNP) haplotypes in the random Caucasian ECACC population. A haplotype tagging program, also using the E.M algorithm, was then used to identify informative SNPs within the haplotypes. A pairwise LD plot was also constructed (section 3.2.4). Combined results permitted confident reduction in the number of SNPs, whilst retaining all information, for genotyping of other populations.

2.2.1.7.2 The E.M algorithm

All haplotype estimation methods used the E.M algorithm. The EM algorithm is used to assign haplotypes to those people for which haplotype assignment is ambiguous.

For example, imagine a sample of people who have been genotyped for two syntenic SNPs, A/a and B/b. There are ten possible haplotype assignments for any diploid individual (although there are only four possible haplotypes) (Table 2.21)

Table 2.21 Possible haplotype assignments using two syntenic loci, A/a and B/b

Genotype	AA	Aa	aa
BB	AB AB	AB aB	aB aB
Bb	AB Ab	AB ab or Ab aB	aB ab
bb	Ab Ab	Ab ab	ab ab

Haplotype assignment is ambiguous for individuals who are double heterozygotes. It is possible to assign haplotypes to all other individuals because they are homozygous at one of the loci. The EM algorithm is used to assign haplotypes to those people for which haplotype assignment is ambiguous. It does this by first calculating the frequency of the other haplotype assignments for those individuals whose phase can be resolved. It then makes a guess at the assignment of haplotypes (which is essentially a random guess), and compares this guess to the likelihood based on the observed haplotypes. The goodness of fit between the algorithm's guess and the most likely assignment is then assessed (often through a likelihood function) and the guess adjusted on this basis. This process is iterated until the change in likelihood

used to assess the goodness of fit is negligible. In this way haplotypes for each individual can be estimated.

2.2.1.7.3 Haplotype trend regression, Haploscore analysis and LD mapping

The verified assays were used to genotype a longitudinal cohort of one month old infants recruited into the Copenhagen study on asthma in childhood (COPSAC). Haplotype trend regression, Haploscore analysis and LD mapping were used to analyse interactions between ADAM33 polymorphism and lung function ($FEV_{0.5}$ and PD_{15}).

The first method used was Haplotype Trend Regression (HTR) [199]. This program was not sufficiently powerful to calculate haplotypes of fourteen SNPs so a moving window of three SNPs was used. This approach starts with the first three SNPs of the group and moves along one SNP for each cycle of analysis. HTR searches for association of haplotype frequencies with continuous (or discrete) traits in samples of unrelated individuals [199]. Haplotype frequencies were estimated using the expectation-maximisation (EM) algorithm. A regression-based approach then compares each phenotypic variable in turn with the estimated haplotypes [199]. The HTR program does not allow for height correction regarding the $FEV_{0.5}$ phenotype, which was a problem since FEV is associated with height. In addition, 'missing' values increase estimation required by the program, which adversely affects the results, necessitating removal before analysis.

The second method, Haploscore [198], uses unambiguous genotypes (i.e. when the haplotypes were directly measured) to derive the first score statistics. Further score statistics are obtained using generalised linear models within the program. Data analysis is presented as estimated frequency of predicted haplotypes, haploscore calculated from the deviation above or below the average of the lung function measurement, and a p-value ($\alpha = 0.05$) denoting significance level of the association of the haplotype with measure of lung function. Thus, a positive score denotes lung function above the mean average and *vice versa.* All missing values were removed as before. $FEV_{0.5}$ was first used uncorrected against height, to allow for direct comparison with the previous HTR analysis. However, since FEV is dependent on

height, the analysis was repeated adjusted for height. PD_{15} has not been reported as dependent on any other co-variant, so did not require adjustment.

The third method used LD map construction and association mapping [201, 202] (section 3.2.4.3), performed in association with Dr. N. Maniatis. The main objective of LD mapping is to facilitate positional cloning by refining the resolution of the candidate region and to estimate location of a causal SNP as accurately as possible [202]. If we consider when a disease mutation first occurs, all alleles flanking the mutant allele are in complete disequilibrium. Disequilibrium decays over time through the action of recombination. Generations of recombination produce a pattern of disequilibrium whereby disequilibrium should be greatest for markers closest to the disease allele. However LD does not decline smoothly with distance. Instead islands of LD are observed, broken by 'recombination hotspots'. The theoretical framework for constructing LD maps is based on the Malecot model [203] which describes the exponential decline of LD in relation to distance between a pair of SNPs. The LD maps developed by Maniatis et al [202, 204] assign locations to markers (SNPs) in Linkage Disequilibrium Units (LDU) and are shown to be more accurate than kb maps, as shown when the LDU map and kb map of the CYP2D6 regions were compared [201]. The LD map was created using LDMAP program [202]. This is modelled with composite likelihood method [205] which depends on three parameters, recombination fraction between the disease locus and arbitrary marker locus (θ), the age of the mutation and a variance parameter. Each single SNP is compared to the arbitrary locus using regression [205]. If an association is observed haplotypes are then put into the model to define the location of the disease-causing SNP [201].

2.2.1.7.4 Univariate regression and haploscore analysis of ADAM33 polymorphism and lung function at three years.

Interaction between *ADAM33* polymorphism and lung function (pre- and post-bronchodilator sRaw) at three years was investigated using univariate regression and haploscore analysis.

2.2.2 Cell culture

Eukaryotic cell lines were propagated for use in transfection experiments to investigate the promoter region of *ADAM33*. ADAM33 is expressed in mesenchymal cells, but not epithelial cells. It is possible that *ADAM33* transcription is inhibited by hypermethylation in epithelial cells. Plasmids are not methylated, so if the reporter plasmid produced a signal in epithelial cells it would suggest the necessary transcription machinery for activation is present and that methylation plays a major part in regulation. The transcription machinery is present in fibroblasts since ADAM33 is expressed in these cells, so reporter expression in these cells would enable definition of the promoter. I therefore decided to use both an epithelial cell line and a fibroblast cell line for the transfections.

MRC5 and WI38 fibroblast cell lines, and H292 and MCF7 epithelial cell lines were chosen. H292 cells are epithelial cells originating a cervical node metastasis of a pulmonary mucoepidermoid carcinoma in a 32 year old Negro female. MCF7 cells are epithelial cells established from the pleural effusion from a 69 year female caucasian suffering from a breast adenocarcinoma. These cells are reported to express the oestrogen and progesterone receptor. MRC5 cells have a fibroblast-like phenotype and originate from normal lung tissue of a 14 week old male fetus. The cells undergo between 60-70 population doublings before senescence. WI38 cells also have a fibroblast-like phenotype and originate from normal embryonic lung of a female fetus.

All media and reagents were pre-warmed to 37°C prior to use. Culture techniques were carried out in a class II biosafety cabinet (Heraeus HERAsafe) using aseptic technique. Used culture media and plastics were disinfected with 2% virkon before discarding.

2.2.2.1 Setting up a cell culture

MRC5 and WI38 cells were purchased from ECACC (passage 16) and arrived frozen and packed on dry ice. MCF7 and H292 cells were removed from cryogenic storage and were donated by Professor D. Davies (University of Southampton). To commence culture, cells were quickly defrosted by adding warm media (in 500 µl aliquots) and decanting into a 25 cm^2 flask (total media volume 5 mls). This allowed

speedy dilution of the cryoprotective DMSO, so minimising cell damage. Fibroblast cells (MRC5 and WI38) were decanted into collagen-coated flasks (section 2.2.2.5) to aid adherence. All cells lines used were adherent. Following transfer to the flask, the cells were incubated for six hours at 37°C to allow cells to adhere. The media was then changed to remove residual cryoprotectant and the flask returned to the incubator for routine culture.

2.2.2.2 Sub-culturing cell line monolayers

The cells used in this study were all incubated at 37°C and 5% CO_2. To sub-culture these cells they first needed to be detached from the plastic of the cell culture vessel. This was done using trypsin-EDTA treatment. Trypsin-EDTA was supplied as concentrate (10x) so needed to be diluted to 1x solution using Hanks balanced salt solution (HBSS), without Ca^{2+} and Mg^{2+} before use. Serum inactivates trypsin so old media was removed from the cells and twice washed with HBSS. Sufficient 1x trypsin/ EDTA to cover the cell monolayer (0.5-2 mls) was pipetted into the flask and the cells incubated until cells 'rounded up' and were released. Tapping the flasks sharply aided detachment of the cells from the surface. Trypsin was inactivated using 5-10 ml of usual culture media. The cell suspension was centrifuged at 1000rpm for four minutes. For routine passage, cell pellets were then resuspended in fresh culture media and reseeded at a density of 1:2.

2.2.2.3 Viable cell counting using trypan blue

A cell count was performed to achieve a specific seeding density. Cells were released from the flask using trypsin-EDTA treatment (2.2.2.2). Following centrifugation the media/trypsin was discarded and the cell pellet resuspended into HBSS to give a single cell suspension. HBSS was used because serum (found in media) prevents trypan blue entering the cells, resulting in possible overestimation of cell viability. Viable live cells exclude trypan blue and dead cells take up the dye. A sample of cells were diluted 1:5 by mixing 20 µl cell suspension, 30 µl HBSS and 50 µl 0.4% Trypan Blue. The diluted cell suspension (10 µl) was pipetted onto a clean Neubauer haemocytometer (depth 0.1mm; 1/400 mm^2) with cover slip. Viable cells were counted using a light microscope (5x magnification) and a mean cell count determined using the following calculation;

Viable cells /ml = (mean cell number /0.1 mm^3) x (dilution factor used (5)) x 10^4

Once the number of cells/ml was estimated the cells were then diluted using the usual cell media to give the required seeding density. An example calculation follows;

For a desired seeding density of 2 x 10^5, with a mean cell count of 25.5;

25.5 x 5 x 10^4 = 1275000 (or 1.275 x 10^6)

2 x 10^5/1.275 x 10^6 = 0.156 ml per 1 ml stock

A 6-well plate requires 2 mls media per well, so 0.156 ml x 2 = 0.314 ml cells in 2 ml media. This amount was then multiplied by the number of wells required. The cell/media mix was then pipetted into the wells.

2.2.2.4 Cryogenic storage of cells

At the conclusion of experiments, and following sub-culture, aliquot(s) of cells were frozen down and put into liquid nitrogen storage to maintain and expand cell stocks. Cells were trypsin-EDTA treated and pelleted by centrifugation. Supernatant was discarded and the cells resuspended in usual cell culture media + 10% DMSO. This was aliquoted into cryovials and cells were slow-frozen overnight to -80°C. Slow freezing prevents formation of ice crystals, which would damage the cells on defrosting. This was achieved by wrapping the aliquots in cotton wool before freezing. The frozen cells were transferred to liquid nitrogen the next day for long-term storage.

2.2.2.5 Collagen coating of plates/wells

Fibroblast and smooth muscle cells were cultured on collagen-coated plastic to improve the adherence of the cells. Vitrogen-100 was diluted 1:500 with sterile water, pipetted into the required cell culture vessel and incubated for one hour. The water/vitrogen mix was discarded immediately prior to inoculation.

2.2.2.6 ADAM33 expression in fibroblast and epithelial cell lines

Expression of ADAM33 in the fibroblast and epithelial cell lines was investigated using reverse transcription, followed by quantitative PCR (qPCR).Cells were harvested in TRIzol and total RNA extracted. A reverse transcriptase assay converted RNA to cDNA which was quantitated using qPCR. Primers and probes for qPCR are listed in table 2.12.

2.2.2.6.1 RNA extraction

To prevent RNase contamination sterile disposable gloves were worn and sterile, disposable plasticware and automatic pipettes reserved for RNA work (to prevent cross-contamination with RNases from shared equipment) were used. A PCR hood (Labcaire PCR workstation, model number: PCR6) was used for RNA work. In the presence of TRIzol Reagent, RNA is protected from RNase contamination but downstream sample handling requires that non-disposable glassware or plasticware be RNase-free. Water was prepared by adding 1 ml of fresh DEPC, an RNase inhibitor, to 1 litre of water to make a final concentration of 0.1%. This was shaken thoroughly to disperse the DEPC before incubatating at 37°C for twelve hours. It was then autoclaved for twenty minutes at 15psi to inactivate the remaining DEPC. Cells were harvested and homogenised in TRIzol reagent (750µl TRIzol/well on a 6-well plate). Chloroform was added according to the amount of TRIzol used for harvest (750 µl TRIzol :150 µl chloroform) and shaken. This centrifuged at 12000rpm and 4°C for fifteen minutes to achieve phase separation. The top layer (aqueous phase) contained RNA. The thin lipid layer contained lipids from the cells and the lower pink layer contained phenol-chloroform, DNA and proteins. The aqueous phase was transferred to a clean RNase/DNase-free eppendorf. To precipitate RNA, 0.5 ml isopropanol (99%) was added per sample. The samples were incubated at room temperature for fifteen minutes and then centrifuged at 4°C, 13000rpm for thirty minutes. Hinges were aligned to the outer edge of the centrifuge to aid identification of the pellet. The supernatant was discarded and 1 ml of 75% ethanol added to each sample before vortexing gently and centrifuging at 7500rpm at 2-8°C for five minutes. The ethanol was removed and the sample pulse-spun. Remaining ethanol was pipetted off and the pellet left to air dry for ten minutes. To prevent any DNA carry-over a DNase mastermix was prepared for each RNA sample containing 1 unit DNase (1 unit/µl), 2 µl 10x buffer and 17 µl H_20 (DEPC). Each dried pellet was resuspended in 20 µl of DNase mix and incubated at 37°C for one hour. Neutralisation buffer (5 µl) was then added. This was mixed thoroughly by tapping and then incubated for two minutes at room temperature followed by a repeat tap-mix and further two minute incubation. The sample underwent pulse-spin for two minutes. An agarose gel (0.75%), using 100 ml 1x TBE (diluted from10x TBE with ultra pure

water), was prepared and 30 ul ethidium bromide added when cool. The gel was poured into a tank pre-cleaned with RNase wipes and dried with paper towel. The gel was left to set for thirty minutes. The RNA was loaded using 8 µl 2x agarose loading buffer mix with 2 µl sample. This was run for twenty to thirty minutes at 100 V, and visualised on the fluorimager (model 595, molecular dynamics).

2.2.2.6.2 Reverse Transcriptase (RT) assay

The RNA was reverse transcribed to create cDNA. This was used as template for the quantitative PCR reactions. The reaction has two stages: annealing, when the primer is bound to the template, followed by extension to form new single-stranded cDNA. The annealing reaction was performed in a sterile, thin 0.2 µl PCR tube. A mastermix containing 1 µl random hexamer primer (3 µg/ml), 1 µl dNTP mix (10 mM) and 3 µl dH_2O was prepared. The mastermix was added to 5 µl RNA (~1 µg). The samples were heated to 80°C for five minutes using the thermocycler then 'snap cooled' using ethanol with dry ice added. For the extension reaction, a 10 µl mastermix was prepared for each sample containing 4 µl 5x M-MLV transcription buffer, 0.5 µl M-MLV transcriptase (100 units) and 5.5 µl dH_2O. This was added to the annealing reaction and returned to the thermocycler for one hour at 37°C. cDNA template was stored at -20°C until required.

2.2.2.6.3 Real time quantitative PCR

Real time quantitative PCR (qPCR) was used to quantitate *ADAM33* expression in the cell lines using cDNA from the RT reaction. qPCR probes are oligonucleotides (about 20-25 bases) containing a fluorogenic reporter dye FAM (6-carboxyfluorescein) and a quencher dye TAMRA (6-carboxy-N,N,N',N'-tetramethylrhodamine). The probe is complementary to a target sequence of cDNA situated between the forward and reverse primers during the annealing stage of PCR. Probes and primers sets are listed in table 2.12. During the cycle, the Taq polymerase binds and extends the cDNA. The Taq polymerase enzyme has 5' exonuclease activity. On reaching the annealed probe it cleaves the FAM from the probe causing an increase in the fluorescent intensity of the reporter dye. Therefore, the faster the fluorescence reaches the threshold (CT value), the higher are the amounts of cDNA encoding the gene of interest in the reaction mix. Each one point

change in CT constitutes a doubling of product. The gene of interest is always measured in relation to a reference gene such as 18s ribosomal RNA. Reference genes are ubiquitously expressed at consistent levels, usually with higher expression than the gene of interest, and act as an internal control. Subtraction of the reference gene CT value (CT_R) from the template gene CT value (CT_T) gives the measure, ΔCT (CT_T-CT_R =ΔCT), thus quantifying expression of the gene of interest in relation to the control gene. To compare expression between cell lines, one of the cell lines is used as a calibrator. This gives a measure ΔΔCT. Since each one-point change in CT constitutes a doubling of product, abundance of product between cell lines, in relation to the calibrator cell line, can be calculated using $2^{\Delta\Delta CT}$.

2.2.2.6.3.1 Setting up a qPCR reaction

Each sample was measured in duplicate along with positive control (18S) and negative controls. A grid was used to map sample positions to the 96-well plate.

Figure 2.1 Example of a qPCR grid

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	1	2	2	3	3	Neg	Neg	ADAM33 probe primer set			
B	1	1	2	2	3	3	Neg	Neg	18S probe primer set			
C												
D												
E												
F												
G												
H												

Figure 2.1 A grid was used to position the duplicate samples on a 96-well plate. In this case three samples are used. Each sample is duplicated both for the gene of interest and the reference gene (18S).

cDNA template was defrosted on ice and diluted to give approximately 25 ng of template. This is 5 µl of a 1:10 dilution of the RT reaction where 1 µg of RNA is reverse transcribed in a total volume of 20 µl. One RT reaction can be used to measure numerous target genes. Sufficient RT reaction was diluted to measure the genes of interest. A Pre-RT RNA negative control was prepared to ensure the qPCR

fluorescence emanated from cDNA and not genomic contamination. DNase-treated RNA (pre RT) was used and diluted accordingly (i.e. RT involved 5 µl RNA in a 20 µl reaction, a 1:4 dilution). The resulting cDNA was then diluted 1:10. So to make a comparable negative control 3 µl DEPC treated water was added to 1 µl RNA. This was diluted 1:10 and used as a negative control for the qPCR. The *ADAM33*/positive control gene assays had been previously verified so standard curves were not included. A 12.5 µl reaction was set up for each sample using 6.5 µl qPCR mastermix and 1 µl probe/primer set (either *ADAM33* or reference gene). The 7.5 µl mix was added to 5 µl diluted cDNA template in the 96-well plate and a clear plate seal applied. The reaction was flick-mixed and underwent a pulse-spin. It was then placed in the i-cycler (Bio-rad i-cycler real time PCR detection system) and amplified using the following cycling conditions: 95°C for eight minutes, 95°C for fifteen seconds and 60°C for one minute over forty cycles. Results were analysed using Bio-rad i-cycler software. The thresholds for fluorescence emission baseline were set just above background levels. Expression levels were calculated as described (section 2.2.2.6.3).

2.2.3 Generation of promoter constructs

Deletion constructs were created containing PCR fragments of the *ADAM33* promoter region, upstream of the translation start site. The fragments were ligated into pGL3 promoter-less plasmids and clones amplified in *E.Coli*. Following purification and verification by Sanger sequencing, plasmid constructs were used in transfection experiments to identify regions involved in *ADAM33* activation.

2.2.3.1 Purification of PCR product

Specific PCR and cycling conditions are detailed in section 5.2.1. All PCRs were run using a ptc-225 DNA engine tetrad and amplicon size observed using electrophoresis (section 2.2.1.6.2).

PCR products were purified using the Qiagen PCR purification kit (table 2.13) according to the protocol. Briefly, 5 volumes of buffer PB was added to 1 volume PCR product. The sample was pipetted into the QIAquick® spin column placed in a 2 ml collection tube. The sample was spun for one minute at 13000rpm in a bench top

centrifuge. After discarding the supernatant the column was washed using 0.75 ml buffer PE. This was spun as before for one minute and the supernatant again discarded. Another one minute spin ensured complete removal of ethanol. The column was then placed in a clean 1.5 ml eppendorf. To elute the DNA, 30 µl distilled water was added to the column and left to stand for one minute before spinning for one minute at 13000rpm. The eluted DNA was then stored at 4°C in a 0.5 ml eppendorf tube. The PCR products underwent electrophoresis as detailed in section 2.2.1.6.2

2.2.3.2 Double restriction digest

Purified PCR product (section 2.2.3.1) and pGl3 plasmid were double digested to expose the restriction sites for ligation. A 50 µl restriction digest was set up by adding 1 µg DNA to a mastermix containing 5 µl NEB2, 0.4 µl Bgl II, 0.4 µl Kpn I, 5 µl Bovine serum albumin (BSA) made up to volume using distilled water. The reagents were mixed and incubated at 37°C for two and a half hours. No inactivation step was necessary for these enzymes. Products were purified using gel extraction.

2.2.3.3 Gel extraction of digested products

The QIAquick® gel extraction kit was used to purify the digest products, according to the manufacturer's protocol. Briefly, following electrophoresis (section 2.2.1.6.2), the gel was placed on a UV light source and correct size bands were excised from the gel. The gel slice was weighed and three volumes buffer QG was added to one volume of gel (100 mg/100 µl). The mixture was incubated at 50°C and briefly vortexed, every two to three minutes, until the gel was dissolved (~ten minutes). For F3X (359 bp) one gel volume of isopropanol was added to the mixture since this increases the yield for DNA fragments of less than 500 bp and greater than 4 kb). A QIAquick® spin column was placed in a 2 ml collection tube. The sample was applied to the QIAquick® column, and centrifuged for one minute. Flow through was discarded and the column replaced in the same collection tube. 0.75 ml buffer PE was added to the QIAquick® column and centrifuged for one minute to wash the sample. The flow-through was discarded and the sample centrifuged for an additional one minute, at 13000rpm, to remove residual ethanol. The QIAquick® column was placed in a clean 1.5 ml eppendorf tube and 30 µl water (pH 7-8.5) applied to the column for elution of

DNA. The column was left to stand for one minute before one minute centrifugation. The eluate was transferred to a clean 0.2 ml PCR tube and stored at 4°C until required. The PCR products underwent electrophoresis as detailed in section 2.2.1.6.2 for concentration estimation.

2.2.3.4 Ligation of PCR products into pGL3 basic vector

Purified, digested PCR products and vector were then used for ligation reactions. In order to maximise the likelihood of ligation the correct ratio of insert (PCR product) to vector (plasmid) is required. First the molar ratio was calculated. The molar ratio is the size of vector (bp) divided by the size of PCR product (bp). Molar ratios are listed in table 2.22.

Table 2.22 Molar ratio of insert: vector

Insert	Insert size	Vector size	Molar ratio
F1/R2	1346	4818	3.57
F2/R2	1087	4818	4.43
F3/R2	637	4818	7.56
F3X/R2	359	4818	13.42

Concentrations (ng) of insert and vector were estimated using electrophoresis (2.2.1.6.2) and the Smartladder (Eurogentec) molecular marker, which estimates concentration as well as size. Ligation reactions (20 µl) were set up on ice using 2 µl Ligase Buffer (10x), 1 µl Ligase, 0.5 µl Vector and insert. Insert quantities were calculated using the following equation;

$$\text{Insert} = \frac{\text{ng vector} \times \text{kb insert} \times \text{Molar ratio insert}}{\text{kb vector}}$$

The volume was made up to 20 µl with distilled water and flick-mixed. The reaction was incubated at 16°C for three hours. Single cut vector (to test ligase activity) and double cut vector (to assess restriction efficiency) controls were included.

2.2.3.5 Transformation of ligation reactions into competent cells

For each transformation 100 µl competent cells were defrosted on ice and transferred to a 1.5 ml eppendorf tube. The ligation reaction (10 µl) was added to the competent cells and flick-mixed. A control transformation containing only competent cells was also set up. The mixture was incubated on ice for ten minutes before heat-shock treatment for fifty seconds in a 42°C water bath. The mixture was incubated on ice for a further two minutes. SOC medium (900 µl) was added and this was incubated for one hour at 37°C. The incubated cell mixture (250 µl) was spread onto an agar plate. Experiment and control plates are summarised below;

- Insert + vector transformation. Growth suggests successful ligation;
- Single cut vector alone. This tests ligase activity. If ligase is active the plasmid should re-ligate and so be ampicillin-resistant;
- Double cut vector alone. This tests restriction. If restriction is complete the plasmid cannot re-ligate and so should not have ampicillin resistance;
- Uncut vector alone. This tests transformation efficiency. This should be ampicillin-resistant;
- Heat-shocked competent cells only, inoculated onto a non-ampicillin plate. This tests competent cell viability following heat-shock treatment. These should grow proliferatively on non-ampicillin plates;
- Heat-shocked competent cells only, inoculated onto an ampicillin plate. This tests ampicillin activity. With no plasmid they will have no ampicillin resistance so will not grow if the ampicillin is active.

Following inoculation and spreading, the plates were left on the bench for thirty minutes to dry and were then inverted and incubated overnight at 37°C.

2.2.3.6 Check PCR of colonies

Any colony growing on the ampicillin plates must contain an intact plasmid with an active ampicillin-resistance gene. Some of the colonies will contain plasmid with ligated insert, others will contain undigested plasmid. The double-cut vector control plate gives an indication of the expected numbers of colonies due to poor restriction.

Colonies were checked for presence of insert with PCR using primers flanking the polylinker region. Vectors containing the insert will yield the correct size PCR product. A sterile p10 pipette tip was used to prick the colony. This was rinsed in 10 µl sterile water and used as template. A 10 µl reaction containing 1x Promega PCR buffer, 500 nM each primer (pGL3 RV3 and pGL3 GL2 - table 2.8), 200 µM dNTP and 1.5 mM $MgCl_2$ was set up. This was amplified over thirty five cycles using the following PCR cycling conditions: 95°C for two minutes, 95°C for twenty seconds, 60°C anneal for thirty seconds, 72°C for thirty seconds and 72°C twenty minutes. The PCR products underwent electrophoresis (section 2.2.1.6.2) for size verification.

2.2.4 Plasmid preparations

Sterile technique was used for plasmid preparations to reduce contamination. Culture flasks were sterilised by autoclaving at 132°C for thirty minutes.

2.2.4.1 Growing up new stocks of competent cells

An agar plate was streaked with XL1-blue *E-coli*. This was extracted from a frozen aliquot of cells using an inoculation loop, and was grown overnight at 37°C. The next morning a colony was picked off, inoculated into 3 ml Luria Broth (LB) and incubated at 37°C overnight in a sterile falcon tube. The next morning 100 mls of sterile LB was inoculated with 1 ml overnight culture into a sterile 500 ml flask and grown up in a shaking incubator until the exponential growth phase was reached (i.e. the optical density was between 0.4-0.6 at a wavelength of 600 nm). Cells were harvested by centrifugation for fifteen minutes at 3000rpm. The supernatant was removed and the cells resuspended in 10 ml of 0.1 M of ice-cold calcium chloride ($CaCl_2$) solution and incubated on ice for thirty minutes. Cells were centrifuged again for five minutes at 3000rpm. The supernatant was discarded and the cells resuspended in 4 ml of 0.1 M ice-cold $CaCl_2$ solution. The cells were incubated on ice for a further two hours. The cell suspension was then mixed with 1.0 ml of 50% glycerol (final 10%) mixed gently on ice. The cells were then aliquoted into eppendorf tubes, snap-frozen in liquid nitrogen and stored at -80°C until required.

2.2.4.2 Preparation of Luria Bertoni broth and Luria Bertoni broth agar plates

Luria Bertoni broth was prepared using 10 g tryptone, 5 g yeast extract and 5 g Sodium chloride in 1 litre distilled water and autoclaved for twenty minutes at 15psi. This was stored at 4°C until required. To prepare agar plates, 1.5 g agar granules was added to 100 ml LB medium (1.5% agar) and autoclaved. This was stored at 4°C until required.

To pour the plates the agar was heated in the microwave until dissolved. Antibiotic was added once the agar had cooled to 42°C, and immediately before pouring into sterile petri dishes. The plates were left to set for ten minutes, then inverted and wrapped in parafilm and stored at 4°C until required. Antibiotic was added to pre-autoclaved media immediately before use. Media and agar plates were warmed to room temperature before use. Ampicillin (final concentration 100 µg/ml) was used for pGL3, *Renilla* luciferase and Sp1 plasmids. Kanamycin (final concentration 50 µg/ml) was used for the green fluorescent protein (pEGFP-N1) plasmid.

2.2.4.3 Preparation of antibiotics

Ampicillin was reconstituted in distilled water to a make a 50 mg/ml stock solution. The solution was filter-sterilised by drawing through a pre-washed 0.22 µm filter before aliquoting into 1 ml eppendorf tubes and freezing at -20°C until required. Kanamycin was supplied at a concentration of 50 mg/ml.

2.2.4.4 Preparation of 0.1 M calcium chloride

Calcium chloride (0.1 M), required for propagation of new competent cell stocks (section 2.2.4.1), was prepared by dissolving 7.35 g $CaCl_2$-$2H_2O$ (MW=147) into 500 ml of distilled water. This was autoclaved for twenty minutes at 15psi and stored at 4°C until required.

2.2.4.5 Preparation of 50% glycerol

Glycerol was used for freezing down plasmid stocks for long term storage (section 2.2.4.7.3). Glycerol (50%) was prepared by adding 50 ml distilled water to 50 ml glycerol. This was autoclaved for twenty minutes at 15psi and stored at 4°C until required.

2.2.4.6 Preparing a starter culture from a discrete colony

Starter cultures were prepared to propagate the discrete colonies containing the correct insert (identified by PCR (section 2.2.3.6). A sterile10 µl pipette tip was used to remove the colony from the plate and inoculate a small volume (3-5ml) of LB containing appropriate antibiotic. A sterile p10 tip with no plasmid was also inoculated into LB (plus antibiotic) as a negative control. The cultures were incubated for eight hours at 37°C. The starter cultures were used to inoculate further cultures.

2.2.4.7 Preparation of plasmid cultures

2.2.4.7.1 Purification of plasmid using Qiagen miniprep kit

A small-volume culture was prepared first, so that the plasmid could be verified by sequencing before growing up a larger volume. Starter culture (1 ml) was inoculated into 15 ml LB, plus antibiotic, and incubated overnight at 37°C. The next morning 1.5 ml of overnight culture was pipetted into a 1.5 ml eppendorf and spun for thirty minutes at 13000rpm. The supernatant was removed and discarded. The pellet was resuspended in 250 µl buffer P1. Buffer P2 (250 µl) was added (alkaline lysis) and the tube inverted four to six times to mix. This was incubated for no more than five minutes. Buffer N3 (350 µl) was then added (neutralisation) and the tube inverted four to six times to mix. This was followed by a ten minute centrifugation at 1300rpm, resulting in a pellet of cell debris and clear supernatant. The supernatant, containing plasmid DNA, was applied to the Qiaprep® spin column and centrifuged for one minute at 13000rpm. The flow through was discarded.

The column was washed with 0.75 ml buffer PE. This was then centrifuged for one minute and the flow through discarded. Residual ethanol was removed with a further one minute centrifugation. Flow-through was discarded, and the column placed in a clean 1.5 ml eppendorf. To elute the DNA, 30 µl water was pipetted into the centre of the column. This was left to stand for one minute before a one minute centrifugation at 13000rpm. The eluate was transferred to a 0.5 ml eppendorf and stored at 4°C until needed. DNA concentration was estimated using a Beckman DU-7000 DNA quantitator (section 2.2.4.7.4).

To verify the presence of the insert the purified plasmid underwent double restriction digest. A digest reaction was set up by adding 1 µg DNA to a mastermix containing 1.2 µl NEB2, 0.1 µl Bgl II, 0.1 µl Kpn I and 1.2 µl BSA (10x) made up to a volume of 12 µl with distilled water. The reaction was incubated at 37°C for two and a half hours. The digest products underwent electrophoresis as detailed in section 2.2.1.6.2. The plasmid was then sequenced to ensure the correct sequence was present (section 2.2.4.8). Once verified, larger stocks of plasmid were prepared.

2.2.4.7.2 Purification of plasmid using Qiagen HiSpeed® midi kit

A transformation reaction was set up using 10 µl of diluted (1:100) purified plasmid mixed with 100 µl competent cells (section 2.2.3.5). This was plated out and grown overnight. A starter culture was prepared and grown overnight (section 2.2.4.6). The starter culture (1 ml) was inoculated into 50 ml LB plus antibiotic and incubated overnight at 37°C. Bacteria were harvested by centrifugation in a Sorvall GSA centrifuge at 6000 rpm. All traces of supernatant were removed by pouring off and inverting the tubes. Bacteria were resuspended in 6 ml buffer P1, ensuring complete resuspension by vortexing or pipetting up and down until no cell clumps remained. Buffer P2 (6 ml) was added (alkaline lysis) and mixed gently but thoroughly by inverting four to six times. This was incubated at room temperature for no more than five minutes. During the incubation period, a QIAfilter midi cartridge was prepared by screwing the cap onto the outlet nozzle. Chilled buffer P3 (6 ml) was added to neutralise the reaction. This was mixed gently but thoroughly by inverting four to six times. A fluffy white precipitate formed and the lysate was immediately transferred to the QIAfilter cartridge, to prevent disruption of the precipitate layer. This was incubated at room temperature for ten minutes. During the incubation time 4 ml buffer QBT was pipetted into a high speed midi tip. The column was allowed to empty by gravity flow. The cap was then removed from the outlet nozzle of the QIAfilter cartridge and the plunger gently inserted into the cartridge. The lysate was filtered into the equilibrated Hi-speed tip. Cleared lysate entered the resin by gravity flow. The tip was washed with 20 ml buffer QC and allowed to flow through by gravity flow. The DNA was eluted by adding 5 ml buffer QF and collected in a tube with a minimum capacity of 15 ml. The DNA was precipitated out of the lysate by adding 3.5 ml room

temperature isopropanol. This was mixed, then incubated at room temperature for five minutes. During incubation the plunger from a 20 ml syringe was removed and a QIAprecipitator midi module attached to the outlet nozzle. This was placed over a waste bottle. The DNA was applied to the precipitator by syringing the eluate under constant pressure through the precipitator. This was then washed by syringing 2 ml 70% ethanol through the precipitator. The DNA was dried by syringing air through the precipitator. The precipitator outlet was dabbed dry with absorbent tissue to ensure no ethanol carry over. DNA was eluted by syringing 1 ml water through the precipitator. The eluted DNA was then passed through the precipitator a second time to ensure maximum yield. DNA concentration was estimated using a Beckman DU-7000 DNA quantitator (section 2.2.4.7.4). Elution of plasmid in water reduced the likelihood of inhibition in downstream experiments. However, the lack of buffering agents meant the DNA would degrade more rapidly, so the plasmid stocks were sub-aliquoted and stored at -20°C.

2.2.4.7.3 Preparation of glycerol stocks of plasmid

A starter culture was prepared (2.2.4.6). 1 ml of starter culture was inoculated into 50 ml media with the appropriate antibiotic (2.2.4.3). The culture was grown up until an optical density of 0.4-0.6 was reached. To 7 ml of culture, 3 ml of 50% glycerol was added and mixed. The culture was divided into aliquots (1 ml) and snap-frozen in liquid nitrogen before storing at -70°C.

2.2.4.7.4 DNA Quantitation

Plasmid concentrations were determined using a Beckman DU-7000 DNA quantitator using DNA diluted 1:200.

2.2.4.8 Sequencing of plasmid

Plasmids were sequenced using pGL3 primers according to the protocol (2.2.1.6.4) using approximately 200 ng plasmid DNA per reaction.

2.2.5 Transfection of eukaryotic cells and the luciferase assay

Transfection is the delivery of foreign molecules such as DNA and RNA into eukaryotic cells and is a powerful tool to study gene expression and gene regulatory elements. There are two types of transfection, transient (used in this study) and stable

transfection. In transient transfection many copies of the plasmid may be present in the cell but the construct does not integrate into the chromosome of the host. This results in high levels of expression of the induced gene but only for a few days. Typically, expression of transiently transfected cells should be analysed within twenty four to ninety six hours of transfection. In a stable transfection the DNA is incorporated into the chromosomal DNA of the host. Once incorporated the transfected gene will be replicated with the genome of the host with each cell division and so expression will continue indefinitely. We used Qiagen Effectene Transfection kit for all transfections. The kit contains a DNA condensation buffer, enhancer reagent and effectene reagent. Briefly, in the presence of condensation (EC) buffer the positively charged enhancer interacts with the negatively charged phosphate groups of the DNA and condenses it. Addition of effectene coats the condensed complex, producing effectene-DNA complexes (micelles) with a slightly positively charged surface. The micelles are mixed with the usual media and added to the cells. The positively charged micelles bind to the negatively charged surface of the cell membrane and are incorporated into the cell by endocytosis. In the first experiments the cells were transfected with two plasmids: the promoter–luciferase reporter prepared in sections 2.2.3 and 2.2.4 and *Renilla* luciferase control. First basal expression was assessed. Secondly the effect of four stimulations on promoter activity was investigated. The Dual-Luciferase® kit (Promega) was used to measure luciferase activity.

2.2.5.1 Transfection of plasmid constructs

Cells were counted, plated out into 6-well plates and grown overnight to reach ~40% confluence by morning (2.2.2.3). Seeding densities and cell passage are stated for specific experiments (section 5.2.3). Transfections were performed in triplicate unless otherwise stated. For each transfection a mastermix was prepared by mixing 1 µg plasmid DNA with 100 µl buffer EC, 300 ng *Renilla* luciferase plasmid and 8 µl enhancer (8:1 ratio)

The reagents were mixed by vortexing and then left at room temperature for three minutes, allowing the DNA in the condensation buffer (EC) to condense by interaction with the enhancer. Effectene reagent was added (10 µl/well) and the reaction again

mixed by vortexing. The reaction was left at room temperature for ten minutes, allowing the effectene to form micelles. The cells were prepared for transfection by removing the cell media and washing twice with phosphate buffered saline (1x) to remove debris. Usual media (1 ml) was then applied to the cells.

Usual media (600 µl per well) was added to the incubating mastermix. This mix was then applied to each well of cells (~650 µl/well) by gentle pipetting, ensuring even distribution of micelles. After six hours incubation at 37°C, the media was removed and 2 mls fresh media applied to each well. Treatments were applied thirty two hours post-transfection and left for sixteen hours. All cells were incubated for a total of forty eight hours post-transfection before harvesting. To harvest, media was removed and the cells washed twice with PBS (1x). Then 1 ml PBS (1x) was added to each well and the cells scraped from the plastic using a cell scraper. The cell solution was pipetted into a 1.5 ml eppendorf and spun down at 8000rpm for twenty seconds to form a pellet. The media was removed and 50 µl of passive lysis buffer (from Dual-Luciferase® kit (Promega)) was added to each pellet. Violent vortexing ensured pellet dispersal before freezing overnight at -20°C.

2.2.5.1.1 Preparation of treatments

IL-13 (lyophilised) was spun briefly and reconstituted in 20 mM HCl to a stock concentration of 10 µg/ml. Once completely dissolved the solution was divided into aliquots and frozen at -20°C. HCl (20 mM) was prepared from 12 M HCl (concentrated) stock. IFN-γ (lyophilised), was spun briefly and reconstituted in 1x PBS, pH8.0 to a stock concentration of 1 mg/ml. Once completely dissolved the solution was divided into aliquots and frozen at -20°C. PBS was prepared by dissolving in sterile water to make a 1x solution, which was then autoclaved. Phorbol myristate acetate was dissolved into DMSO to make a 5 mM stock solution. This was divided into aliquots and frozen at -20°C. Dexamethasone (water soluble) was dissolved in water to make a 25 mM stock solution. This was divided into aliquots and frozen at -20°C. Further dilutions to achieve required treatment concentrations are specified in section 5.2.3.1.1.

2.2.5.2 The Dual-Luciferase® reaction

The Dual-Luciferase® kit contains passive lysis buffer (5x) which is diluted to 1x with sterile distilled water before using, lyophilised Luciferase Assay Substrate which is reconstituted with Luciferase Assay Buffer II to make Luciferase Assay Reagent II (LAR II), Stop & Glo® buffer, and Stop & Glo® substrate (50x) (980 µl buffer:20 µl substrate). Firefly and *Renilla* luciferases have distinct evolutionary origins and so have dissimilar enzyme structures and substrate requirements. The differences allow independent measurement of stable luminescence from two reporter genes in a single sample. The lysed cell sample is mixed with LAR II, which contains luciferase substrate. Firefly luciferase does not require post-translational processing for enzymatic activity so functions as a reporter immediately upon translation. Photo emission is achieved through oxidation of beetle luciferin in a reaction that requires ATP, Mg^{2+} and O_2. The oxidation results in a flash of light that decays after substrate and enzyme are mixed. Three readings are recorded, using a luminometer, following addition of LAR II. The activity of the primary reporter (pGL3 promoter construct) is correlated with the activity of the co-transfected control reporter (*Renilla*) which provides an internal control to normalise results and reduce sample variability. Addition of Stop & Glo® reagent stops the Firefly reaction and commences the *Renilla* reaction. Two readings are taken following addition of Stop & Glo reagent.

2.2.5.3 Assessing luciferase activity

Kit reagents required reconstitution before use. To prepare the Luciferase Assay Reagent II (LAR II), the Luciferase Assay Substrate was mixed with the Luciferase Assay Buffer II according to manufacturer's instructions. LAR II does not tolerate repeated freeze thaw so it was divided into 1 ml aliquots and stored until required at -20°C.

The cells in lysis buffer (2.2.5.1) were removed from the freezer and defrosted on ice. The LAR II was defrosted on the bench until it reached room temperature. Stop & Glo® reagent was prepared by adding 980 µl Stop & Glo® buffer to 20 µl Stop & Glo® 50x reagent, according to the manufacturer's instructions.

For each sample LAR II (100 µl) was pipetted into a cuvette and placed in the luminometer. Cells in lysis buffer (20 µl) were mixed with this in the cuvette and three readings recorded. *Renilla* Stop & Glo® mix (100 µl) was added and two readings recorded. An average of the firefly luciferase, and *Renilla* luciferase was recorded. The ratio of firefly to *Renilla* luciferase was calculated to give a measure of relative luminescence.

2.2.5.4 Over expression of Sp1 plasmid

Results of luciferase assays, *ADAM33* homology and putative transcription binding sites (sections 5.2.4-5.2.5) suggested several possible binding sites for transcription factors, including six Sp1 sites. Cells were therefore transfected with promoter reporter plasmid and Sp1 expression vector to see if a change in expression occurred. Details are found in section 5.2.6.

Chapter 3

Generation of 5' nuclease assays for SNP genotyping of ADAM33

3.1 Introduction

The latest data from Ensembl (v36 Dec 2005) lists fewer genes than originally estimated, suggesting that gene regulation, and alternative splicing, is far more important than gene number in determining the diversity of protein forms and regulation of activity necessary for cell function. The genome contains around three billion base pairs. A variation in one of the bases, existing at a frequency of greater than 1% in the population, constitutes a single nucleotide polymorphism (SNP). These are estimated to occur at one in every 1 to 1.9 kb, with an estimated total of 60,000 SNPs falling within exons (coding and 3' untranslated regions) in the human genome [206].

Although more than 99% of human DNA sequences are the same across the population, variations in DNA sequence can have a major impact on how humans respond to disease; environmental insults (e.g. bacteria, viruses, toxins, chemicals), and also to drugs and other therapies. This makes identification of SNPs of great value for biomedical research and for developing pharmaceutical products or medical diagnostics. SNPs are generally evolutionarily stable (i.e. changing little from generation to generation) making them relatively easy to follow in population studies. The position of a SNP may have phenotypic consequences. Diseases (usually rare and monogenic) have now been identified where one SNP results in a change directly causing the disease phenotype. For example, in cystic fibrosis an allele change in the coding region results in a defective protein that affects transport of chloride ions across the cell membrane. This results in a change in salt balance affecting the

constitution of mucous in the lung [207]. An individual carrying two copies of the disease gene will display the disease phenotype. However, there are many different SNPs of the CF gene, all of which will produce a similar though not always identical clinical phenotype. It could be that a lone SNP does not cause disease *per se*, but helps determine the likelihood that someone will develop a particular disease. Often SNPs in several genes will affect the likelihood and severity of a disease. The polygenic nature of complex disorders such as asthma, diabetes, and heart disease makes genetic testing for them complicated. *ADAM33* is a highly polymorphic gene associated with asthma, though none to date have been identified as the true disease-causing SNP. Possibly SNPs are acting in synergy to confer susceptibility, or the true causative SNP is lying as yet undetected within the gene. It is also possible that the observed association is due to *ADAM33* SNPs being in LD with a causative allele completely outside of this gene region, although several successive studies support the original findings [37, 38, 181-184].

When studying complex diseases it is beneficial to look beyond the individual SNP to the haplotype. The word haplotype derives from 'haploid genotype' and refers to the particular set of alleles or SNPs present on a strand of DNA between two loci on the same chromosome, thereby defining the genetic constitution of an individual. Each individual has two copies of each chromosome and so has two haplotypes, one maternally, and one paternally inherited. The haplotype transmitted to offspring may be exact replicas of the parental haplotype, or may differ due to reshuffling during recombination events. Recombination events (also termed crossing over) occur in the pachytene stage (prophase) of meiosis I. Crossover involves the physical breakage of the double helix in one paternal and one maternal chromatid and joining of the maternal and paternal ends. Recombination is not uniform across the genome, with much variation between chromosomes and between sexes. For example, the frequency of recombination in female autosomes is 1.65 times higher than in males, with more recombination likely towards the centromeres in females and towards the telomeres of chromosomes in males [208]. Regions of high recombination are called recombination hotspots. SNPs that are close together have a reduced likelihood of separation by recombination so are likely to remain on the same strand. These

'linked' genes or markers are said to be in linkage disequilibrium. *ADAM33* displays high linkage disequilibrium (i.e. little recombination occurring). Among sequences of the same length the number of common haplotypes is minimal under low recombination, (high LD) [202]. Thus haplotype analysis of *ADAM33* should identify a minimal number of common haplotypes accounting for the majority of the population. Analysis of the haplotypes for linkage with disease phenotype such as BHR or atopy may highlight haplotypes, and alleles within the haplotype, which prevail in a given phenotype aiding identification of a causal SNP.

An efficient and reproducible genotyping method is required to investigate interactions between gene polymorphism and phenotype. Many methods of genotyping are available and some of these are discussed in section 1.5. For this work 5' allelic discrimination assay (Taqman®) was chosen. This method demonstrates high specificity and reproducibility (section 3.2.1.2). It uses small reaction volumes with minimal DNA concentrations (reducing reagent costs), and permits high-throughput analysis (384-well plate).

3.1.1 Aim

The aim of this work was to:

- Design Taqman® 5' nuclease discrimination assays for SNPs spanning *ADAM33*;
- Test the assays on DNA from a random Caucasian population to assess and compare observed allele frequencies with known frequencies. This ensures the observed (inferred) genotypes (strictly speaking these are phenotypes interpreted as genotypes) are in Hardy Weinberg equilibrium (HWE) suggesting the assays are accurate. Accuracy will be further verified using Sanger dideoxy sequencing; (but this would not detect null alleles)
- Identify common *ADAM33* haplotypes and the most informative SNPs marking those haplotypes, to reduce the number of SNPs required for genotyping studies.

3.2 Methods

3.2.1 Design of Taqman® 5' nuclease discrimination assays

3.2.1.1 Choice of SNPs

One hundred and five SNPs are listed on the dbSNP database (www.ncbi.nlm.nih.gov/projects/SNP/) at time of writing, though not all these have been experimentally verified. Thirty seven *ADAM33* SNPs had been investigated in the original study [79] and twenty three of these were chosen for 5' nuclease assay design. SNPs were chosen according to location (covering as much of the gene as possible), SNP frequency and with reference to previously observed association with BHR and asthma [79] (Table 3.1). Colleagues at the University of Manchester (Dr. Sally John, personal communication) had identified a further SNP, Bp1. This had been submitted by them to AbD for assay manufacture and was included in this study. Figure 1.3 illustrates the SNPs chosen for assay design and their location in the *ADAM33* gene. Table 3.1 lists these polymorphisms. For details of probe and primer sequence please refer to tables 2.3 and 2.4.

Table 3.1 ADAM33 Single Nucleotide Polymorphisms chosen for Taqman® 5' nuclease assay design

RSnumber	SNP	Location	Change	Case control analysis [79]
rs487377	Bp1	Intron	G/A	n/a
rs3918391	D1	ORF	T/C	-
rs3918392	F1	ORF	A/G	-
rs511898	Fp1	Intron	G/A	UK p=0.03
rs2271511	I1	ORF	C/T	US p=0.01
rs2280092	Lm1	Intron	G/A	US p=0.01
rs3918394	L1	ORF	C/T	-
rs3918395	Mp1	Intron	G/T	US p=0.01
rs612709	Qm1	Intron	C/T	UK p=0.04, Combined UK and US p=0.02
rs3918396	S1	ORF	G/A	UK p=0.03, Combined UK and US p=0.02
rs528557	S2	ORF	G/C	UK p=0.004
rs44707	STp4	Intron	A/C	UK p=0.02, Combined UK and US p=0.02
rs597980	STp5	Intron	C/T	-
rs574174	STp7	Intron	G/A	Combined UK and US p=0.02
rs2280091	T1	ORF	T/C	US p=0.003
rs2280090	T2	ORF	C/T	US p=0.02
rs630712	Tp2	Intron	T/G	-
rs628977	Vm3	Intron	C/T	-
rs628965	Vm2	Intron	A/G	-
rs543749	Vm1	Intron	G/T	UK p=0.01, Combined UK and US p=0.006
rs3918401	V1	ORF	A/T	-
rs3918400	V2	3' UTR	G/A	-
rs2787094	V4	3' UTR	C/G	UK p=0.03, Combined UK and US p=0.03
rs3746631	V5	3' UTR	T/C	-

Table 3.1 SNP names correspond to exon designation followed by SNP location relative to the nearest exon (p=plus m=minus). SNP Bp1 was identified as significant in a subsequent study (S. John, personal communication) and the assay was designed for this polymorphism using Assay by Design

(ABI) in collaboration with colleagues at University of Manchester. Case control analysis results are taken from the study by Van Eerdewegh et al [79] that found polymorphism in ADAM33 to be associated with BHR and Asthma, supporting choice of SNPs for assay development.

3.2.1.2 Principles of Taqman® TM 5' allelic discrimination assay

The 5' nuclease allelic discrimination is a high-throughput genotyping method combining fluorescence resonance energy transfer (FRET) with the polymerase chain reaction. Forward and reverse primers are designed, producing a product (ideally between 75-150 bp) which incorporates the SNP of interest. Two allele specific probes are also designed with one probe recognising one allele of the polymorphism and the other probe the second allele. Each probe has a reporter dye linked to the 5' end (Vic or 6-Fam in this case) and a quencher at the 3' [209]. Energy emitted from the excited reporter fluorophore (Vic or 6-Fam) is absorbed by the acceptor (quencher) This occurs up to distances of 70-100 A [119] and is known as fluorescence resonance energy transfer (FRET). Whilst the reporter and quencher are together on the probe (within this distance), no fluorescence is emitted. During the reaction, allele specific probes bind firmly to the complementary sequence whilst mismatched probes bind far less efficiently. The annealing temperature of the probe must be sufficiently higher than the annealing temperature of the primers, so that the probes anneal to their target site before primer extension occludes the probe-binding site [210]. On the 3' end there is a 'block' to prevent extension of the probe by polymerase. The probes had a minor groove binder (MGB) attachment at the 3' end. This binds to the minor groove of the DNA double helix, between the probe and the DNA template, increasing the melting temperature (TM) and so allowing reduction of probe length. This increases specificity, making the probe more accurate [211]. As the primers extend they come into contact with the probes. If the probes are firmly annealed, the 5' exonuclease activity of the polymerase displaces the probe and hydrolysis occurs at the phosphodiester bond joining the displaced (forked) region with the base paired portion of the strand. This results in cleavage of the probe [210]. The reporter and quencher are separated and so energy transfer between them stops, resulting in an increase in fluorescent emissions of the reporter dye. If the probe is incompletely annealed (mismatch) the polymerase dislodges the probe,

exonuclease activity does not occur, and the reporter and quencher remain in close proximity with no resulting fluorescence. The change in fluorescence measured is normalised against a background fluorophore (Rox). In the ABI7700HT instrument, measurement of fluorescent change is via light from an argon ion laser, which is sequentially directed to each well. The laser excites the fluorescent dyes present. The resulting fluorescent emission (500-650 nm) is collected via a series of lenses, filters and a dichroic mirror focus, on to a grate. The grating separates the light based on wavelength. SDS software (ABI, Warrington, UK) collects the fluorescent signal patterns via a camera and applies data analysis algorithms. The different dyes in the reaction emit different wavelengths of light and so can be distinguished from each other by changes in fluorescence [209, 210]. This can be viewed as a multi-component plot (real-time PCR) (e.g. figures 3.9 a-c) or at endpoint only. The endpoint data is viewed as a scatterplot (e.g. figure 3.9 d).

3.2.1.2.1 Assay by design and custom design assays

The reference sequence used was based on a 23574 bp sequence, accession AF466288 from BAC RP11-1098L22, and can be found at www.ncbi.nlm.nih.gov/. We used Applied Biosystems' automated system for design and manufacture of Taqman® 5' nuclease assays (Assay by Design (AbD)). Files were prepared and submitted to ABI as specified (section 2.2.1.1). Those failing QC1 or QC2 were custom designed using Primer select (Lasergene) software (section 2.2.1.2).

3.2.2 Testing of the assays

3.2.2.1 Choice of DNA

Human random Caucasian DNA (HRC) was obtained from the European Collection of Animal and Cell Culture (ECACC). There are five DNA arrays panels available, consisting of commercially produced high quality purified DNA from cell lines. The cell lines originate from a collection of around 500 lymphoblastoid cell lines derived from randomly selected healthy, Caucasian blood donors whose parents and grandparents were born in the UK or Ireland. Commercially produced DNA was used to provide uniformity of concentration and quality and, with derivation from a random Caucasian population, allele frequencies should be similar to that observed in the control

population used by Van Eerdewegh *et al [79].* In this study we used two panels (HRC-1 and HRC-2) providing 192 samples.

3.2.2.2 Assessing required DNA concentration

In order to achieve clear definition of genotype whilst minimising use of DNA, fourteen AbD assays, plus one custom assay, Tp2, were first tested using ten to sixteen random samples of ECACC DNA at concentrations of 4 ng, 8 ng and 16 ng using the protocol stated in section 2.2.1.4. This was to assess activity of the assays and to define the minimum DNA concentration required for accurate genotyping of the 192 samples. When investigating the effect of change in DNA concentrations on an assay, the same individual samples were used to allow direct comparison. Eight further custom design assays were tested on 8 ng DNA only.

3.2.2.3 Assessing the effect of dilution of the probe/primer sets

In order to reduce costs, an experiment was carried out that investigated how reducing the amount of probe/primer used in a reaction affected discrimination. Assay Fp1 was chosen because it was shown to give good discrimination. Probe/primer mix underwent dilution (1:2, 1:4, and 1:8). Reactions were set up as described (section 2.2.1.5.) using undiluted (control) and diluted probe/primer.

3.2.2.4 Testing the Taqman® 5' allelic discrimination assays using ECACC random samples

Following tests on small numbers of samples, the assays were used to test 192 DNA samples from ECACC (Human Random Control (HRC)-1 and HRC-2) using 8 ng dried DNA as described (section 2.2.1.3). Two 5' nuclease assays were performed on each 384-well plate and run on the ABI 7900HT instrument (section 2.2.1.5).

3.2.3 Validation of the assays

On completion of genotyping the assays were validated. First, observed genotype frequencies were compared with expected frequencies (observed in UK, US and combined populations by Van Eerdewegh *et al* [79]) and Pearson's Chi-squared test was used to detect significant differences. The observed ECACC allele frequencies were then used to work out expected genotype frequencies if random distribution of

alleles had occurred (i.e. in HWE). Observed and expected frequencies were compared using Pearson's Chi-squared test. Secondly, random samples, covering the three genotypes from each assay, were sequenced using Sanger dideoxy chemistry to verify that the assays were working accurately. Design of primers for amplification of the region and sequencing is described in section 2.2.1.6. The region containing the SNP was first amplified using PCR (section 2.2.1.6.1) and purified using SAP/Exo I purification technique (section 2.2.1.6.3). A sequencing reaction was set up according to section 2.2.1.6.4 using primers listed in table 2.5.

3.2.4 Identification of common haplotypes and informative *ADAM33* SNPs

Estimation of haplotype frequencies may be achieved by four methods

- Inferred on basis of allele frequencies e.g. using Expectation Maximisation algorithm (EM algorithm-phase unknown) as implemented in SNPHAP [200];
- Inferred from parental genotyping;
- Molecular haplotyping using for example Amplification Refractory Mutation System (ARMS) [118, 120];
- Sperm typing [212, 213].

EM will sometimes produce very low frequency haplotypes that are not actually present but for research questions based on which common haplotypes are important this method is adequate and is the method chosen here. SNPHAP was chosen for haplotype analysis in this case due to ease of use. Where knowledge of rare haplotypes is critical, molecular haplotyping is necessary to determine linkage phase (i.e. origin of transmission).

3.2.4.1 Identification of common haplotypes

The data was formatted into SPSS and imported into the SNPHAP program as described in section 2.2.1.7. Twenty SNPs were entered in the first instance. Following analysis using SNPtagger (section 3.2.4.2), and construction of an LD map (section 3.2.4.3), the number of SNPs was reduced to fourteen. The genotype data

from fourteen SNPs was then resubmitted to SNPHAP to define common fourteen-SNP haplotypes.

3.2.4.2 Identification of informative SNPs within the haplotypes

Haplotypes defined by common SNPs have important implications for mapping human disease genes. Haplotype tagging SNPs (htSNPs) are markers which capture most of the haplotypes in a region in LD. A preliminary sample is used to detect tagging SNPs, and the selected SNPs used to genotype a larger sample. Thus it is possible to retain most of the haplotype information by retaining a reduced subset of markers. This reduces the scale and cost of genotyping. The data set of twenty SNPs haplotypes were imported into the haplotype tagging program, SNPtagger [214], to identify haplotype tagging (ht) SNPs. Analysis was performed in association with S.J.Barton.

3.2.4.3 Linkage disequilibrium plot of the ADAM33 gene

The ECACC genotyping data provided allele frequency information to construct a pairwise LD plot (using D') of the *ADAM33* region. This use of D' is often justified on the grounds that it is related to the recombination rate. A D' value of 1 denotes complete disequilibrium, i.e. markers are not independent, and D' value of 0 denotes the markers to be independent of each other. A disadvantage of this measure is that it is upwardly biased in small samples. Markers which are not independent (i.e. in LD) will give duplicate information, making it unnecessary to type both markers in a study.

3.3 Results

3.3.1 Design, manufacture and testing of Taqman® 5' nuclease discrimination assays

Twenty three SNPs were submitted to ABI for design and manufacture using the Assay by Design (AbD) automated method. One SNP, Tp2, failed QC1, and eight further SNPs failed QC2. These were custom designed using primer express® software and submitted for manufacture.

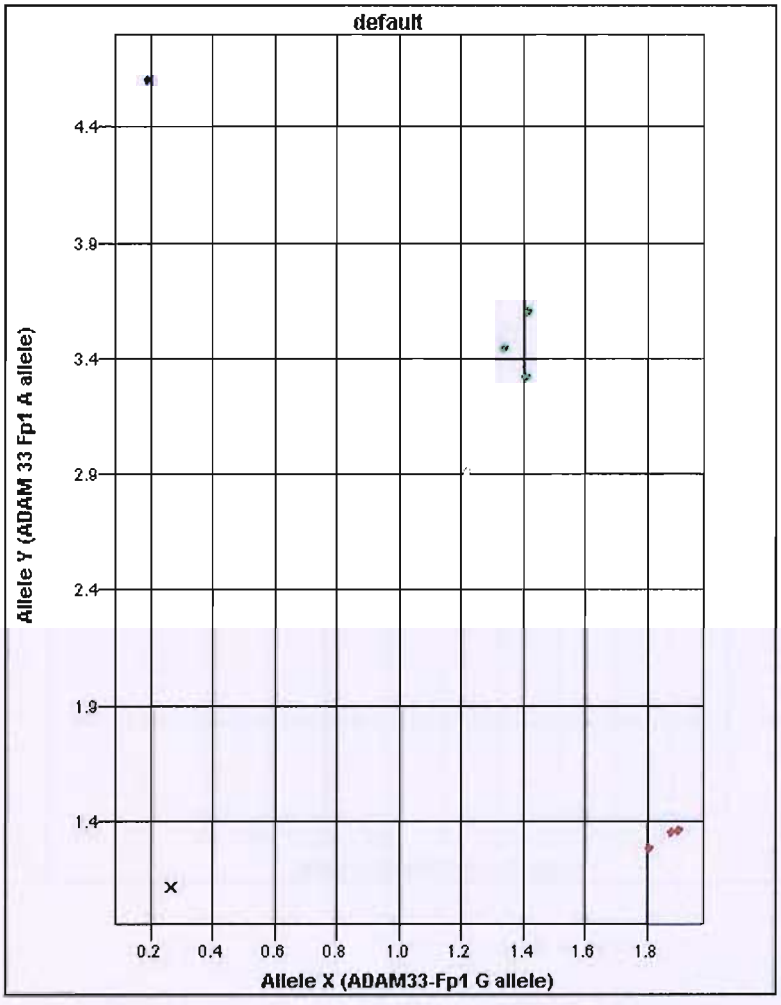
3.3.1.1 Assessing required DNA concentration and preliminary testing of the assays

The fourteen AbD assays, plus one custom assay (Tp2), were tested on ten to sixteen samples using three DNA concentrations (4 ng, 8 ng and 16 ng). The same individual samples were used when investigating the effect of a change in DNA concentrations on an assay to allow direct comparison. The aim was to assess assay activity and find an optimal concentration at which all assays performed well. This was needed so that DNA could be plated out using a robotic liquid handling machine, achieving consistency in pipetting and high-throughput.

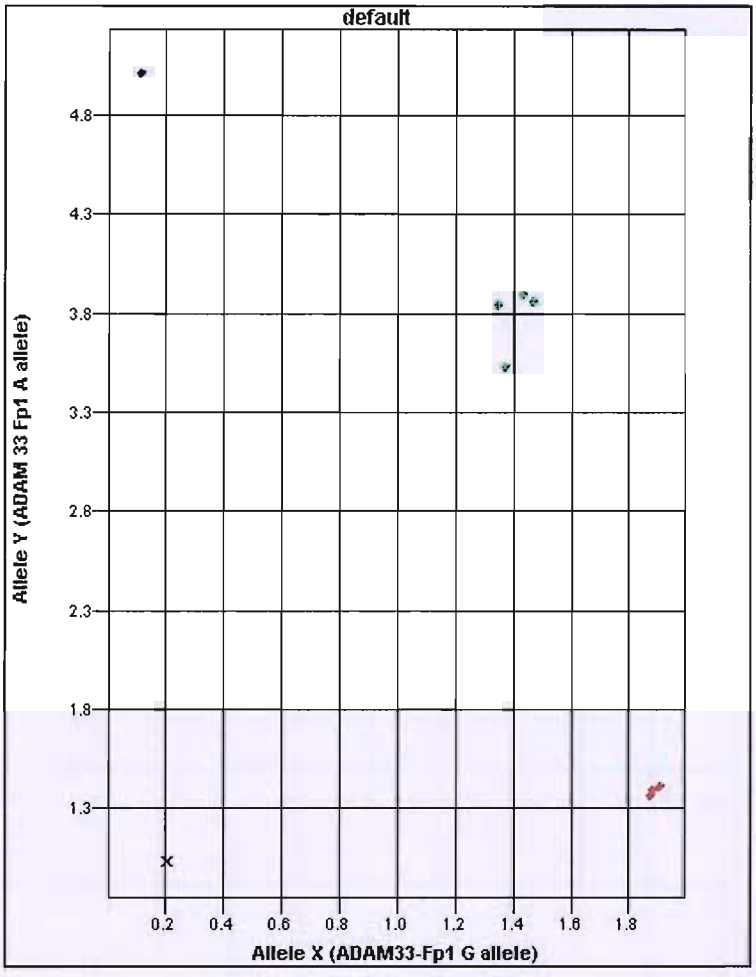
Increasing DNA concentration from 8 ng to 16 ng made little difference or was detrimental. Several assays (F1, L1, STp5, V2 and Tp2) showed reduced discrimination at 4 ng concentration. Figure 3.1 illustrates the effect of changing DNA concentration on the Fp1 assay.

Figure 3.1 The effect of a change in DNA concentration on discrimination of the Fp1 5' nuclease assay.

a) Discrimination of the Fp1 5' nuclease assay at 4ng DNA



b) Discrimination of the Fp1 5' nuclease assay at 8ng DNA



c) Discrimination of the Fp1 5' nuclease assay at 16ng DNA

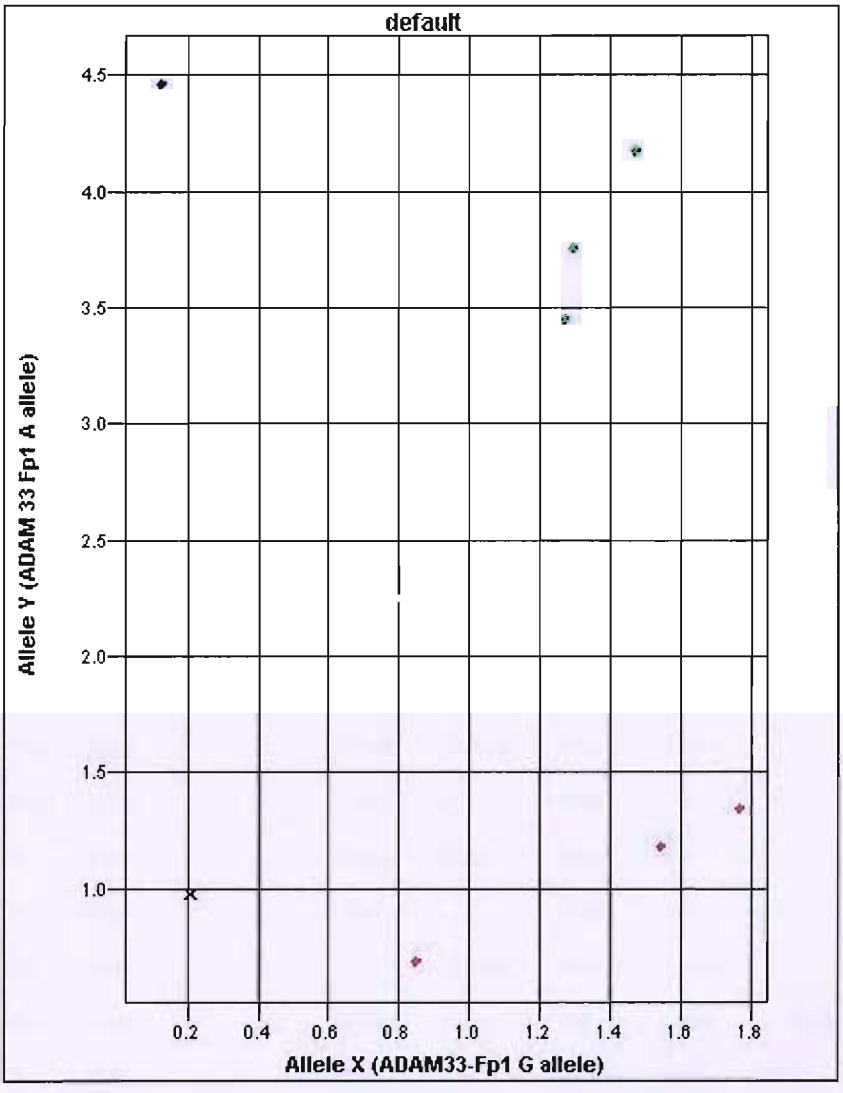


Figure 3.1 illustrates the Fp1 assay using a) 4 ng, b) 8 ng and c) 16 ng ECACC DNA. The assays use the same DNA samples for direct comparison. The clusters are seen to be tighter and more discrete in the 8 ng assay. The change in Vic fluorescence is comparable between a and b, but 6-Fam fluorescence is slightly increased in the 8 ng assay effectively tightening the clusters. The 16 ng assay has a greater spread of fluorescent signal with Vic ranging from 0.9 to 1.75 units and 6-Fam fluorescence at 4.5 units, resulting in a spread of the clusters. Allele x, red; allele y, blue; heterozygote, green. The negative control (NTC) is marked by a black cross and is comparable in all cases.

It was decided the most suitable concentration for genotyping of the ECACC population was 8 ng/µl. The remaining custom assays were tested using 8 ng of DNA. Table 3.2 summarises assay performance tests.

Table 3.2 Summary of assay performance

SNP	AbD QC1	AbD QC2	Test at 4 ng	Test at 8 ng	Test at 16 ng	Remanufacture by ABI	Custom design test at 8 ng
Bp1	University of Manchester		-	Pass	-	-	-
D1	Pass	Pass	Pass	Pass	Pass	-	-
F1	Pass	Pass	Unclear	Pass	Pass	-	-
Fp1	Pass	Pass	Pass	Pass	Pass	-	-
I1	Pass	Fail	-	Unclear	nt	-	Unclear
Lm1	Pass	Fail	-	Unclear	nt	-	Unclear
L1	Pass	Pass	Unclear	Pass	Pass	-	-
Mp1	Pass	Pass	-	Pass	Pass	-	-
Qm1	Pass	Pass	Pass	Pass	Pass	-	-
S1	Pass	Pass	Pass	Pass	Pass	-	-
S2	Pass	Fail	-	Unclear	nt	-	Unclear
STp4	Pass	Fail	-	Pass	nt	-	Pass
STp5	Pass	Pass	Unclear	Pass	Pass	-	-
STp7	Pass	Fail	-	Pass	nt	-	Pass
T1	Pass	Pass	Pass	Pass	nt	-	-
T2	Pass	Fail	-	Pass	nt	-	Pass
Tp2	Fail	-	Unclear	Pass	Pass	-	Pass
Vm3	Pass	Pass	Fail	Fail	Fail	Failed	Pass
Vm2	Pass	Fail	-	Pass	nt	-	Pass
Vm1	Pass	Pass	Pass	Pass	Pass	-	-
V1	Pass	Pass	Fail	Fail	Fail	Pass 8 ng	-
V2	Pass	Pass	Unclear	Pass	nt	-	-
V4	Pass	Pass	Pass	Pass	nt	-	-
V5	Pass	Fail	-	Pass	nt	-	Pass

Table 3.2 Fifteen assays (all AbD assays, plus Tp2 custom design) were tested to optimise DNA concentration. Comparable results were achieved at 8 ng and 16 ng but several of the assays (F1, L1, STp5, V2 and Tp2) showed reduced discrimination when 4 ng DNA was used. When investigating the effect of change in DNA concentrations the same individual samples were used for direct comparison.

In addition the assay submitted to AbD by the University of Manchester (Bp1) was tested at 8 ng and found to perform well. nt=not tested

Two of the fourteen AbD assays were unacceptable at any concentration (Vm3 and V1). These assays showed poor discrimination and low fluorescence. These were remanufactured and retested. Vm3 remained poor so was custom designed, making ten assays in total using custom design. Seven of the custom design assays worked adequately and three were unclear. Assay D1 was originally included as it encoded an amino acid change. However the expected frequency was less than 1%, so the assay was subsequently excluded. Another assay (Bp1), designed by ABI (assay by design) via University of Manchester was included.

3.3.1.2 Assessing the effect of diluting the probe/primer sets to reduce costs

Assay Fp1 (figure 3.2) was used to assess the effect of dilution on the probe/primer sets. A single homozygous sample was used at 8 ng DNA per well.

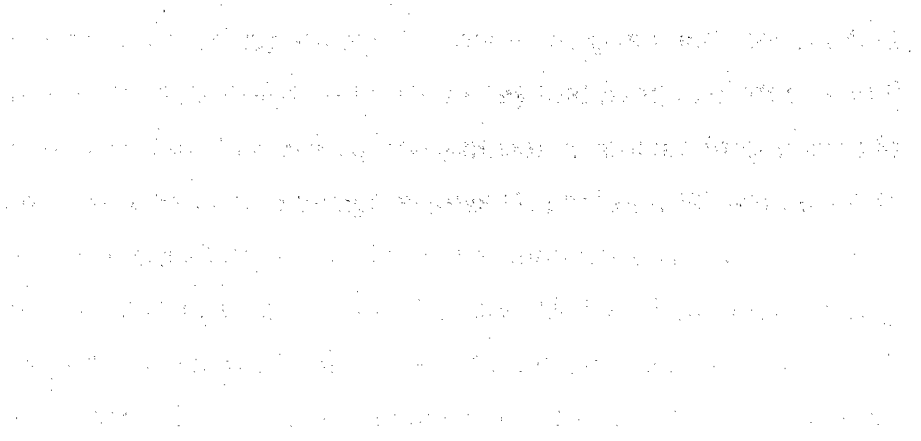


Figure 3.2 The effect of serial dilution on assay Fp1

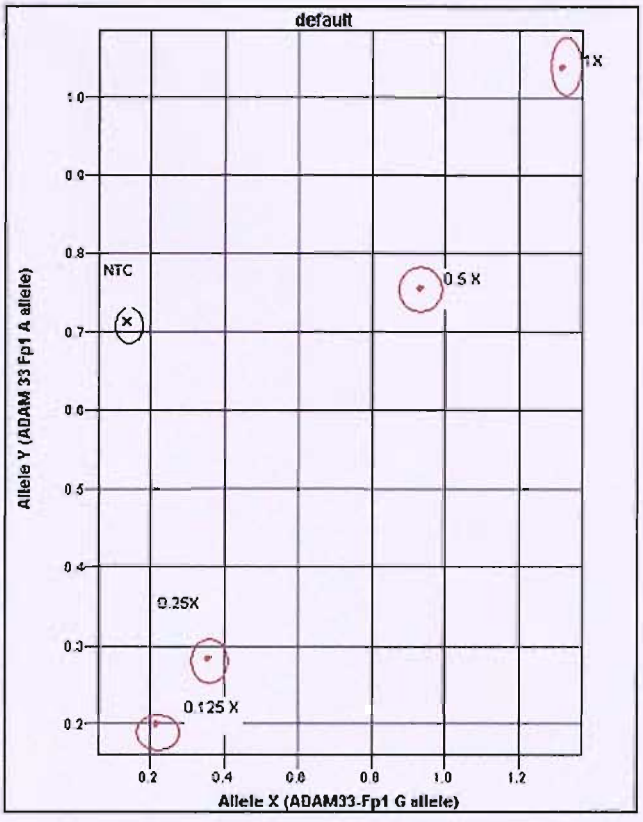


Figure 3.2 illustrates how serial dilution of the probe and primer sets reduces the efficiency of the assay. A dilution of 1:4 (0.25x) and 1:8 (0.12x) renders the assay unusable. However a 1:2 dilution, whilst reducing the signal does appear to give a reasonable signal. The negative control (NTC) is marked by a black cross.

In summary, twenty working assays had been designed and manufactured, three assays showed unclear results and one assay had been excluded. A further assay was included from the University of Manchester. Expected frequencies for the rare alleles of the unclear custom design assays I1, Lm1 and S2 were 0.17, 0.13 and 0.27 respectively. The possibility existed that the numbers used in the assay validation experiments were too low to observe the rarer alleles. Experiments using different DNA concentrations showed 8 ng DNA to be the best concentration for the majority of the assays. Probe/primer concentration would be used at 1x to maintain discrimination in the weaker assays. Therefore, all assays except D1 were used to

genotype the ECACC DNA samples. Figure 3.3 summarises the pathway of design and testing of the assays.

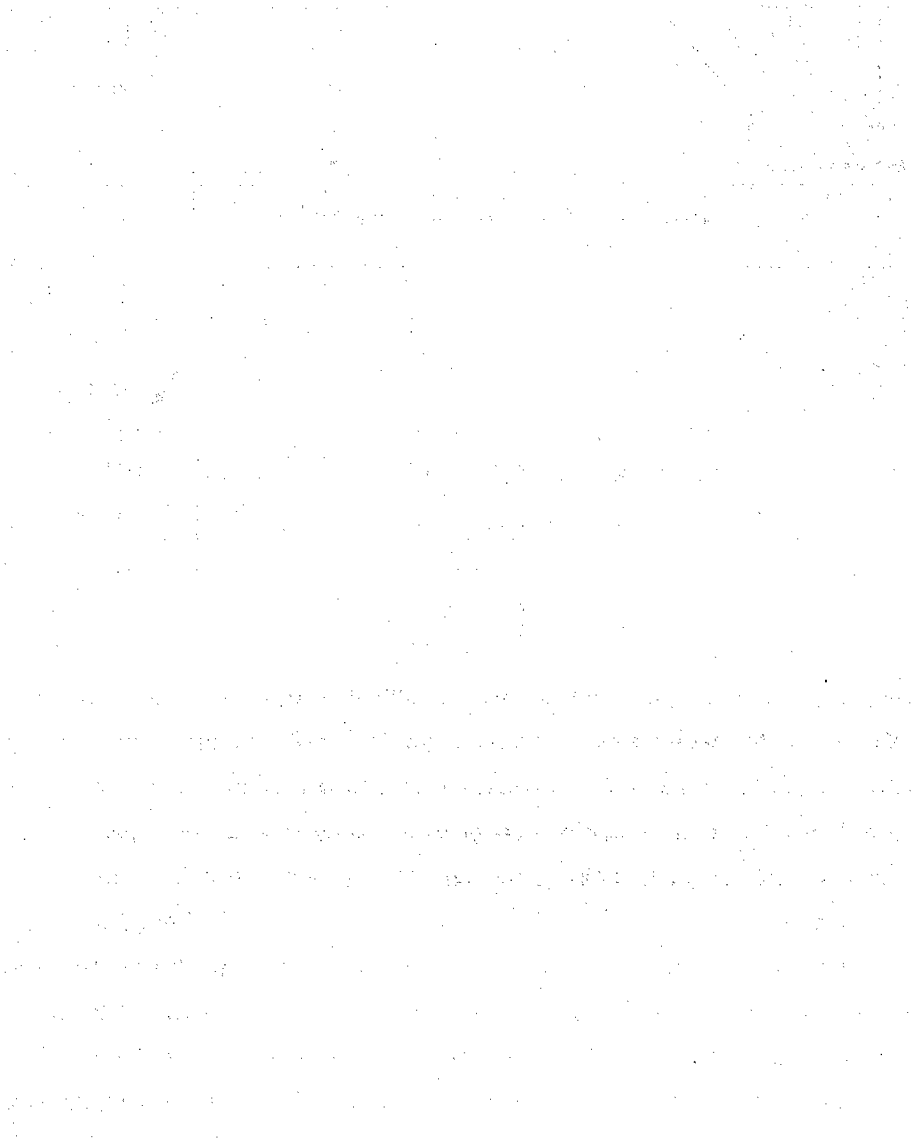


Figure 3.3 Summary of 5' nuclease assay manufacture route.

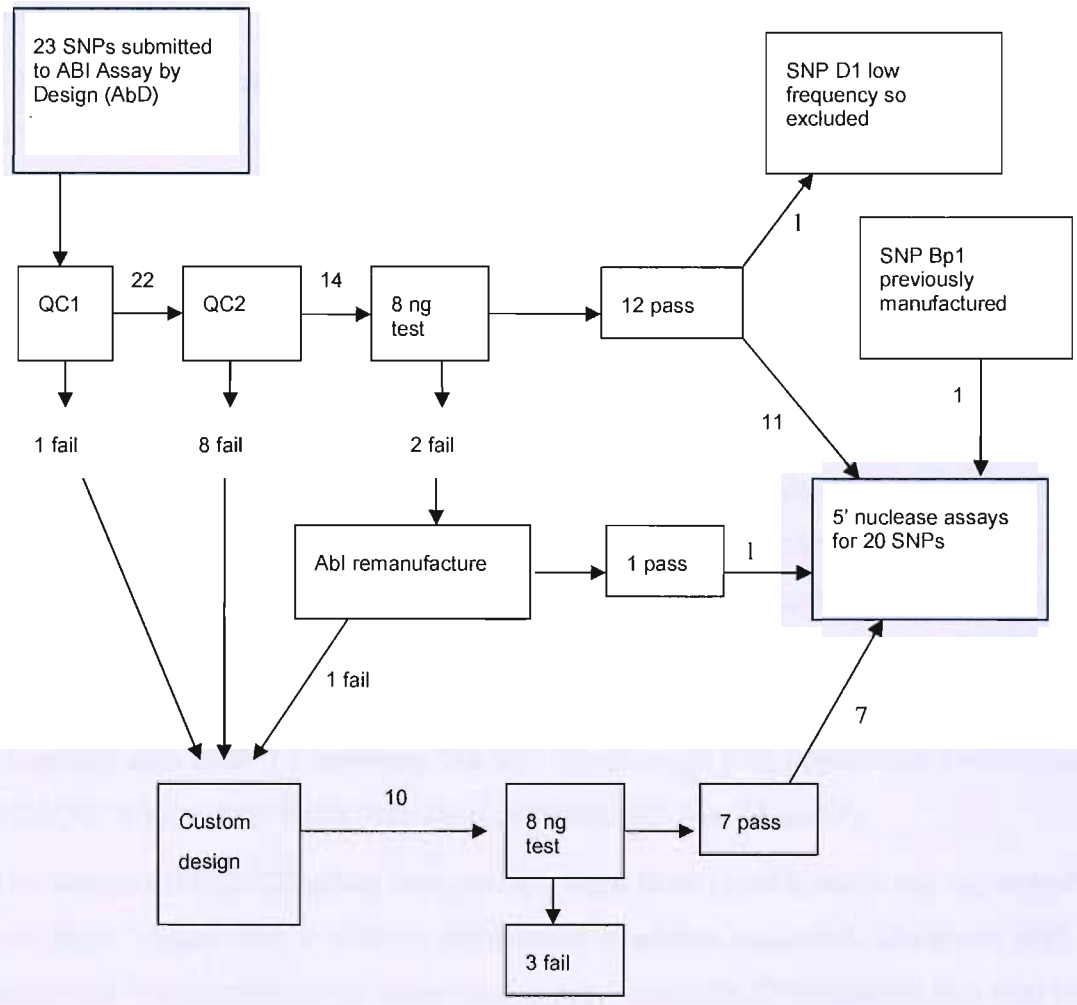


Figure 3.3 Twenty three SNPs were submitted to ABI for design and manufacture using the Assay by Design (AbD) automated method. SNP Tp2 failed QC1 and eight further SNPs failed QC2. These were custom designed and submitted for manufacture. Fourteen assays were manufactured via AbD. Two of these, (Vm3 and V1) were unacceptable showing poor discrimination and low fluorescence. These were remanufactured by ABI and retested. V1 was then passed as acceptable but Vm3 remained poor, so was custom designed. Ten assays were designed in total by custom design. Of these, seven worked adequately and three were unclear. However, since only ten to sixteen samples were used in the preliminary tests it was decided to include these assays in the random population screen (192 samples). SNP D1, encoding an amino acid change, was included in the original choice of SNPs but its very low frequency (<1%) makes it less informative, so it was excluded. Bp1, an assay designed by colleagues at University of Manchester, was included. This gave a total of twenty acceptable assays and three assays which were unclear but worth testing on a larger cohort.

3.3.2 Genotyping of a random Caucasian population using twenty three 5' nuclease assays

192 ECACC samples were genotyped using all the assays listed in table 3.2, except D1, as described (section 2.2.1.5). Custom design assays I1, S2 and Vm2 did not give readable results.

3.3.2.1 Validation of the assays using HWE

The observed genotypes were used to calculate observed allele frequencies. Allele frequencies cited by Van Eerdewegh et al [79], were used to calculate expected allele frequencies when HWE is observed in UK, US and combined populations used in that study. Using Pearson's Chi-squared test these values were compared with observed ECACC frequencies to see if the populations were significantly different. Results are shown in table 3.3 (columns 1, and 8-10). Several significant differences were observed between the UK, US and combined populations (SNPs F1, Fp1, Qm1, STp4, T1 and T2), when compared to the ECACC data. A significant difference was observed with SNP T1 between the Van Eerdewegh [79] population frequencies and ECACC, suggesting there may be a problem with the T1 assay.

The observed ECACC allele frequencies were then used to work out expected genotype frequencies if random distribution of alleles occurred. Observed and expected frequencies were compared using Pearson's Chi-squared test and results are listed in table 3.3 (columns 1-7). SNP T1 was significantly different from expected ($p=0.02$, χ^2 8.08), but not greatly. The scatterplot of assay T1 shows discrete clustering and good fluorescence (figure 3.10) and sequencing suggested the calls were accurate (figure 3.5).

Table 3.3 ECACC genotyping results, HWE calculation and comparison of observed ECACC frequencies with UK, US and combined populations used by Van Eerdewegh et al [79]

1	2	3	4	5	6	7	8	9	10
SNP	Total alleles/ samples	Allele frequency ECACC	Observed genotype	Observed calls ECACC	Expected calls if in HWE	χ^2 (2df)	UK expected	US expected	Combined expected
Bp1	380/190	G 0.79	G 301	GG 122	GG 119	1.77	-	-	-
		A 0.21	A 79	GA 57	GA 63	P=0.41			
				AA 11	AA 8				
F1	380/190	A 0.96	A 364	AA 174	AA 175	0.077	AA 183	AA 172	AA 179
		G 0.04	G 16	AG 16	AG 15	p=0.96	AG 7	AG 18	AG 11
				GG 0	GG 0		GG 0	GG 0	GG 0
							χ^2 **13.94**	χ^2 0.259	χ^2 2.64
							p=0.009	p=0.88	p=0.27
Fp1	382/191	G 0.70	G 267	GG 98	GG 93	2.17	GG 86	GG 86	GG 81
		A 0.30	A 115	GA 71	GA 80	p=0.34	GA 84	GA 84	GA 87
				AA 22	AA 18		AA 21	AA 21	AA 23
							χ^2 3.73	χ^2 3.73	χ^2 **6.55**
							p=0.15	p=0.15	**p=0.04**
I1	-	Failed assay	-	-	-	-	-	-	-
LM1	376/188	C 0.90	C 338	CC 150	CC 152	0.79	CC 142	CC 159	CC 149
		T 0.10	T 38	CT 38	CT 34	p=0.67	CT 43	CT 28	CT 37
				TT 0	TT 2		TT 3	TT 1	TT 2
							χ^2 2.62	χ^2 3.40	χ^2 0.51
							p=0.27	p=0.18	p=0.78
L1	382/191	G 0.99	G 378	GG 187	GG 187	0.00	GG 187	GG 187	GG 187
		A 0.01	A 4	GA 4	GA 4	P=1	GA 4	GA 4	GA 4
				AA 0	AA 0		AA 0	AA 0	AA 0
							χ^2 0.00	χ^2 0.00	χ^2 0.00
							p=1.0	p=1.0	p=1.0

Table 3.3 ECACC genotyping results, HWE calculation and comparison of observed ECACC frequencies with UK, US and combined populations used by Van Eerdewegh et al [79] (continued)

1	2	3	4	5	6	7	8	9	10
SNP	Total alleles/ samples	Allele frequency ECACC	Observed genotype	Observed calls ECACC	Expected calls if in HWE	χ^2 (2df)	UK expected	US expected	Combined expected
Mp1	378/189	C 0.90	C 340	CC 153	CC 153	0.00	CC 143	CC 160	CC 150
		A 0.10	A 38	CA 34	CA 34	P=1	CA 43	CA 28	CA 37
				AA 2	AA 2		AA 3	AA 1	AA 2
							χ^2 2.92	χ^2 2.59	χ^2 0.30
							p=0.34	p=0.27	p=0.859
Qm1	380/190	C 0.87	C 331	CC 147	CC 144	3.9	CC 140	CC 131	CC 137
		T 0.13	T 49	CT 37	CT 43	P=0.14	CT 46	CT 54	CT 49
				TT 6	TT 3		TT 4	TT 5	TT 4
							χ^2 3.11	χ^2 **7.5**	χ^2 4.67
							p=0.21	**p=0.02**	p=0.09
S1	378/189	G 0.90	G 339	GG 150	GG 153	1.03	GG 150	GG 153	GG 153
		A 0.10	A 39	GA 39	GA 34	P=0.60	GA 37	GA 34	GA 34
				AA 0	AA 2		AA 2	AA 2	AA 2
							χ^2 0.52	χ^2 1.029	χ^2 1.029
							p=0.77	p=0.59	p=0.59
S2	-	Failed assay	-	-	-	-	-	-	-
STp4	366/183	A 0.57	A 209	AA 60	AA 59	0.028	AA 42	AA 59	AA 50
		C 0.43	C 157	AC 89	AC 90	P=0.99	AC 91	AC 90	AC 91
				CC 34	CC 34		CC 50	CC 34	CC 42
							χ^2 **12.88**	χ^2 0.028	χ^2 3.57
							p=0.002	p=0.99	p=0.17
STp5	382/191	G 0.51	G 196	GG 49	GG 50	0.20	GG 60	GG 48	GG 56
		A 0.49	A 186	GA 98	GA 95	P=0.90	GA 94	GA 95	GA 95
				AA 44	AA 46		AA 37	AA 48	AA 40
							χ^2 3.51	χ^2 0.45	χ^2 1.37
							p=0.17	p=0.80	p=0.50

Table 3.3 ECACC genotyping results, HWE calculation and comparison of observed ECACC frequencies with UK, US and combined populations used by Van Eerdewegh et al [79] (continued)

1	2	3	4	5	6	7	8	9	10
SNP	Total alleles/ samples	Allele Frequency ECACC	Observed genotype	Observed calls ECACC	Expected calls if in HWE	χ^2 (2df)	UK expected	US expected	Combined expected
STp7	374/187	C 0.80	C 299	CC 121	CC 120	0.73	CC 117	CC 108	CC 114
		T 0.20	T 75	CT 57	CT 60	P=0.69	CT 62	CT 68	CT 64
				TT 9	TT 7		TT 8	TT 11	TT 9
							χ^2 0.66	χ^2 3.71	χ^2 1.19
							p=0.72	p=0.16	p=0.55
T1	382/191	T 0.88	T 338	TT 154	TT 148	**8.08**	TT 144	TT 162	TT 150
		C 0.12	C 44	TC 30	TC 40	**P=0.02**	TC 44	TC 28	TC 38
				CC 7	CC 3		CC 3	CC 1	CC 3
							χ^2 10.48	**χ^2 36.53**	**χ^2 7.12**
							p=0.005	**p=0.0001**	**p=0.03**
T2	380/190	C 0.89	C 340	CC 154	CC 151	2.73	CC 152	CC 163	CC 156
		T 0.11	T 40	CT 32	CT 37	P=0.25	CT 36	CT 26	CT 32
				TT 4	TT 2		TT 2	TT 1	TT 2
							χ^2 2.47	**χ^2 10.88**	χ^2 2.03
							p=0.29	**p=0.004**	p=0.36
Tp2	382/191	A 0.88	A 336	AA 149	AA 148	2.73	AA 146	AA 153	AA 149
		C 0.12	C 46	AC 38	AC 41	P=0.33	AC 42	AC 36	AC 39
				CC 4	CC 2		CC 3	CC 2	CC 3
							χ^2 0.78	χ^2 2.22	χ^2 0.36
							p=0.68	p=0.33	p=0.80
Vm3	380/190	C 0.62	C 235	CC 72	CC 73	0.03	CC 73	CC 78	CC 75
		T 0.38	T 145	CT 91	CT 90	P=0.99	CT 90	CT 87	CT 89
				TT 27	TT 27		TT 27	TT 25	TT 26
							χ^2 0.025	χ^2 0.80	χ^2 0.20
							p=0.99	p=0.67	p=0.90
Vm2	-	Failed assay	-	-	-	-	-	-	-

Table 3.3 ECACC genotyping results, HWE calculation and comparison of observed ECACC frequencies with UK, US and combined populations used by Van Eerdewegh et al [79] (continued)

1	2	3	4	5	6	7	8	9	10
SNP	Total alleles/ samples	Allele Frequency ECACC	Observed genotype	Observed calls ECACC	Expected calls if in HWE	χ^2 (2df)	UK expected	US expected	Combined expected
Vm1	380/190	G 0.88	G 336	G 148	G 147	0.34	G 140	G 131	G 138
		T 0.12	T 44	GT 40	GT 40	P=0.84	GT 46	GT 54	GT 48
				TT 2	TT 3		TT 4	TT 5	TT 4
							χ^2 2.24	**χ^2 7.63**	χ^2 3.058
							p=0.33	**p=0.02**	p=0.22
V1	380/190	A 0.94	A 357	AA 167	AA 168	0.05	AA 183	AA 168	AA 175
		T 0.06	T 23	AT 23	AT 21	P=0.97	AT 7	AT 21	AT 15
				TT 0	TT 1		TT 0	TT 1	TT 0
							χ^2 44	χ^2 0.054	χ^2 4.94
							p=0.0001	p=0.97	p=0.08
V2	378/189	C 0.96	C 364	CC 175	CC 174	0.08	CC 180	CC 167	CC 174
		T 0.04	T 14	CT 14	CT 15	P=0.96	CT 9	CT 21	CT 15
				TT 0	TT 0		TT 0	TT 1	TT 0
							χ^2 3.26	χ^2 3.43	χ^2 0.08
							p=0.19	p=0.18	p=0.96
V4	372/186	C 0.78	C 290	CC 109	CC 113	2.92	CC 106	CC 117	CC 109
		G 0.22	G 82	CG 72	CG 64	P=0.23	CG 69	CG 61	CG 67
				GG 5	GG 9		GG 11	GG 8	GG 10
							χ^2 3.48	χ^2 3.66	χ^2 2.87
							p=0.17	p=0.16	p=0.24
V5	368/184	A 0.96	A 353	AA 170	AA 170	0.08	AA 173	AA 166	AA 170
		G 0.04	G 15	AG 13	AG 14	P=0.96	AG 11	AG 8	AG 14
				GG 1	GG 0		GG 0	GG 0	GG 0
							χ^2 0.95	χ^2 1.04	χ^2 0.08
							p=0.62	p=0.59	p=0.96

Table 3.3 The Chi square (χ^2) test (2df) was used to compare frequencies of known data (Van Eerdewegh et al [79] columns 8-10) with the observed frequencies using ECACC DNA (columns 2-7). P-values statistically different from expected frequencies are in bold type. Custom design assays I1, S2 and Vm2 did not give readable results. df = degrees of freedom

3.3.2.2 Validation of the assays using sequencing

Sanger dideoxy chemistry was used to verify the accuracy of the Taqman® genotyping assays. Representative samples for each genotype were sequenced for each SNP where possible. Two examples are illustrated: figure 3.4 shows an example of sequencing for the Fp1 assay, and figure 3.5 the T1 assay. Sequencing confirmed the assays as reliable and accurate. The deviation from HWE that was observed in assay T1 was not large and may or may not be important. However, taken together with the elevation of the observed homozygote frequencies from those previously observed by Van Eerdewegh *et al* [79], this potentially indicates a bias in allele calling by the T1 assay. It is important to note that T1 is only 29 bp from the T2 polymorphism. This could potentially affect probe/primer binding. However, checks of the probe and primer sequences for the T1 assay show that they do not overlap with the T2 polymorphism.

Figure 3.4 Fp1 sequence verification

a

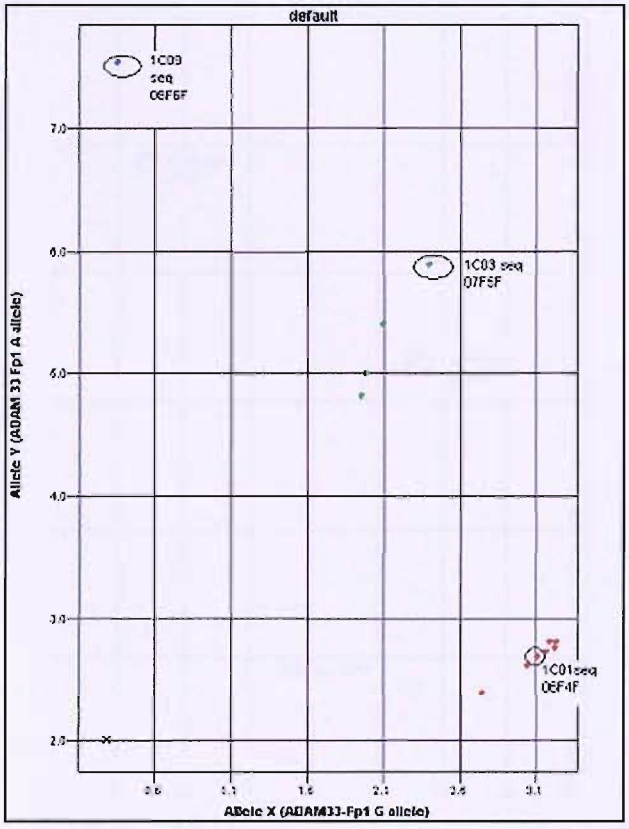


Figure 3.4 The three circled samples were sequenced according to the protocol. The scatterplot a) illustrates allelic discrimination using 5' nuclease assay. Allele x, red; allele y, blue; heterozygote, green; NTC is marked by a black cross. Sequencing of these samples is shown in b). The SNP position is marked by an arrow at position 275. The top sequence (064F4) reads G homozygote, the second (08F6F) reads A homozygote and the fourth sequence (07F5F) reads "n" (heterozygote). The top sequence written in black is the reference sequence taken from NCBI AF244688.

b

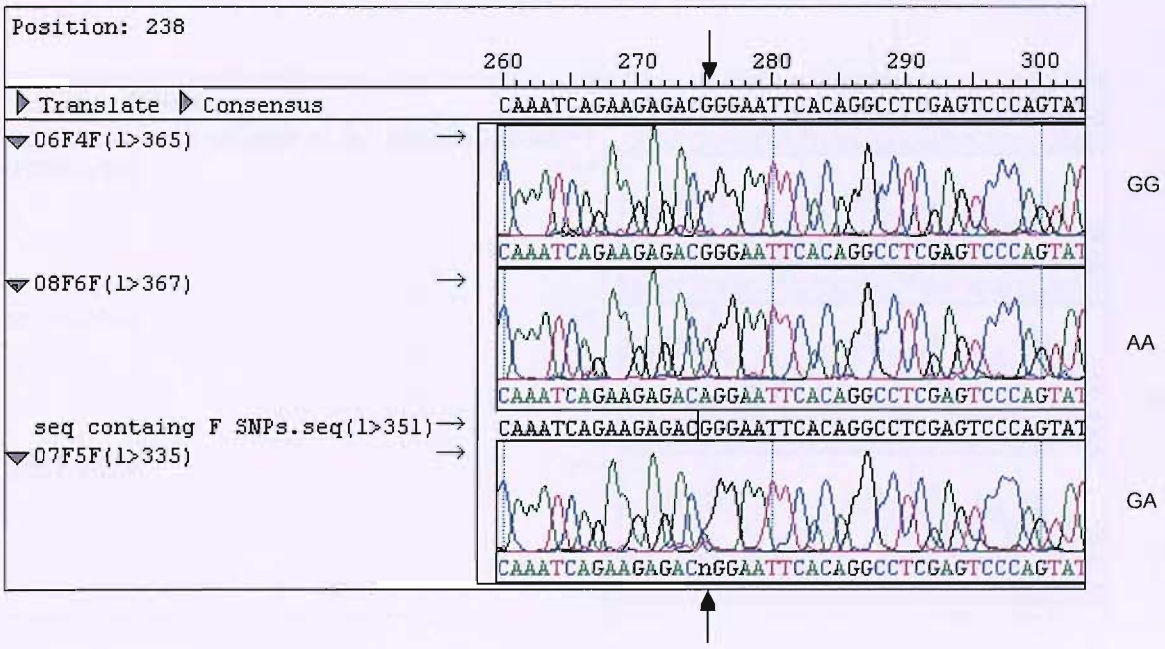


Figure 3.5 T1 sequence verification

a

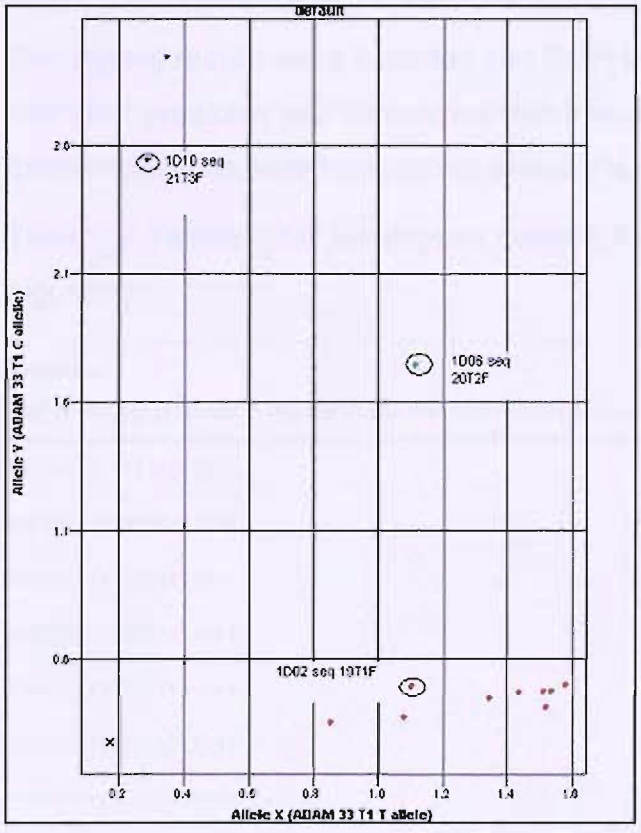
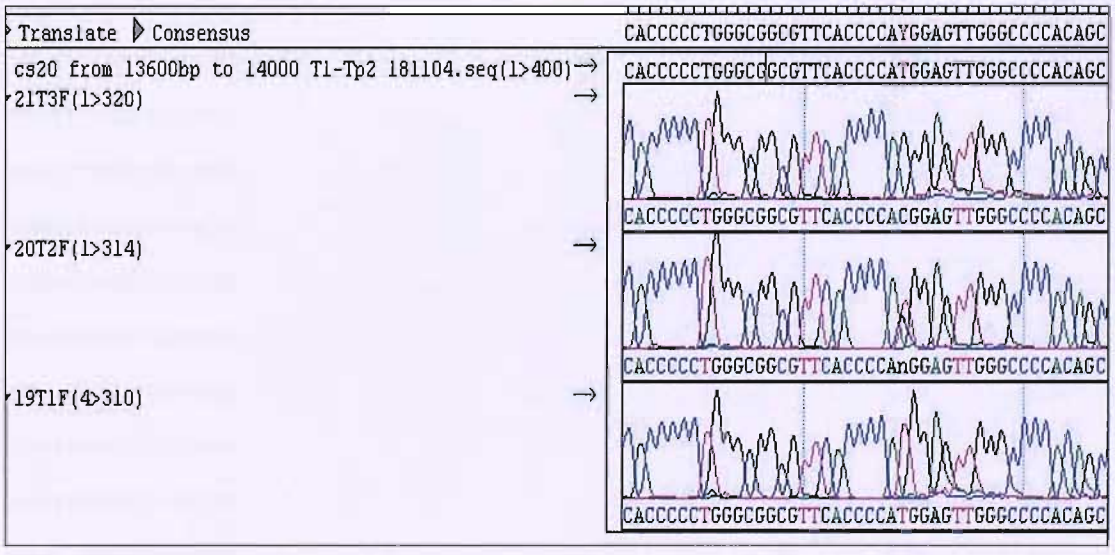


Figure 3.5 The three circled samples were sequenced according to the protocol. The scatterplot a) illustrates allelic discrimination using 5' nuclease assay. Allele x, red; allele y, blue; heterozygote, green; NTC is marked by a black cross. Sequencing of these samples is shown in b). An arrow marks the SNP position. The top sequence is the reference sequence taken from NCBI AF244688. The first sequence trace (21T3F) reads C homozygote, the second (20T2F) reads "n" heterozygote and the third sequence (07F5F) reads T homozygote.Sequencing suggests the assay is reliable

b



3.3.3 Identifying common haplotypes and informative SNPs

3.3.3.1 Identification of common twenty-SNP haplotypes

Genotyping results were imported into SNPHAP as described (section 2.2.1.7). SNPHAP predicted two haplotypes with frequencies of 5% and over, with twenty three haplotypes, with frequencies above 1%, accounting for 67% of the population.

Table 3.4 Twenty-SNP Haplotypes inferred from observed genotypes using an EM algorithm

Haplotype, **SNP order:** Bp1 F1 Fp1 Lm1 L1 Mp1 Qm1 S1 STp4 STp5 STp7 T1 T2 Tp2 Vm3 Vm1 V1 V2 V4 V6	**Predicted frequency**	**Cumulative frequency**
11111111121111111111	0.22	0.22
11111111211111211111	0.07	0.29
11211111111211111111	0.03	0.32
11221211111221111111	0.03	0.35
11111111111111111111	0.03	0.38
21111111211111211111	0.02	0.40
11111111221111211111	0.02	0.42
11211122122111121121	0.02	0.44
11111112122111111111	0.02	0.46
11111211121121111111	0.02	0.48
21111111222111211111	0.02	0.50
21111111221112211111	0.02	0.52
11211111111111111111	0.02	0.54
11211121122111121121	0.02	0.56
21111111211112211121	0.02	0.58
11111111211111212121	0.02	0.58
11111111121111112111	0.02	0.60
11211121121111121121	0.02	0.62
11111111211112211121	0.01	0.63
11211111221111211111	0.01	0.64

Table 3.4 Twenty-SNP Haplotypes inferred from observed genotypes using an EM algorithm (continued)

Haplotype, **SNP order:** Bp1 F1 Fp1 Lm1 L1 Mp1 Qm1 S1 STp4 STp5 STp7 T1 T2 Tp2 Vm3 Vm1 V1 V2 V4 V5	**Predicted frequency**	**Cumulative frequency**
21211111212111211111	0.01	0.65
21121211121121111111	0.01	0.66
21111111211112212121	0.01	0.67

Table 3.4 includes haplotype estimations of greater than 1% accounting for 67% of the population. 1 denotes the major allele, 2 denotes the minor allele. Allele frequencies are listed in table 3.3 column 3.

3.3.3.2 Identification of informative SNPs using SNPtagger

Results from htSNP analysis identified six ht-SNPs which are listed below;

Bp1

STp5

STp7

*(T2, Mp1, Lm1)

*(Vm3, STp4)

V4

Several SNPs offered the same information and these were designated corresponding ht-SNPs (* bracketed groups denote corresponding SNPs). Any SNP in a corresponding group offers the same information, so if assay design for one assay fails, the other SNP can be used instead. Only one SNP in a corresponding group was tagged. Analysis suggests that the information from examining six SNPs should be sufficient for genetic analysis across *ADAM33*, since these can be used to construct and represent all common haplotypes with a frequency of above 2%.

Table 3.5 Selection of haplotype-tag SNPs (ht-SNPs) from twenty SNP haplotypes using SNPtagger haplotype tagging software.

Haplotypes from htSNPs	Represented haplotypes from 20 SNPs SNP order: Bp1 F1 Fp1 Lm1 L1 Mp1 Qm1 S1 STp4 STp5 STp7 T1 T2 Tp2 Vm3 Vm1 V1 V2 V4 V5	Frequency	Cumulative frequency
121111	1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	0.22	0.22
111121	1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1	0.05	0.27
121121	1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1	0.03	0.30
111122	1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1	0.03	0.33
211122	2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1	0.03	0.36
111211	1 1 2 2 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1	0.03	0.39
121112	1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 2 1	0.03	0.42
122111	1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1	0.02	0.44
212121	2 1 2 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1	0.02	0.46
221121	2 1 1 1 1 1 1 1 2 2 1 1 1 2 2 1 1 1 1 1	0.02	0.48
111111	1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1	0.02	0.50

Table 3.5 Six-SNP haplotypes were constructed using ht-SNPs. These can be used to construct and represent all common haplotypes with a frequency above of 2%. ht-SNPs are underlined and in bold type.

3.3.3.3 Investigation of linkage disequilibrium across ADAM33

The Linkage Disequilibrium (LD) plot (figure 3.6), illustrates LD using the measure D'. Regions of complete LD are in black (D' =1) and it can be seen that a high level of LD exists across *ADAM33* with a few islands of lower LD.

3.3.3.4 Reducing the number of SNPs

Results from SNPtagger and LD were examined together to reduce the number of SNPs necessary for further genotyping. The efficiency of the 5' nuclease assays were taken into account when choosing between corresponding SNPs. Six SNPs were excluded from further analysis. These are listed, with rationale, in table 3.6.

Table 3.6 Rationale behind reduction in SNP numbers

Excluded SNP	Reasons for exclusion	Reference
L1	High LD. Low frequency	Figure 3.6, Table 3.3
Lm1	Corresponding with Mp1 and T2. Mp1 better assay	Section 3.3.3.2, Table 3.2
T2	Complete LD with Mp1 and Tp2 Mp1 better assay	Section 3.3.3.2, Figure 3.6
Tp2	Complete LD with Mp1 and T2. Mp1 better assay	Figure 3.6
V1	High LD. Low frequency	Figure 3.6, Table 3.3
V2	High LD. Low frequency	Figure 3.6, Table 3.3

Figure 3.6 Pairwise D' Linkage Disequilibrium plot across the ADAM33 gene

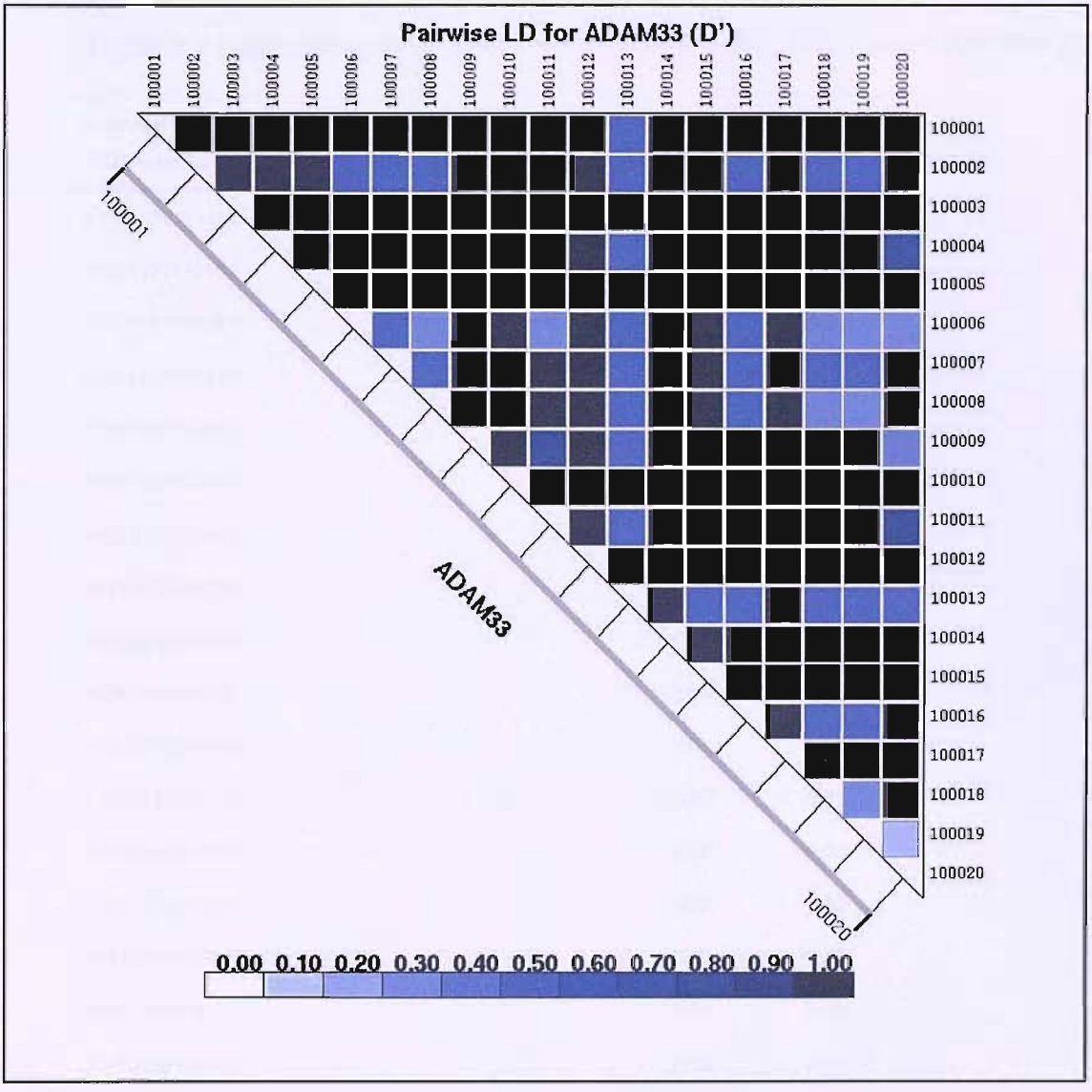


Figure 3.6 The map was constructed using LDPlotter (hhtp://innateimmunity.net//IIPGA2/Bioinformatics). This LD plot uses D' as a measure of LD. A value of 1 (black) denotes complete disequilibrium, i.e. markers are not independent, and 0 (white) denotes the markers are independent of each other.

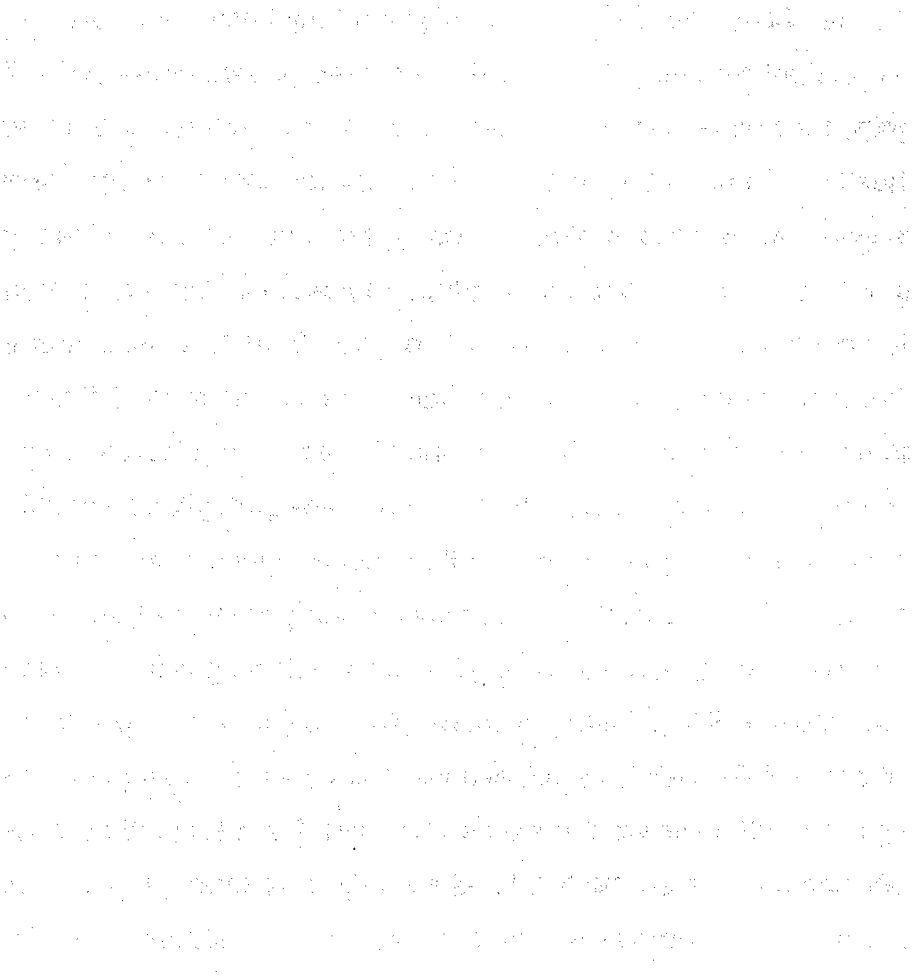
3.3.3.5 Haplotype analysis on fourteen SNPs using SNPHAP.

Table 3.7 Haplotypes observed in ECACC from 14 SNPs using SNPHAP [200]

Haplotype **SNP order:** Bp1 F1 Fp1 Mp1 Qm1 S1 STp4 STp5 STp7 T1 Vm3 Vm1 V4 V5	**Cumulative frequency**	**Predicted frequency**
11111112111111	0.22	0.22
11111121112111	0.07	0.29
21111121112121	0.06	0.35
11211111121111	0.05	0.40
11111121112121	0.05	0.45
11111122112111	0.03	0.48
11211111111111	0.03	0.51
21111122112111	0.02	0.53
11121112111111	0.02	0.55
11212112111221	0.02	0.57
11111212211111	0.02	0.59
11221111121111	0.02	0.61
11212112211221	0.02	0.63
21111121212111	0.02	0.65
11111212111111	0.02	0.67
11211121112111	0.02	0.69
21111112111111	0.02	0.71
21111122212111	0.02	0.73
11111111111121	0.02	0.75
21211121212111	0.02	0.77
11121121112111	0.01	0.78
11212212211221	0.01	0.79
11211121122111	0.01	0.80
11111211211111	0.01	0.81
21121112111111	0.01	0.82

Table 3.7 includes haplotype estimations of greater than 1% accounting for up to 82% of the population. Allele frequencies are listed in table 3.3, column 3. major allele=1, minor allele=2.

SNPHAP predicted five haplotypes with frequencies of 5 % and over, with twenty five haplotypes with frequencies over 1%, accounting for 83% of the population.



3.4 Discussion

3.4.1 Taqman® 5' nuclease allelic discrimination assay

The 5' nuclease assay (Taqman®) is a high-throughput method for identifying single SNPs in a large population [209]. In the past the cost has been relatively high when compared to other allelic discrimination assays and so use has been minimized. Recent improvements include use of reduced experimental volumes, thus reducing costs of reagents and generating renewed interest in the technique. Applied Biosystems (ABI) has developed a fully automated system called Assay by Design (AbD). The fully automated approach of AbD virtually removes the need for 'manpower' in the manufacture of these assays. The sequence containing the SNP of interest is submitted in a format which can be uploaded on to the ABI software. The rest of the design and manufacture process is automated involving only machinery. This reduces costs and increases throughput. In addition assays can be ordered in three different 'scales' of 1000, 3000 or 10000 reactions. Prices at time of writing are listed in table 3.8. Another (cheaper) option is to ABI's Assay on Demand (AoD). They have ready-manufactured probe and primer sets for many markers across the genome. Many investigators wishing to look at a particular region and not a specific SNP will find this very useful. However, this work involved looking at specific SNPs, none of which had an assay previously designed, so AoD could not be used. With so many probes containing similar sequences short, specific probes were needed, which would anneal only to one region. The probes supplied by ABI have an attachment at the 3' end of the probe called a minor groove binder (MGB). This binds to the minor groove between the probe and the DNA template, increasing the melting temperature (TM) and so allowing reduction of probe length whilst maintaining specificity, making the probe more accurate. It was decided to use MGB probes supplied by ABI.

Table 3.8 Prices and scales of AbD assays

Scale	Number of reactions	Cost	Cost per reaction
V scale	1000	£216	0.22p
S scale	3000	£404	0.13p
A scale	12000	£766	0.06p

Table 3.8 Prices listed above were correct at time of ordering (March 2004). Cost per reaction reduces as scale increases.

The assays are designed by a fully automated program which adheres strictly to a set of criteria dictating primer and probe length, percentage GC and melting temperature (Tm). Although significantly reducing costs, the scope for optimisation of assays is removed. If the program cannot find a set of probes and primers within the set criteria, the sequence is rejected (Failure at quality control 1 (QC1)). Custom design probe primer sets, using primer design software such as Primer express® or DNAstar, have a greater failure rate than those designed by AbD and are more expensive. Costs for Assay by design and custom design reactions are listed in tables 3.9 and 3.10.

Table 3.9 Cost of Taqman® 5' allelic discrimination assay using Assay by Design probe and primer set

Item	Cost	Order number	Cost per 384-well plate*
384-well optical plate	£106 for 50 plates	4309849	£0.90
Optical lid	£90.20 for 20 covers	4311971	£2.12
ABI qPCR mastermix (no Amp erase UNG)	£161.15 for 5 mls	4324018	£33.84
SNP specific probe and primers mix.	£404 for 3000 reactions	(S scale see section 2.2.1)	£56.82
		Total cost	£93.68
		Per 5 ul reaction	24p

Table 3.9 All reagents and plastics were supplied by Applied Biosystems.

**Costs include 10% extra mastermix to allow for pipetting errors and residual volumes and assume use of all probe/primer mix.*

Table 3.10 Cost of Taqman® 5' allelic discrimination assay using custom design probe and primer set

Item	Cost	Details	Cost per 384-well plate*
384-well optical plate	£106 for 50 plates	Order 4309849	£0.90
Optical lid	£90.20 for 20 covers	Order 4311971	£2.12
ABI qPCR mastermix (no Amp erase UNG)	£161.15 for 5mls	Order 4324018	£33.84
SNP specific probe	5000rxn £554		£92.84
Forward primer 100umol 1168ul (x10)	£9.40/25955	900 nmol/5 ul rxn needed 10 umol=10000 nmol (900/10000 nmol) x 5 ul =0.45 ul	£0.15
Reverse primer 100umol 1429ul (x10)	£8.46/31755	0.45ul/11680 ul=25955rxn 0.45ul/1429 ul=31755rxn	£0.11
		Total cost	£129.96
		Per 5 ul reaction	34p

*Table 3.10 All reagents and plastics were supplied by Applied Biosystems, except the custom primers which were supplied by Sigma-Genosys. *Costs include 10% extra mastermix to allow for pipetting errors and residual volumes and assume use of all probe/primer mix. rxn reaction*

The recent development of AbD and a reduction in reaction volume has made the 5' nuclease assay more attractive for general use. Despite the excellent results, Taqman® remains relatively expensive. The probe and primer sets make up 54% of the costs of each reaction. Assay Fp1 (figure 3.2) was used to assess the effect of dilution on the probe primer sets. The results suggest that dilution of the probe/primer set would be possible for robust assays with good discrimination, so reducing costs. However, some of the assays did not discriminate as clearly as Fp1, and dilution could result in diminished discrimination. It was therefore decided to use 1x probe/primer mix for genotyping.

3.4.1.1 Success of AbD and Custom probe/primer sets

A total of twenty three sequences were submitted to ABI for Assay by Design (not including Bp1, which was submitted by University of Manchester). Of these, one was rejected as not being suitable for design using the ABI software parameters (QC1).

Twenty two sequences passing quality control 1 (QC1) were accepted for design and manufacture by AbD. These were tested before dispatch on twenty random good quality DNA samples. Fourteen assays showed a change in fluorescence above 0.5 units and passed quality control 2 (QC2) (figure 3.3). Twelve of the fifteen assays that had passed QC2 worked well. Two assays, V1and Vm3 were unacceptable. Both assays were tested on 192 samples of ECACC DNA. Assay V1 had a good A allele (6-Fam) fluorescent signal, but poor T allele (Vic) signal (figure 3.7a). This assay was returned to ABI who remanufactured the probe and primer set under the guarantee. Figure 3.7b illustrates the new assay and shows a good change in fluorescence from both probes. Observed frequencies were statistically different from expected frequencies when compared to a U.K population (χ^2 44, p=0.001), but not in the US or combined populations [79]. The assay was in Hardy Weinberg equilibrium suggesting the assay was working and the difference could be due be population differences (table 3.3).

Figure 3.7 Assay V1

a) Scatterplot of failed assay V1

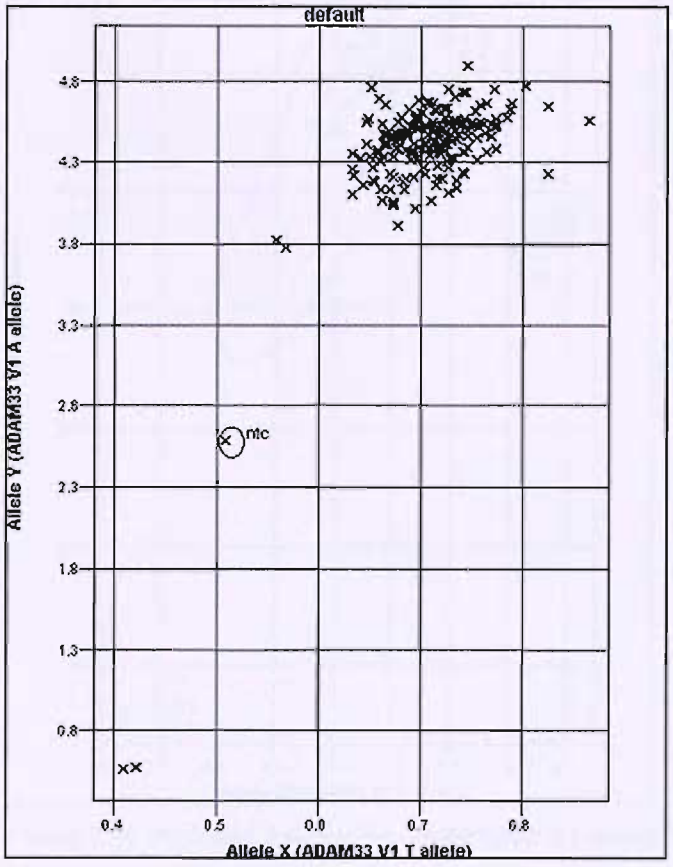


Figure 3.7a shows results from the original V1 assay. The A allele probe (Fam, y axis) fluorescence signal gives an acceptable emission but the T (Vic, x axis) allele is poor with a change in fluorescence not exceeding 0.4 units. The no template control (ntc) is further up the Y axis than expected, signifying either contamination with DNA, or an unstable probe. No calls could be made.

b) Assay V1 following remanufacture

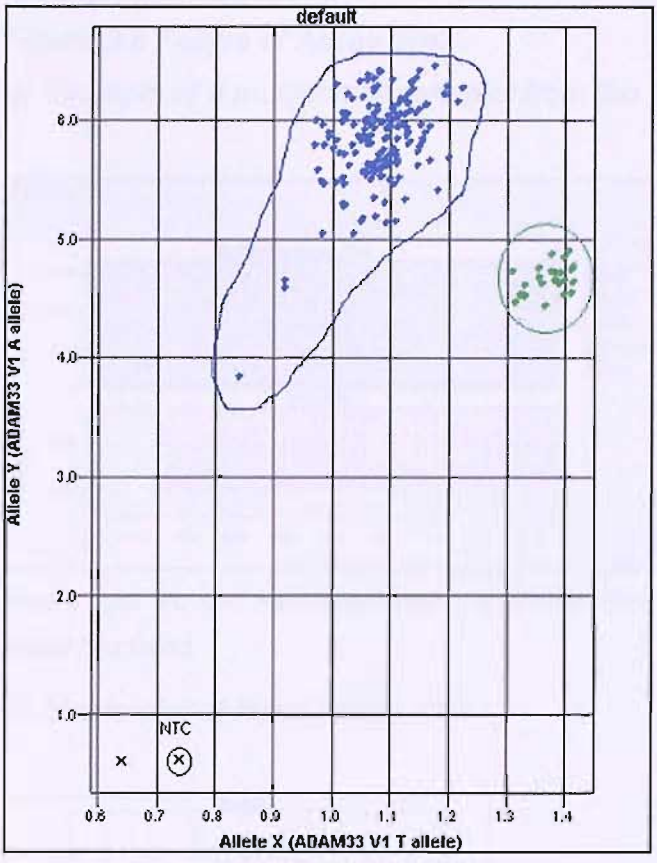


Figure 3.7b illustrates the newly manufactured V1 assay. There are no TT (Vic) homozygotes. This is a little surprising as the T allele has an expected frequency of 2% and 192 samples were genotyped. However, the AT heterozygote and AA homozygotes are tightly clustered with no template control (NTC), marked as a circled black cross, near the origin. A chi square test to compare observed frequencies with expected frequencies suggested that frequencies observed in this assay were statistically different to those expected in a UK population (χ^2 44, p=0.0001). When compared to a U.S. population and combined population observed frequencies were not statistically different. HWE was observed, so the discrepancy between frequencies could be due to population differences rather than assay failure. Allele x, red; allele y, blue; heterozygote, green; undetermined samples, black cross.

Assay Vm3 showed less than the minimum of 0.5 units change in fluorescence required for acceptance by ABI quality control, resulting in no allelic discrimination (figure 3.8). ABI re-tested their assay and found no problem with the design, and re-manufactured the assay under the guarantee. The new assay was no better so Vm3 probes were redesigned using the original primer set and manufactured by ABI at no

extra charge. The new assay was tested on the same DNA samples shown in figure 3.8, with much improved results (figures 3.9 a-d).

Figure 3.8 Failure of Assay Vm3

a) Example of a multicomponent plot from the failed Vm3 assay

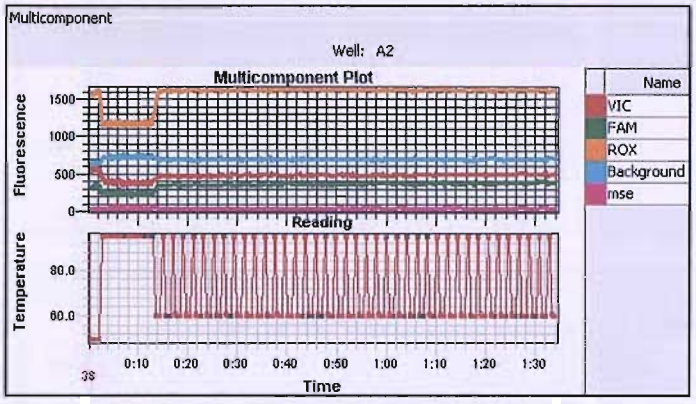


Figure 3.8a Vic and Fam amplification is absent. Rox remains constant but at the usual level. This assay has failed.

b) Scatterplot of failed assay Vm3

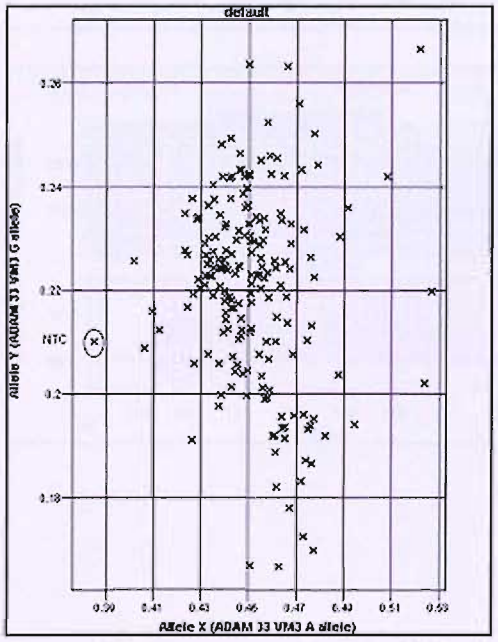
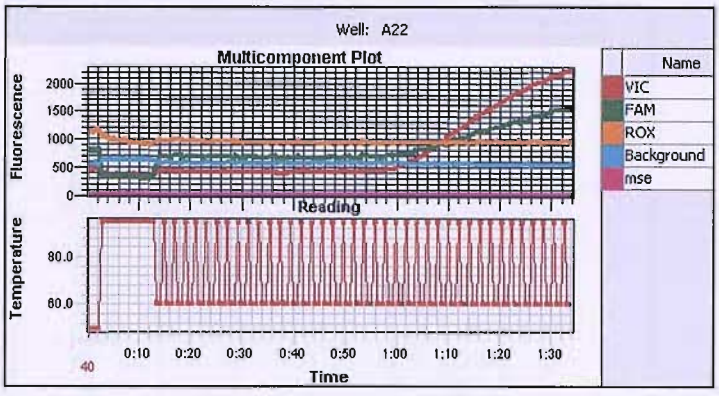


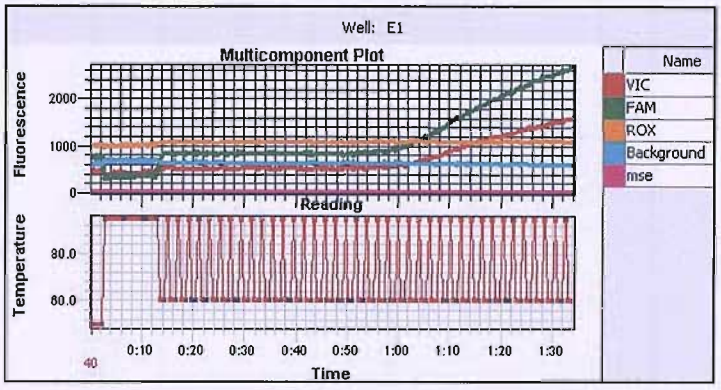
Figure 3.8 b shows a scatterplot of failed assay Vm3. Neither probe exhibits a fluorescence equal to or exceeding the 0.5 unit minimum change required to pass QC2. The no template control (NTC) is not near the origin. No discrete clusters are apparent.

Figure 3.9a-c Example of multicomponent plots from re-manufactured assay Vm3 illustrating genotypes a) heterozygote CT, b) homozygote TT (Fam) and c) homozygote CC (Vic)

a



b



c

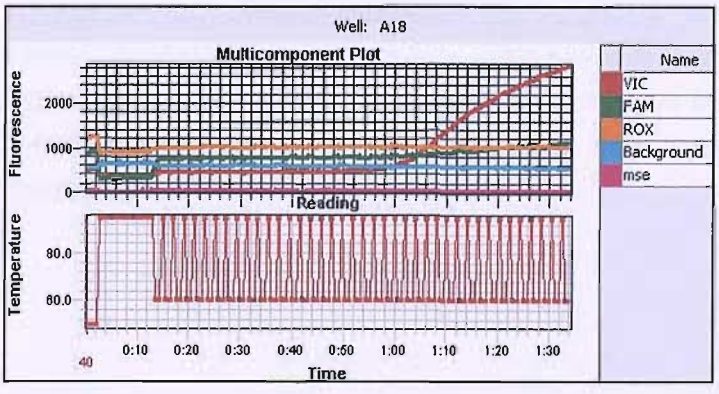


Figure 3.9. On receipt the newly designed assay was tested with the same DNA samples as before and both probes amplified successfully. The CT heterozygote (Figure 3.9 a) shows Vic (C) probe to have a higher gradient of amplification than Fam (T) resulting in fluorescence crossing the threshold at the same point, with the line plots crossing as the cycles continue. Figure 3.9 b (Fam TT displayed in green) illustrates a change in fluorescence in Vic and Fam with Fam amplification exceeding Vic. Vic is amplifying but with a lesser gradient than that seen in 3.9a. Figure 3.9 c illustrates CC homozygote (Vic (CC) displayed in red) and shows a far greater amplification gradient than figure 3.9b. Fam (T allele) amplification is nearly absent.

Figure 3.9d. Scatterplot of the remanufactured assay Vm3

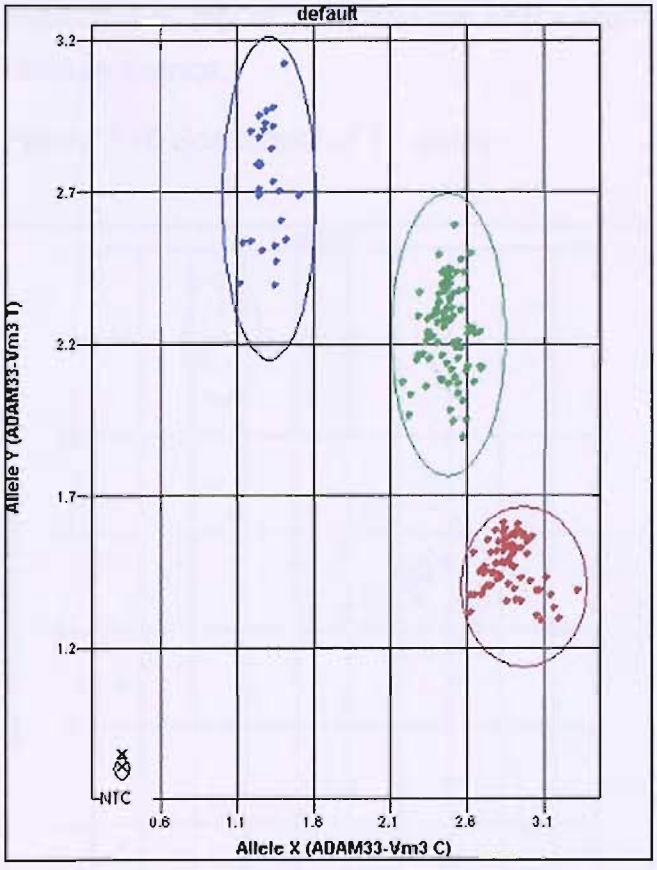


Figure 3.9d A scatterplot of the new assays reveals a change in Vic and Fam fluorescence far exceeding the ABI minimum (0.5 units). NTC, circled in black, is near the origin with the samples forming discrete clusters. Allele x, red; allele y, blue; heterozygote, green; undetermined samples, black cross.

Another assay, SNP T1, was significantly different from expected (p=0.02, χ^2 8.08). However, the scatterplot of assay T1 shows discrete clustering and good fluorescence (figure 3.10) and the so the slight deviation from HWE may be due to random chance.

Figure 3.10 Scatterplot of T1 assay

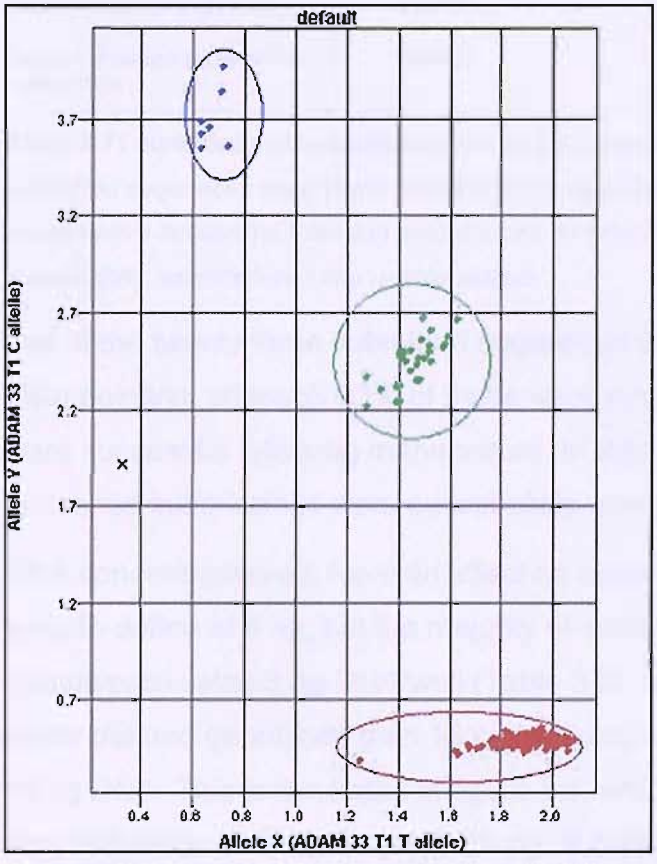


Figure 3.10 The scatterplot of assay T1 shows discrete clustering and good fluorescence. The negative control (marked x) shows higher fluorescence than expected suggesting possible contamination. Allele x, red; allele y, blue; heterozygote, green; NTC is marked by a black cross.

Only three out of the ten custom design assays did not achieve good results. The probe and primers were supplied in a 40x mix, so separate primer and probe optimisation was not an option. Assay D1 was excluded because of low frequency and another (Bp1) was included, making a total of twenty working assays. Figure 3.3 illustrates the pathway of design and table 3.11 summarises the success of the manufactured assays.

Table 3.11 Assay by design summary of submitted and received assays

SNPs submitted	AbD	Custom
Submitted	23 (100%)	10 (100%)
Passed QC1	22 (96%)	-
Passed QC2	14 (61%)	-
Success of assays submitted	13 (56%)	7 (70%)
Success of assays accepted for manufacture	13 (93%)	-

Table 3.11 summarises the success of the assays designed by AbD and custom design. 61% of the submitted sequences were manufactured and 93% of these were successful. 70% of the custom assays were successful following manufacture. In total 87% of the original submission sequences were successfully manufactured into usable assays.

Out of the twenty three submitted sequences 61% were manufactured, which was disappointing, although 93% of these were successful. Of the custom assays, 70% were successful following manufacture. In total, twenty of the twenty three original sequence submissions were successfully manufactured (87%).

DNA concentration did have an effect on assay reliability. Some assays were not easy to define at 4 ng, but the majority of assays undefined at 4 ng could be called (genotyped) using 8 ng DNA/well (Table 3.2). In some cases 8 ng/well assays had better defined genotypes than 16 ng/well and no assay was markedly improved using 16 ng DNA. This is illustrated in figure 3.1, which compares assay Fp1 using DNA concentrations of 4 ng, 8 ng and 16 ng. A possible reason behind this may be a consequential increase in inhibitors present in the reaction when the amount of DNA is increased. TE (Tris, EDTA) for example is often used to dissolve DNA following extraction. EDTA chelates magnesium ions from the reaction. Magnesium ions are required to aid unwinding of DNA strands to allow attachment of polymerase. A reduction in available magnesium ions would reduce the yield of the PCR reaction, so affecting the fluorescent signal. An experiment was performed to see if reducing the amount of probe/primer mix was an option, in order to reduce costs. Results suggested that in some of the assays (such as Fp1) it would be possible to reduce the amount of probe by half and still maintain adequate discrimination (as long as

DNA concentration remained consistent). Some of the assays were not as robust as Fp1 and so, to maintain consistency in our genotyping, the probe/primer mix was used at manufacturer's recommended concentrations (section 2.2.1.4).

3.4.1.2 Reliability of Taqman® 5' nuclease reaction.

The 5' nuclease assay is reported to be a repeatable, sensitive and specific assay giving consistent results using small amounts of DNA template (1-20 ng) [209]. Assays worked best at 8 ng and gave reliable results (table 3.11). Pipetting inaccuracies and well-to-well discrepancies between reporters were accounted for using a passive reference dye called Rox, which is a component of the mastermix. During the real-time run, spectral data (fluorescent signal of between 500 nm and 660 nm), is collected by the ABI 7900HT software (SDS). Multicomponenting is the term used for distinguishing the contribution each individual dye and background component makes to the composite fluorescent spectrum detected by the 7900HT instrument. The composite spectrum is made up of Vic, 6-Fam, Rox and background fluorescence. The reporter signal, i.e. Vic and 6-Fam, is calculated by dividing by the fluorescent signal of the passive reference dye (Rox), thus normalising the data. The multicomponent data could be used to identify problems with the assays, i.e. if one of the probes was not active or was performing poorly. Amplification of the individual probes was viewed to identify how many cycles were needed before crossing the threshold (CT). Each assay had a unique pattern. Comparison of signals from a definite call with those of the dubious call may aid genotype calling. Figure 3.11 illustrates an example of a series of multicomponent plots with the corresponding scatterplot following end-point analysis. It highlights the distinct pattern seen in this assay.

Figure 3.11a-c Multicomponent plots and scatterplot from STp5 assay illustrating the 'pattern of amplification' observed in samples falling within the main cluster

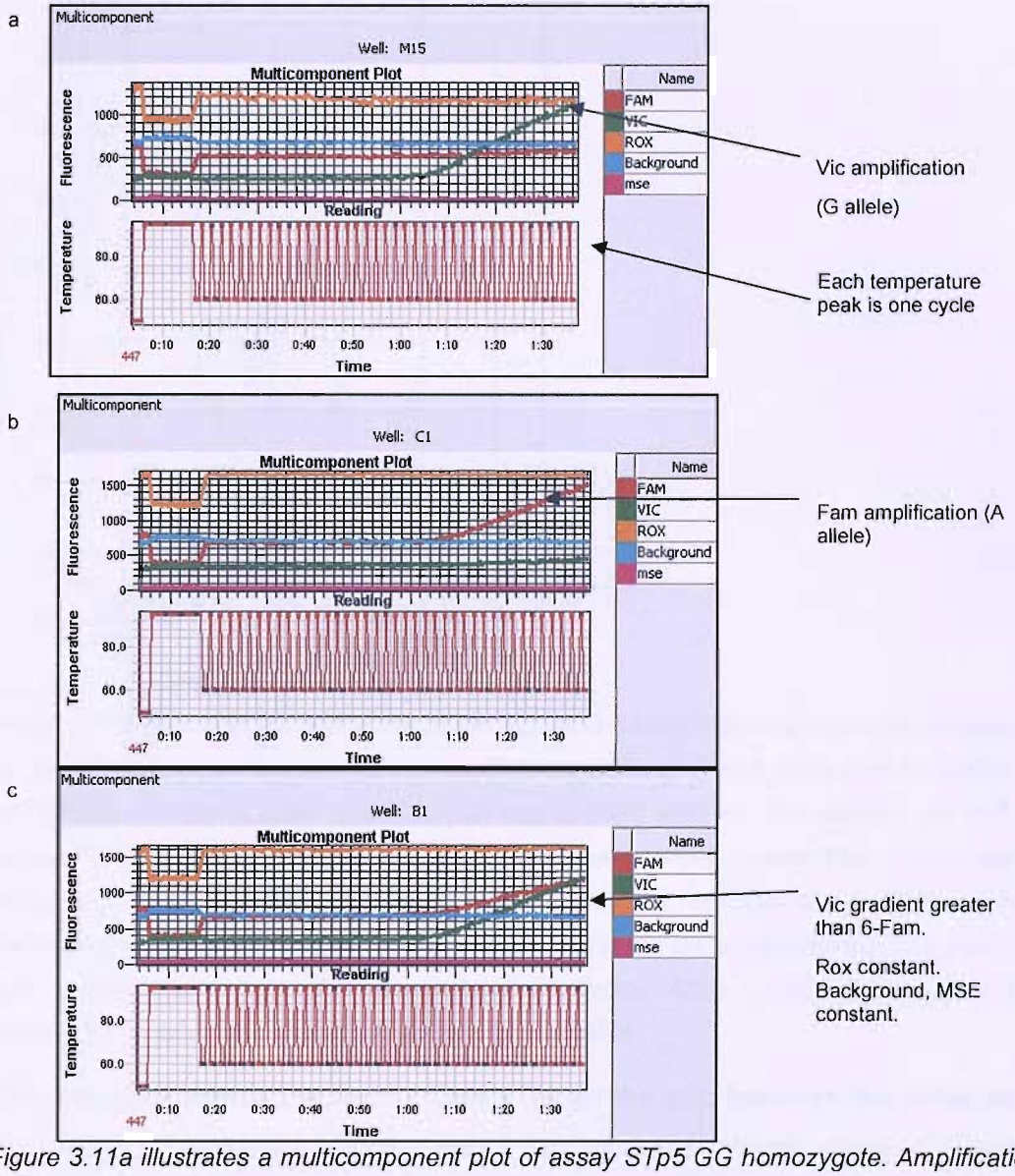


Figure 3.11a illustrates a multicomponent plot of assay STp5 GG homozygote. Amplification of the Vic signal (G allele) is in green. 6-Fam (A allele), in red, remains flat. The internal control (Rox) is constant. Figure 3.11b illustrates a multicomponent plot of assay STp5 AA homozygote. Amplification of 6-Fam signal (A allele) is in red and Vic (G allele) is in green and remains flat. The internal control (Rox) is again constant. Figure 3.11c illustrates a multicomponent plot of assay STp5 GA heterozygote. Amplification of 6-Fam signal (A allele) is in red and Vic (G allele) in green. Both signals amplify, with Vic having a slightly higher gradient.

Figure 3.11 d Scatterplot of 5' nuclease allelic discrimination assay, STp5

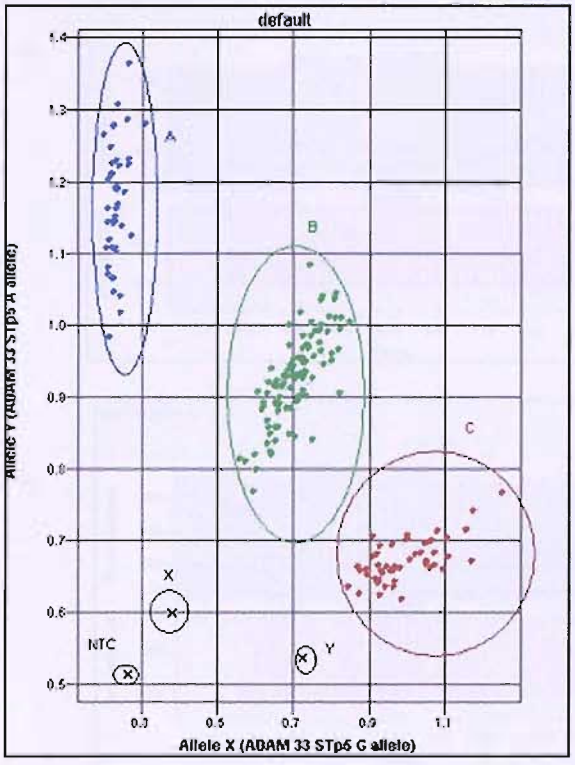
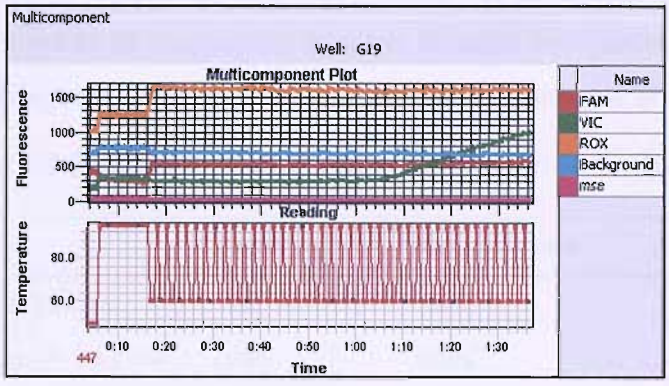


Figure 3.11d illustrates the scatterplot of the ADAM33 STp5 (G>A intronic polymorphism) probe/primer set (multicomponent plots are illustrated in Figures 3.10a-c). The A allele specific probe was labelled with 6-Fam and the G allele specific probe was labelled with Vic. The clusters are well defined with cluster A identifying homozygote AA (high 6-Fam fluorescence), cluster B identifying heterozygote AG (~equal 6-Fam/Vic fluorescence) and cluster C identifying homozygote GG (high Vic fluorescence). Two samples (X and Y), fall outside the main clusters and are undetermined. This may be a result of poor amplification or lower DNA template concentration. Allele x, red; allele y, blue; heterozygote, green. The negative control (NTC) is circled and labelled.

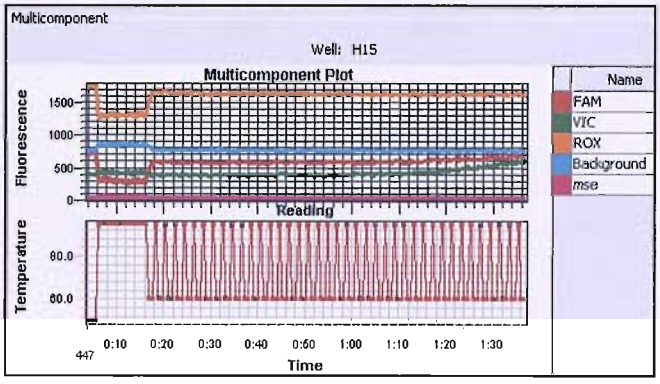
The samples outside the main clusters fall on the axis between the negative template control (NTC) and the main clusters (X with cluster B, Y with cluster C), suggesting the genotype to be called. The multicomponent plots for these two samples (figure 3.12 a-c) were observed for similar amplification patterns to those observed for the definite calls. This allowed y to be called whilst sample x remained ambiguous.

Figure 3.12 Multicomponent plots of calls outside of the main cluster for assay STp5

a)



b)



c)

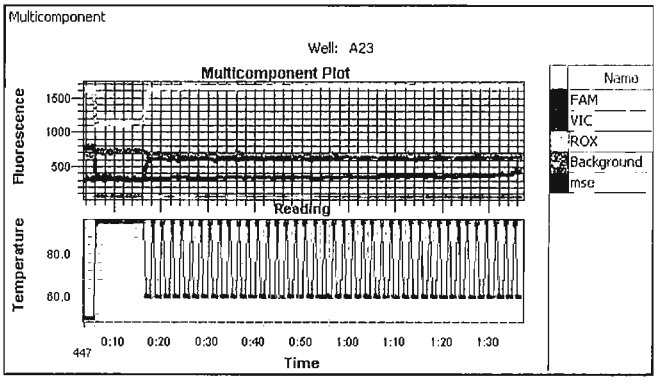


Figure 3.12 illustrates multicomponent plots of points X, Y and NTC from the previous scatterplot (figure 3.11). 3.12a shows point X. Only Vic fluorescence crosses the threshold, suggesting that X is GG homozygote. Note the totally flat 6-Fam (A allele) signal. Figure 3.8 b shows point Y, with very poor amplification of both signals. This may be a heterozygote, but cannot be called with confidence. Figure 3.11c shows the negative control with little or no fluorescent signal.

All samples that were sequenced confirmed genotyping results (examples in figures 3.3 and 3.4). This suggests 100% accuracy for called samples. Over 95% of samples were called in all successful assays. Results are summarised in table 3.12. Samples with ambiguous calls were excluded from haplotype analysis.

Table 3.12 Summary of allele discrimination in individual assays

SNP	Total calls	Successful calls
Fp1 L1 T1 Tp2 STp5	191	100%
Bp1 F1 Qm1 T2 Vm3 Vm1 V1	190	99.4%
Mp1 S1 V2	189	98.9%
LM1	188	94.7%
STp7	187	97.9%
V4	186	97.4%
V5	184	96.3%
STp4	183	95.8%
Vm2	62 (fail)	32.4%
I1 S2	Fail	0%

3.4.2 Common haplotypes, informative SNPs and linkage disequilibrium

SNPHAP predicted two major twenty-SNP haplotypes in the ECACC data at 22% and 7% frequency respectively. Individual haplotypes as well as the haplotype frequencies are predicted for each sample. Individual calls have not been included here, since the information will not be used for phenotypic analysis. The haplotypes included in the results are those with an estimated frequency above 1% (i.e. in 192 samples 1 to 2 samples would be expected to have that haplotype). Widely used methods to measure LD include D' and r^2. Both these measures depend on the degree of association between pairs of markers but differ according to the way they depend on marginal allele frequencies. Despite this, it would be expected that, in a given population, D' and r^2 would give comparable results. This is not found to be so, in particular there seems to be much more random variation in values of D' with increasing distance [215]. Possibly this is due to varying recombination rates across

the genome, making genetic distances and physical distances incomparable. Rho (ρ) is a better measure of LD, since it takes into account physical distances as well as recombination rates [201]. The extent of linkage disequilibrium was assessed and an LD map constructed using Rho (ρ). The measure of LD was used in conjunction with SNPtagger to identify SNPs which were important in defining the commonest haplotypes. A high level of LD is shown across *ADAM33*, with several islands of reduced LD. Identification of ht-SNPs and corresponding SNPs correlated with the LD pattern observed, allowing confident exclusion of SNPs without losing genetic information. Six SNPs were subsequently excluded, leaving fourteen SNPs. SNPHAP predicted five fourteen-SNP haplotypes with frequencies of 5% and over, and twenty five haplotypes with frequencies over 1%, accounting for 83% of the population.

3.5 Concluding remarks and future work.

Twenty Taqman® 5' nuclease assays were successfully manufactured and tested for *ADAM33* genotyping. Reliability of the assays was assessed by the frequency observed (compared with those observed by Van Eerdewegh et al [79]), whether Hardy Weinberg equilibrium was obeyed and by Sanger dideoxy sequencing. Sequencing confirmed the high reliability of this genotyping method with 100% accuracy for well clustered calls. Assays tended to work very well, with discrimination of 95-100% of samples (table 3.12). Three worked very poorly with less than 33% discrimination. The SNPHAP haplotype determination program was used to define possible haplotypes of *ADAM33* and identified two common twenty-SNP haplotypes. Both were at a low frequency, probably due to the level of linkage disequilibrium observed in this region. The possibility exists that the association seen between *ADAM33* and asthma phenotypes could emanate from a polymorphism outside *ADAM33*, perhaps in one of the flanking genes Sialoadhesin (SN) and GDNF family receptor alpha 4 (GFRA4), and be in LD with *ADAM33* polymorphisms. Future work should include design of 5' nuclease assays outside the *ADAM33* gene, extending into these neighbouring gene regions to confirm the linkage signal observed by Van Eerdewegh et al [79]. These SNP assays will be used in further genotyping studies (Chapter 4).

Chapter 4

Genetic Polymorphism of the A Disintegrin and Metalloprotease gene (*ADAM 33*) and early life origins of Asthma

4.1 Introduction

4.1.1 Implication of *ADAM33* with early life origins of asthma

Polymorphisms in *ADAM33* have been associated with asthma, bronchial hyperresponsiveness (BHR) and related phenotypes in several populations [79, 181-184, 186]. ADAM33 is a transmembrane protein belonging to the family of matrixins and adamalysins (section 1.6.1). This family are the principle agents responsible for extracellular matrix degradation and remodelling, playing important roles in development, wound healing, and the pathology of diseases such as arthritis and cancer [125]. Functions of the ADAMs family include proteolytic activity, cell adhesion and signalling (section 1.6.3). However, the role of many ADAMs, including ADAM33, is not clearly defined. The selective expression of ADAM33 in mesenchymal cells [79, 171], its association with BHR [79, 182] and expression studies [161, 172], strongly suggests that alterations in its activity may underlie abnormalities in the function of airway smooth muscle cells and fibroblasts, leading to pathological airway remodelling in asthma, perhaps explaining its genetic association with BHR. Expression of the asthmatic phenotype is believed to be dependent on genetic predisposition and environmental interactions. Possible risk factors for asthma include amongst many, environmental tobacco smoke (ETS), allergen exposure, breastfeeding/nutritional factors, obesity, family size and socioeconomic factors [216]. The combination of these influences in concert with genetic make-up ultimately determines whether asthma develops or not, and at what age it manifests. There is

increasing evidence to support the notion that certain windows of opportunity exist during life that allow the action of environmental factors. The unification of Germany in 1989 provided a unique opportunity to study populations living under markedly different economical and environmental circumstances, but with similar genetic properties. For example, at unification (1989) the prevalence of asthma, allergic rhinitis and atopic sensitisation was significantly higher in West Germany than East [217]. A follow up study in 1996 revealed that atopic sensitisation, but not asthma, had increased significantly in the East. The children in the second study were born three years before unification and were therefore exposed to western living conditions only after their third birthday. Thus factors operating in early life may be important for the acquisition of childhood asthma, whereas atopic sensitisation and hay fever may also be affected by environmental factors beyond infancy [218]. Airway remodelling and inflammation have been shown to exist in children with asthma, [32, 33] indicating that pathological changes occur early in life. It is feasible that variation in *ADAM33* alone, or in combination with environmental factors, may lead to early abnormal development of the airway structure. This transfers the origin of asthma into the intrauterine or early postnatal development period.

4.1.2 Defining asthma in children

Paediatricians have long since recognised that several 'types' of asthma exist, and that these have different pathogenesis and prognoses [219]. Recurrent wheeze, a common symptom in infants and young children, has been associated with diminished airway function in early life [220-224]. Occurrence of wheeze has been used to define three phenotypes of childhood asthma: transient infant wheeze, non atopic wheeze of the toddler and pre-school child, and IgE-mediated wheeze [225, 226].

Transient wheezers may have several episodes of wheeze during the first two to three years of life, but this usually resolves in early childhood (age three to five). It is not associated with family history of asthma or allergic sensitisation [56, 226]. The primary risk factor seems to be reduced lung function before occurrence of any lower respiratory tract infection [227]. Observed reduced level of lung function in this group has been associated with smoking during pregnancy [194], and low birth weight [228,

229]. Wheeze in this group occurs because reduced small airway calibre produces turbulent airflow thereby generating vibration in air columns. It will be highlighted when inflammation occurs in association with virus infections. Smaller airways may possibly increase the likelihood of chronic obstructive airways disease in later life, especially if such individuals commence smoking [225].

Non atopic wheezers are children likely to wheeze beyond the third year of life following early viral lower respiratory tract infection (LRTI). Nearly all children contract RV and/or RSV in the first three years of life, and it is likely that the development of bronchiolitis (as opposed to uncomplicated upper respiratory tract infection) with early virus infection is a marker for the later development of atopy and asthma [230, 231]. Following acute viral respiratory infections lung function abnormalities (e.g. BHR) often persist [232-234]. It could be that infection is instrumental to airway damage or a pre-disposed genetic abnormality is exposed as a result of infection [232]. It seems that initial airway diameter, airway length and characteristics of the lung parenchyma may predispose certain infants to wheezing during, and following, common viral infections [220]. Children who present with mild disease or wheeze in infancy and those with no significant family history of asthma and /or allergies are most likely to remit after the first decade of life with no increased risk of asthma or allergies later in life [190]. No significant loss of function is apparent in those with mild symptoms [235]. However, the interaction between asthma and atopy is not well defined [56, 236, 237]. The presence of atopy and a family history of atopy confers an increased risk of severe bronchiolitis with RSV infection [238]. Also it is likely that the atopic future asthmatic has an increased susceptibility to rhinovirus infection [239], with wheeze caused by RV infection being more strongly associated with subsequent asthma than RSV infection [240].

In a substantial minority of infants wheezing episodes are probably related to a predisposition to asthma [190] Many of these children have associated allergic sensitisation with increased levels of IgE [190, 226]. Children with atopy have been shown to have lower lung function measurements by three years of age [27]. Among school age children with persistent asthma, the onset of symptoms before three years is associated with increased severity of disease, and increased BHR continuing into

adulthood [56]. Subjects who persistently wheeze during childhood, and continue wheezing into adulthood, have consistently lower lung function measurements than subjects who never wheeze [191]. Significant loss of lung function often occurs after birth but before six years of age [226]. This decline in lung function does not appear to continue past the age of fourteen, although reduced levels of function track into adulthood [235]. So children with significant symptoms in the first years of life, especially those with a family history of asthma are likely to develop persistent respiratory symptoms later in life, suggesting that changes in airway physiology start early in life [190] and that outcomes of adult asthma may be defined in childhood [191]. Figure 4.1 summarises risk factors associated with these wheeze phenotypes.

Figure 4.1 Definition of childhood asthma based on wheeze phenotypes suggested by the Tuscon study [225] and associated risk factors.

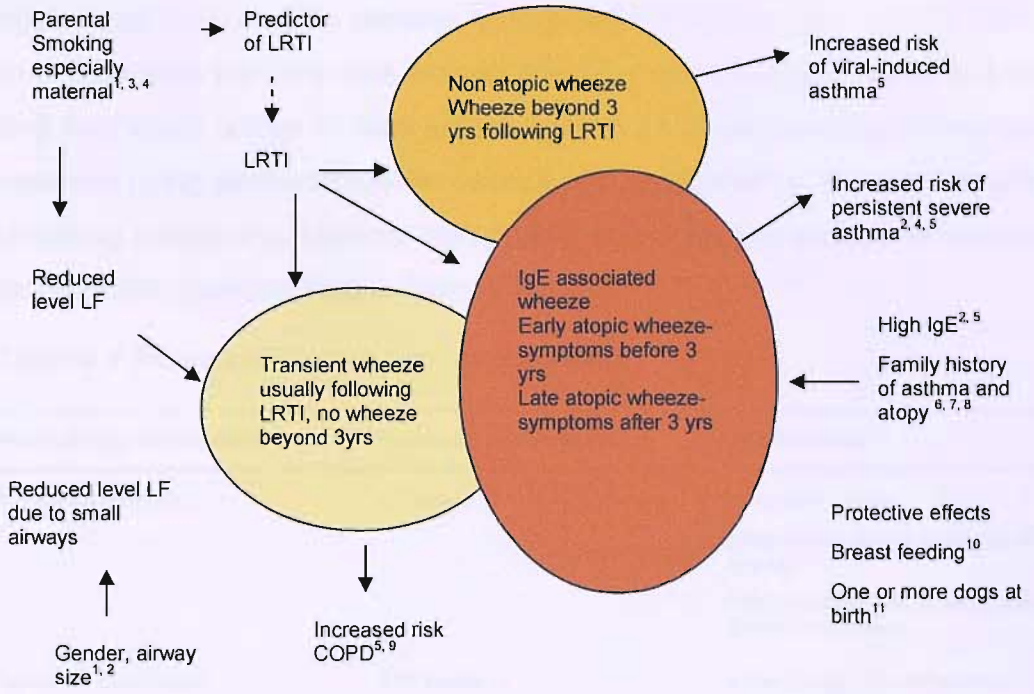


Figure 4.1 Definition of childhood asthma and associated risk factors.

LRTI Lower respiratory tract infection, LF lung function

[1] Parental factors affecting respiratory function during the first year of life [241], [2] Initial airway function is a risk factor for recurrent wheezing respiratory illnesses during the first three years of life [221], [3] Asthma and wheezing in the first six years of life [190], [4] Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants [194], [5] Tuscon Children's Respiratory Study [225], [6] Evidence for genetic associations between asthma, atopy, and bronchial hyperresponsiveness: a study of 8- to 18-yr-old twins [242], [7] Factors in childhood as predictors of asthma in adult life [243], [8] Paternal history of asthma and airway hyperresponsiveness in children with asthma [244], [9] Role of respiratory infection in the onset of asthma and chronic obstructive pulmonary disease [56], [10] The role of breastfeeding in the development of allergies and asthma [245], [11] Exposure to dog in infancy decreases the subsequent risk of frequent wheeze but not of atopy [246]

4.1.3. Development of the human lung

To develop normally, the lungs need space in which to develop, secretion of intraluminal fluid, and the stimulus of fetal breathing movements [247]. During embryogenesis the lung is an actively secreting organ, producing up to 5 ml/kg/hr of lung fluid which is vital for fetal lung development. Fetal breathing movements can be observed using ultrasound in the second half of pregnancy. Impairment of fetal breathing results in pulmonary hypoplasia. There are four phases of fetal human lung development (summarised in table 4.1).

Table 4.1 Phases of human lung development

Phase of lung development	No. weeks post conception	Development
The embryonic period	1-7 weeks	Lung buds appear
		Mesenchyme-driven branching of airways
		First appearance of primitive pulmonary arteries and veins
Pseudoglandular stage	5-17 weeks	Development of bronchial tree
		Cellular differentiation (cartilage, neuronal tissue, ciliated cells, and smooth muscle) commences.
		Pre-acinar airways develop
Canalicular stage	16-26 weeks	Early acini become visible under the light microscope
		Capillaries form a 'mesh' within the mesenchyme
Saccular or terminal sac stage	24 weeks to term	Blind saccules start to develop and alveolarisation commences
		The capillary network becomes closer together and the walls between the sacs contain a double capillary network
Postnatal alveolar stage	Birth up to 2 years	Formation of alveoli, most in first 6 months. 80% complete by 2 years, continuing for up to 8 years.
Postnatal microvascular maturation	Birth	Formation of new double capillary layers, followed by remodelling to form a mature single layer
Postnatal smooth muscle	From birth	Changes in smooth muscle deposition continue into maturity. Increase in muscle layer in asthmatics

The embryonic phase occurs from conception to seven weeks' gestation, starting with lung bud formation and progressing to mesenchyme-driven branching of the airways and blood vessel connection to the heart. This is followed by the pseudoglandular phase between five and seventeen weeks' gestation. During this phase the bronchial tree, including pre-acinar airways and blood vessels, develop. The respiratory (intra-acinar) region develops during the canicular phase (sixteen to twenty six weeks). During this period the peripheral epithelium and mesenchyme become thinner. From twenty seven weeks until birth, the alveolar phase sees the development of saccules and then alveoli. By two years, over eighty per cent of alveolar development is complete but there is continued formation up to eight or nine years of age [248]. Smooth muscle mass increases rapidly in the first few weeks postnatally and by eight months is close to adult levels [249]. During infancy and early childhood the airways are large relative to total volume. Girls have wider and shorter airways than boys which may explain the lesser tendency of wheeze in girls [225, 250]. During childhood alveolar enlargement, peripheral airway elongation and enlargement of the central airways occurs, but there is no further multiplication of the airways. All structures continue to increase in size until adulthood [247, 249].

4.1.4 Dynamics of respiration and measurement of lung function.

Pulmonary ventilation, or breathing, is the process by which air flows into the lungs during inspiration (inhalation) and out of the lungs during expiration (exhalation). Air flows because of pressure differences between the atmosphere and the gases inside the lungs. Air, like other gases, flows from a region with higher pressure to a region with lower pressure. Muscular breathing movements and recoil of elastic tissues create the changes in pressure that result in ventilation. Pulmonary ventilation involves three different pressures:

- Atmospheric pressure (pressure of the air outside the body);
- Intra-alveolar (intrapulmonary) pressure (the pressure inside the alveoli of the lungs);
- Intra-pleural pressure (the pressure within the pleural cavity).

Inspiration (inhalation) is the active phase of ventilation resulting from contraction of the diaphragm causing an increase in volume of the thoracic cavity. This decreases the intra-alveolar pressure and air flows into the lungs. Relaxation of the diaphragm and elastic recoil of tissue decrease the thoracic volume and increase the intra-alveolar pressure causing exhalation. The measurement of lung volume provides important information on the state of the respiratory system. Indeed other parameters of lung function, such as resistance, compliance and forced expiratory flow, are dependent upon the lung volume at which they are measured. Common terms used to sub-divide lung volume measurements are illustrated in figure 4.2.

Figure 4.2 Common lung function measurements.

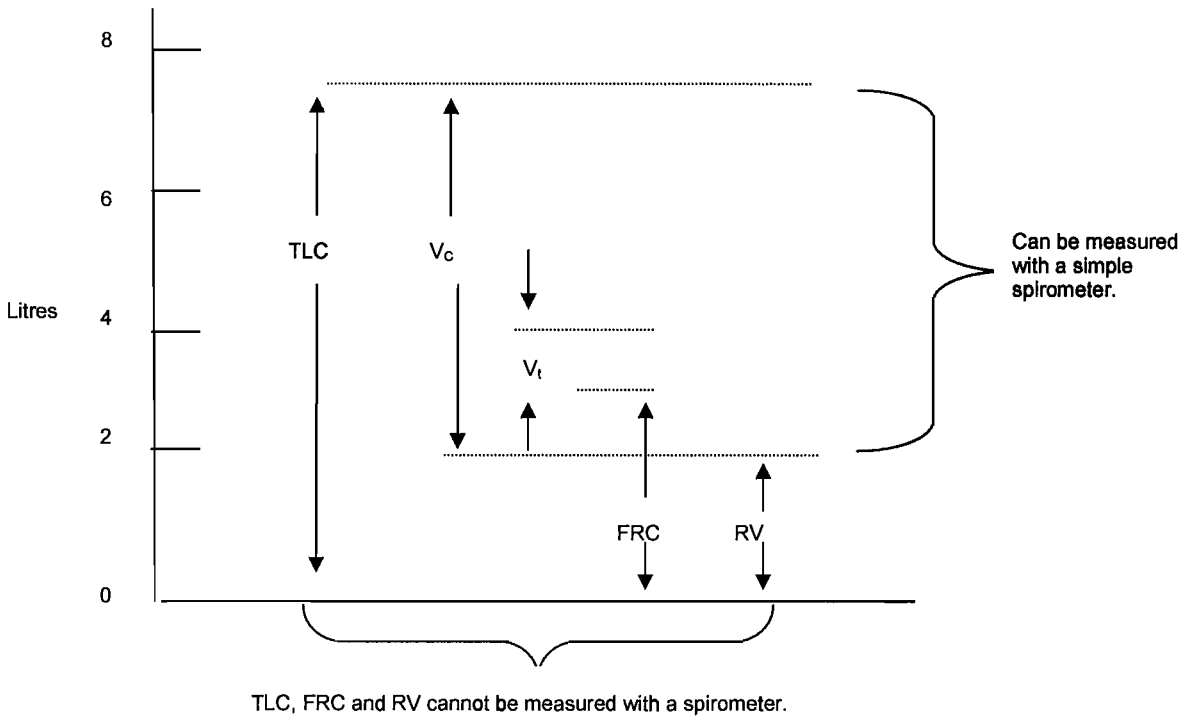


Figure 4.2 illustrates common measurements of lung function. It is not possible to measure TLC, FRC or RV with a simple spirometer so gas dilution technique or whole body plethysmography may be used for this. TLC Total Lung Capacity, FRC Functional Residual Capacity, RV Residual volume, V_c vital capacity, Vt tidal volume.

Functional residual capacity (FRC) is the volume of air remaining in the lungs at the end of a normal quiet expiration. During normal tidal breathing FRC (also called end expiratory volume) coincides with the elastic equilibrium volume (EEV). EEV occurs

where inward elastic recoil of the lungs balances outward elastic recoil of the lungs. EEV is the volume the respiratory system would assume if all the breathing muscles were relaxed during passive expiration. This normally occurs at 40% of vital capacity (V_C) in adults and older children. Vital capacity is the volume of gas that can be expelled from the lungs from a position of full inspiration with no time limit on expiration. Tidal breathing begins from this 'relaxation' volume (40% V_C). Tidal volume (V_t) is the volume of air inspired or expired during each normal quiet respiratory cycle. This volume is thought to be an efficient volume from which to breathe with respect to the length/tension of the inspiratory muscles. During exercise or when lung disease is present FRC can fall below EEV making breathing less efficient. Mild to moderate asthma is rarely associated with changes in lung volumes but as disease progresses the volume of gas in the lung increases resulting in hyperinflation. Total lung capacity (TLC) is the volume of air in the lungs at the end of maximal inspiration. TLC usually remains unchanged but an increase in RV and FRC results in a reduction in V_C. The mechanisms behind this are unclear [251]. The elastic properties of the respiratory system are determined by its structure. When a force is applied to an elastic structure it resists deformation by an opposing force known as the elastic recoil pressure (Pel). The pressure required to overcome the elastic recoil of the lung and chest wall depends on the lung volume above (or below) EEV (ΔV). Elastic recoil pressure divided by lung volume equals elastance (E) $E = Pel / \Delta V$. Elastance is also known as airway resistance (R_{aw}) and is a measure of the elastic properties of the respiratory system measuring airway calibre rather than lung volume. Narrowness of the airway results in elevated R_{aw} values. The reciprocal of resistance is compliance (G_{aw}).

4.1.5 Assessing pulmonary lung function in infants and young children

Measurement of lung function can be carried out in children of all ages, though two to three year old toddlers remain a difficult group to assess [252]. Children between three and six years are unable to do forced expiratory manoeuvres. However, resistance can be measured during tidal breathing and therefore, provided the child can tolerate a face mask, lung function measurement can be performed in this age group. Measurements are carried out in a body plethysmograph. This is a sealed box in which the subject is seated. The patient breathes normally (tidal volume) using a

mouthpiece or mask and pressure changes are recorded. Airway resistance is calculated as the pressure difference between the alveoli and the mouth divided by the flow rate. Mouth pressure can be measured using a manometer and alveolar pressure deduced from measurements made in the body plethysmograph. Specific airways resistance (sRaw) is calculated as the relationship between simultaneous variations in respiratory flow and variations of the volume recorded by a constant-volume whole-body plethysmograph (R_{aw} x FRC). The calculation is based on the maximum changes in plethysmographic volume during inspiration and expiration. Compensation for body temperature and barometric pressure at water vapour saturation (BTPS) is achieved electronically. When needed, an adult may accompany the child [253, 254].

sRaw is a measure of airway calibre independent of lung size with narrowing of the airways resulting in elevated values. This method allows reliable assessment of standard lung function in pre-school children and has been shown to be useful in identifying pre-school children with asthma, with sensitivity at least as good as spirometry in older children [255]. Infants (under two years) cannot offer the degree of cooperation required for this method to be useful and need to be sedated to permit measurement of lung function. This is usually achieved using oral chloral hydrate. Sedation induces a period of 'quiet sleep' which induces minimal change in respiratory rate and establishes a stable end expiratory level. It should be noted that infant pulmonary function tests are obtained in the supine position. This will influence the position of the diaphragm, efficiency of respiratory muscles, FRC, lung mechanics and distribution of ventilation. These factors must be taken into account when interpreting results, especially in longitudinal studies [252].

Forced expiratory flow can be obtained in infants using the rapid thoracoabdominal compression (RTC) technique [256]. An inflatable vest is placed on a sedated infant. Breathing is monitored via a facemask linked to a flowmeter. At the end of inspiration the vest is rapidly pressurised, causing a sudden forced expiration leading to a measure of partial forced expiratory flow (FEF). This technique uses tidal volume and flow limitation is difficult to ascertain [257].

RTC has been developed further to produce the raised volume rapid thoracoabdominal compression technique (RVRTC). The test may take place inside a whole body plethysmograph. The infant wears an inflatable vest and is sedated as before. The infant's lungs are inflated to a given airway pressure, using an external gas source, before forced expiration begins. The lung volume from which the forced expiration occurs can be standardised to a given pressure, allowing comparison between individual measurements or between infants. Forced expiration gives a measure of forced expiratory volume (FEV) and in adults is measured over one second (FEV_1). FEV occurs more quickly in younger subjects so $FEV_{0.5}$, $FEV_{0.4}$, $FEV_{0.75}$ and forced expiratory flow (FEF_{25-75}) are more commonly used measurements than FEV_1 in children. RVRTC produces flow-volume curves similar to that seen in older children and adults [257] using spirometry. Methacholine challenge may then be performed to assess BHR, giving a measure of provocative dose (PD) inducing a drop in baseline FEV of 15% or 20% (PD_{15} or PD_{20} respectively). Using lung function measurements to define which children are at risk of continuing disease could improve prognosis through earlier and more effective medical intervention and better management of symptoms.

4.1.6 Longitudinal studies investigating environmental modifiers of lung function during early life

Several longitudinal studies are in progress aiming to investigate the complexities of lung function during early life. The Tuscon Children's Respiratory Study (TCRS) is a long-term longitudinal study which began in 1980. It has followed 1246 subjects from birth, together with their family members, to delineate the complex interrelationships between potential risk factors, acute lower respiratory tract illnesses and chronic lung disorders (especially asthma) later in life [220, 221, 225, 258, 259]. The study has resulted in better classification of asthma in children. Findings suggest it is events occurring in early life, not in utero, that may be important determinants of subsequent asthma, and that initial diminished airways function, although associated with wheezing illness, does not usually continue into adulthood.

The association between wheezing and prenatal smoking in the Tuscon study was independent of birth weight and the effect was only significant before three years of

age [260]. Infants with smoke-induced transient wheeze start with reduced lung function, which is normal by age six years. Children with raised IgE, and those who have family members with asthma, have a higher risk of developing persistent asthma in adulthood despite having normal lung function in early life [225].

The Western Australia Pregnancy cohort consists of infants randomly recruited before birth from a metropolitan hospital during the years 1987 and 1991.This study found that the effects of *in utero* smoke exposure, family history of asthma and maternal hypertension during pregnancy were associated with reduced respiratory function after birth [194]. The adverse effects of smoking are predicted to start in the fetus, continuing into adulthood [195]. Further studies supported the idea that lung function was adversely affected by *in utero* passive tobacco smoke exposure [241] and that increased airway responsiveness at one month was associated with decreased forced expiratory volume in one second (FEV_1), physician-diagnosed asthma and lower respiratory tract symptoms, at six years of age [261]. The most striking finding from studies using this cohort was that at age six years the initial measure of airway hyperresponsiveness in infancy was the best predictor of respiratory outcome [197]. In a follow up study, reduced airway function present in early infancy was associated with persistent wheeze at 11 years of age, and this relationship was independent of the effect of increased airway hyperresponsiveness and atopy in childhood [192].

The National Asthma Campaign Manchester Asthma and Allergy Study (^{NAC}MAAS) is a study of the development of asthma and allergies in childhood [262]. Pregnant women were recruited at the antenatal unit and atopic status of the mother and father was assessed using skin prick testing and a questionnaire regarding allergies. After delivery, the children were assigned to risk groups according to parental atopic status. Environmental exposures were measured and controlled for and included house dust mite, cat and dog allergen exposure during pregnancy, pet ownership and exposure, childcare arrangements, number of siblings, vaccination uptake, dietary intake and endotoxin exposure. This study has demonstrated that sensitisation to indoor allergens is strongly associated with asthma and that the number of positive skin prick test allergens was correlated to disease [262]. These findings were supported by another longitudinal study on a New Zealand population [191]. In this unselected birth

cohort more than one in four children had wheezing that persisted from childhood into adulthood, or relapsed after remission. The factors predicting persistence or relapse were sensitisation to house dust mite, airway hyperresponsiveness, female sex, smoking, and early age onset. Together with persistently low lung function the results suggest outcome in adult asthma may be predicted in early childhood [191].

The ^{NAC}MAAS cohort was recently genotyped for polymorphism of *ADAM33* using the assays developed in chapter 3 [184]. SNPs in *ADAM33* were associated with reduced lung function (increased specific airways resistance (sRaw)). In children aged three years, carriers of the rare allele of F+1 SNP had reduced lung function (p=0.003). When the recessive model was considered, four SNPs, F+1, S1, ST+5 and V4, showed association with sRaw at age five years (p<0.04). Using LD mapping, evidence of a significant causal location between B+1 and F1 SNPs, at the 5' end of the gene was found. Four SNPs were associated with lower FEV_1 (F+1, M+1, T1 and T2, p<0.04) and V-3 was associated with a positive challenge test (p=0.02). In addition, children homozygous for the rare allele of F+1 were more likely to have wheezed in the first year of life (OR 2.6 95% CI, p=0.004). There was no association between any SNP and allergic sensitisation or physician-diagnosed asthma. This implies *ADAM33* may have a role in alteration of lung function in early childhood.

The work presented in this chapter utilises another longitudinal cohort, the Copenhagen Study of Asthma in Childhood (COPSAC). This cohort was initiated to investigate the relationship between the clinical presentation of asthma, atopic dermatitis, atopic rhinitis and allergy in young children with immunologic, environmental, lifestyle factors and genetics. The aim was to identify risk factors in order to target primary prevention, limit disease progression and improve clinical management [263]. This longitudinal population cohort comprises 410 children born to asthmatic mothers, recruited at birth. Lung function (and other risk factors) were observed from four weeks of age and repeated at two, three, four, five and six years. The four week lung function test was performed using RVRTC (section 4.1.5) producing the lung function measurements $FEV_{0.5}$ and PD_{15}. At three years lung function measurements were performed producing measurements of baseline and post-bronchodilator specific airways resistance (sRaw, section 4.1.5). Allergy and

dermatological testing was also performed and risk factors, including maternal heredity (lung function, events during pregnancy, socio-economics, feeding method, endocrine disruption, microbiology), history of tobacco exposure, serum cotinine levels, and air quality in the home were measured. Gender, biometrics (infant birth weight and length) and parental asthma status were also recorded.

4.1.7 Aim and hypothesis

The aim of this chapter was to test the hypothesis: **Polymorphism of the *ADAM33* gene leads to altered lung development and consequently altered lung function in early life**. The 5' nuclease assays (chapter 3) were used to genotype the *ADAM33* polymorphisms in the COPSAC longitudinal cohort. Lung function measurements were used to test for interaction between *ADAM33* polymorphism and lung function in one month old babies ($FEV_{0.5}$ and PD_{15}) and at 3 years of age (pre- and post-bronchodilator sRaw).

4.2. Methods

4.2.1 COPSAC DNA preparation

A sample of blood was taken from each infant. DNA was extracted and quantified by colleagues in Denmark and shipped to the University of Southampton for genotyping. The DNA was normalised to 5ng/µl into four 96-well arrays. These were then pipetted at 10ng/well into a 384-well plate. DNA was dried overnight as described (section 2.2.1.3) and then frozen at -20°C until required.

4.2.2 Taqman® 5' allelic discrimination assays using COPSAC samples.

The principles of the 5' nuclease allelic discrimination assay are discussed in section 3.2.1.2. The fourteen SNPs were chosen out of twenty available assays following analysis using LD, SNPHAP and SNPtagger programs (section 3.3.3). All assays used in the analysis had been verified by sequencing and obeyed HWE (section 3.3 2). The fourteen SNPs listed in Table 4.2 were genotyped in the COPSAC cohort using Taqman® 5' nuclease assays. One 5' nuclease assay was performed on each 384-well plate using 10ng dried DNA. The assays were run according to methods described in section 2.2.1.5. Samples that did not show adequate discrimination were repeated. The fourteen SNPs were tested for deviation from Hardy Weinberg equilibrium using a χ^2 test.

4.2.3 Collection and analysis of data

The data was exported as described in section 2.2.1.7.

4.2.3.1 Haplotype determination

Common haplotypes were defined using the SNPHAP haplotype determination program [200]. The data was formatted into SPSS and imported into the SNPHAP program (www.litbio.org) as described (section 2.2.1.7). Haplotypes were compared against those observed in the ECACC population (chapter 3).

4.2.3.2 Phenotype data

Parametric tests assume that the dependent variable ($FEV_{0.5}$ or PD_{15}) is continuous and the underlying population follows a normal distribution. The data distribution was analysed using SPSS. The $FEV_{0.5}$ data followed a near-normal distribution but PD_{15} and sRaw (pre- and post-bronchodilator) required log-transformation.

4.2.3.3 Interaction of ADAM33 with lung function measurements

ADAM33 SNPs were investigated for interaction with lung function at one month ($FEV_{0.5}$ and PD_{15}) and at three years (sRaw pre-bronchodilator and post-bronchodilator) using the COPSAC cohort. At one month three different methods of analysis were employed to determine haplotypic association with lung function; haplotype trend regression [199], haploscore [198] and LD map construction and association mapping [201, 202] (section 2.2.1.7). At three years univariate regression and haploscore were used.

4.3 Results

4.3.1 Genotyping the COPSAC cohort

Fourteen 5' nuclease allelic discrimination assays were used to genotype the COPSAC cohort. Results are summarised in table 4.2.

Table 4.2 Fourteen SNPs genotyped in the COPSAC cohort

SNP	Number of alleles/individuals	Observed genotype	Observed alleles	Observed allele frequency	Expected genotype	HWE
Bp1	740/370	GG 247	G 602	G 0.81	GG 245	χ^2 0.47
		GA 108	A 138	A 0.19	GA 112	P = 0.79
		AA 15			AA 13	
F1	750/375	AA 346	A 721	A 0.96	AA 345	χ^2 1.00
		AG 29	G 29	G 0.04	AG 29	P = 0.61
		GG 0			GG 1	
Fp1	738/369	GG 159	G 492	G 0.67	GG 166	χ^2 1.44
		GA 174	A 246	A 0.33	GA 163	P = 0.49
		AA 36			AA 40	
Mp1	740/370	CC 276	C 635	C 0.87	CC 280	χ^2 4.24
		CA 83	A 105	A 0.13	CA 84	P = 0.12
		AA 11			AA 6	
Qm1	740/370	CC 277	C 643	C 0.87	CC 280	χ^2 0.99
		CT 89	T 97	T 0.13	CT 84	P = 0.61
		TT 4			TT 6	
S1	740/370	GG 308	G 677	G 0.91	GG 306	χ^2 1.35
		GA 61	A 63	A 0.09	GA 61	P = 0.51
		AA 1			AA 3	
STp4	740/370	AA 142	A 444	A 0.60	AA 133	χ^2 3.80
		AC 160	C 296	C 0.40	AC 178	P = 0.15
		CC 68			CC 59	
STp5	710/355	GG 116	G 387	G 0.55	GG 107	χ^2 5.26
		GA 155	A 323	A 0.45	GA 176	P = 0.07
		AA 84			AA 72	

Table 4.2 Fourteen SNPs genotyped in the COPSAC cohort (continued)

SNP	Number of alleles/individuals	Observed genotype	Observed alleles	Observed allele frequency	Expected genotype	HWE
STp7	740/370	CC 250	C 609	C 0.82	CC 249	χ^2 0.09
		CT 109	T 131	T 0.18	CT 109	P = 0.96
		TT 11			TT 12	
T1	724/362	TT 271	T 623	T 0.86	TT 268	χ^2 1.73
		CT 81	C 101	C 0.14	CT 87	P = 0.42
		CC 10			CC 7	
Vm3	740/370	CC 156	C 472	C 0.64	CC 152	χ^2 1.10
		CT 160	T 266	T 0.36	CT 170	P = 0.58
		TT 53			TT 48	
Vm1	740/370	GG 279	G 645	G 0.87	GG 281	χ^2 0.87
		GT 87	T 95	T 0.13	GT 83	P = 0.65
		TT 4			TT 6	
V4	740/370	CC 225	C 579	C 0.78	CC 226	χ^2 0.30
		CG 129	G 161	G 0.22	CG 126	P = 0.86
		GG 16			GG 18	
V5	744/372	AA 344	A 716	A 0.96	AA 343	χ^2 0.039
		AG 28	G 28	G 0.04	AG 29	P = 0.98
		GG 0			GG 0	

Table 4.2 Polymorphisms genotyped in the COPSAC cohort. The table lists SNP ID, allele frequencies, genotyping results. Observed and expected genotypes were compared using χ^2 (with 2 degrees of freedom). All assays observed HWE. A p-value of 0.05, is considered statistically significant and would suggest the assay is not working properly.

HWE was confirmed using χ^2 (section 1.4.4.4.3) suggesting that the assays were performing well. 370 samples were genotyped for all polymorphisms except assays F1 and V5. These assays used 375 samples following a further shipment of DNA. Samples not successfully called were repeated.

4.3.2 Haplotype estimation

Genotyping results were exported, formatted and uploaded into SNPHAP. The program was run 150 times in random order. Five haplotypes with frequencies of 7% or above were observed. This compares to five major haplotypes (5% or above) observed in the ECACC population (Table 3.7). Table 4.3 lists haplotypes with estimated frequencies of above 1% predicted in the COPSAC population.

Table 4.3 Haplotypes observed in COPSAC from 14 SNPs using SNPHAP

Haplotype	Estimated frequency	Cumulative
11111112111111	0.34	0.34
11111121112111	0.15	0.49
11221111121111	0.10	0.59
21111121112121	0.08	0.67
11212212211221	0.08	0.75
21211121212111	0.04	0.79
11111121112121	0.04	0.83
12212121211212	0.03	0.86
21221111121111	0.02	0.88
21111121112111	0.02	0.90
21111112111111	0.01	0.91
11211112111111	0.01	0.92

Table 4.3 The most common haplotypes (estimations greater than 1%) observed using fourteen SNPs accounted for up to 92% of the COPSAC population. The SNPs are listed in gene order (figure 1.3) as follows: Bp1 F1 Fp1 Mp1 Qm1 S1 STp4 Stp5 STp7 T1 Vm3 Vm1 V4 V5. Major allele = 1 Minor allele = 2.

4.3.3 Phenotype data

In order to for parametric tests to be valid the data has to follow a normal distribution. Figure 4.3 illustrates that the $FEV_{0.5}$ data followed a near-normal distribution.

Figure 4.3 Distribution curve of $FEV_{0.5}$ data.

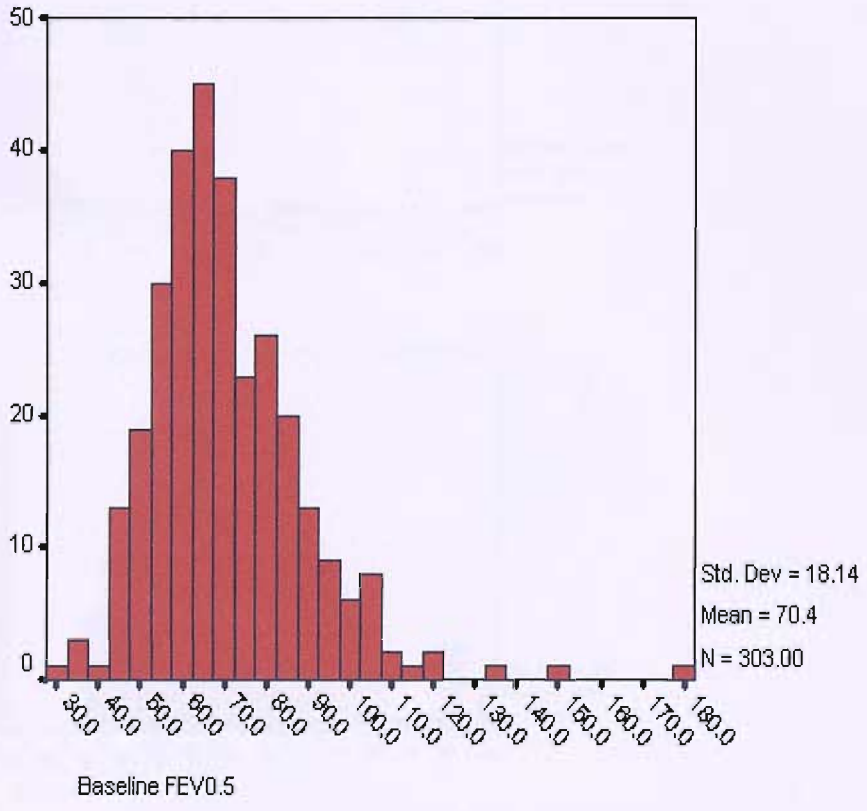


Figure 4.3 A normal distribution of data points reveals a 'bell shape' curve with an equal number of data points each side of the mean when plotted on a histogram. This data set shows a near-normal distribution (n=303).

Figure 4.4a shows distribution of the PD_{15} data. The x axis shows dose of methacholine required to induce a drop in baseline FEV by 15% and the y axis the number of subjects (n=292). The data does not follow a normal distribution with the majority of values skewed towards zero. This signifies that the majority of cases had a drop in FEV of 15% at a low dose of methacholine.

Figure 4.4 Distribution curves of PD_{15} data

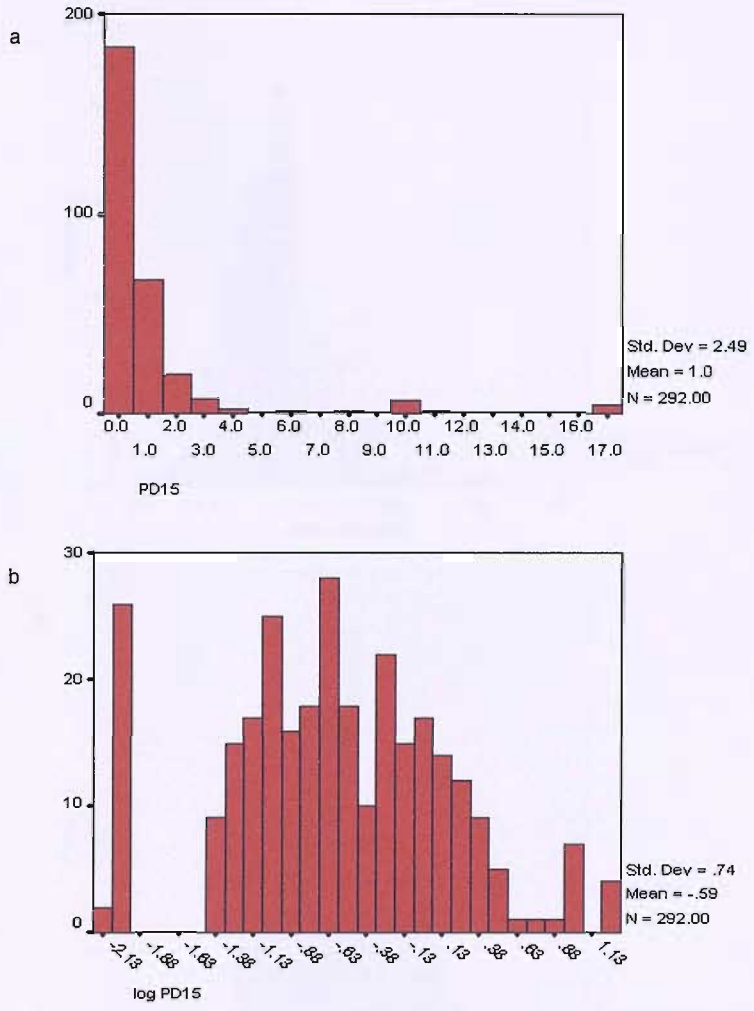


Figure 4.4 a) The histogram of PD_{15} data set shows the data is not normally distributed. It is skewed with most values occurring near zero. This signifies that the majority of cases had a drop in FEV of 15% at a low dose of methacholine. The data cannot be used in this form to run a parametric statistical test. b) Distribution of PD_{15} data after log-transformation appears more normally distributed.

Log-transformation is a common method of transforming non-normally distributed data to allow parametric statistical analysis. Following log-transformation the PD_{15} data showed a nearer-normal distribution (figure 4.4b).

Figures 4.5 and 4.6 show pre-bronchodilator sRaw and post-bronchodilator sRaw data respectively. Both data sets required log-transformation in order to run a parametric statistical test.

Figure 4.5 Pre-bronchodilator sRaw data before and after log transformation

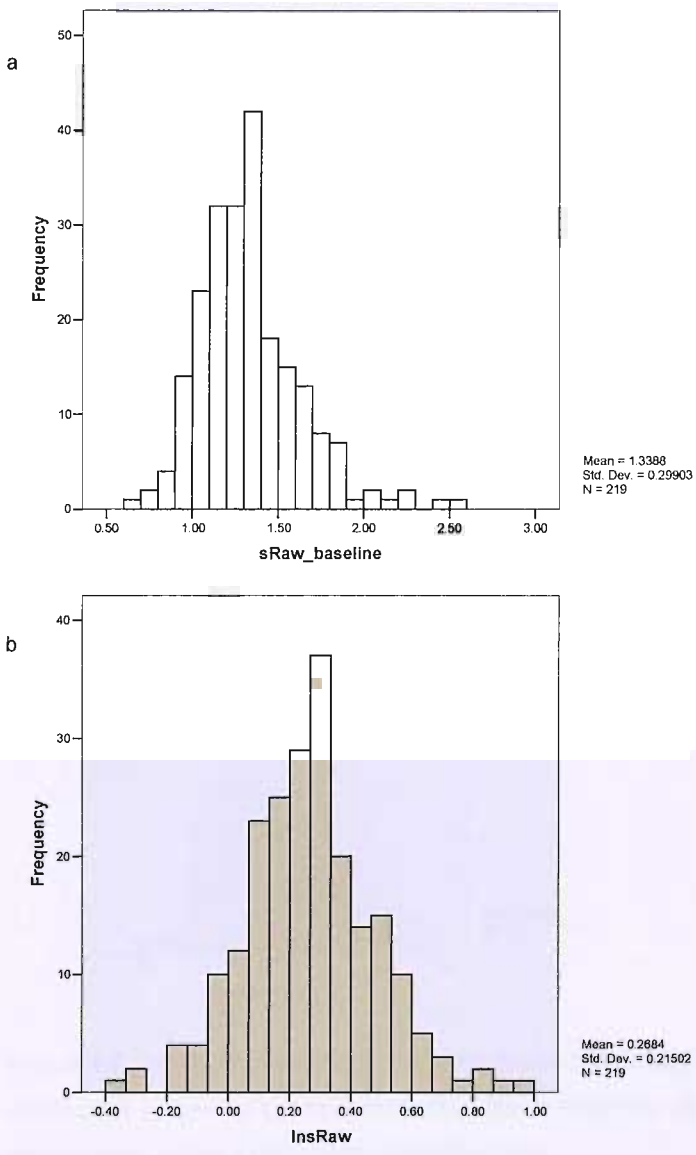


Figure 4.5 The histogram of pre-bronchodilator sRaw data (figure 4.5a) shows the data is not normally distributed. Following log-transformation (figure 4.5b) the data appears more normally distributed and can be used to run a parametric statistical test.

Figure 4.6 Post-bronchodilator sRaw data before and after log transformation

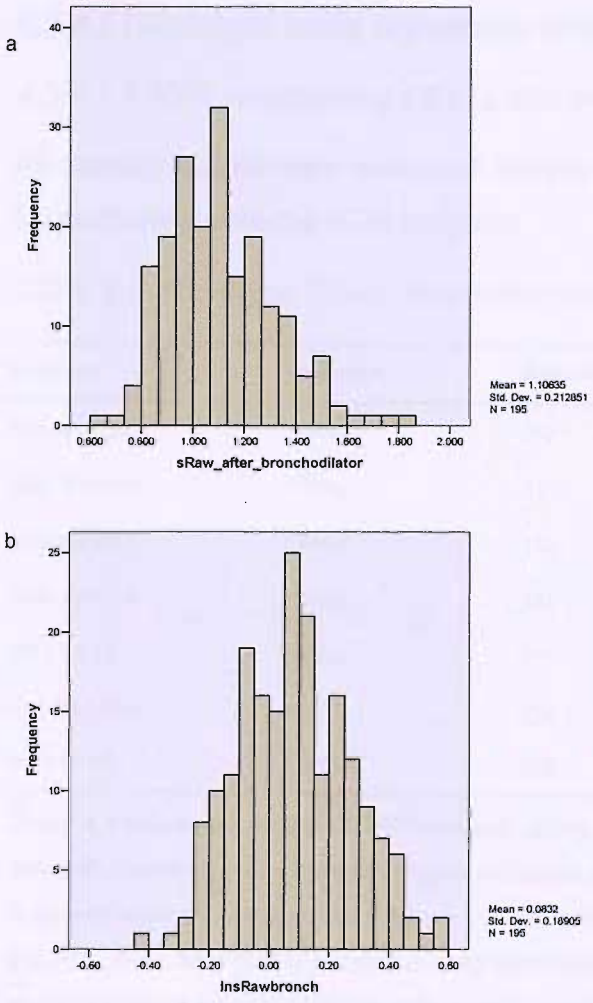


Figure 4.6 The histogram of post-bronchodilator sRaw data (figure 4.6a) shows the data is not normally distributed. Following log-transformation (figure 4.6b) the data appears more normally distributed and can be used to run a parametric statistical test.

4.3.4 Interaction of *ADAM33* polymorphism with lung function at one month

The *ADAM33* haplotypes were investigated for association with lung function ($FEV_{0.5}$ and PD_{15}). Three methods were used: haplotype trend regression (HTR) [199], haploscore [198] and LD association mapping [201, 202, 204].

4.3.4.1 Haplotype trend regression (HTR)

4.3.4.1.1 HTR investigating $FEV_{0.5}$ with three-SNP moving window haplotypes.

All missing values were excluded. Height correction for the $FEV_{0.5}$ phenotype cannot be performed with the HTR program.

Table 4.4 Haplotype Trend Regression analysis of the COPSAC cohort

Loci set	Phenotype	Estimated haplotypes	Estimated frequency	P-value
Fp1 Mp1 Qm1	$FEV_{0.5}$	212	0.13	0.03
Stp7 T1 Vm3	$FEV_{0.5}$	211	0.13	0.05
T1 Vm3 Vm1	$FEV_{0.5}$	112	0.13	0.06
Vm3 Vm1 V4	$FEV_{0.5}$	122	0.09	0.06
Vm1 V4 V5	$FEV_{0.5}$	111	0.75	0.04
Fp1 Mp1 Qm1	log PD_{15}	112	0.004	0.06
Vm1 V4 V5	log PD_{15}	222	0.009	0.12

Table 4.4 illustrates results of HTR analysis using moving windows of 3 SNPs, incorporating 14 SNPs. Several haplotypes are approaching significance using $FEV_{0.5}$. Only loci sets with p-values less than 0.06 and estimated frequencies above 1% are shown. The most significant haplotypes observed using log PD_{15}.(Fp1 Mp1 Qm1) is approaching significance but has a frequency of less than 0.5%.The next most significant was (Vm1V4V5) with a p-value of 0.12 and a frequency of less than 1%.

4.3.4.1.2 HTR investigating log PD_{15} with three-SNP moving window haplotypes

All missing values were excluded. Haplotype trend regression revealed no significant association between SNPs and log-transformed PD_{15}. Table 4.4 lists the two most significant estimated haplotypes with frequencies of 0.4% and 0.9%.

4.3.4.2 Haploscore analysis

This program incorporated all fourteen SNPs and was performed uncorrected for height to allow comparison with HTR. It was then repeated correcting for height.

4.3.4.2.1 Haploscore analysis comparing $FEV_{0.5}$ (uncorrected for height) with 14 SNP haplotypes

Analysis revealed no highly significant association with haplotypes and $FEV_{0.5}$ uncorrected for height (table 4.5 co-variant a) although one reached borderline significance (p=0.07). The two most significant estimated haplotypes had frequencies of 0.06% and 2% respectively. All other haplotypes were less significant (p>0.14) When the analysis was repeated with $FEV_{0.5}$ corrected for height, weight and gender (table 4.5, co-variant b), the borderline haplotype showed reduced significance. All other haplotypes showed p-values >0.14. When $FEV_{0.5}$ was adjusted for height alone (table 4.5, co-variant c), the significance of the haplotype increased to p= 0.06, confirming the correlation of height with FEV.

Table 4.5 Haploscore analysis comparing $FEV_{0.5}$ with 14 SNP haplotypes

Bp 1	F 1	Fp 1	Mp 1	Qm 1	S1	Stp 4	Stp 5	Stp 7	T1	Vm 3	Vm 1	V4	V5	Hap Freq	p-val	Co-var
1	2	2	1	2	1	2	1	2	1	1	2	2	2	0.006	0.07	a
2	1	1	1	1	1	2	1	1	1	2	1	1	1	0.02	0.09	a
1	2	2	1	2	1	2	1	2	1	1	2	2	2	0.006	0.09	b
1	2	2	1	2	1	2	1	2	1	1	2	2	2	0.006	0.06	c

Table 4.5 Haploscore analysis revealed no significant associations with $FEV_{0.5}$ uncorrected for height (a). The previously borderline significant haplotype reduced in significance when corrected for height, weight and gender (b), though it remained the most significant. On correcting for height alone (c) significance increased. All other haplotypes showed p-values >0.14 Co-var; co-variants

4.3.4.2.2 Haploscore analysis comparing log PD_{15} with 14 SNP haplotypes

Estimated haplotypes were analysed using PD_{15}. The results revealed no significant haplotypes. The most significant values are illustrated in table 4.6 with p>0.15.

Table 4.6 Haploscore analysis using log PD_{15}

Bp1	F1	Fp1	Mp1	Qm1	S1	Stp4	Stp5	Stp7	T1	Vm3	Vm1	V4	V5	Hap-Freq	p-val
1	1	1	1	1	1	1	2	1	1	1	1	1	1	0.34	0.18
2	1	1	1	1	1	1	2	1	1	1	1	1	1	0.01	0.22
1	1	2	1	2	2	1	2	2	1	1	2	2	1	0.07	0.27

Table 4.6 displays the most significant results from haploscore analysis using PD_{15}. All other haplotypes have p-values exceeding 0.3.

4.3.4.3 Comparison of HTR with Haploscore analysis ($FEV_{0.5}$, uncorrected for height).

One haplotype was identified as approaching significance using haploscore. This was compared with the HTR results for $FEV_{0.5}$ (uncorrected for height). Table 4.7 shows that the most significant haplotypes observed by HTR analysis were present within the most significant haplotype observed by haploscore analysis.

Table 4.7 Comparison of Haploscore (HS) and HTR analyses using $FEV_{0.5}$ (uncorrected for height)

Haplotype														HS Freq	HS P-value	HTR Freq	HTR P-value
Bp1	F1	Fp1	Mp1	Qm1	S1	STp4	STp5	STp7	T1	Vm3	Vm1	V4	V5				
1	2	2	1	2	1	2	1	2	1	1	2	2	2	0.006	0.07		
		2	1	2												0.13	0.03
								2	1	1						0.13	0.05
									1	1	2					0.13	0.06
										1	2	2				0.09	0.06
											1	1	1			0.75	0.04

Table 4.7 Haplotype 12212121211222 (most significant est. haplotype) was compared to the results from HTR 3 moving window results. The HTR analysis correlated with the 5 most significant 3 moving window SNP haplotypes.

4.3.4.4 LD association analysis

An LD map of the region was plotted using HAPMAP markers (based on Caucasian European population (CEU) within the Centre d'Etude Polymorphism Humaine (CEPH) population, phase 1 release). The CEU consists of thirty U.S trios, with Northern and Western European ancestry. LDU location for the CEU population was obtained from LDDB (http://cedar.genetics.soton.ac.uk/public_html/). Kb was plotted against LDU across the *ADAM33* region (figure 4.7).

Figure 4.7 LD map of the flanking region of ADAM33

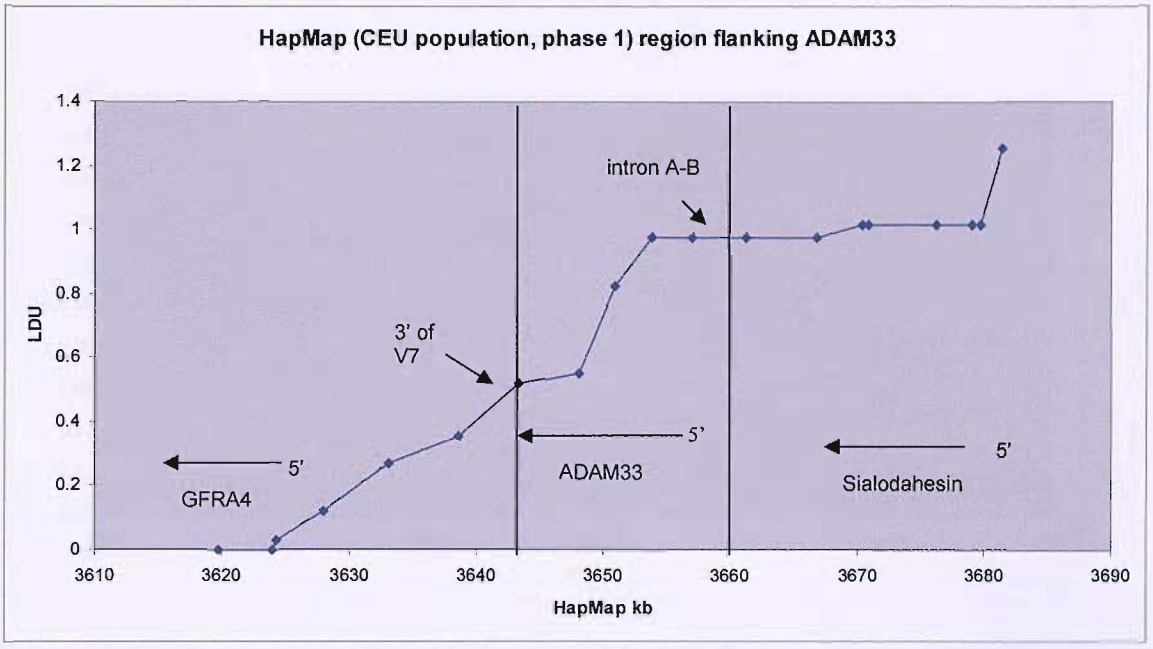


Figure 4.7 LD map illustrating linkage disequilibrium of ADAM33 and its flanking region. The ADAM33 region is located between the two vertical lines. The most 5' and 3' markers within the ADAM33 region are labelled for clarity. Gene direction is indicated by the arrow. Flanking genes are also marked. kb=kilobases; LDU = linkage disequilibrium units. LD data used HAPMAP markers based on Caucasian European population (CEU)

An area of recombination (break in LD) between markers at 3643 bp and 3653 bp is apparent. Further definition was carried out using *ADAM33* markers to map LD across *ADAM33* (figure 4.8).

Figure 4.8 LD map of immediate ADAM33 region

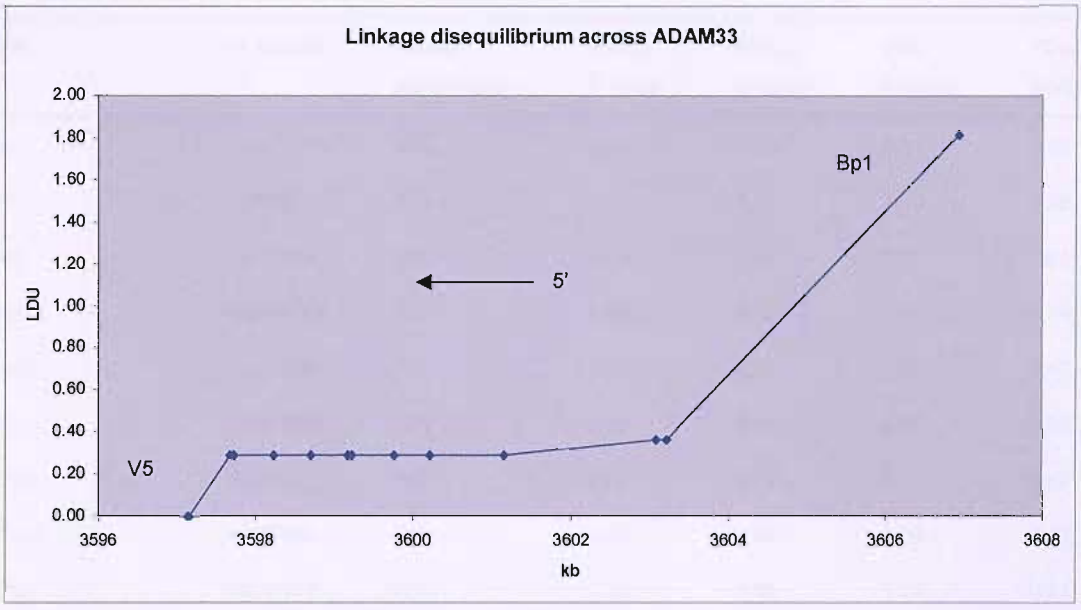


Figure 4.8 A higher resolution LD map of the ADAM33 region illustrates linkage disequilibrium across ADAM33. Kb positions were obtained using rs numbers (http://genome.ucsc.edu/). For clarity, only the end markers are labelled. The arrow denotes gene direction. Two breaks in LD are apparent, between V4 and V5, and between Bp1 and F1.

The SNP markers, used for genotyping, were used to generate a higher resolution map based on genotype data derived from the COPSAC cohort. This revealed two breaks in LD, between V4 and V5, and between Bp1 and F1. LD association analysis uses regression to detect association between single SNPs of *ADAM33* and lung function measures ($FEV_{0.5}$ of PD_{15}). F-values obtained, and the probability of association (p-value) are shown in table 4.8. None of the values reached significance suggesting *ADAM33* SNPs are not associated with lung function at one month of age. Further analysis to discover the position of the causative SNP was not performed.

Table 4.8 LD association analysis

SNP	rs number	Alleles major/minor	$Fev_{0.5}$ F value	$Fev_{0.5}$ p-value	PD_{15} F value	PD_{15} p-value
Bp1	rs487377	G/A	0.06	0.80	0.41	0.52
F1	rs3918392	A/G	1.15	0.28	1.19	0.28
Fp1	rs511898	G/A	0.32	0.57	0.22	0.64
Mp1	rs3918395	G/T	0.15	0.70	2.15	0.14
Qm1	rs612709	C/T	1.38	0.24	0.53	0.47
S1	rs3918396	G/A	0.83	0.36	0.19	0.66
STp4	rs44707	A/C	0.09	0.76	0.03	0.87
STp5	rs597980	G/A	0.00	0.97	0.66	0.42
STp7	rs574174	C/T	3.00	0.08	1.22	0.27
T1	rs2280091	T/C	0.15	0.70	2.15	0.14
Vm3	rs628977	G/A	0.00	0.98	0.03	0.87
Vm1	rs543749	G/T	0.82	0.36	0.11	0.73
V4	rs27877094	C/G	0.87	0.35	1.00	0.32
V5	rs3746631	A/G	0.89	0.34	1.00	0.32

Table 4.8 $FEV_{0.5}$ values are adjusted for height and weight. An F value of >4, and a p-value <0.05 suggests the values X (SNP) and Y (lung function) are related.

4.3.5 Interaction of *ADAM33* polymorphism with lung function at three years

General linear model univariate analysis (SPSS version 14) was used to analyse association of single SNPs with lung function (sRaw pre- and post-bronchodilator). Small numbers were observed for some genotypes so the heterozygote and minor homozygote genotypes were grouped. Regression analysis testing for the effect of gender suggested no effect, so gender was removed from the model.

4.3.5.1 Univariate regression analysis

Using pre-bronchodilator sRaw, univariate analysis (one SNP at a time) showed that the regression model was significantly different from the null model (a model with only a constant term) for F1 (p=0.042), Qm1 (p=0.011), STp4 (p=0.006), STp7 (p=0.038) and T1 (p=0.037). For SNPs F1, Qm1, STp4 and STp7 the major homozygote was associated with a higher sRaw. For the T1 SNP the minor homozygote/heterozygote group was associated with higher sRaw.

Using post-bronchodilator sRaw univariate analysis (one SNP at a time) showed that the regression model was significantly different from the null model (a model with only a constant term) for S1 (p=0.026). For SNPs S1 the minor homozygote and heterozygote group was associated with higher sRaw.

Analysis showed several SNPs to be associated with lung function. Bonferroni correction was not applied since it assumes variables are independent of each other. Considering the extent of LD across *ADAM33* it was felt this correction would be too stringent. However, fourteen SNPs were tested against both pre- and post-bronchodilator sRaw using a significance level of p=0.05, so it would be expected that one false positive would occur in each phenotype.

4.3.5.2 Haploscore analysis

The *ADAM33* haplotypes were investigated for association with lung function using haploscore [198].

Table 4.9 Haploscore analysis of ADAM33 haplotypes and pre- and post-bronchodilator sRaw

hap no.															pre hap freq	pre hap score	pre p value	post hap freq	post hap score	post p value
1*	2	1	2	1	1	1	2	1	2	1	2	1	1	1	0.046	-2.32	**0.02**	0.049	-1.03	0.299
2*	1	2	2	1	2	1	2	1	2	1	1	2	1	2	0.040	-2.01	**0.044**	0.040	-1.71	0.086
3	2	1	1	1	1	1	2	1	1	1	2	1	2	1	0.094	-1.61	0.107	0.091	-1.26	0.204
4	2	1	1	1	1	1	2	1	1	1	2	1	1	1	0.019	-1.50	0.133	0.018	0.12	0.903
5	1	1	1	1	1	1	2	1	1	1	2	1	1	1	0.123	-1.18	0.236	0.125	-0.64	0.519
6	1	1	1	1	1	1	2	1	1	1	1	1	1	1	0.006	-1.13	0.258	0.007	-0.29	0.768
7+	1	1	2	1	2	2	1	2	2	1	1	2	2	1	0.080	-1.05	0.292	0.082	1.89	**0.058**
8	1	2	2	1	2	1	2	1	2	1	1	2	2	2	0.005	-0.49	0.617	0.005	-0.65	0.513
9	1	1	1	1	1	1	2	1	1	1	2	1	2	1	0.045	-0.35	0.726	0.045	-0.01	0.991
10	1	1	2	1	1	1	2	1	2	1	2	1	1	1	0.012	0.35	0.719	0.013	0.14	0.888
11	1	1	2	1	1	1	1	2	1	1	1	1	1	1	0.008	0.89	0.367	0.008	-0.65	0.509
12	2	1	2	2	1	1	1	1	1	2	1	1	1	1	0.030	1.32	0.187	0.034	1.40	0.161
13	2	1	1	1	1	1	1	2	1	1	1	1	1	1	0.016	1.52	0.126	0.018	1.40	0.161
14#	1	1	1	1	1	1	1	2	1	1	1	1	1	1	0.327	1.76	0.077	0.322	-0.09	0.925
15*	1	1	2	2	1	1	1	1	1	2	1	1	1	1	0.101	2.46	**0.013**	0.096	1.19	0.230

Table 4.9 A positive hap score denotes a higher mean sRaw indicating increased specific airways resistance, and so reduced lung function. A negative score denotes lower mean sRaw indicating reduced airways resistance and better lung function. Hap freq denotes the predicted frequency of the haplotype in the population. The SNPs are listed in gene order as follows: Bp1 F1 Fp1 Mp1 Qm1 S1 STp4 Stp5 STp7 T1 Vm3 Vm1 V4 V5. Major allele = 1 Minor allele = 2. For clarity significant p values are in bold type.

Haploscore analysis using pre-bronchodilator loge sRaw revealed haplotypes 1, 2 and 15 (*) to be associated with lung function. Haplotype 14(#) reached borderline significance. Haplotypes 1 and 2 were associated with improved lung function (haploscore -2.32 and -2.01 respectively) and 15 with reduced lung function (haploscore 2.46). The most significantly associated haplotype (15, p=0.013) was predicted at a frequency of 10%.

Analysis using post-bronchodilator sRaw revealed a borderline association of haplotype 7 (p=0.058, 8% frequency) with reduced lung function.

4.4 Discussion

4.4.1 Taqman® 5' nucleases assays using COPSAC DNA

Taqman® 5' nuclease assay was shown to be a reliable, accurate and repeatable genotyping method. Fourteen of the assays successfully discriminated >98% of samples. This was comparable with the ECACC genotyping (table 4.10). The least successful assay was STp5 with 96% of the samples successfully genotyped in total.

Table 4.10 Success of the Taqman® assays on COPSAC DNA compared with commercial DNA (ECACC)

SNP	Number of samples	1st run no. of dropouts	Successful calls after repeats (%)	ECACC successful calls (%) no repeats
Bp1	370	0	100	99.4
F1	375	7	100	99.4
Fp1	369	9	99.7	100
Mp1	370	9	100	98.9
Qm1	370	0	100	99.4
S1	370	5	100	98.9
STp4	370	56	100	95.8
STp5	355	25	96	100
STp7	370	6	100	97.9
T1	362	8	97.8	100
Vm3	370	31	100	99.4
Vm1	370	2	100	99.4
V4	370	17	100	97.4
V5	375	10	99	96.3

Table 4.10 lists the number of non calls for the COPSAC cohort with the total calls following repeat genotyping. Performance was comparable with ECACC genotyping.

4.4.2 Comparison of haplotypes between ECACC and COPSAC.

Haplotype estimation was performed using SNPHAP [200] using the genotyping results of the fourteen SNPs (table 4.2). COPSAC haplotypes (table 4.11), with

frequencies greater than 4% and accounting for 83% of the population, were compared with the results obtained from the random UK population (ECACC, table 3.7). The two most common haplotypes are identical in both populations although frequencies differ. The most frequent haplotype occurs at a slightly higher frequency in the COPSAC population. The second most common haplotype is estimated at twice the frequency in the COPSAC population than in the ECACC, signifying less recombination in the COPSAC population. The COPSAC population has a third haplotype with an estimated frequency of 10% which occurs at an estimated frequency of only 2% in the ECACC population. The fourth listed haplotype has an estimated frequency of 8% in COPSAC and a frequency of 6% in the ECACC population. There are more haplotypes estimated in the ECACC population but the frequencies are lower, with only three haplotypes occurring at frequencies over 5%, accounting for 35% of the population. In the COPSAC population five haplotypes are estimated with frequencies over 5%, accounting for 63% of the population. Although both populations are of Caucasian origin, a higher number of haplotypes were expected in the ECACC population, since the COPSAC population has had less mixing of the gene pool, resulting in a more homogeneous population.

Table 4.11 Common haplotypes observed in COPSAC and ECACC populations

Haplotype	COPSAC estimated frequency	ECACC estimated frequency
11111112111111	0.34	0.22
11111121112111	0.15	0.07
11221111121111	0.10	0.02
21111121112121	0.08	0.06
11212212211221	0.08	0.01
21211121212111	0.04	0.02
11111121112121	0.04	0.05

Table 4.11 The SNPs are listed in gene order as follows: Bp1 F1 Fp1 Mp1 Qm1 S1 STp4 Stp5 STp7 T1 Vm3 Vm1 V4 V5. Major allele = 1 Minor allele = 2.

4.4.3 Statistical methods

4.4.3.1 Analysis of haplotypic association with lung function at one month

Three different methods of analysis were employed to determine haplotypic association with lung function at one month. Haplotype Trend Regression (HTR) [199] is not sufficiently powerful to allow all SNPs to be entered at one time, so a three-SNP moving window was used. The HTR program did not allow for height correction for the $FEV_{0.5}$ phenotype. It is known that FEV is linked to height (taller people have greater FEV). Adult height is influenced by many factors during fetal development, growth during pregnancy and adolescence. Factors resulting in reduced birth weight, such as maternal smoking and poor nutrition may therefore affect adult height and subsequent lung function. Failure to adjust for co-variant factors may therefore give misleading results [264]. Haploscore [198] allows all SNPs to be analysed concurrently and it allows for height correction when using $FEV_{0.5}$. The HTR analysis revealed several three-SNP haplotypes marginally associated with decreased lung function. If HTR could correct for height, it is possible that these associations would be lost [264]. Haploscore analysis can correct for height, so analysis was repeated using this method. However, to allow direct comparison with the HTR results, haploscore was first performed using the fourteen-SNPs set with $FEV_{0.5}$ uncorrected for height. A further analysis using $FEV_{0.5}$ (corrected for height) did not reveal any significant association between lung function and polymorphism of *ADAM33* in one month old infants. It was of interest that the most significant fourteen-SNP haplotype contained all the significant HTR three-SNP haplotypes, suggesting the causal SNP may reside within this haplotype.

PD_{15} is a measure of the responsiveness or irritability of the lung (BHR). In the study by Van Eerdewegh *et al* [79] *ADAM33* was more strongly associated with the BHR phenotype than the other asthma phenotypes. HTR analysis of *ADAM33* polymorphism and PD_{15} revealed no association with BHR. This was confirmed by haploscore analysis.

Linkage disequilibrium and association analysis [201, 202, 204] confirmed LD to be as observed by Simpson *et al* [184] but failed to detect any association between *ADAM33* polymorphism and lung function. This suggests the causal SNP(s) does not

reside on any of the haplotypes tested, that gene environmental interaction occurs after four weeks postnatally or the measurement of phenotype was inadequate.

4.4.3.2 Association of ADAM33 polymorphisms with lung function at three years

Univariate regression was used to investigate *ADAM33* polymorphism and association with lung function, and to test for the effect of gender. Gender was removed from the model as results suggested no effect. Five SNPs were significantly associated with lung function (pre-bronchodilator sRaw) but results did not duplicate those seen in the study by Simpson *et al* [184]. Correction for the number of tests was not applied, even so, only one positive association would be expected at p=0.05 significance level for each phenotype.

Haploscore analysis using pre-bronchodilator loge sRaw, showed two haplotypes (table 4.9) were associated with improved lung function (haplotypes 1 and 2; haploscore -2.32 and -2.01 respectively) and one haplotype with reduced lung function (haplotype 15; haploscore 2.46). The most significant haplotype (haplotype 1, p=0.013) was predicted at a frequency of 10% in the study population.

It is possible the association of S1 with bronchial hyperresponsiveness (post-bronchodilator sRaw) may be a false positive since we would expect one association by chance with this number of tests. In addition the haplotype analysis did not corroborate the association with post-bronchodilator sRaw.

4.4.4 Comparison with the collaborative study 'Polymorphism in *ADAM33* predicts impaired early life lung function' [184]

The Taqman® 5' nuclease assays were developed and verified to genotype both the COPSAC cohort and also for a duplicate collaborative study. Colleagues at the University of Manchester used these assays to genotype the ^{NAC}MAAS cohort [184, 262]. SNPs were denoted differently in the two studies thus ^{NAC}MAAS denotations are in brackets. The results showed carriers of the rare allele of Fp1 (F+1) had reduced lung function at age three years (p=0.003, n=285). A recessive model showed that Fp1 (F+1), S1, STp5 (ST+5) and V4 were associated with reduced lung function (sRaw) at five years (p<0.04, n=470). Further multivariate analysis identified Fp1

(F+1), Mp1 (M+1), T1 and T2 to be associated with reduced FEV_1. It could be concluded that *ADAM33* polymorphism is associated with reduced lung function (increased specific airways resistance (sRaw)) in children aged three and five and the causal SNP may lie between Bp1 and F1 [184]. The COPSAC study did not replicate these findings at one month, but did find association at three years, although not with the same SNPs.

There were similarities in the study design of the ^{NAC}MAAS study and the COPSAC study. Both studies involve Caucasian populations, although ^{NAC}MAAS may be a more heterogeneous population than COPSAC. Both cohorts were recruited before birth. ^{NAC}MAAS recruited parents on the basis of allergy questionnaires and skin prick tests. The COPSAC cohort recruited on the basis of maternal asthma comprising a high asthma risk group rather than allergy *per se*, so investigates a different sub-section of the asthma population. Data was unavailable for comparison of factors such as environment, diet and maternal smoking rates. The ^{NAC}MAAS study genotyped seventeen SNPs based on allele frequency and gene coverage whilst COPSAC reduced the number of SNPs to fourteen using SNP tagging and LD mapping. These methods reduce the number of SNPs needing to be typed without losing information (section 3.2.4.). Allele frequencies observed from genotyping these cohorts were similar (table 4.12).

Table 4.12 Allele frequencies observed in [NAC]MAAS and COPSAC

SNP	rs number	Alleles major/minor	*MAF COPSAC	*MAF MAAS
Bp1	rs487377	G/A	0.19	0.2
F1	rs3918392	A/G	0.04	0.04
Fp1	rs511898	G/A	0.33	0.32
Mp1	rs3918395	G/T	0.14	0.11
Qm1	rs612709	C/T	0.13	0.14
S1	rs3918396	G/A	0.09	0.09
S2	rs528557	C/G	-	0.25
STp4	rs44707	A/C **(C/A MAAS)	0.40	0.43
STp5	rs597980	G/A	0.45	0.44
STp7	rs574174	C/T	0.18	0.17
T1	rs2280091	T/C	0.14	0.12
T2	rs2280090	C/T	-	0.12
Vm3	rs628977	G/A **(A/G MAAS)	0.36	0.38
Vm1	rs543749	G/T	0.13	0.12
V2	rs3918400	C/T	-	0.04
V4	rs27877094	C/G	0.22	0.23
V5	rs3746631	A/G	0.04	0.04

*Table 4.12 *MAF minor allele frequency*

4.4.4.1 Reflections on phenotype measurements

In the COPSAC study no association was seen between *ADAM33* polymorphism lung function using $FEV_{0.5}$ and PD_{15} in one month old infants. It is known that airway inflammation and remodelling of the airway wall plays an important role in the pathophysiology of asthma. Study of the developmental progress of the lungs makes it clear that different components of the lung develop at different rates and reach functional maturity at different time points, both ante and post natally. Thus any insult, causing impediment to growth and/or development, will have differing impacts depending on the timing and duration of the insult [248, 265]. Structural changes, observed following bronchoscopy of a group of mildly asthmatic children (n=10, age 9.3 years), were seen to be similar to that observed in adults by age nine years [32]. Biopsy studies indicate that the airways of asthmatic children are inflamed although the infiltrate may be more lymphocytic than eosinophilic [32]. The continued development after birth, and rapid growth in the first few years, means that the environmental interactions may have a more profound physiological impact in the

early years of life than in later life. Thus a critical window may exist, perhaps between birth and lung maturation, in which environmental challenges interact with genetic loci, so altering lung pathology and predisposing the individual to future lung disease. This may explain why we see no association at one month.

An alternative explanation may be postulated when methods of lung function measurement of are considered. In the ^{NAC}MAAS cohort, lung function was assessed using plethysmographic measurement of specific airways resistance (sRaw) at age three years and again at five years. sRaw has been shown to correlate with $FEV_{0.5}$ [266]. In the COPSAC study infants at one month required sedation, inducing a period of 'quiet sleep' but this has been shown to have no detrimental effect on lung function measurement [252]. This initially suggests the different methods of assessing lung function in these studies should not influence the outcome. However, a report by Ranganathan *et al* [267] suggests $FEV_{0.3}$ and $FEV_{0.4}$ are the only timed volume parameters that are reliably calculated in infants during the first months of life. The report recommends that $FEV_{0.4}$ be routinely reported in infants less than 3 months of age [267], This view is not universally accepted [268], yet it remains possible that $FEV_{0.5}$ is not a sufficiently sensitive or reliable measure of lung function at this age and could explain the lack of association.

Sensitivity of the airways to methacholine was assessed at one month using measurements of PD_{15}, i.e. the dose or concentration required to provoke a drop in FEV of 15%. On reviewing the literature available it appears the use of PD_{20} is more commonly used than PD_{15} [269, 270] so it is possible PD_{15} is too sensitive. Indeed the distribution curves seen in figure 4.4 supports this. The histograms showing PD_{15} data illustrates that many infants had a drop in FEV of 15% at a very low dose of methacholine. This could explain the lack of association seen between *ADAM33* polymorphism and BHR (PD_{15}) at one month. The analysis of the same group of children aged 3 years using univariate regression analysis (sRaw) showed association with ADAM33 polymorphism, conceivably supporting the theory of inadequate phenotype measurements at one month.

The analysis has revealed no significant association between *ADAM33* polymorphisms and lung function (measured as $FEV_{0.5}$ or PD_{15}) in a Danish

population of one month old infants. This could be a true result or could be due to incongruous phenotype measurement. At three years (using sRaw) several SNPs were significant, replicating previous studies, but not identifying a causal SNP. Haplotype analysis supports the associations seen between pre-bronchodilator sRaw and *ADAM33*.

4.4.5 Other considerations

Several other outcomes were measured in the COPSAC data set. Fractional concentration of nitrous oxide (FeNO) was measured simultaneously with innate lung function at four weeks. NO is an easily accessible marker of airway inflammation in allergic asthma [252], increasing when inflammation is present. Also eosinophils were measured at six months and eighteen months. It would be interesting to look at markers of inflammation at one month and see whether this correlates to eosinophilia, prevalent in asthmatic airways [271] later in life.

ADAM33 polymorphisms may only be associated with changes in lung function at four weeks in association with pre-existing environmental insults such as smoking during pregnancy, nutrition etc. This needs to be compared between the cohorts and used in multiple regression analysis. At one month the inflammatory status of the infant may still be modified by prenatal influences or breast milk, although the contribution of breast feeding to development of allergy and asthma is not clear [245]. Outside environmental effects may be attenuated by the extent of maternal influence. In any multivariate analysis it would be logical to include co-variants, such as maternal medication during pregnancy or whether the child was breast fed, to detect any contributory effect on inflammation and lung function.

4.4.6 Conclusions

To summarise, this study supports previous association studies [37, 38, 79, 181-183, 186] implicating *ADAM33* polymorphism in lung function variation. The causal SNP remains illusive. The function of *ADAM33* within the remodelling process remains unclear. It is necessary to define the promoter region of this gene and define the signalling pathway(s) in which *ADAM33* resides before we can elucidate its ultimate function, in both normal airway biology and in disease.

Chapter 5

Characterisation of the *ADAM33* promoter and regulation of transcription

5.1 Introduction

5.1.1 Rationale for promoter analysis of *ADAM33*

Positional cloning has identified *ADAM33* as a gene associated with asthma and BHR [79]. Subsequent studies have confirmed this association [181-184], and shown *ADAM33* polymorphism is also associated with progression of asthma [37], accelerated decline in lung function [38], and in predicting a reduced lung function in young children [184], but not in early infancy (chapter 4). Its potential multiple molecular actions [21] and the preferential expression of *ADAM33* in mesenchymal cells in adults [171] provide additional evidence implicating this molecule in the pathogenesis of asthma. Expression studies in human embryonic lung support a role for *ADAM33* in smooth muscle, bronchial, vascular and neuronal development suggesting a role in airway 'modelling' of the human embryonic lung [161]. Western blotting of lysates from embryonic tissues revealed an isoform (25 kD) not observed in adult mesenchymal cells [161]. In adult asthmatics and non asthmatics *ADAM33* total mRNA and splice form ratios did not differ. It is interesting to speculate which isoforms, and in what ratios, they occur in the fetus. It is possible that a different isoform ratio in the fetus affects the modelling of smooth muscle, before the gene is regulated to its stable adult expression profile. Differences in ratios of splice form expression during development could predispose the individual to an increased risk of developing an asthmatic phenotype. It is not known what transcription factors activate *ADAM33* during embryogenesis. The COPSAC study (chapter 4) suggests that a critical window exists in which gene-environmental interaction occurs. It may be that 'modelling' effects of *ADAM33* variants during development are not obvious until the

necessary environmental trigger has been met later in life. It is necessary to define the promoter region of this gene to define which molecules regulate *ADAM33*. Definition of the transcriptional pathway in which *ADAM33* resides will help discover its ultimate function.

5.1.2 Gene regulatory mechanisms in eukaryotes

Patterns of eukaryotic gene expression and regulation are complex. Timing of expression and amount of transcript produced are paramount to the correct functioning of every cell. The vast majority of genes in a eukaryotic cell are turned off (silenced) at any given time. Gene regulatory mechanisms must be able to switch on necessary genes when they are needed, and switch off when the 'task' is complete. Regulation of gene expression can be achieved by adjusting the rate of transcription of DNA into RNA, or of the translation of RNA into protein. Within this simplistic view are many other levels of gene regulation, including epigenetic factors (e.g. DNA methylation), factors affecting mRNA stability and splicing, and post-translational protein modification such as glycosylation. Most regulation, however, is thought to be achieved at the transcriptional level.

5.1.3 Mechanism of transcription

The basic mechanism of transcription is instigated when molecular signals from outside or inside the cell lead to the localised unwinding of chromatin and bonding of regulatory *trans*-acting proteins (also called transcription factors) to specific DNA sites adjacent to the protein-coding region. Transcription factors possess one or more of the following functional domains: a domain that recognises the correct regulatory element (DNA docking site); a domain that interacts with one or more proteins of the basal transcription apparatus (RNA polymerase or RNA polymerase associated protein); a domain that interacts to proteins bound to nearby docking sites, so they can act cooperatively to regulate transcription; a domain that influences chromatin condensation and/or a domain that acts as a sensor of physiologic conditions within the cell [272]. Binding of transcription factors to the regulatory sites directly, or indirectly, assists binding of RNA polymerase to its transcription initiation site, or may repress transcription by preventing the binding of RNA polymerase. The nuclei of eukaryotic cells contain three different RNA polymerase enzymes which carry out

distinct functions. RNA polymerase I (pol I) predominantly transcribes ribosomal RNA genes (rRNA), RNA polymerase II predominantly transcribes protein-coding genes into messenger RNA (mRNA) and RNA polymerase III predominantly transcribes transfer RNA genes (tRNA). Eukaryotic RNA polymerases cannot initiate transcription by themselves. Instead, combinations of short sequence elements in the immediate vicinity of the gene (*cis*-acting elements) act as recognition signals for *trans*-acting transcription factors to bind DNA in order to guide the RNA polymerase to the correct site for transcription initiation. The *cis*-acting elements responsible for the initiation of transcription constitute the proximal promoter region.

Examples of *cis*-acting elements recognised by ubiquitous transcription factors include the TATA box, CAAT box and GC box. The TATA box (sequence TATAAA), for example, is usually located 25-30 bp upstream of the transcription initiation site. It is recognised and bound by the general transcription factor (GTF) TFIID. The TFIID-TATA box complex is then bound by TFIIA via protein-protein interactions, which stabilises the new pre-initiation complex (PIC). Other factors then bind to the complex positioning the polymerase at the correct site for transcription initiation. The PIC contains six general multiprotein complex transcription factors (GTFs), and RNA polymerase II. The promoters of many genes lack TATA boxes but have either a GC box (containing variants of the sequence GGGCGG), usually found ~100 bp upstream of the transcription initiation site, or a CAAT box (sequence CCAAT) usually found ~80 bp upstream of the transcription initiation site. GC boxes and CAAT boxes appear to be able to work in either orientation, despite their sequences being asymmetrical.

Diverse patterns of gene expression are afforded by modularity (interactions of different combinations of a small set of transcription factors) and cooperativity (the ability of proteins to interact). Thus different combinations of transcription factors can display unique interactions, resulting in distinct patterns of gene expression. In addition to the general upstream transcription elements which are recognised by ubiquitous transcription factors, more specific recognition elements are known which are recognised by tissue-restricted transcription factors affording tissue-specific gene expression [110]. A further group of *cis*-acting elements, called enhancers, occur in

variable positions and can exert an effect up to 50 kb upstream or downstream of the promoter they control [272]. Their function is independent of orientation. Gene-regulatory proteins appear to bind to these regions, causing the DNA to loop out, which allows interaction between the proteins bound to the proximal promoter and the enhancer region, increasing transcriptional activity. Silencers are equivalent regulatory elements that can inhibit the transcriptional activity of specific genes [110].

5.1.4 Previous promoter analysis of *ADAM33*

Previous work had identified the translation initiation and termination sites, polyadenylation sites and exon/intron structure of *ADAM33*. *ADAM33* was confirmed as having one transcription start site using 5' rapid amplification of cDNA ends (RACE). Sequencing of the putative promoter region showed the absence of a TATA box or CAAT box [79, 169].

5.1.5 Use of reporter genes to investigate regulatory sequences

To test the assertion that a particular DNA sequence is involved in the regulation of gene expression, it is necessary to introduce these putative regulatory sequences into a cell and determine their activity. This is done by combining the sequence of interest with a reporter sequence or gene whose product can be readily monitored. There should be a minimal background level of expression of the gene, and little interference from other genes expressed in the cell. Common reporter genes used in animal cells include chloramphenicol acetyltransferase (CAT), luciferase and green fluorescent protein (GFP). CAT is the enzyme product of the cat gene from the *E.Coli* transposon Tn9. This enzyme transfers acetyl groups from acetyl CoA on to the antibiotic chloramphenicol. Cell lysates are mixed with ^{14}C-labelled chloramphenicol, incubated and then separated. CAT activity is determined by measuring the relative amounts of acetylated and non-acetylated chloramphenicol using a phosphorimager or scintillation counter. CAT is a very stable protein and can efficiently determine when a gene is switched on; however, the protein products persist after the gene has switched off. An alternative reporter is the *luc* gene from the American firefly (*Photinus Pyralis*) encoding the enzyme luciferase. In the presence of oxygen, ATP and magnesium ions, the enzyme catalyses oxidation of luciferin [273]. The reaction releases a flash of light, the intensity of which is proportional to the level of enzyme

activity. This is detected using a luminometer. The emitted light signal then decays rapidly. The luciferase assay is over 100 times more sensitive than CAT assays [273, 274]. Luciferase genes from other organisms (e.g. *Renilla reniformis*) have different activities with the reactions releasing light at different wavelengths. This allows simultaneous monitoring of several genes and forms the basis of the Dual-Luciferase® Reporter Assay system used in this work (sections 2.2.5.2). A third reporter, the *gfp* gene, from the jellyfish *Aequoria victoria* encodes green fluorescent protein (GFP), a bioluminescent marker that emits bright green fluorescence when exposed to blue or ultraviolet light [275]. GFP activity can be measured by quantifying light emission, but requires no substrate, so can also be used to assay cellular processes in real time.

Reporter genes are inserted into a vector (or plasmid) containing other elements, such as a gene for selection following transformation into cells and a polylinker region in which the sequence of interest may be readily inserted.

5.1.6 Commercially available reporter vectors

Bacterial plasmids are small circular DNA molecules that are distinct from and additional to the main bacterial chromosome. They replicate their DNA independently of the bacterial chromosome, so are an efficient means of amplifying cloned DNA since there are so many copies made per cell. A plasmid modified for use in genetic manipulation studies, by addition of a reporter gene for example, is termed a vector. Plasmids carrying genes for drug resistance were chosen for vector development since drug resistance allows selection for cells transformed by the vector.

5.1.6.1 pGL3 reporter vectors

There are many vectors available commercially. This study used one of a group of four pGL3 vectors supplied by Promega Corporation®. The four pGL3 vectors are structurally identical except for differences in promoter and enhancer regions. Distinguishing features of the pGL3 family are listed in table 5.1.

Table 5.1 The pGL3 vector family

Vector	Feature	Function
pGL3-Basic	Lacks eukaryotic promoter and enhancer sequences.	Identification of functional promoter elements
pGL3-Enhancer	Contains an SV40 enhancer downstream of luc+	Aids verification of functional promoter elements as presence of the enhancer will often result in transcription of luc+ at high levels.
pGL3-Promoter	Contains an SV40 promoter upstream of luciferase gene	Putative enhancer elements can be inserted up/down-stream of the promoter-luc+ transcriptional unit
pGL3-Control	Contains an SV40 promoter and enhancer sequences	Results in strong luciferase expression in many mammalian cells. Useful in monitoring transcription efficiency and can be used as a control compared against pGL3 recombinants.

Each vector contains a high copy number prokaryotic origin of replication for maintenance in *E.coli*, an ampicillin resistance gene (β-lactamase) for selection and a filamentous phage origin of replication (f1 ori) for single-stranded DNA (ssDNA). Restriction sites for insertion of DNA are located in the polylinker site. Downstream is the luciferase gene sequence (luc+) and a polyadenylation sequence conferring mRNA stability. Primer binding sites (RV3 and GL2 table 2.8) are found flanking the polylinker site for sequence verification of the insert. Figure 5.1 shows a plasmid map illustrating the main features of the pGL3 basic vector.

Figure 5.1 The pGL3 basic plasmid (adapted from Promega Corporation technical manual 033 [276]).

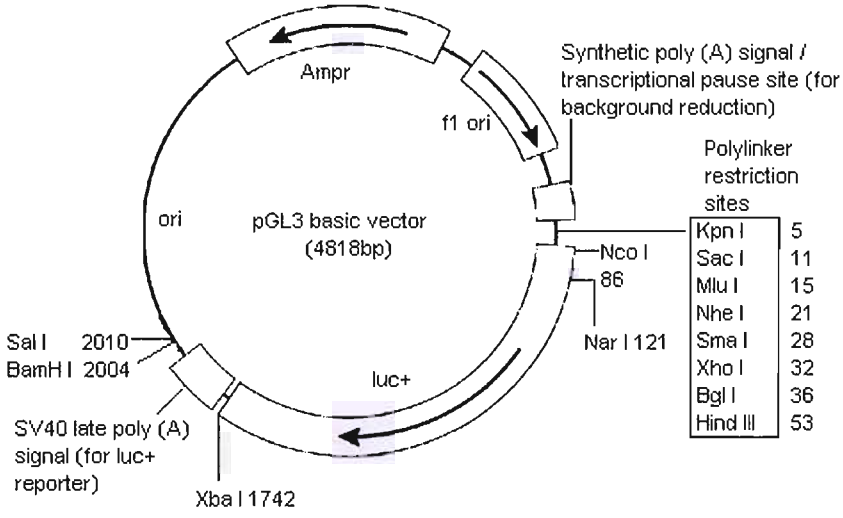


Figure 5.1 pGL3 basic vector plasmid map. Numbers denote base pair (bp) position. luc+ (88-1740 bp) denotes the gene encoding the modified firefly luciferase. Ampr (3080-3940 bp) denotes the β-lactamase gene conferring ampicillin resistance in E.coli. Arrows within luc+ and Ampr genes indicate the direction of transcription. Kpn I and Bgl II restriction sites lie within the polylinker site (position 1-58 bp) and were used in this work for ligation of putative promoter sequences into the plasmid. The luc+ gene sequence lies immediately upstream with a SV40 late poly (A) signal which confers mRNA stability (1772-1993 bp). The constructs were sequenced using primers GL primer 2, binding site 89-111 bp and RV primer 3, binding site 4760-4779 bp (not marked on map)

5.1.6.2 Renilla luciferase reporter vectors

The phRL and pRL families of *Renilla* luciferase reporter vectors contain cDNA encoding *Renilla* luciferase cloned from the anthozoan coelenterate *Renilla Reniformis*. pRL vectors encode the native DNA sequence whilst phRL encodes a synthetic sequence that has been modified to improve use in mammalian cells. However, both families of vectors code essentially the same protein. In this work phRL *Renilla* luciferase reporter vector was used as a control within Promega's Dual-Luciferase® reporter assay system. The vector contains a T7 promoter which drives activation of the *Renilla* gene and an SV40 poly (A) sequence for mRNA stability. Figure 5.2 illustrates the phRL plasmid map.

Figure 5.2 phRL Renilla reporter vector plasmid map

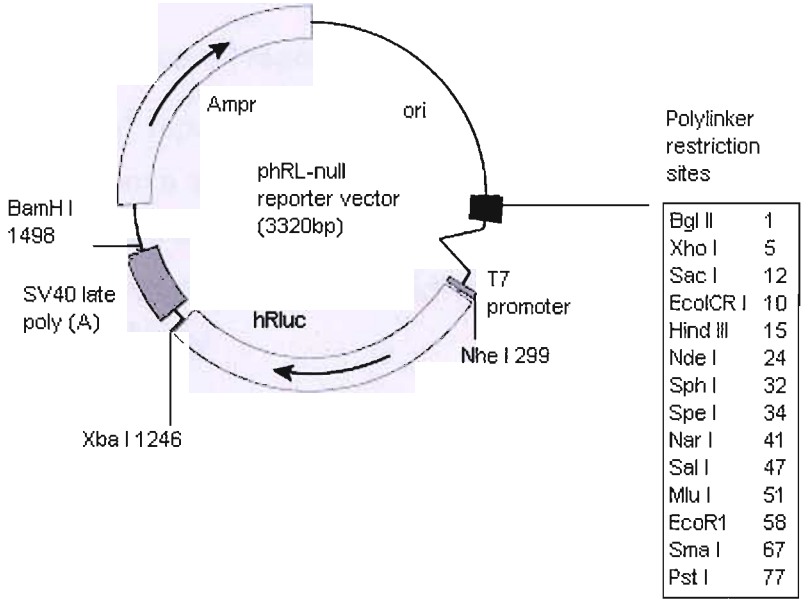


Figure 5.2 Features of the phRL-null reporter vector. Numbers denote base pair (bp) position. hRluc (309-1244 bp) denotes the gene encoding the Renilla luciferase. Ampr (1634-2494 bp), denotes the β-lactamase gene conferring ampicillin resistance in E.coli. Arrows within hRluc and Ampr genes indicate the direction of transcription. The hRluc gene sequence lies immediately upstream of a SV40 late poly (A) signal, which confers mRNA stability (1266-1487 bp).

5.1.6.3 Expression vectors

pEGFP-N1 is a kanamycin-resistant vector encoding a variant of the wild type GFP, which has been optimised for brighter fluorescence and higher expression in mammalian cells. In this work GFP was used as an expression vector to assess transfection efficiency of the cells prior to luciferase reporter transfection (section 5.2.2.2). It can also be used as a reporter.

Sp1 is a sequence-specific DNA binding protein that activates RNA polymerase II transcription from promoters that contain properly positioned GC boxes [277]. Sp1 is the founder member of the Sp1-like transcription factor family [278, 279]. It is ubiquitously expressed and it displays versatile functions by regulating expression of many housekeeping genes, probably acting as a basal transcription factor and helping to recruit the TATA-binding protein. An expression plasmid (pPac Sp1) was

kindly supplied by Dr I. Clarke (University of Norwich) to investigate how over-expression of Sp1 protein affects *ADAM33* promoter activation.

5.1.7 The Dual-Luciferase® reporter assay

The Dual-Luciferase® reporter assay sequentially measures the activity of firely and *Renilla* luciferase from a single sample. The activity of the primary reporter (firefly luciferase) is correlated with the inserted DNA sequence whilst the activity of the co-transfected control reporter (*Renilla*) provides an internal control to normalise results, reducing sample variability

5.1.8 Defining the promoter region of *ADAM33* using deletion analysis

If all the necessary elements are present in a reporter vector for expression of a human gene, high expression of the reporter gene will result when transfected into a cultured human cell. A series of progressive deletion constructs can then be made using restriction enzymes or PCR. The fragments are cloned into reporter vectors and transfected into mammalian cells. In this way it is possible to see which region of the original sequence is important for promoter activity. Choice of cell line is discussed in section 2.2.2.

5.1.9 Comparative genomics

Comparative genomics is a relatively new field of biological research which compares the genome sequences of different species. By comparing the finished reference sequence of the human genome with genomes of other organisms, researchers can identify regions of similarity and difference. A region of conservation suggests the sequence has an important function which has been conserved through evolutionary processes. Likewise, comparison of a human gene of unknown function with the whole human genome may show similarity with a well-defined gene. This can point the researcher towards the new gene's ultimate function, and although it cannot take the place of experimentation, it offers a way of focussing the often minimal resources in the right direction. Software tools such as VISTA (http://www-gsd.lbl.gov/vista) [280, 281], allow large-scale sequence alignments expressing the results in graphical format.

They can compare human-against-human sequence and human-against-non-human sequence. The program allows 2–100 sequences to be submitted at any one time. Results are displayed on Vista graphs for each pair of submitted sequences and show information including input sequences, alignments and conserved sequence statistics. Conservation and visualisation parameters, that are used to calculate conserved regions, can be adjusted. Parameters include calculation window size; conservation identity; minimum conservation width and the lower boundaries of the graph (minimum y axis). The calculation window is the size of sliding window used to calculate conservation scores of each base. The VISTA curve is plotted using a score calculated by sliding a variable sized sliding window across the alignment. A score is calculated at each base in the sequence, counting the number of conserved bases within that window. The default window size is 100 bp. The minimum conservation width is the minimum width a conserved region must be (default 100 bp) before it is recognised as such, and coloured accordingly on the graph. The conservation identity is the minimum conservation identity that must be maintained over the window (min cons width) for the region to be considered conserved (default 70%). The y axis of the graph is set at 50%-100% and can be reduced to lower the visual boundaries of the graph.

5.1.10 Identification of putative transcription factor binding sites

MatInspector (http://www.genomatix.de/) is a software tool that utilises a large library of matrix descriptions determined by automatic evaluation of pubmed abstracts containing co-citation of transcription factors. Its purpose is to identify possible transcription factor binding sites in imported DNA sequences. Similar and/or functionally-related transcription factor binding sites are grouped into matrix families that share the same anchor portion of sequence. There are several sub-sets, including fungi, insects, plants, vertebrates, miscellaneous (e.g. Bacteria) and other functional elements (e.g. PolyA signal) to refine the search. It is also possible to filter for tissue-specific transcription factors within the vertebrate sub-set [282].

5.1.11 Use of expression plasmids to investigate putative binding sites

It is possible to test putative binding sites in a given sequence using an expression vector that expresses the putative transcription activating factor alongside the reporter

plasmid. Luciferase assays are then performed as described (section 2.2.5) to observe change in expression.

5.1.12 Aim

Most basal transcription apparatus is situated within 100 bp-200 bp of the transcription initiation site. Other *cis*-acting elements such as enhancer and silencer elements, can be situated much further away [272]. The aim of this chapter is to:

- identify the core promoter region necessary for basal expression of *ADAM33* and define any positive or negative regulatory regions using deletion constructs of the *ADAM33* promoter region. The luciferase assay is used to detect relative activity of each construct;
- comparative genomic tools are used to identify conserved regions of the *ADAM33* promoter region. Conserved regions suggest biological activity, so may offer beneficial information regarding *ADAM33* activation;
- information from the luciferase work and comparative genomics will identify possible transcription binding sites important for *ADAM33* activation.

5.2 Methods

A 1650 bp fragment, upstream of and including exon 1, of the *ADAM33* promoter region was amplified by PCR. Four smaller PCR fragments of 1346 bp, 1087 bp 637 bp and 359 bp upstream of the translation start site (ATG) were then generated from the first PCR template, and were cloned into pGl3 basic vectors containing the luciferase gene. These were transiently transfected into fibroblast and epithelial cell lines using effectene (Qiagen). Luciferase activity was measured using Dual–Luciferase® reporter assays (Promega). Transfected fibroblast cells were treated with IL-13, IFN-γ, dexamethasone and a phorbol ester to see if this changed expression. Comparative genomics and transcription factor site studies were employed to detect conserved regions and putative transcription binding sites. Figure 5.3 shows an overview of the methods used for *ADAM33* promoter analysis.

Figure 5.3 Overview of methods used for ADAM33 promoter analysis

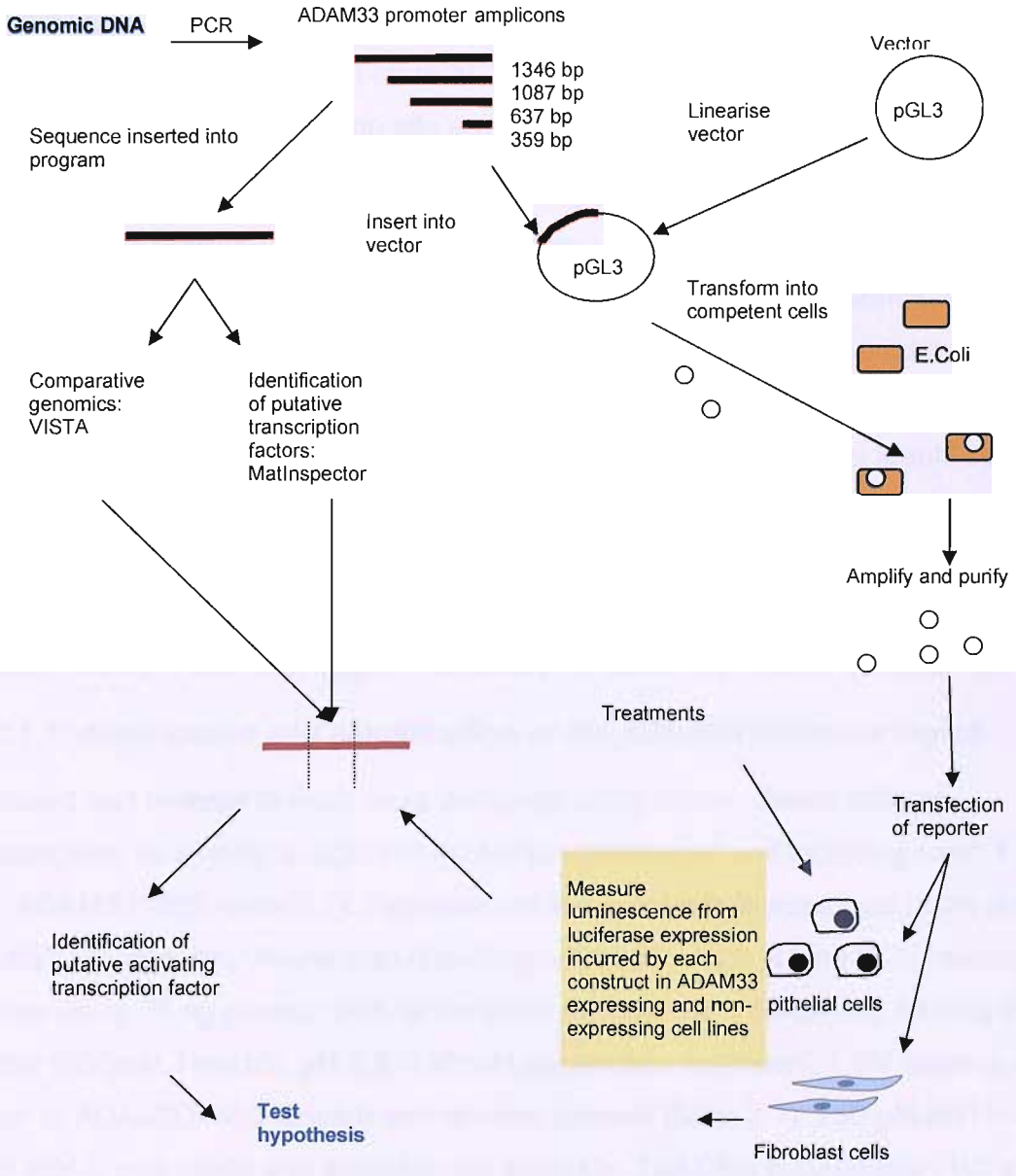


Figure 5.3 Methods used for ADAM33 promoter analysis. Emboldened blue script denotes start and end points.

5.2.1 Generation of promoter constructs

The basal promoter of a gene is usually situated within 200 bp from the transcription initiation site. A 1650 bp region, upstream of and incorporating exon 1, of *ADAM33* was selected for amplification by PCR. This is a sufficiently large region to incorporate the basal transcription initiation site and immediate *cis* elements.

The next stage was to divide the amplified region into four equal segments to make the deletion constructs. A reverse PCR primer was designed which would omit the translation initiation site (ATG) but retain maximum upstream sequence, including the transcription initiation site. The 3' ends of the deletion fragments would thus be identical for cloning into the vectors. The promoter sequence is very GC-rich, so design of the forward primer, and size of amplicon, was dictated by identification of a suitable forward primer site. Reporter constructs were generated by cloning the PCR-amplified fragments of the putative promoter region into a promoter-less vector containing a luciferase reporter gene (section 5.1.6.1). The primary aim was to identify the core promoter region necessary for basal expression of *ADAM33*.

5.2.1.1 Identification and amplification of the ADAM33 promoter region

Forward and reverse primers were designed using Primer Select software (Lasergene) to amplify a region of *ADAM33* upstream of and including exon 1 (primer set ADAM331650, table 2.7). Specificity of the primers was assessed using the NCBI BLAST program http://www.ncbi.nlm.nih.gov/BLAST. A 20 µl long PCR reaction was set up using 75 ng pooled DNA as template. A mastermix containing 1x long PCR buffer (500mM Tris.HCl, pH 8.9, 140 mM ammonium sulphate), 1.3M betaine, 400 nM each of ADAM331650 forward and reverse primers (table 2.7), 250 µM dNTP and 2 mM $MgCl_2$ was made and added to the template. Taq DNA polymerase (1U) and Pwo DNA polymerase (0.008U) was added. The reaction was amplified over 35 cycles using the following PCR cycling conditions: 94°C for two minutes, 94°C for twenty seconds, 59.1°C for thirty seconds, 68°C for two minutes (one min/kb) and 68°C for twenty minutes. PCR products (3 µl) underwent electrophoresis as detailed in section 2.2.1.6.2. The remainder was purified using Qiagen PCR purification kit (2.2.3.1)

5.2.1.1.1 Restriction digest of the 1619 bp amplification product to confirm ADAM33 promoter amplicon

Lasergene software was used to identify a restriction enzyme (HphI) that would give a specific restriction pattern along this amplicon. The purified PCR product was digested using restriction endonuclease HphI according to the following protocol: using a 2 µl PCR tube, 6 µl PCR product was added to a mastermix containing 1.2 µl 10x NEB4, 0.02 µl Hph1 enzyme (5000U/ µl) made up to 12 µl with distilled water. The contents were flick-mixed and then briefly spun down before incubating at 37°C for eight hours, followed by a twenty minute deactivation step. The PCR products underwent electrophoresis as detailed in section 2.2.1.6.2, to observe fragment size.

5.2.1.2 Generation of the deletion constructs

Four forward and one reverse primers were designed using Primer Select (Lasergene) software to amplify four consecutively smaller regions of *ADAM33* with the position of the reverse primer remaining constant. The reverse primer was designed to be immediately adjacent to, but did not include, the translation start site (ATG). Positions of primers in relation to the ATG and transcription start site and known polymorphisms are illustrated in figure 5.4. The NCBI BLAST program http://www.ncbi.nlm.nih.gov/BLAST was used to confirm specificity of the primers. To aid cloning, primers were designed to incorporate restriction sites present in the polylinker region of the pGL3 basic vector (section 5.1.6.1). The forward primers were tagged with KpnI restriction site sequence and the reverse primer with Bgl II restriction site sequence.

Figure 5.4 Primer positions and known polymorphisms within the putative ADAM33 promoter sequence

TATGCTTTCTGTACTTTTTGATTTTGAGATATGTGAATGTAGGTTTCTCTCACTGCTCGAACTTTCACTAACCAAATTACTACAT
TCCAAATTCTCAAAAACAAATAGATTTACTTAAAAGTAGGCTGGGTGCGGTGTCTCACGCCTGTAATTCCAGCGCTTTGGGA
GGCCGAGGCGGGCAGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAACATGGAGAAACCCCATCTCTACTAAAA F1
ATACAAAATTAGCCAGGCGTGGTGGCGAATGCCTGTAAT**[G/a]**[4]CCAGCTACTCGGGAGGCTGAGGCAGAAGAATCACTTGA
ATCTGGGAGGCAGAGGTTGCAGTGAGCCCAGATCATGCCATTGCACTCCAGTCTGGGTAACAAGAGAGAAACTCTGTCTCA
AAAAAAAAAAAAAAAAAAAAAGATTTGCTTAAAAGTTAACATCTCCGGCCGGGCGCGGTGGCTCATGCCTGTAATCCCAGCGC F2
TTTGAGAGGCCGAGGCGGGTGGATCACGAGATCAGGAGATTGAGACCATCCTGGCCAAAATGGTGAAACCTCGTCTCTGCT
AAAAATACAAAAGTTAGCTGGGGGTGGTAGCGCGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCT
TGAACCAGGGAGTCGGAGGTTGCAGTGAGCCAAGATCGCGCCGCTGCACTCCAGCCTGGCGACAGAGGGAGACTCCATCT
CAAAAAAAAAAAAAAAAAAAAAGTTAACATCTCATCCAAATTTGCACCGAGTAGGAAAACAAAAGTTTAAAACA**[T/a]**[3]GAAACA
GATGTTACTGAGGCCGAAAGGGTCTCCCAGGCCTGGGAGTCTGCAGCTTTTATGCAATTCTGCCCTCTGGCCACCGCCAGG
GAAGAAAGGTTGTCTCCGTCTGCTGCATCGCCTTTGCCCAGCAATGAAGCCCCCAAGACAGCGGCAGCCGGTTGCCTGAA F3
CCTTCCTATCCTTGGGGGCACCCAGTGCAGGTGGATGACCCGACTCAACCTCCGCCAGGGCACCCTCGGGGCAGGACGG
GTAGCAAGGAGGGGACAGAGATCGGCCCCAGGAGACCACGGAAGATCGCGCTCCTGGGGCCAACTTCAGCAGCGAGAGG
CGGCCTTTGCCCACCGCCTCATCCCACCACGCCGCGGTCCTCCAAGAACCTTCCCAGCGGTTCTCCTCCTCTCAGGAGT F3X
AGAGGC**[C/t]**[2]CTCTGAGACCGACGGGG**[A/t]**[1]GGGACGGCTCGGGCCGGTCATCCGAGGGGCCGCACGGATTCCCTCCTCC
GCCCAGCTCCACCCCCTCGAGGGGCGGCGGTCCGGGAGTGGCGACCCGGCTCCCCCATGGCGCGCGCCGTCGGGGCCC
CTGGCCAGGCTCCGAGCGGGGTTGGCGGGGAGGGGAGGCGGGAGCGAGGGCGGGCGGTGGGAGGTGGGGGCGGGAA
GGTCCGAAGGCGGCGGCCTGAG*gctgcaccGGGCACGGGTCGGCCGCAATCCAGCCTGGGCGGAGCCGGAGTTGCGAGCC
GCTGCCTAGAGGCCGAGGAGCTCACAGC*atgggctggaggccccggagagctcgggggaccccgttgctgctgctgctactactgctgctgctc* Rev2

*Figure 5.4 Forward primers are denoted in red. Amplicon sizes, when used with the reverse primer (Rev2) in blue, are as follows: F1 1346 bp, F2 1087 bp, F3 637 bp, F3X 359 bp. The transcription start site is denoted by * followed by lower case letters, the start of the coding region (ATG into exon 1) is in lower case letters. Known promoter polymorphisms are underlined and in square brackets [major upper case/minor lower case] with superscript numbering for identification. [1]Polymorphism rs3918390 position -313 from ATG, A/t Major allele frequency 0.99 identified by Van Eerdewegh et al [79]. [2]Position -330 from ATG, C/t Major allele frequency 0.93, novel SNP identified by Chae et al [283]. [3]Position -753 from ATG, T/a Major allele frequency 0.97, novel SNP identified by Chae et al [283].[4]Position -2225 from ATG, G/a Major allele frequency 0.84, novel SNP identified by Chae et al [283]. Further polymorphisms were identified by Chae et al lying outside this sequence at positions -2154 [G/a] freq 0.70 and -1263 [C/g] freq 0.82. Polymorphism positions are measured from the ATG.*

5.2.1.2.1 Generation of the F1 deletion construct

Purified 1619 bp PCR product (~100 ng/µl) was diluted 1:25 and used as a template in a 20 µl PCR reaction containing 500 nM each of forward primer F1, and reverse primer Rev2 (table 2.7), with 500 µM dNTP and 2 mM $MgCl_2$. The reaction was amplified over 35 cycles using the following PCR cycling conditions: 94°C for two minutes, 94°C for twenty seconds, 55°C for twenty seconds, 68°C for ninety seconds (one minute/kb) and 68°C for twenty minutes. PCR product (3 µl) underwent electrophoresis (2.2.1.6.2) to observe amplicon size. The remainder was then purified using Qiagen PCR purification kit (2.2.3.1). The purified PCR product and pGL3 plasmid underwent double restriction digest (section 2.2.3.2), followed by gel extraction and concentration estimation (section 2.2.3.3). A ligation reaction was set up (section 2.2.3.4), followed by a transformation reaction (section 2.2.3.5). The following morning check PCR (section 2.2.3.6) of resulting colonies was performed using primers F1/Rev2 (table 2.7), to observe whether the insert was present. Positive colonies were inoculated into a starter culture and incubated for eight hours at 37°C (2.2.4.6), then purified using Qiagen miniprep kit (section 2.2.4.7). Double restriction digest was performed to check the correct size insert was present (section 2.2.3.2) before sequencing using pGL3 basic primers (RV3 and GL2, table 2.8), according to the protocol (section 2.2.1.6.4).

5.2.1.2.2 Generation of F2 deletion constructs using purified F1 plasmid as template

F1 plasmid (~80 ng/µl) was diluted 1:25 in sterile water to give an approximate concentration of 3.2 ng/µl. The diluted plasmid (1 µl) was used as template in a long PCR reaction. A mastermix containing 1x long PCR buffer (500mM Tris.HCl, pH 8.9, 140 mM ammonium sulphate), 1.3M betaine, 400 nM each of ADAM331650 forward and reverse primers (table 2.7), 400 µM dNTP and 2.5 mM $MgCl_2$ was made and added to the template. Taq DNA polymerase (1U) and Pwo DNA polymerase (0.008U) was added. The reaction was amplified over 35 cycles using the following PCR cycling conditions: 95°C for two minutes, 95°C for twenty seconds, 58.4°C for thirty seconds, 68°C for ninety seconds and 68°C for twenty minutes. PCR product (3 µl) underwent electrophoresis (2.2.1.6.2) to observe amplicon size.

5.2.1.2.3 Generation of F3 (637 bp) and F3X (359 bp) amplicons

The F3 amplicon was generated using 1 µl (~3 ng) F1 plasmid as template in a PCR reaction containing 500 nM each of forward primer F3 and reverse primer R2 (table 2.7), with 2 mM magnesium and an anneal temperature of 67.4°C. The F3X amplicon was generated using 1 µl (~3 ng) F1 plasmid as template in a PCR reaction containing 500 nM each of forward primer F3X and reverse primer R2 (table 2.7) with 1 mM magnesium and an anneal temperature of 62.8°C. The reactions were amplified over 35 cycles using the following PCR cycling conditions: 94°C for two minutes, 94°C for twenty seconds, anneal for thirty seconds, 68°C for two minutes (1 minute/kb) and 68°C for twenty minutes. PCR product (3 µl) underwent electrophoresis (2.2.1.6.2) to observe amplicon size.

5.2.1.2.4 Insertion of the PCR amplicons into pGl3 vector

The PCR reactions were then purified using Qiagen PCR purification kit (2.2.3.1). The purified PCR products underwent double restriction digest (section 2.2.3.2), followed by gel extraction and concentration estimation (section 2.2.3.3). A ligation reaction was set up (section 2.2.3.4) using previously prepared pGL3 vector (section 2.2.3.2), followed by a transformation reaction (section 2.2.3.5). The following morning a check PCR (section 2.2.3.6) on resulting colonies was performed on F2, F3 and F3X colonies, using primers F2/Rev2, F3/Rev2 and PGL3 (RV3 and GL2) respectively (tables 2.7 and 2.8), to observe whether the insert was present. Positive colonies were inoculated into a starter culture and incubated for eight hours at 37°C (2.2.4.6), then purified using Qiagen miniprep kit (section 2.2.4.7). Double restriction digest was performed to check the correct-size insert was present (section 2.2.3.2) before sequencing using pGL3 basic primers (RV3 and GL2, table 2.8), according to the protocol (section 2.2.1.6.4).

5.2.2 Choosing cell lines for transfection

5.2.2.1 Assessing ADAM33 expression in cell lines

ADAM33 is selectively expressed in airway fibroblasts and smooth muscle, but not in epithelial cells. Two fetal lung fibroblast lines (MRC5 and WI38) were available to us. It was also desirable to use a cell line that did not express *ADAM33*. Two human breast epithelial cell lines (H292 and MCF7) were available. The four cell lines were assessed for *ADAM33* expression using Reverse Transcription followed by Taqman® quantitative PCR (qPCR). RNA was extracted from the cell lines (section 2.2.2.6.1) and reverse transcribed (section 2.2.2.6.2) to create cDNA which was used in quantitative Taqman® PCR reactions (section 2.2.2.6.3).

5.2.2.2 Testing transfection efficiency of the cell lines using green fluorescent protein (pEGFP-N1)

MRC5, and WI38 fetal fibroblasts (both passage 18), and H292 (passage 40) and MCF7 epithelial cell lines (passage 3 following defrost from liquid nitrogen) were seeded at $3x10^4$, $3x10^4$, $2x10^4$ and $2x10^5$ respectively, into 6-well plates and grown overnight to reach ~40% confluence by the following morning (2.2.2.3).

Transfections were performed in duplicate. For each well a mastermix was prepared using 1 µg of DNA (pEGFP-N1 plasmid), 100 µl Buffer EC and 8 µl enhancer (8:1 ratio). Effectene reagent was added, according to the protocol (section 2.2.5.1), and the transfection mix applied to the cells. The media was changed six hours post-transfection and the cells left for forty eight hours before observing under a Zeiss axiovert microscope. Transfection efficiency was estimated by counting green and non-green cells.

5.2.3 Assessing activity of deletion constructs

5.2.3.1 Observing baseline and stimulated activity of deletion constructs

The aim of this experiment was to look at promoter activity of *ADAM33* using the deletion constructs manufactured in section 5.2.1, firstly to look at basal activity, and secondly see if reporter activity could be regulated up or down. Interleukin-13 (IL-13) and interferon-γ (IFN-γ) were chosen as stimulants based on their known roles in

asthma and immune regulation. Allergic diseases are associated with Th_2 immune responses, which are characterised by high levels of interleukins (IL)-4, IL-5, IL-9 and IL-13. Interleukin-13 is a central immunoregulatory cytokine produced by activated lymphocytes in allergic asthma. The biological activities of IL-13 include B cell proliferation and immunoglobulin production. It has the potential to modulate airway inflammation and induces airway hyperresponsiveness independently of IgE and eosinophilia in animal models [284, 285]. Previous stimulation of epithelial cell lines (MCF7 and H292) with IL-13 did not result in a change in *ADAM33* expression (data not shown). Interferon-γ is a cytokine secreted by Th_1 cells, which was reported to down-regulate *ADAM33* in cultures of human airway smooth muscle cells [286]. A transcription factor recognition program, MatInspector [282], was used to identify potential binding sites along the promoter region of *ADAM33*. A potential steroid response element (SRE) was observed. For a steroid hormone to turn gene transcription on, its receptor must first bind to the hormone, bind to a second copy of itself to form a homodimer, move from the cytoplasm to the nucleus and then bind to a steroid response element (SRE) in order to activate other transcription factors to start transcription. Dexamethasone, a steroid used in asthma treatment for its anti-inflammatory effects, was chosen as a third stimulant to test whether the SRE was implicated in *ADAM33* regulation. The final stimulant was phorbol ester. Phorbol esters are proteins which contains at least one phorbol ester/diacylglycerol binding domain. Phorbol ester (Phorbol Myristate Acetate or PBA) is an analogue of diacylglycerol (DAG) which activates the protein kinase C (PKC) signal transduction pathway via pathways including the AP1 (activator protein 1) response element [274], map kinase pathways and Sp1-dependent mechanisms [287]. PBA was therefore used to investigate whether the PKC pathway was important in *ADAM33* activation.

5.2.3.1.1 Transfection of plasmid constructs

Experiments were performed in triplicate. H292 cells (passage 42-46) were plated out at $2x10^4$ into 6-well plates and grown overnight to reach ~40% confluence by the following morning (2.2.2.3). Treatments were not performed on these cells. MRC5 cells (passage 20-22) were plated out at $3x10^4$. For each construct enough MRC5 cells were plated out to allow four different treatments and a no-treatment control.

Cells were transfected with 1 µg of either F1, F2, F3 or F3X deletion construct plus *Renilla* luciferase (300 ng/well) as an internal control, according to the protocol (section 2.2.5.1). Treatments were applied to cells ~thirty two hours post-transfection. Treatment stocks were prepared according to the protocol (section 2.2.5.1.1) and diluted to the following concentrations using the cell's usual media, IL-13 10 ng/ml, IFN-γ 10 ng/ml, dexamethasone 1 µM and PBA 10 nM, before applying to the cells. At this time the no-treatment control cells had the media changed. All cells were harvested sixteen hours post-treatment according to the protocol (section 2.2.5.1).

5.2.3.1.2 Measurement of luciferase activity

Luciferase activity was measured according to the protocol (2.2.5.2). Briefly, lysed cells (20 µl) were added to 100 µl luciferase substrate and luminescence measured in triplicate on a 20/20 Turner luminometer. *Renilla* luciferase substrate (100 µl) was then added and luminescence measured in duplicate. A ratio of average luciferase to average *Renilla* was calculated.

5.2.4 *ADAM33* homology

Then comparative genomic tool, mVista, was used to identify homologous and conserved sequences in the promoter regions of the closest relatives of human (h) *ADAM33* (*hADAM19*, *hADAM15*, *hADAM12* and *hADAM8*) and in the putative orthologs of *ADAM33* in mouse (*Mus musculus*), rat (*Rattus norvegus*), dog (*Canis familiaris*) and chimp (*Pan troglodytes*). Sequences were taken from Ensembl Genome Browser (http://www.ensembl.org) and included 1500 bp of 5'UTR plus exon one. Sequence details are in listed in table 5.2

Table 5.2 Ensembl sequence used for comparative genomics

Gene name / study name	Organism	Ensembl gene identity	Chromosome location
ADAM33 / hADAM33	Homo sapiens	ENSG00000149451	Chr 20 location 3,596,618 – 3,611,337
ADAM19/ hADAM19	Homo sapiens	ENSG00000135074	Chr 5 location 156,840,033 – 156,935,318
ADAM15 / hADAM15	Homo sapiens	ENSG00000143537	Chr 1 location 151,836,835 – 151,848,325
ADAM12 / hADAM12	Homo sapiens	ENSG00000148848	Chr 10 location 127,693,831 – 128,067,014
ADAM8 / hADAM8	Homo sapiens	ENSG00000151651	Chr 10 location 134,965,542 – 134,979,211
ADAM33 / mADAM33	Mus musculus	ENSMUSG000000027318	Chr 2 location 130,564,679 – 130,577,423
- / *rADAM33*	Rattus norvegus	ENSRNOG00000021242	Chr 3 location 118,728,328 – 118,764,698
- / *cADAM33*	Canis familiaris	ENSCAFG00000006234	Chr 24 location 20,686,339 – 20,699,136
ADAM33 / pADAM33	Pan troglodytes	ENSPTRG00000013200	Chr 20 location 3,575,024 – 3,590,347

Table 5.2 Ensembl sequence details. Rattus and Canis ortholog sequence have no gene name assigned so a study name has been assigned to all sequences for clarity.

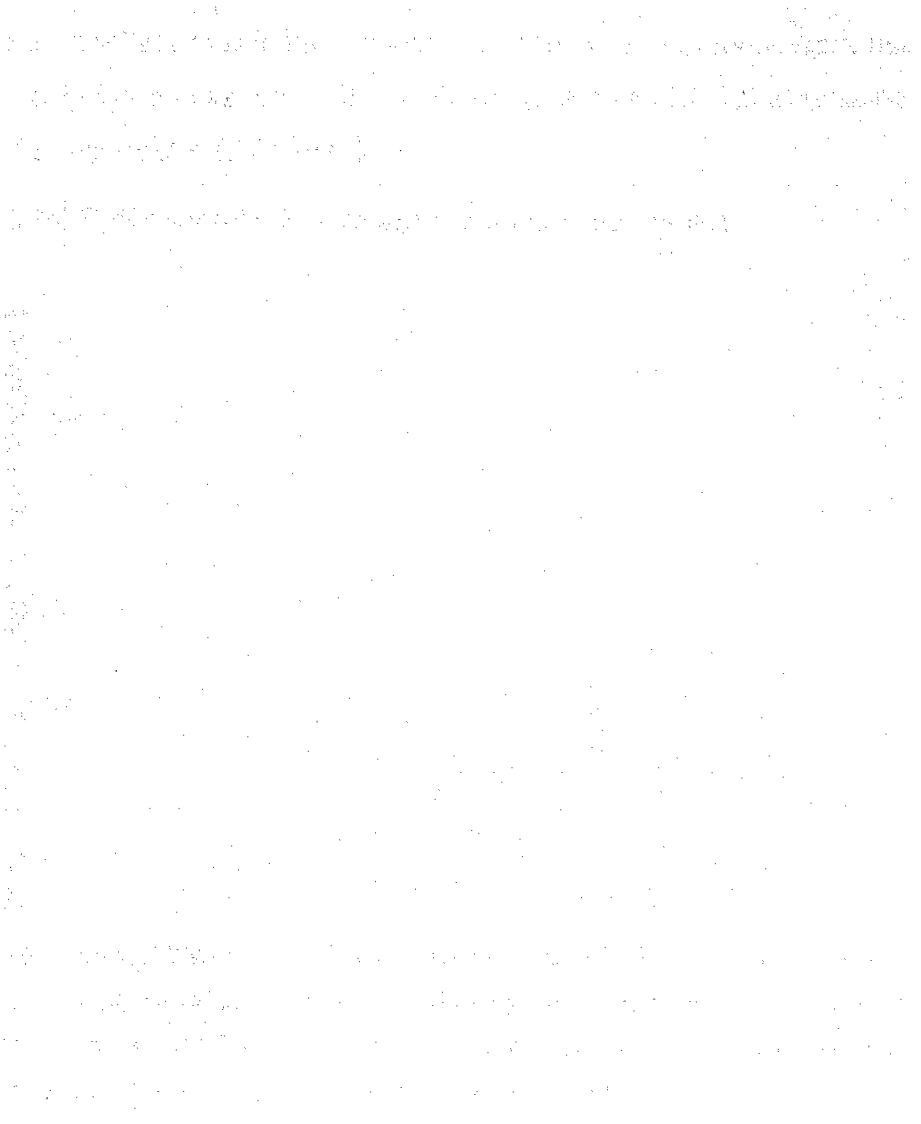
5.2.5 Putative transcription factor binding sites

The human *ADAM33* promoter sequence was loaded into the MatInspector software [282] and transcription binding sites identified. The sequence was run again using tissue filters to identify transcription factors specific to certain tissues. Filters were chosen based on known ADAM33 expression [171] and included respiratory tissue, smooth muscle, lung, heart, digestive system, urogenital and embryonic structures.

5.2.6 Over expression of Sp1

The luciferase experiments were repeated, using an Sp1 expression plasmid (pPac Sp1, kindly supplied by Dr I. Clarke, University of Norwich) which was co-transfected with the promoter reporters F2, F3 and F3X to assess the activity of the Sp1 sites. MRC5 fetal fibroblasts (passage 18) and MCF7 cells (passage 4 following defrosting from liquid nitrogen) were prepared according to the protocol (section 2.2.1). Transfections were performed in triplicate. For each transfection a mastermix was prepared by mixing 1 µg plasmid DNA, 1 µg Sp1 expression plasmid with 100 µl buffer EC. This was flick-mixed and left for forty five minutes on ice, to ensure even distribution of plasmid. *Renilla* luciferase plasmid (300 ng) and 8 µl enhancer (8:1

ratio) was then added and the protocol continued as described (section 2.2.5.1). Cells were incubated for six hours at 37°C, when the media was removed and 2 ml fresh media applied to each well. The cells were incubated for a total of 48 hours post-transfection before harvesting (section 2.2.5.1).



5.3 Results

5.3.1 Generation of promoter constructs

5.3.1.1 Generation of a 1619 bp amplicon spanning the ADAM33 promoter region

The ADAM331650 primers were shown to bind to genomic region 20p13. The forward primer showed no complementarity with other regions. The reverse primer also showed complementarity to a region of chromosome 18 but since this would not result in exponential amplification, it was retained. These primers were used with the PCR protocol stated in section 5.2.1.1. Electrophoresis (2.2.1.6.2) revealed a single product of the correct size (figure 5.5).

Figure 5.5 PCR of genomic DNA using ADAM331650 primers

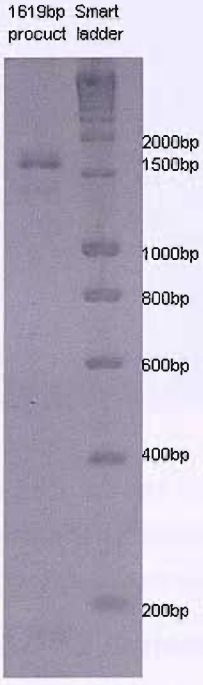


Figure 5.5 75 ng pooled DNA was used as template in a 20 µl PCR reaction using ADAM331650 primers. Product (3µl) was loaded onto a 2% agarose gel with a quantifying molecular marker ladder (Smartladder (Eurogentec)). This was run at 100 V for twenty minutes and stained using EtBr (0.5 µg/ml) for fifteen minutes. Electrophoresis of PCR products revealed a single amplicon of 1619 bp.

After purification (2.2.3.1), the product was digested using the restriction enzyme HphI. Three fragments of expected size (1083 bp, 346 bp and 181 bp) were observed

(figure 5.6). Blast analysis of the primers, molecular marker sizing of the amplicon and restriction digest suggested the product to be to the correct amplicon.

Figure 5.6 Electrophoresis of the restriction digest products

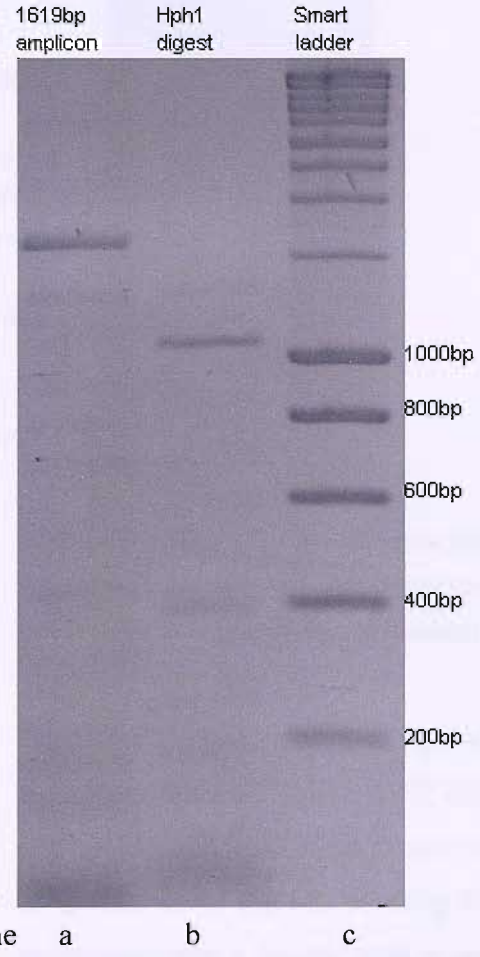


Figure 5.6 The 1619 bp PCR product (lane a) underwent Hph1 restriction digest. Electrophoresis of the restriction digest products (lane b) revealed the expected size fragments of 1083 bp, 386 bp and 181 bp. Sizes were estimated (lane c) using Smart ladder molecular marker (Eurogentec).

5.3.1.2 Generation of the F1 deletion construct using purified 1619 bp amplicon as template

The purified 1619 bp PCR product (section 2.2.3.1) was used as template to generate a 1346 bp amplicon (F1). Electrophoresis of the product (figure 5.7) revealed a 1346 bp amplicon.

Figure 5.7 F1 1346 bp amplicon

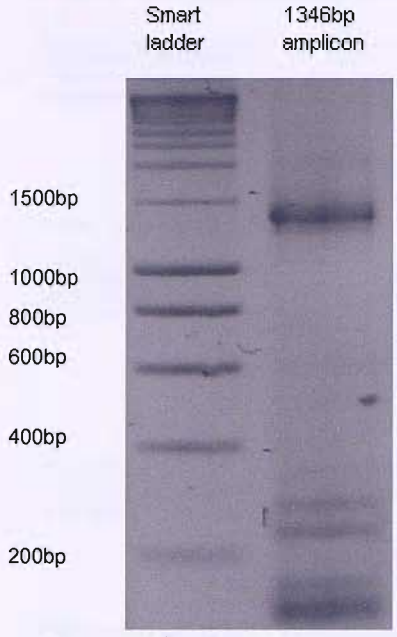


Figure 5.7 A 20 µl PCR reaction was set up using F1/rev2 primer set. Product (3 µl) was loaded onto a 2% agarose gel with a quantifying molecular marker ladder (Smartladder (Eurogentec)). This was run at 100 V for twenty minutes and stained using EtBr (0.5 µg/ml) for fifteen minutes. Electrophoresis revealed an amplicon of 1346 bp.

The 1346 bp amplicon was ligated into pGl3 basic vector (sections 2.2.3.1-2.2.3.4) and transformed into *E-Coli* (2.2.3.5). Reactions were plated onto agar plates containing ampicillin and grown overnight. Colonies growing overnight were checked for presence of the insert using PCR (section 2.2.3.6). Positive colonies were inoculated into a starter culture and incubated for eight hours at 37°C (2.2.4.6). The next day the plasmid culture was purified using Qiagen miniprep kit (section 2.2.4.7.1). The purified F1 plasmid product was quantified using a Beckman DU-7000 DNA quantitator (section 2.2.4.7.4) and a double restriction digest performed (section 2.2.3.2), confirming the insertion of the 1346 bp sequence into the pGL3 plasmid. The sequence of the plasmid was verified using pGL3 basic primers (RV3 and GL2, table 2.8), according to the protocol (section 2.2.1.6.4). This confirmed the correct insert sequence was present (figure 5.8).

Figure 5.8 Sequencing of F1 plasmid

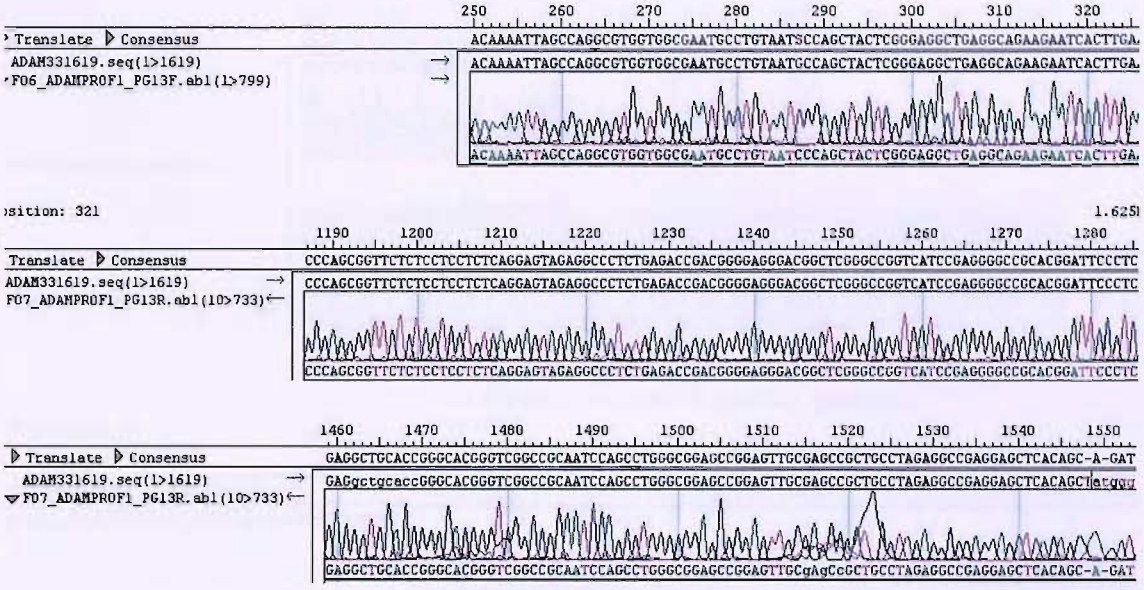
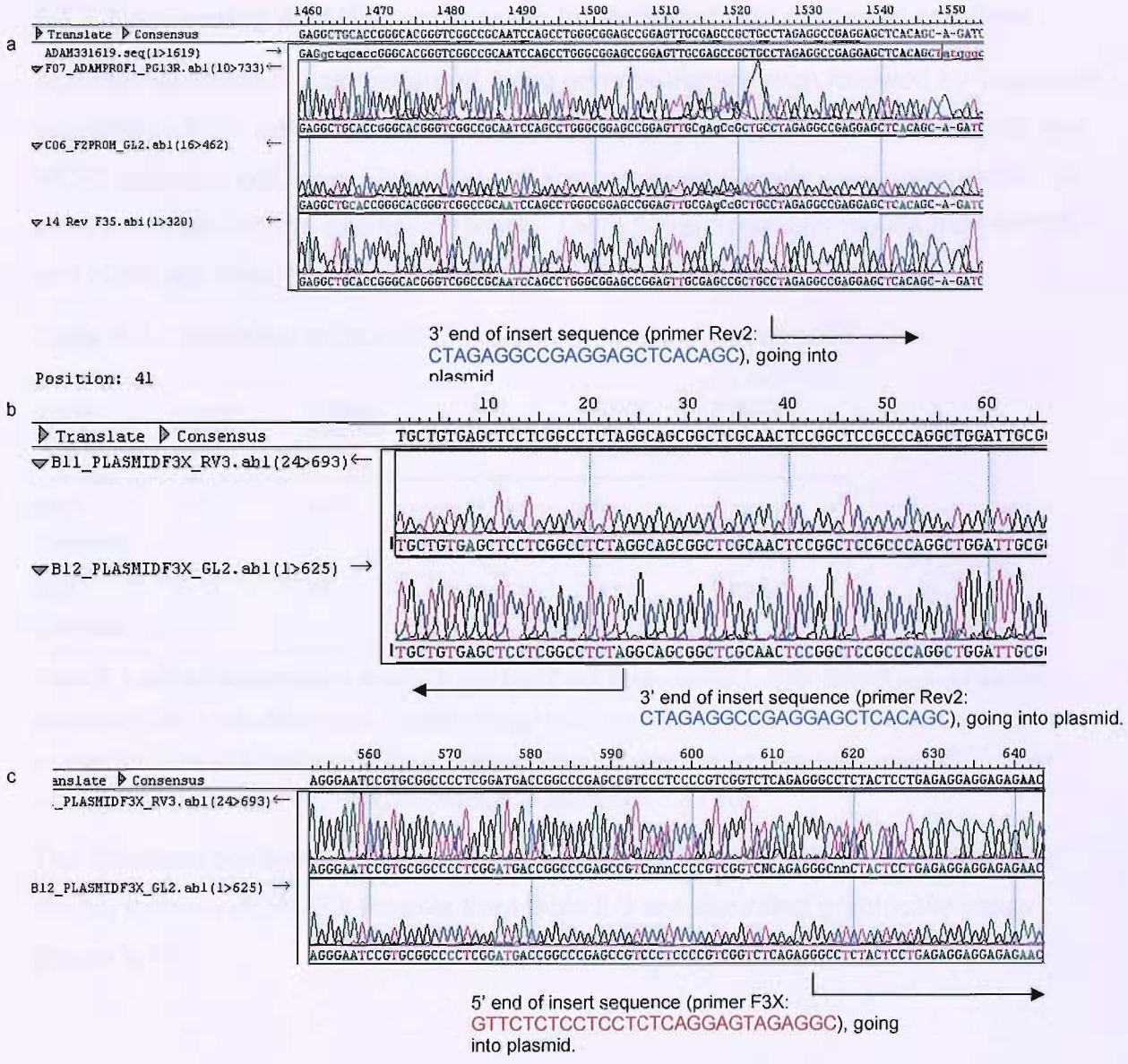


Figure 5.8 illustrates examples of the F1 plasmid sequence. The consensus sequence was imported from NCBI. Arrows define direction of sequence (green forward, red reverse). Sequence a) shows a C/G polymorphism at position 287 previously reported as -1263 by Chae et al [283]. Sequence b) shows complete sequence agreement between the plasmid insert and consensus sequence. Sequence c) shows that at position 1547, agreement between the plasmid insert sequence and consensus sequence ends. This is the start of the pGL3 sequence.

5.3.1.3 Generation of F2, F3 and F3X deletion constructs using F1 plasmid

Amplification of the F1 plasmid using primer sets F2/Rev2, F3/Rev2, F3X/Rev2 resulted in a 1087 bp amplicon, a 637 bp amplicon and a 359 bp amplicon respectively. The PCR products were processed as for F1 (5.3.1.2) and then the sequences of the three plasmids, F2, F3, F3X, were verified using pGL3 basic primers (RV3 and GL2), and Pro C (table 2.8), according to the protocol in section 2.2.1.6.4. This confirmed the correct insert sequences were present.

Figure 5.9 Examples of sequencing of F2, F3 and F3X plasmids



a) F2 and F3 sequences run into plasmid sequence where consensus breaks b) F3X sequence verification showing the 3' end of the insert sequence matching the primer Rev2. c) F3X sequence verification showing the 5' end of the insert sequence matching the primer F3X.

5.3.2 Choosing cell lines for transfections

5.3.2.1 Assessing ADAM33 expression in fibroblast and epithelial cell lines

ADAM33 expression was measured using reverse transcription followed by Taqman® quantitative PCR (qPCR) in MRC5 and WI38 fetal fibroblast cell lines, and H292 and MCF7 epithelial cell lines. Fibroblast cell line expression levels were comparable, as were epithelial cell line expression levels. Table 5.3 summarises results from MRC5 and H292 cell lines.

Table 5.3 Calculation of ADAM33 expression in two different cell lines

Cell line	Average 18S	Average ADAM33	ΔCT (CT_T-CT_R)	ΔΔCT (MRC5 calibrator)	2^ΔΔCT
MRC5 (2 samples)	14.8	30.25	15.45	0	1
H292 (2 samples)	13.15	40	26.85	-11.4	0.00037

Table 5.3 ADAM33 expression in MRC5 and H292 cell lines relative to 18S. MRC5 is used as the calibrator (ΔΔCT =0). Since each 1 point change in CT constitutes a doubling of product, abundance of product between cell lines, in relation to the calibrator cell line, can be calculated using $2^{\Delta\Delta CT}$. Thus when MRC5 expression = 1, H292 expression is seen to be 3.7×10^{-4}.

The fibroblast cell line (MRC5) expressed *ADAM33* and the epithelial cell line H292 did not express *ADAM33*. Results from table 5.3 are illustrated graphically below (figure 5.10).

Figure 5.10 Comparison of ADAM 33 expression in fibroblast (MRC5) and epithelial (H292) cell lines

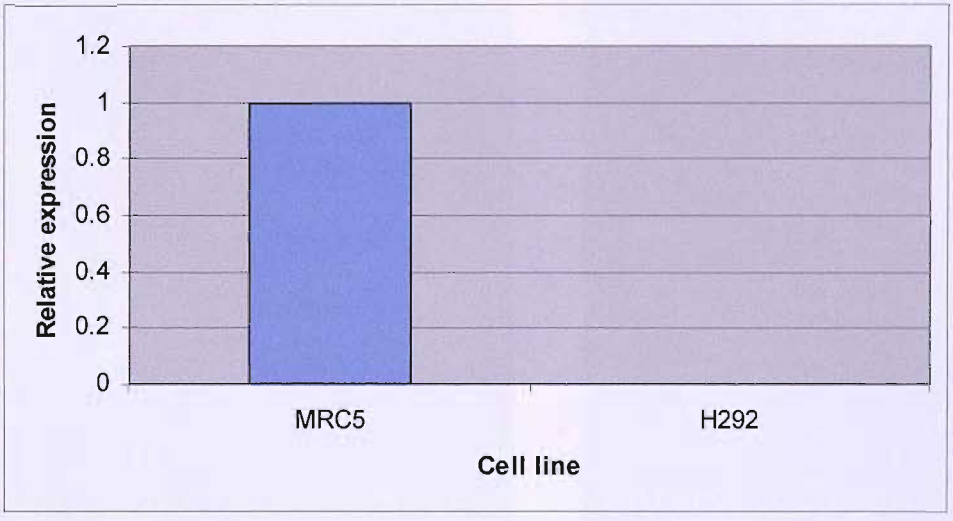


Figure 5.10 qPCR results are calculated using 18s as the reference gene. Results confirm expression of ADAM33 in fetal lung fibroblast cell line (MRC5) but not in the epithelial cell line (H292).

5.3.2.2 Transfection of GFP plasmid to test transfection efficiency

The MRC5 and WI38 fetal fibroblasts (both passage 18) and H292 (passage 40) and MCF7 (passage 3 following defrost from nitrogen stocks) epithelial cell lines were transfected with GFP. In all cases approximately 30% of cells exhibited green fluorescence forty eight hours post-transfection. Figure 5.11 illustrates GFP transfection of the MRC5 cell line. It was noted that WI38 fibroblasts were less robust post-transfection, with higher cell death and a slowing down of growth. The epithelial cell lines were comparable in *ADAM33* expression, transfection ability, sensitivity to transfection and growth rate.

Figure 5.11 GFP transfection of the fetal fibroblast cell line, MRC5

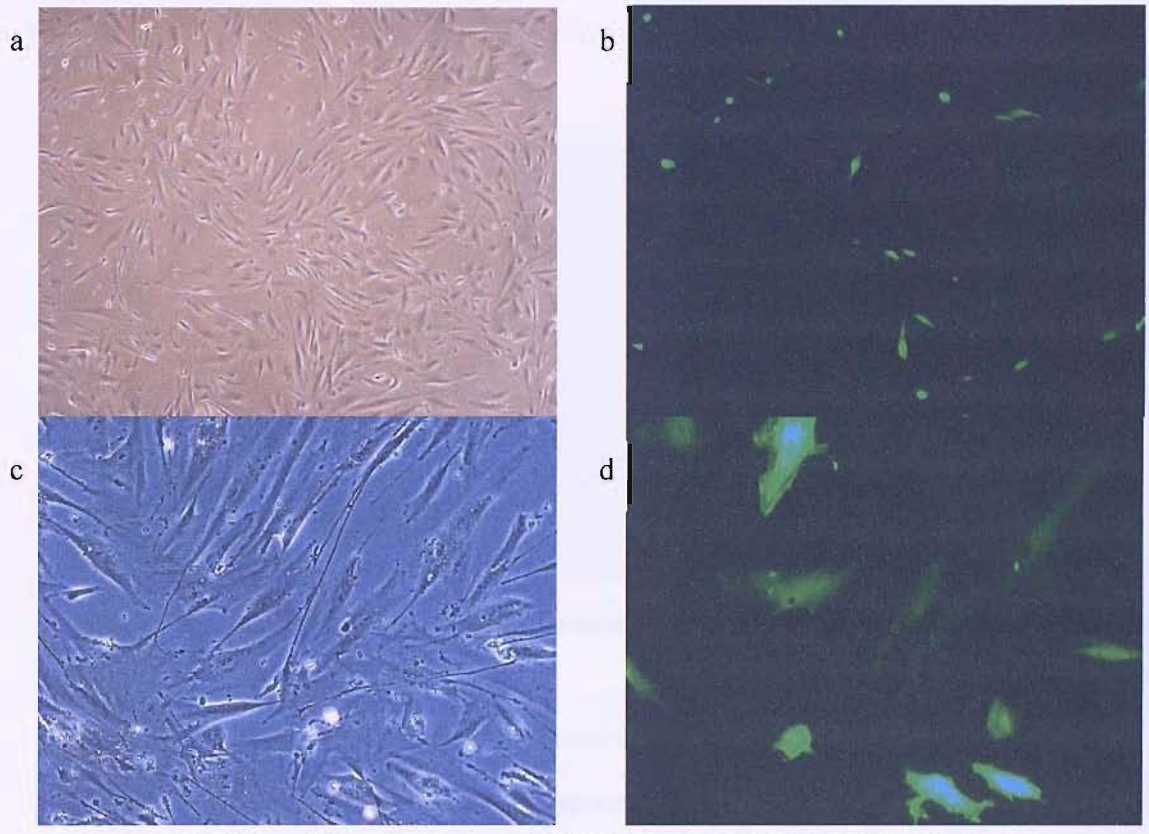


Figure 5.11 a and b illustrate MRC5 cells (passage 18) 48 hours after transfection with GFP. 5x magnification a) white light and b) FITC filter. The same cells are shown at 20x magnification c) white light and d) FITC filter. Transfection efficiency was estimated at approximately 30%

5.3.3 Transfection of deletion constructs

5.3.3.1 Baseline activity of deletion constructs in MRC5 fetal fibroblast cells and H292 epithelial cells

Cells were transfected and harvested according to the protocol (section 2.2.5.1).

H292 and MRC5 cells were transiently transfected with F1, F2, F3, F3X and no-insert (NI) constructs. H292 cells show minimal expression. MRC5 cells show a low (comparable) basal expression when transfected with constructs F1, F2 and F3. When MRC5 cells were transfected with the F3X construct, a marked increase in expression was observed.

Figure 5.12 Basal reporter luciferase expression in H292 and MRC5 cell lines

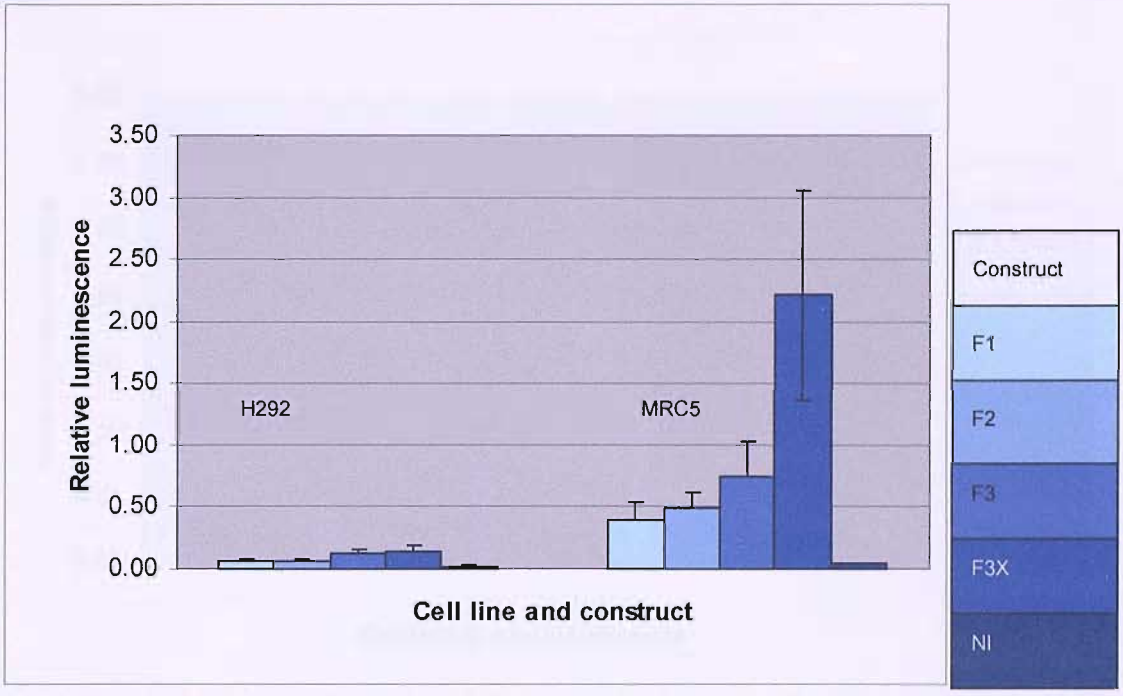


Figure 5.12 Cell type is denoted above the columns. Constructs are colour coded. NI: No insert.

5.3.3.2 Do stimulation treatments increase activity of the ADAM33 promoter constructs in transfected fetal fibroblast cells?

Transiently transfected MRC5 fibroblast cells treated with PBA, IL-13, IFN-γ or dexamethasone showed no significant change in expression (figure 5.13).

Figure 5.13 PBA, IL-13, IFN-γ and dexamethasone treatments on transfected MRC5 cells

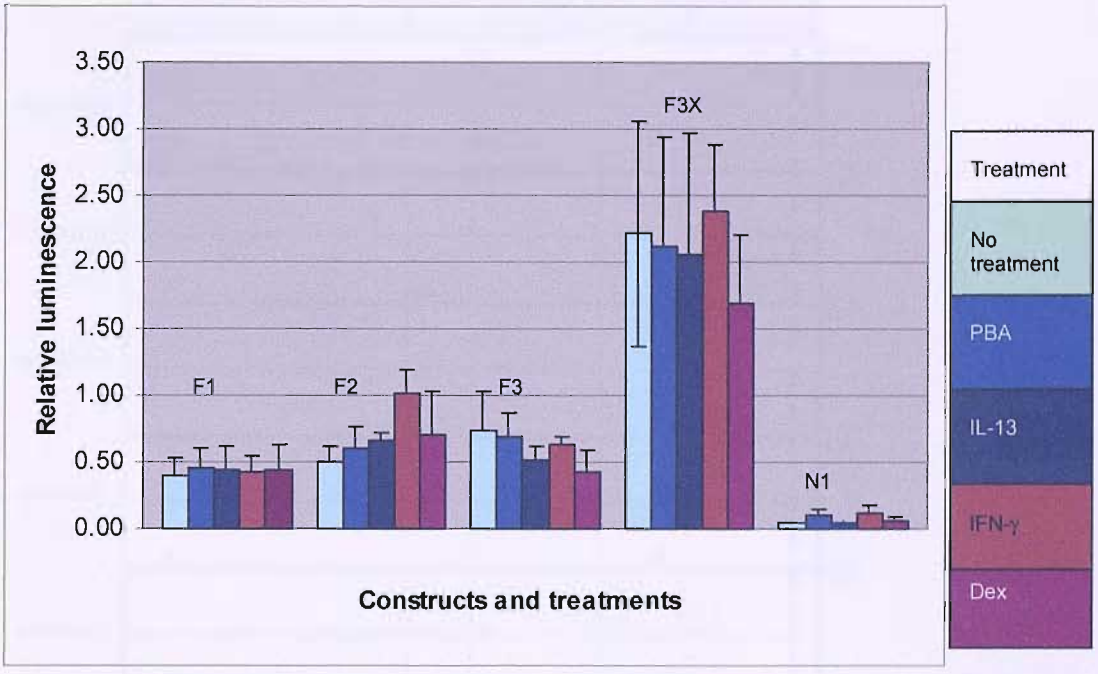


Figure 5.13 Treatments are colour-coded according to the key. Constructs are denoted above the columns.

5.3.4 Comparative genomics

Sequences from the promoter regions of the genes listed in table 5.2 were compared. Using default settings (minimum identity 70%, minimum length 100 bp) results revealed significant homology to *Pan troglodytes* and some conservation to *Canis familiaris (*data not shown). However, by reducing the parameters (minimum identity 40%, minimum length 10 bp) several minimally conserved regions were identified in other orthologs (figure 5.14). Related human ADAMs showed little conservation.

Figure 5.14 Conserved regions of selected ADAMs

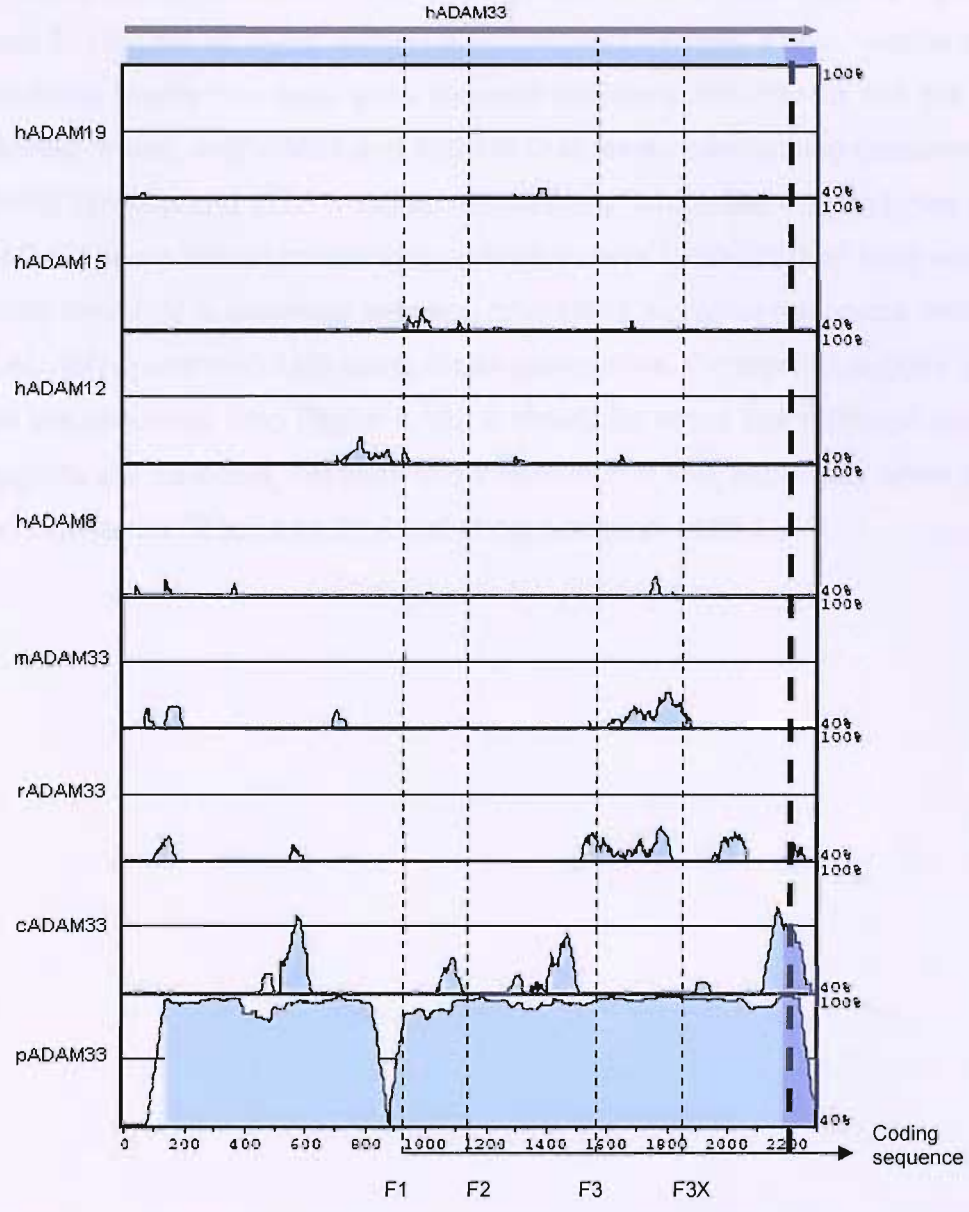


Figure 5.14 Human ADAM33 promoter sequence is compared against most related human ADAMs (hADAM Homo sapien ADAMs 19, 15, 12,and 8), and orthologs of other species (mADAM33 Mus musculus ADAM33, rADAM33 Rattus norvegus ADAM33, cADAM33 Canis familiaris ADAM33 and pADAM33 Pan troglodytes ADAM33. Parameters: minimum y axis 40%; minimum identity 40%; minimum sequence length 10 bp. Pale blue regions denote conserved areas, non-coloured peaks are <40% conserved. Mid blue denotes the start of coding sequence. The dotted lines illustrate the position of cloned fragments in relation to the whole sequence.

Results show 93% homology between *hADAM33* and *pADAM33*. This level of conservation gives little useful information. *hADAM33* and *cADAM33* share several

conserved regions at positions 513-605 bp (upstream of the cloned region), 1055-1100 bp (lying just within F1 spreading into F2), and 1420-1496 bp (lying within F2), and 2113-2261 bp (lying within F3X). *hADAM12* shows a small region of conservation covering twenty four base pairs between positions 766-790 bp, but this is outside the cloned region. *mADAM33* and *rADAM33* show an overlapping conserved region 1760-1850 bp and 1750-1800 bp respectively, which lies within clones F3 and F3X. *rADAM* has a further conserved region between 1950-2050 bp lying within clone F3X. Little similarity is observed between *hADAM33* promoter sequence and *hADAM19*, *hADAM15*, and *hADAM8* using these parameters. Conserved regions are highlighted on the sequence map (figure 5.15). It should be noted that although conserved regions are detected, the level of conservation is low, especially when considering the window size (10 bp) and the level of conservation (40%).

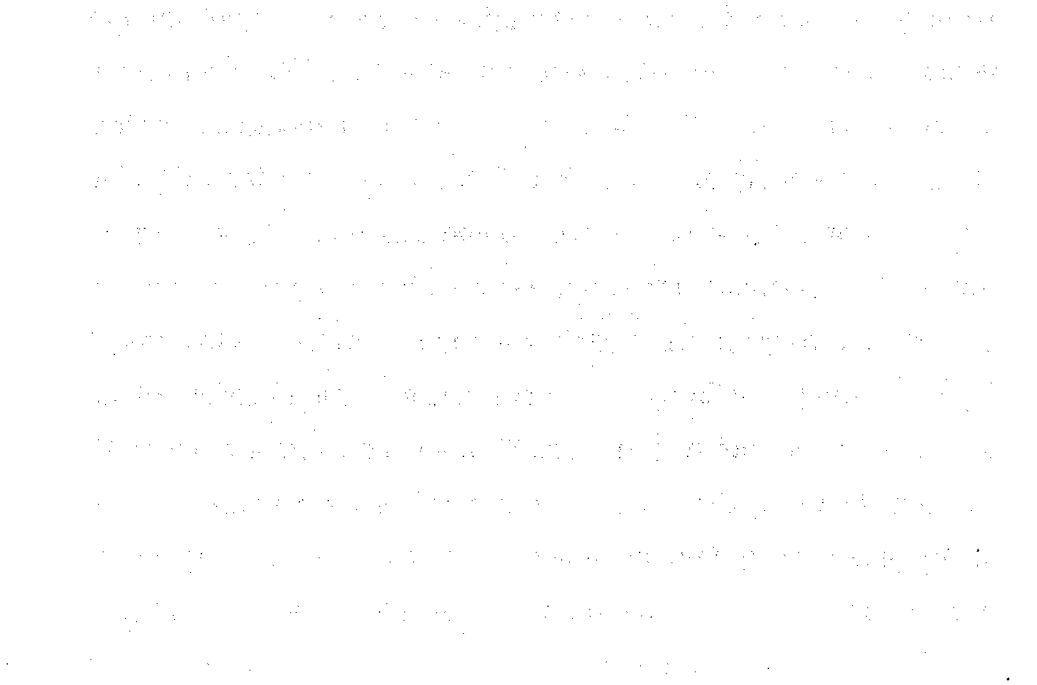


Figure 5.15 Sequence map of ADAM33 conserved regions

TTCGAGACCAGCCTGACCAACATGGAGAAACCCCATCTCTACTAAAAATACAAAATTAGCCAGGCGTGGTGGCGAATGCCT
GTAATGCCAGCTACTCGGGAGGCTGAGGCAGAAGAATCACTTGAATCTGGGAGGCAGAGGTTGCAGTGAGCCCAGATCAT
GCCATTGCACTCCAGTCTGGGTAACAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAGATTTGCTTAAAAGTTAA
CATCTCCGGCCGGGCGCGGTGGCTCATGCCTGTAATCCCAGCGCTTTGAGAGGCCGAGGCGGGTGGATCACGAGATCAG
GAGATTGAGACCATCCTGGCCAAAATGGTGAAACCTCGTCTCTGCTAAAAATACAAAAGTTAGCTGGGGGTGGTAGCGCGC
GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCAGGGAGTCGGAGGTTGCAGTGAGCCAAGA
TCGCGCCGCTGCACTCCAGCCTGGCGACAGAGGGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAGTTAACATCTCATCC
AAATTTGCACCGAGTAGGAAAACAAAAGTTTAAAACA[T/a]³GAAACAGATGTTACTGAGGCCGAAAGGGTCTCCCAGGCCTG
GGAGTCTGCAGCTTTTATGCAATTCTGCCCTCTGGCCACCGCCAGGGAAGAAAGGTTGTCTCCGTCTGCTTGCATCGCTTT
GCCCAGCAATGAAGCCCCCAAGACAGCGGCAGCCGGTTGCCTGAACCTTCCTATCCTTGGGGGCACCCAGTGCAGGTGGA
TGACCCGACTCAACCTCCGCCAGGGCACCCTCGGGGCAGGACGGGTAGCAAGGAGGGGACAGAGATCGGCCCCCAGGAGA
CCACGGAAGATCGCGCTCCTGGGGCCAACTTCAGCAGCGAGAGGCGGCCTTTGCCCACCGCCTCATCCCACCACGCCGC
GGTCCTCCAAGAACCTTCCCAGCGGTTCTCTCCTCCTCTCAGGAGTAGAGGC[C/t]²CTCTGAGACCGACGGGG[A/t]¹GGGA
CGGCTCGGGCCGGTCATCCGAGGGGCCGCACGGATTCCCTCCTCCGCCCAGCTCCACCCCCTCGAGGGGCGGCGGTCC
GGGAGTGGCGACCCGGCTCCCCCATGGCGCGCGCCGTCGGGGCCCCTGGCCAGGCTCCGAGCGGGGTTGGCGGGGAG
GGGAGGCGGGAGCGAGGGCGGGCGGTGGGAGGTGGGGGCGGGAAGGTCCGAAGGCGGCGGCCTGAG*gctgcaccGGGC
ACGGGTCGGCCGCAATCCAGCCTGGGCGGAGCCGGAGTTGCGAGCCGCTGCCTAGAGGCCGAGGAGCTCACAGCatgggct
ggaggccccgg

*Figure 5.15 Conserved regions in the promoter sequence based on comparative genomic analysis (figure 5.14). The hADAM33 sequence starts at the beginning of the F1 primer (forward primers are in red and reverse in blue). Conserved sequences are highlighted or underlined: Canis familiaris pale blue highlight; mouse underlined and rat dark green highlight (some overlap with mouse). The transcription start site is denoted by * and lower case letters, the start of the coding region (ATG into exon 1) is in lower case letters.*

5.3.5 Identification of putative transcription binding sites using MatInspector

Results from the luciferase assays suggested a possible inhibitor within sequence F3 which when removed (F3X) afforded a marked up-regulation of expression. There was little difference between sequence F1, F2 and F3. Conserved regions occurred mainly within sequences F3 and F3X. The F3 clone sequence was imported into MatInspector to identify putative binding sites. The screen for all vertebrate transcription factors revealed ninety one putative transcription binding sites (data not shown). A group filter, incorporating transcription factors specific to respiratory tissue, smooth muscle, lung, heart, digestive system, urogenital and embryonic structures, revealed twenty three matches in the F3 sequence. A further screen using a filter for respiratory tissue alone revealed one transcription binding site, MyoD, situated within F3 sequence (figure 5.16). The MyoD site was not recognised using the smooth muscle filter. *ADAM33* is expressed in smooth muscle, suggesting MyoD may not be important for activation, but could be involved in repression [288, 289]. The positions of the putative sites were considered in relation to the conserved regions identified (section 5.3.4). Several transcription factor binding sites were recognised within the conserved region that had ubiquitous or unknown tissue expression. Of note were

Sp1 binding sites within F3X sequence and two mini-muscle initiator sites upstream of and within the murine-conserved region (figure 5.16).

Figure 5.16 Putative transcription binding sites and conserved regions

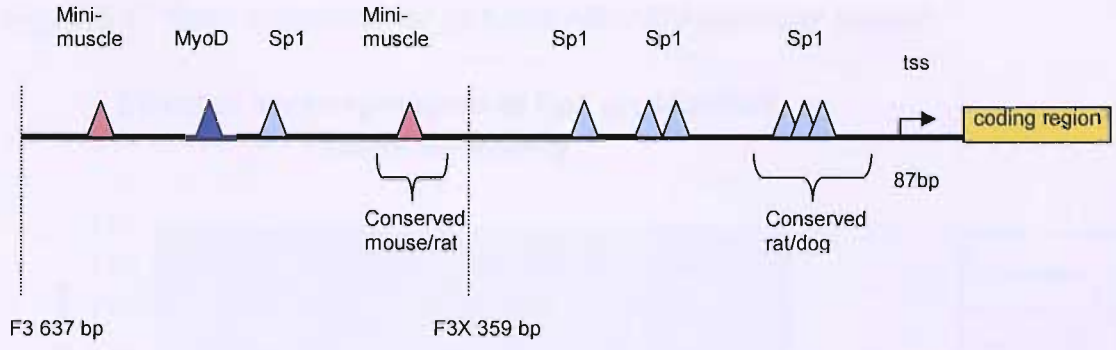


Figure 5.16 Putative binding sites of the ADAM33 promoter based on deletion analysis, comparative genomics and searches for putative binding sites. Coloured triangles denote positions of putative binding sites in relation to conserved regions and deletion constructs. tss: transcription start site.

5.3.6 Over expression of SP1 and ADAM33 promoter activity

The results of co-transfection of SP1 expression plasmid with the promoter reporters are illustrated in figure 5.17.

Figure 5.17 Over expression of Sp1 and ADAM33 promoter activity

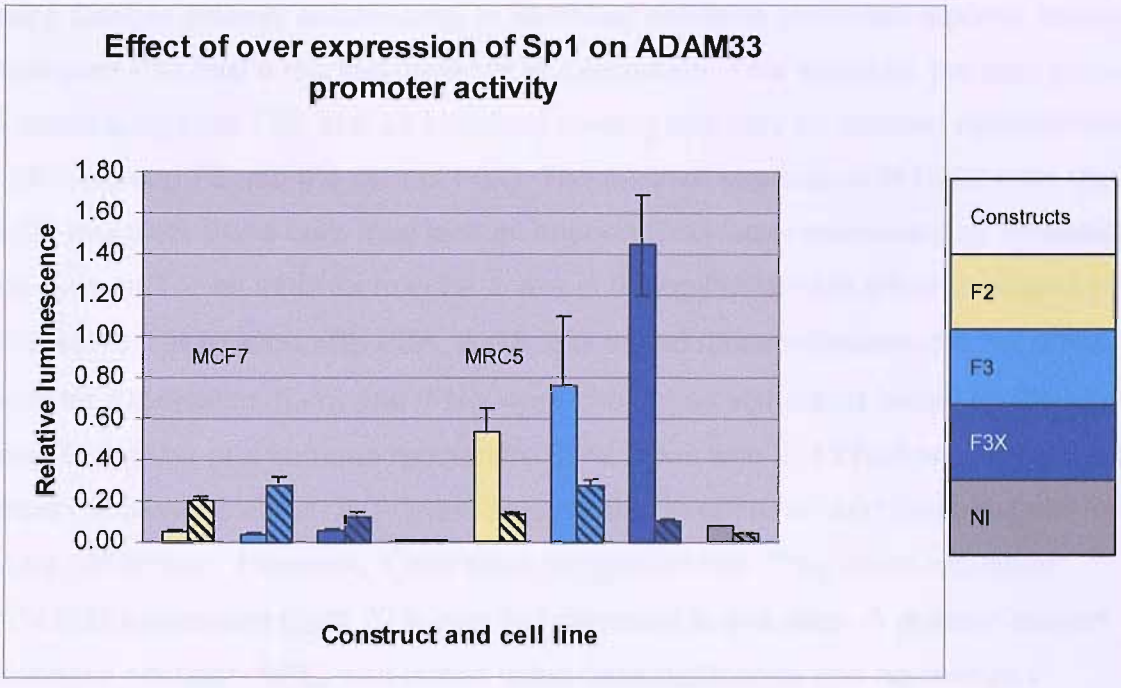


Figure 5.17 Constructs are colour-coded according to the key. Block colour denotes transfection with promoter reporter, diagonal stripes denote cells additionally transfected with Sp1 plasmid. Renilla luciferase control was used as a control in all transfections. NI denotes no insert control.

In the epithelial cells (MCF7), over-expression of Sp1 appears to increase expression, whereas over-expression in fibroblasts (MRC5) cells has the opposite effect, reducing levels to that seen in the SP1 transfected epithelial cells.

5.4 Conclusions

The first aim of this chapter was to identify the core promoter region necessary for basal expression of *ADAM33* and define any positive or negative regulatory regions. Four deletion constructs were manufactured and verified as the correct sequences using Sanger dideoxy sequencing. In fibroblast cell lines luciferase reporter assays illustrated F3X had a marked increase in expression. This suggests the core promoter is within sequence F3X and an inhibitory binding site may be situated upstream of F3X (between F3 and the start of F3X). The minimal expression in H292 cells (figure 5.12) suggests these cells may lack an important co-factor necessary for activation of the construct or an inhibitor may be active in the epithelial cells which is absent in fibroblasts. Stimulation with PBA, IL-13, IFN-γ, and dexamethasone did not affect reporter expression. IL-13 and IFN-γ were chosen as stimulants based on their known roles in asthma and immune regulation. Stimulation with IL-13 had previously been shown to have no effect on *ADAM33* expression in epithelial and fibroblast cell lines (data not shown). However, it had been suggested that IFN-γ down-regulated *ADAM33* expression [286]. This was not observed in this case. A putative steroid response element (SRE) was tested using dexamethasone and no change in expression was observed, suggesting it is not implicated in *ADAM33* regulation. The final stimulant was phorbol ester. Phorbol esters activate the PKC pathway of signal transduction via mechanisms including Sp1-dependent mechanisms [287]. It was thought that PMA would up-regulate reporter expression. This was not the case suggesting the PKC pathway may not be important in *ADAM33* activation.

The second aim was to identify possible conserved biologically active regions of the *ADAM33* promoter region, using comparative genomic tool, mVISTA. The F1 sequence (1346 bp) was used, as it incorporated all promoter constructs. Results were inconclusive. Homology between *hADAM33*'s closest relatives, *hADAM19*, *hADAM15*, and *hADAM8* was minimal. The conserved region observed between *hADAM12* and *hADAM33* was upstream of the sequence under study, but again was minimal. The *ADAM33* ortholog in mouse (*Mus musculus*) and rat (*Rattus norvegus*) shared some regions of low conservation (40% similarity using a 10 bp window), situated within F3 and F3X reporter regions. Homology between *hADAM33* and dog

(*Canis familiaris*) was higher than that seen in rat or mouse (figure 5.14), and occurred in different regions of sequence. Dogs are phylogenetically more distant from the human than from mouse [274]. However, considering all levels of conservation remained below 50%, perhaps it merely confirms the lack of conservation seen between *ADAM33* and its closest relatives and orthologs. Conservation between *hADAM33* and chimp (*Pan troglodytes*) was too high to be informative.

The third aim was to identify transcription binding sites important in *ADAM33* activation. The F3 sequence (639bp), incorporating F3X, was investigated as a consequence of the luciferase experiments. MatInspector revealed 91 putative transcription binding sites in the region. Application of a tissue filter specific for the respiratory system revealed a putative MyoD binding site (figure 5.16). MyoD is a muscle-specific transcription factor which is known to act as a trigger of (skeletal) muscle differentiation in non-muscle cells. Differentiation induces a complete block of cell proliferation, possibly brought about by the binding of MyoD to transcriptional activation sites of growth genes, thereby preventing transcription [289]. *ADAM33* is expressed in smooth muscle but, interestingly, a smooth muscle filter did not reveal MyoD as a recognised transcription factor. MatInspector matrices are determined by automatic evaluation of pubmed abstracts containing co-citation of transcription factors so absence of MyoD may be due to incomplete information or perhaps it has not been identified as an active transcription factor in smooth muscle. It would be useful to establish whether MyoD is expressed in smooth muscle cells. This could be achieved using reverse transcription followed by qPCR. It is worth noting that MyoD has been implicated as a down-regulator of Sp1 expression during myogenesis (in skeletal muscle). This directly results in down-regulation of the Sp1 activated gene, Glut1 [290]. Several putative Sp1 sites were predicted in the F3X region of *ADAM33*, so perhaps a similar mechanism may exist to control *ADAM33* activation via Sp1 binding sites. MyoD has been shown to be expressed in myofibroblast cell lines from kidney and liver [291] but has not been characterised in lung myofibroblasts to date. *ADAM33* is implicated in the remodelling mechanisms thought to underlie BHR. Remodelling is a term used to describe the deposition of new matrix in the airway wall [292] and is proposed to be a result of proliferation and activation of sub-mucosal

myofibroblasts [293]. Myofibroblasts have an increased capacity to generate extra cellular matrix deposition and mitogens for smooth muscle proliferation, leading to airway wall thickening [167]. Immunohistochemistry and laser confocal microscopy of adult bronchial biopsies showed that α-SMA and ADAM33 immunoreactivity were co-localised to smooth muscle and isolated cells in the sub-mucosa [161]. Perhaps the differentiation from fibroblast to myofibroblast induces MyoD, which then acts as a negative regulator of *ADAM33*. It would be interesting to investigate whether there is a difference between MyoD and *ADAM33* expression levels and ratios in fibroblasts and myofibroblasts. The mini-muscle binding sites situated upstream of, and within the murine conserved region, are putative muscle initiator sites, that are, as yet, poorly defined [282]. These sites may or may not, be important but since one lies within the murine conserved region, they may be of interest.

The final experiment investigated whether the putative Sp1 sites were important in regulation of ADAM33 expression. Over-expression of Sp1 was used to observe the effect on *ADAM33* promoter activation. Epithelial cells (MCF7) and fibroblasts (MRC5) were transfected with a Sp1 expression vector. The results were surprising, in that reporter expression increased in epithelial cells, whereas in fibroblasts (MRC5) cells expression was reduced to levels seen in the SP1 transfected epithelial cells. Sp1 is a promoter-selective transcription factor that binds to GC rich elements. It belongs to a family of eight Sp1-like transcription factors, of which Sp1-Sp4 are the best characterised. Sp1, Sp2 and Sp3 are ubiquitously expressed and Sp4 is expressed predominantly in brain [277, 294, 295]. Sp1 proteins are involved in the regulation of several hundreds of genes [296] and have been reported to regulate genes via activation [277], enhancement [297, 298] and repression [299]. It is widely reported that Sp1 and Sp3 act together to regulate transcription. In a study by Hagen *et al* [300] Sp3 suppresses SP1-mediated transcription by competing for binding to the GC element [300]. This mechanism could explain the up-regulation of reporter expression in the epithelial cells. If Sp1 activity is usually repressed in epithelial cells by an inhibitory factor such as Sp3, then over-expression of Sp1 could out-compete the inhibitor and increase expression of the *ADAM33* reporter. The down-regulation of expression in fibroblasts following Sp1 over-expression cannot be explained this way. Another regulatory function of Sp1, first described by Li *et al* [299] may explain this

observation. Their study of the proximal promoter of the human adenine nucleotide translocase 2 (ANT2) gene suggested that activity of the ANT2 proximal promoter is dependent on three Sp1 binding sites, named A, B and C. Two (A and B) are adjacent to each other (positions -81 to -76 and -71 to -66 respectively) and interact synergistically to activate transcription, the third Sp1 site lies adjacent to the transcription start site. Sp1 bound to site C was found to lower transcription initiation efficiency [299]. This was verified in four different cell lines [299]. The *ADAM33* proximal promoter region has three Sp1 binding elements (figure 5.16) although the third lies close to, but not adjacent to, the transcription start site. If a similar mechanism exists for *ADAM33*, it could explain the inhibition observed following over-expression of Sp1 in fibroblasts.

In conclusion, activation of *ADAM33* is likely to be via Sp1 mediated mechanisms. However this is based on the result of one experiment (performed in triplicate) and needs to be repeated. If the results concur, further investigation of the regulatory effects of the Sp1 binding sites should be performed. Sp1 expression has been shown to be variable in tissues during developmental processes [295], so it is possible that *ADAM33* exerts its 'modelling' effects before lung maturity. A possible inhibitory effect by MyoD transcription factor is worth investigating further. Future investigations to elucidate this pathway and identify other ligands which may regulate *ADAM33* expression will be discussed in chapter 6.

Chapter 6

Final Discussion

6.1 Polymorphism of ADAM33 and asthma

ADAM33 is a highly polymorphic gene that has been shown to be associated with susceptibility to asthma and BHR in several studies [37, 38, 79, 181-183, 186], and marginal [187] or no association [188] in others. In this study association of ADAM33 polymorphisms with lung function at three years supports previous studies. However, associated SNPs are not consistent between populations (table 1.5) and it remains unknown which, if any, of the polymorphisms studied is the causal SNP(s). The finding that *ADAM33* polymorphism is not associated with reduced lung function in one month old infants suggests that gene-environmental interaction occurs after four weeks or the methods employed for the phenotyping were inadequate for the age of population.

6.1.1 Identification of the causal SNP

Results of association studies are often poorly replicated in subsequent studies [301, 302] and this is often due to study design. A good study design should include cases with clearly defined disease phenotypes. This is not easy to achieve in multifactorial diseases such as asthma, which has a spectrum of overlapping phenotypes. Controls must be completely free from the disease under investigation. Genetic heterogeneity (different populations with different genes acting to regulate the phenotype) could mask true association effects in poorly defined populations. Confounding by stratification occurs when individuals are selected from two genetically different populations, in different proportions, in cases and controls (1.4.4.4.2). Cases and controls should therefore be well matched with regard to ethnicity, and also with regard to age and gender, since genes can display age-specific and gender-specific

effects [303, 304]. A number of environmental factors may be necessary for the expression of the disease phenotype, e.g under-activity, obesity, pollution, smoking, hygiene and nutrition, so cases and controls should come from similar environments and comparable demographic areas.

The study population must be of sufficient size to detect an increase in risk of disease associated with the presence or absence of the genetic variant in question. Lack of power may lead to a negative association when there is an effect (type II error, 1.4.4.4.4). Power depends on several factors, including magnitude of effect, sample size (N), allele frequency and required level of statistical significance (1.4.4.3). Multivariate analysis can be used to detect whether co-variates, such as height, weight or gender, modulate the gene-phenotype interaction. However, the higher the number of markers or alleles tested, the more likely a significant result will occur by chance (type I error, 1.4.4.4.4) and this should be taken into consideration when planning a study. Applying a correction, such as Bonferroni, which divides the observed level of significance (p-value) by the number of tests performed (1.4.4.4.4), is one way of controlling for multiple testing. However, it is overly conservative in most situations as it assumes independence of all tests, an assumption that is not valid for interrelated phenotypes or SNPs in LD.

Despite the difficulties in replication of association studies, association of *ADAM33* with BHR and asthma has been confirmed in this study and has been shown in several other populations [37, 38, 79, 181-183, 186]. Therefore, although the causal SNP has not been located it is reasonable to continue the search for the causal allele(s).

The original study by Van Eerdewegh [79] included one promoter polymorphism (A-1) which showed no association with asthma or its associated phenotypes. Five further promoter polymorphisms have since been identified [283], two of which lie outside the promoter region investigated in this work (figure 5.4). None of these has yet been incorporated into association studies. The majority of studies performed to date suggest that the causal SNP lies toward the 3' end of the gene. However, excluding the original study by Van Eerdewegh et al [79], only two studies (the COPSAC study (chapter 4) and Simpson *et al* [184]), have included SNPs upstream of Fp1 (figure

1.3). The results from Simpson *et al* [184] suggested the causal SNP resides between Bp1 and F1. No other SNPs have been located between Bp1 and F1 to date.

SNP Fp1 (134 bp downstream of F1) was shown to be in LD with STp7, Qm1, VM1, T1, and T2 [184]. High LD between these SNPs was also observed in a German population [182] and in UK/US Caucasian populations [79]. Haplotype analysis results suggest haplotypes containing Fp1 [184] and S2 [79, 182] may be important whilst S2, T1 and T2 have been associated with accelerated lung function decline in asthma [37, 38]. These results suggest the true causal SNP has not been located but is in LD with Fp1, Qm1, STp7, S2, T1, T2 and Vm1. Future studies should therefore span the entire gene, incorporating a more detailed coverage of the 5' end. Re-sequencing of *ADAM33* to locate unmapped SNPs will identify other potential variants that can be tested for association and for functionality. Such an approach is already underway (personal communication, J.W. Holloway).

LD studies of the five genes flanking *ADAM33* showed that LD extended up to 130 kb to the left and 40 kb to the right of *ADAM33*. Within this 185 kb interval, twenty-four SNPs in three genes, *ADAM33*, *GFRA4* and *Sn*, (figure 4.6) were significant in the association analysis. Ten of the associated SNPs were not located in *ADAM33*. The extent of LD across the region, and association of SNPs outside of *ADAM33* (section 1.4.4.5.1) suggest the causal SNP could potentially lie outside of *ADAM33*, either in a different gene, or perhaps in an intergenic region. Future work focusing on localisation of the causal SNP, should therefore extend the search to SNPs flanking the *ADAM33* region and into the neighbouring genes.

At time of writing there are 105 SNPs reported within *ADAM33*. Extending the search outside the gene and inclusion of new markers within the gene will further increase the number of markers. The number of potential markers could be reduced by genotyping a random subset of samples from a larger study population, using a large number of markers. SNPtagger [214] (section 3.2.4.2) could be used to identify informative SNPs and the resulting subset of informative markers could then be used to genotype the complete study population. The LD mapping method [201, 202, 204] could be used to pinpoint the location of the causal locus across this greater distance.

6.1.2 Polymorphism and gene expression

Polymorphism along a DNA sequence can affect conformation of chromatin. Unwinding of closed chromatin is necessary for access of transcription activation apparatus, so any change which affects conformation or ability of a transcription factor to bind to regulatory sequences along that sequence can affect whether or not a gene is transcribed. An example of intergenic effects is described by Takemoto *et al* [305], who investigated the 5q31-33 gene cluster containing the genes encoding IL-4, IL-5 and IL-13. The study showed that transcription factor GATA-3 may regulate chromatin conformation at the IL-4/ IL-13 intergenic regulatory region, underlying coordinate expression of this gene cluster. Such enhancer sequences can act up to 50 kb away from the promoter. Intergenic SNPs between *GFRA-4*, *ADAM33* and *SN* should therefore be included in further genetic studies.

Another regulatory mechanism controlling gene expression is DNA methylation. In general, the level of methylation correlates with the transcriptional state of a gene, where active genes are less methylated than inactive genes. DNA methylation is a post-synthetic modification of DNA that normally occurs after each replication. Only cytosine-guanine (CG) base pairs are methylated. Since DNA replication is semi-conservative, one strand of the new DNA is already methylated and the other strand remains to be methylated by DNA methyltransferases (DNMTs) [306]. Preliminary results of investigations into *ADAM33* methylation identified several CpG islands spanning *ADAM33*, one lying within the putative basal promoter region of *ADAM33*. This region of putative methylation incorporates the F3X construct that showed marked up-regulation of *ADAM33* expression (figure 6.1). It is possible that methylation pattern affects *ADAM33* expression. Other CpG islands were noted upstream of the promoter sequence investigated in this work (personal communication, Y.Yang). Longer promoter constructs should be therefore manufactured incorporating this region.

Figure 6.1 Position of the CpG island in the promoter region of ADAM33

TATGCTTTCTGTACTTTTTGATTTTGAGATATGTGAATGTAGGTTTCTCTCACTGCTCGAACTTTCACTAACCAAATTACTACAT
TCCAAATTCTCAAAAACAAATAGATTTACTTAAAAGTAGGCTGGGTGCGGTGTCTCACGCCTGTAATTCCAGCGCTTTGGGA
GGCCGAGGCGGGCAGATCACCTGAGGTCGGGAGTTCGAGACCAGCCTGACCAACATGGAGAAACCCCATCTCTACTAAAA
ATACAAAATTAGCCAGGCGTGGTGGCGAATGCCTGTAATGCCAGCTACTCGGGAGGCTGAGGCAGAAGAATCACTTGAATC
TGGGAGGCAGAGGTTGCAGTGAGCCCAGATCATGCCATTGCACTCCAGTCTGGGTAACAAGAGAGAAACTCTGTCTCAAAA
AAAAAAAAAAAAAAAAAAGATTTGCTTAAAAGTTAACATCTCCGGCCGGGCGCGGTGGCTCATGCCTGTAATCCCAGCGCTTT
GAGAGGCCGAGGCGGGTGGATCACGAGATCAGGAGATTGAGACCATCCTGGCCAAAATGGTGAAACCTCGTCTCTGCTAA
AAATACAAAAGTTAGCTGGGGGTGGTAGCGCGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTG
AACCAGGGAGTCGGAGGTTGCAGTGAGCCAAGATCGCGCCGCTGCACTCCAGCCTGGCGACAGAGGGAGACTCCATCTCA
AAAAAAAAAAAAAAAAAAAAGTTAACATCTCATCCAAATTTGCACCGAGTAGGAAAACAAAAGTTTAAAACATGAAACAGATGT
TACTGAGGCCGAAAGGGTCTCCCAGGCCTGGGAGTCTGCAGCTTTATGCAATTCTGCCCTCTGGCCACCGCCAGGGAAG
AAAGGTTGTCTCCGTCTGCTGCATCGCCTTTGCCCAGCAATGAAGCCCCCAAGACAGCGGCAGCCGGTTGCCTGAACCTTC
CTATCCTTGGGGGCACCCAGTGCAGGTGGATGACCCGACTCAACCTCCGCCAGGGCACCCTCGGGGCAGGACGGGTAGC
AAGGAGGGGACAGAGATCGGCCCCAGGAGACCACGGAAGATCGCGCTCCTGGGGCCAACTTCAGCAGCGAGAGGCGGCC
TTTGCCCACCGCCTCATCCCACCACGCCGCGGTCCTCCAAGAACCTTCCCAGCGGTTCTCTCCTCCTCTCAGGAGTAGAGG
CCCTCTGAGACCGACGGGGAGGGACGGCTCGGGCCGGTCATCCGAGGGGCCGCACGGATTCCCTCCTCCGCCCAGCTCC
ACCCCCTCGAGGGGCGGCGGTCCGGGAGTGGCGACCCGGCTCCCCCATGGCGCGCGCCGTCGGGGCCCCTGGCCAGGC
TCCGAGCGGGGTTGGCGGGGAGGGGAGGCGGGAGCGAGGGCGGGCGGTGGGAGGTGGGGGCGGGAAGGTCCGAAGG
CGGCGGCCTGAG*gctgcaccGGGCACGGGTCGGCCGCAATCCAGCCTGGGCGGAGCCGGAGTTGCGAGCCGCTGCCTAGA
GGCCGAGGAGCTCACAGCatgggctggaggccccggagagctcgggggaccccgttgctgctgctgctactactgctgctgctc

*Figure 6.1The CpG island is highlighted in yellow. Red sequence denotes forward primer position and blue sequence denotes reverse primer position. The transcription start site is denoted by * and lower case letters, the start of the coding region (ATG into exon 1) is in lower case letters. Position of known promoter polymorphisms are underlined (details figure 5.4).*

Methylation can affect enhancer and silencer components of the gene's transcription mechanism some distance from the basal promoter. Work on methylation patterns of *ADAM33* and how SNPs along the gene may disrupt this pattern is ongoing, and will not be further discussed.

A number of intronic SNPs, including Fp1, Qm1, STp4, STp7 and Vm1, have shown positive association to BHR and asthma. However, none of the associated intronic SNPs appeared instrumental in disruption of splicing [154]. Ratios of splice variants observed in human airway fibroblasts of asthmatic and non-asthmatic samples were compared, but no difference was observed. It was noted that the majority of transcripts lacked the catalytic domain, suggesting *ADAM33* may not have proteolytic activity. Alternatively, it may be that very little proteolytically-active *ADAM33* is needed to exert an effect in these cells. Western Blot analysis of fibroblast cell lysates confirmed the existence of multiple protein isoforms. These included a faint protein band at 120 kD, consistent with unprocessed ADAM33, and a band at 100 kD, equivalent to processed ADAM33. Several smaller bands were noted, at 50 and 60 kD, consistent with the spliced isoforms containing a variable N-terminal region but lacking the MP domain, and expressing domains downstream of the disintegrin domain (figure 1.3). A smaller band at 37 kD was the most abundant protein, and was

consistent with a transcript expressing half the cysteine-rich domain and the downstream C-terminal portion of ADAM33 [154]. Further work, using adult bronchial biopsies (BB), smooth muscle (SM), human embryonic lung (hEL), and bronchial brushings confirmed the presence of previously identified isoforms (molecular weights: 37, 55 and 65 kD) and identified a smaller band at 22 kD in BB, SM and hEL. In addition, a band at 25 kD was noted in hEL [161]. ADAM33 expression was not detected in bronchial brushings (epithelial cells) [161]. In both these studies western blots were performed using antibody raised against the cytoplasmic domain, the sequence of which was present in all transcripts except the soluble form [154, 161].

A recent study by Lee *et al* [173] identified the presence of ADAM33 protein in bronchoalveolar lavage fluid (BAL). BAL fluid is obtained by irrigation or washing out of the bronchi and alveoli and is used in performing measurements to ascertain the levels of various mediators and substances. It can contain creola bodies, eosinophilic cationic protein, alveolar macrophages [174] and activated fibroblasts [175]. The study by Lee *et al* [173] compared ADAM33 expression in BAL fluid, between asthmatics and non-asthmatics. Results showed that increased levels of ADAM33 in the asthmatics correlated with reduced FEV_1. Western blotting was used to identify ADAM33, using a rabbit polyclonal affinity-purified antibody raised against the catalytic domain [173]. Detection of the metalloprotease domain, which was absent in the majority of isoforms in previous studies [154, 161], could be due to the use of antibody specifically raised against this domain. However, it is thought the expression of ADAM33 found in this study may derive from activated fibroblasts [175] or alveolar macrophages [171] found in BAL fluid. These cells may have different splice-form profiles from the bronchial biopsies and fibroblast cell lines used in previous studies. Indeed, important differences between myofibroblast phenotypes of mild asthma and scleroderma, a fibrotic disease affecting the skin, have been observed, suggesting motile BAL myofibroblasts may play distinct roles within a given disease [175]. It is reasonable, then, to suggest that *ADAM33* expression profiles may differ between activated myofibroblasts of asthmatic and non-asthmatics, and this should be investigated. In asthma, injury to the epithelium causes the release of growth factors such as TGF-β, promoting proliferation and differentiation of underlying fibroblasts into myofibroblasts. Thus, it is possible that the increase in ADAM33 expression seen

in asthmatics observed by Lee *et al* could be due to an increase in myofibroblast numbers in BAL fluid [173].

Expression of full length *ADAM33* in mammalian cell lines has shown that the metalloprotease domain is functional [165, 166]. ADAM33 can cleave peptides derived from beta-amyloid precursor protein (APP), Kit-ligand-1 (KL-1), tumour necrosis factor-related activation-induced cytokine, and insulin B chain, but for all these substrates cleavage is relatively inefficient, suggesting these are not the natural substrates [166]. Many transmembrane growth and differentiation factors, that play a crucial role in development, require ectodomain shedding for proper action *in vivo*. For example, members of the EGF family, a group of growth factors responsible for remodelling processes seen in asthmatic airways [307], are processed by *ADAM17* (TACE), *ADAM12*, *ADAM10* and *ADAM9* [133, 135]. A functional MP domain requires contiguous FGHI exons, found in less than 5% of transcripts in fibroblast cell lines [154]. SNP Fp1 lies in the intronic region between exon F and G and has been associated with asthma and lung function in several studies [38, 79, 182, 184, 186], although no effect has been elucidated. Several coding SNPs reside in the catalytic domain, including F1 (non-synonymous, G/A, Alanine – Threonine) and I1 (synonymous, A/G, Glycine-Glycine) but to date there is no evidence to suggest functional effects of these SNPs. However, a functional metalloprotease domain, which is inactivated by mutation of the catalytic site [166] implicates ADAM33 in shedding of membrane proteins. Identification of the natural substrate(s) of ADAM33 is needed to clarify its role in asthma pathogenesis. It will then be possible to test how SNPs in the region affect function.

The 3' portion of the gene, encoding the transmembrane domain, cytoplasmic domain and 3' UTR is one of the key regions linked to asthma and BHR [79]. The transmembrane domain (exon S) and cytoplasmic domain (exons T, U and V) contain several SNPs associated in different studies. Associated coding SNPs within exon S include S1 (non-synonymous, A/G, Isoleucine-Valine) [38, 79, 181, 184] and S2 (synonymous, C/G, Glycine-Glycine) [37, 38, 79, 181, 186]. Exon S codes for the transmembrane domain (figure 1.3) and disruption could affect the protein's anchoring ability, perhaps giving rise to the soluble form of ADAM33.

Exons T, U and V (5'), encode the relatively short cytoplasmic tail of ADAM33. It contains a putative SH3 binding site which may be important in ADAM33 signalling [169]. It has been reported that the cytoplasmic tails of ADAMs play a role in signal transduction to regulate the activity of the extracellular regions [308]. It follows, then, that polymorphism in the signalling domain could affect metalloprotease activity.

Functional work on the cytoplasmic tails could identify important signalling pathways, helping to ascertain the targets and functions of ADAM33. To investigate the interaction of SH3 binding motifs with cytoplasmic proteins the yeast-two-hybrid system could be used. This system is used primarily for initial identification of interacting proteins and is based on the modular organisation of many transcription factors. These proteins often have two or more discrete structural and functional units, or domains. The first protein to be used for this was the yeast protein GAL4 which has a DNA-binding domain (or DBD) and an activation domain (or AD). Two types of hybrids are made. The DBD hybrid contains the DBD fused to a protein of interest, or bait, i.e. cytoplasmic ADAM33. This fusion protein can bind to the DNA or a reporter construct, but cannot activate transcription because the bait does not contain an activation function. A second hybrid, containing the AD fused to another protein (or prey) is prepared, usually from a recombinant DNA "library" in which genes for many different proteins, from a specific tissue type, are fused to the AD. Then both DBD and AD hybrid proteins are expressed in the same cell containing a reporter gene (such as β-galactosidase). Those expressing the reporter gene are identified and purified for further characterisation. The DBD hybrid can be mutated to incorporate polymorphic changes, and subsequent alterations in binding observed [309].

6.1.3 Gene-environmental interaction

This study revealed no significant association between *ADAM33* polymorphisms and lung function (measured as $FEV_{0.5}$ or PD_{15}) at one month old in a Danish population.

This may indicate that *ADAM33* exerts its effect after 1 month of age, or that the infants had not yet encountered the environmental trigger needed for expression of the phenotype. Alternatively, it may be a result of inappropriate phenotype measurement methods for the age of population.

Lung function measurements (sRaw) were performed on the COPSAC population at three years using plethysmographic measurement of specific airways resistance (sRaw) (section 4.1.5). This offered an opportunity to perform a close replication of the ^{NAC}MAAS study using the COPSAC cohort, since lung function measurement method, age of population, ethnicity and size of population were very similar. In addition the work by Simpson *et al* used the Taqman® assays set up in chapter 3 to genotype the population (National asthma campaign, Manchester asthma and allergy study (^{NAC}MAAS)) from the South of England. The two populations may differ slightly in atopy status since ^{NAC}MAAS were recruited from atopic parents while COPSAC were recruited from asthmatic parents.

The positive association at three years old in the COPSAC cohort strengthens the evidence that *ADAM33* is associated with decreased lung function in children aged three years [184]. Although both studies found association, the associated SNPs were not the same. This is likely to be because the true causal SNP is in LD with the identified SNPs, either within the gene, or outside the gene region. In both studies associated haplotypes were identified that again did not agree between studies.

The period between birth and lung maturation (two to three years) may be important for *ADAM33*-environment interaction. What aspect of the environment is not yet known, but lower respiratory infections have been associated with wheezing and subsequent asthma [225, 237, 310]. The infants included in the COPSAC study had not contracted lower respiratory tract infection before the first measurement of lung function. Environmental and life-style assessments were measured in the COPSAC study. These included identification of respiratory bacteria, and viruses at one month and twelve months of age in asymptomatic children and also following lower respiratory symptoms of over three days' duration [263]. Investigation of correlations between viral/bacterial colonisation, reduction in lung function (sRaw) and *ADAM33* polymorphism could be performed using regression analysis. It is possible that viral infection could be an important environmental modifier of *ADAM33* expression. Alternatively, viral infection may simply expose those with a pre-disposition to asthma.

It is clear that inflammation alone does not explain all of the features of chronic asthma [311]. *ADAM33* polymorphism has been implicated in structural modification

rather than inflammation, but these factors will interact. Other outcomes in the COPSAC cohort of potential interest include data on lung symptoms (monitored in diary-cards day-by-day for three years by the mothers) and "wheezy" episodes, medication use and asthma diagnoses. To give an indication of inflammatory status, FeNO levels were measured at four weeks (simultaneously with the innate lung function measurements), and eosinophils were measured at six months and eighteen months. These and other co-variates such as diet (infant milk, introduction of solid food) and inoculation record should be considered for inclusion in a regression analysis.

6.2 Activation of ADAM33

The second part of this work investigated the transcriptional activation of *ADAM33*. Conservation was minimal between the promoter region of ADAM33 and its closest relatives and orthologs. Analysis of the deletion constructs of the *ADAM33* promoter region suggested the basal promoter lies within sequence F3X and activation of ADAM33 is likely to be via Sp1 mediated mechanisms (section 5.4). An inhibitory binding site may be situated upstream of F3X (figure 5.12). Stimulation with PBA, IL-13, IFN-γ, and dexamethasone did not affect reporter expression (figure 5.13). These cytokines were chosen because they are known to be involved in inflammatory pathways. Previous stimulations of epithelial cell lines (data not included) with epidermal growth factor (EGF), tumour necrosis factor alpha (TNF-α), transforming growth factor beta (TGF-β), IL-4, IL-13 and cigarette smoke extract also failed to up-regulate *ADAM33* expression. EGF and TGF-β have the potential to cause an epithelial-mesenchymal transition. Cell lines used have been shown to have the relevant receptors and signal transduction mechanisms to be able to respond to the stimulants employed [312-317] .

6.2.1 Investigating mechanisms of *ADAM33* regulation

The Sp1 expression work (figure 5.17) was only performed once (in triplicate) so needs to be repeated to confirm the results. In addition, it is desirable to repeat the investigations on primary lung smooth muscle cells and primary epithelium to check that the observations are the same as that seen in cell lines. Cell lines are a useful model for investigating possible cellular functions, but they have been transformed to

immortalise them, and so may not act the same as primary cells. It is difficult to dissect a pure smooth muscle cell population from fibroblasts out of biopsy material. A mixture of the two populations may give confusing results, so cells should be stained with a muscle cell marker, such as alpha-smooth muscle actin (α-SMA), to assess population purity. If results confirm that Sp1 is important in *ADAM33* transcriptional activation several other hypotheses can be considered.

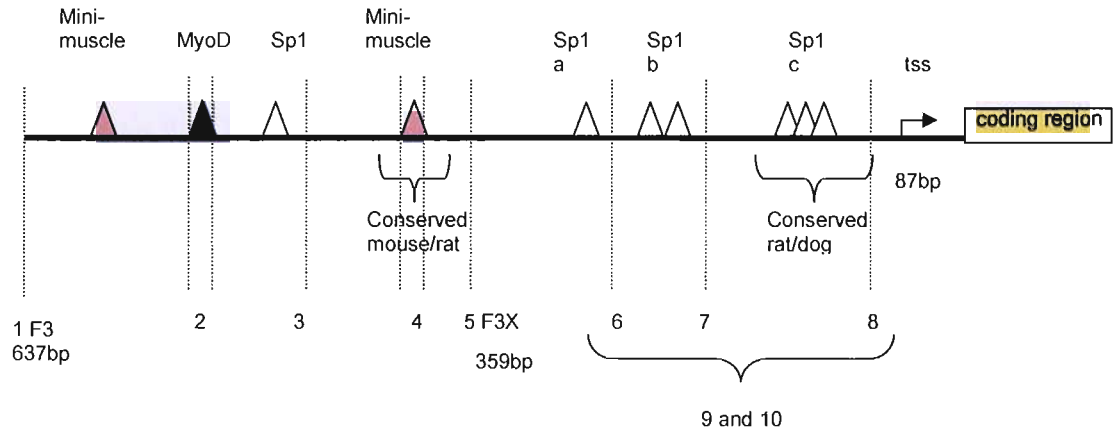
6.2.1.1 Over expression of Sp1 induces ADAM33 expression in epithelial cell lines and attenuates ADAM33 expression in fibroblasts

Epithelial and fibroblast cell lines, or primary cells could be transfected with Sp1 expression plasmids, and reverse transcription followed by qPCR used to detect and quantify the relative amounts of transcript in non-transfected and SP1 transfected cells. Primers should be used that detect all transcripts to detect variability in splice forms. Western blot could be used to confirm protein isoforms.

6.2.1.2 ADAM33 expression is regulated by Sp1 transcription factor binding sites

It is necessary to identify which of the Sp1 sites are important for *ADAM33* activation. Further deletion constructs should be made deleting each Sp1 site in turn. In addition, the putative MyoD and mini-muscle binding sites (figure 5.16) within the F3 sequence should be deleted to test their role in *ADAM33* regulation. There is evidence to suggest MyoD activation during myogenesis causes the downregulation of Sp1 [290], and this could have implications for ADAM33 expression. Of the two mini-muscle binding sites, the one lying in the murine conserved region is of potential interest and should be included in future investigations. Suggested constructs are illustrated in figure 6.2. Using the F3 and F3X original constructs as controls, the new constructs could be transfected as before and activity measured using luciferase assays (section 2.2.5).

Figure 6.2 Proposed constructs to test the role of putative transcription binding sites in ADAM33 activation



1) F3 original construct, 2) Delete MyoD only, 3) Delete unconserved mini-muscle, MyoD, and Sp1, leave conserved region containing mini-muscle, 4) As for 3 but delete conserved mini-muscle, 5) F3X original construct, 6) Delete Sp1a, 7) Delete Sp1a and b, 8) Delete Sp1 a b and c, 9) Delete/mutate Sp1 b, 10) Delete/mutate Sp1 a and c. tss transcription start site.

Electromobility shift assays (EMSA) could be used to confirm the binding of Sp1 sites by transcription factors. EMSA uses radio-labelled fragments of DNA containing the Sp1 site sequence. The electrophoretic mobility of the DNA is determined in the presence and absence of proteins extracted from cell lysates. Protein binding to the labelled DNA retards it in the gel causing a shift in the location of the fragment band, which is detected by autoradiography. Addition of Sp1 antibody will bind to Sp1 protein bound to DNA causing a 'supershift' in the band. An alternative method is an Enzyme-linked immunosorbant assay (ELISAs) kit manufactured by Active Motif®. A biotinylated oligonucleotide is synthesised, or PCR product generated, that contains the transcription factor binding site of choice. The oligo is incubated with cell nuclear-protein extract, and is pipetted onto a streptavidin-coated plate. The biotinylated oligonucleotide binds to the streptavidin plate. A primary antibody recognising the transcription factor of choice (Sp1) is added, followed by an anti-IgG horseradish peroxidase (HRP) conjugate. Developing and stop solution is than applied and the sequences that have bound to the transcription factor of interest identified and amount of binding quantified. A kit such as this offers a 100-fold increased sensitivity than that of gelshift assays and use of radiation is eliminated.

6.2.1.3 Sp1 and Sp3 ratios are implicated in regulation of ADAM33 expression

If Sp1 is important for regulation of *ADAM33*, the mechanism of interaction, and the co-factors involved, need to be elucidated. Many mechanisms of Sp1 regulation have been reported (section 5.4) including Sp3/Sp1 antagonism [300]. To test the hypothesis that Sp1 and Sp3 interact to regulate *ADAM33*, reverse transcription followed by qPCR could be used to quantitate mRNA expression of Sp1 and Sp3 in epithelial and fibroblast cell lines. If results suggest an antagonistic interaction, expression vectors could be used to observe how changing the Sp1/Sp3 ratio affects *ADAM33* expression. Methods for discovering protein-protein interactions include the yeast-two-hybrid system described previously (section 6.1.2).

6.2.1.4 Other directions

The activation of *ADAM33* by Sp1 has implications for the human developing lung. ADAM33 is expressed in smooth muscle bundles around the bronchi in human embryos, strongly suggesting it may play a role in smooth muscle development and function. In addition, the occurrence of ADAM33 in undifferentiated mesenchymal cells, which are more abundant in embryonic lung than adult, suggests it may have a role in the development of bronchial, vascular and neuronal structures of the airways, as well as in differentiation of mesenchymal cells such as fibroblast and smooth muscle cells. The occurrence of a distinct 25 kD isoform in embryonic lung tissue that is not found in adult lung may be of particular interest [161] and needs to be investigated. Murine studies show Sp1 is ubiquitously expressed in mouse embryos and its expression level changes at different developmental stages and in different cell types [295]. The roles of Sp1 in the formation and differentiation of mouse lung could be addressed by targeted knock-out of the gene, and then assessing expression of *ADAM33*. Alternatively, *ADAM33* and *SP1* knockouts could be attempted, in cell lines or in primary human cells, using RNA interference.

6.3 ADAM33 and the pathogenesis of asthma

Asthma is a disorder of the airways in which the conducting airways contract too much and too easily, either spontaneously, or as a result of exposure to stimuli. Several longitudinal cohort studies suggest more severe asthma has its origins in early life, and can be predicted by impaired lung function and BHR [264]. Conversely, it also been shown that reduced airway function present in early infancy was associated with persistent wheeze at 11 years of age, but this relationship was independent of the effect of increased airway hyperresponsiveness and atopy in childhood [192]. This suggests that asthma, once considered to be a single disease driven by allergic sensitisation, is a spectrum of distinct diseases, sharing overlapping symptoms.

The mechanism underlying BHR is not fully understood, but an increase in smooth muscle and alterations in its physicochemical properties are considered important [25]. *ADAM33* is an asthma susceptibility gene implicated in BHR and remodelling of the lung. Possible roles of *ADAM33* in the development of bronchial hyperresponsiveness are summarised in figure 6.3.

Figure 6.3 Involvement of ADAM33 in Bronchial Hyperresponsiveness.

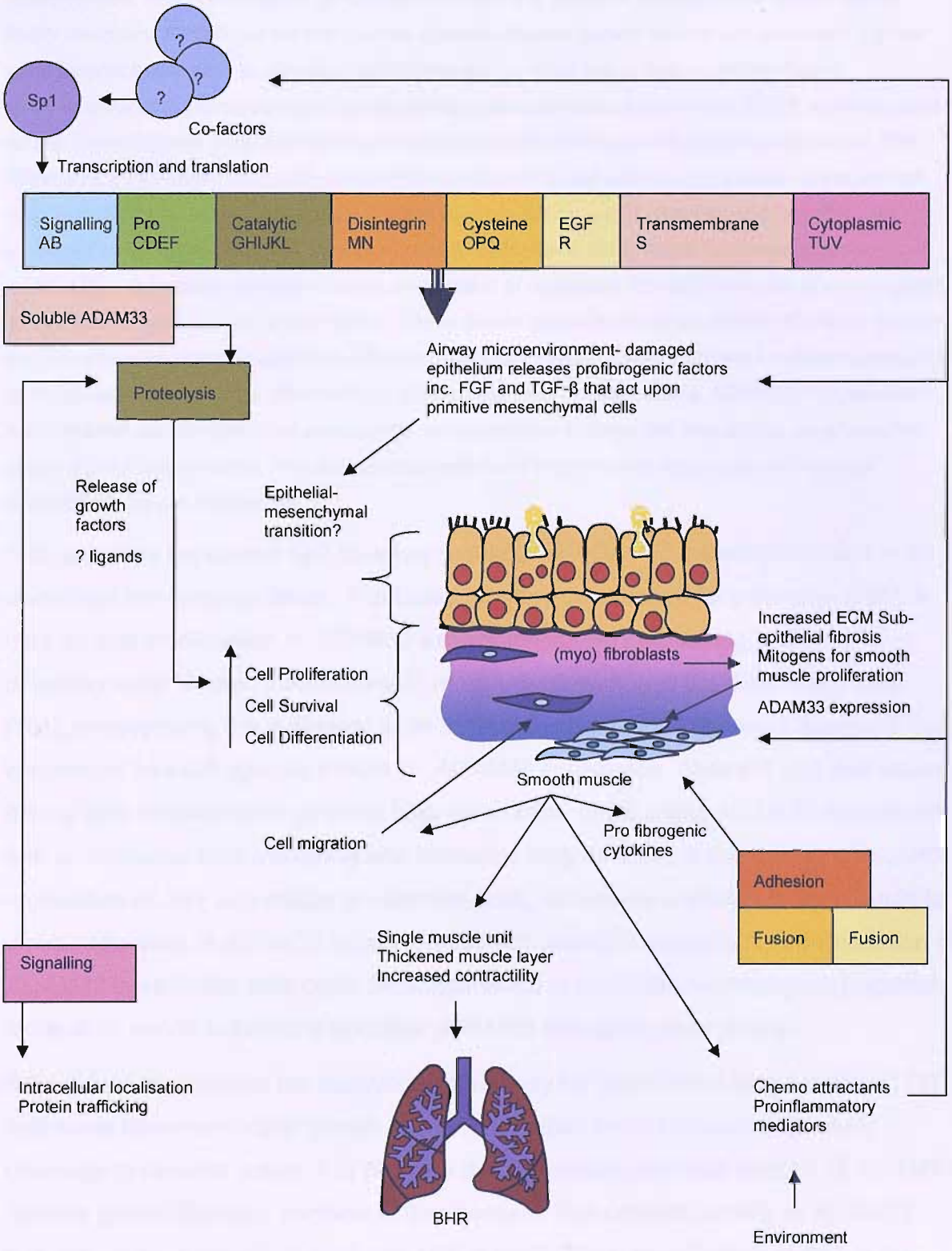


Figure 6.2 Evidence suggests ADAM33 is regulated in part by Sp1, although regulatory cofactors are undetermined. Putative functions of ADAM33 domains are based on knowledge of related ADAM family members. Proteolysis via the catalytic domain releases growth factors (yet unknown) that act upon mesenchymal cells to increase proliferation and survival and to induce differentiation. Mesenchymal cells (fibroblasts and myofibroblasts) release extracellular matrix (ECM) proteins causing sub-epithelial fibrosis. They also produce mitogens, which induce smooth muscle proliferation. The disintegrin, cysteine-rich and EGF domains are implicated in cell adhesion and fusion of the smooth muscle cells. Increase in smooth muscle mass leads to airway wall thickening, and together with increased contractility of asthmatic smooth muscle cells, leads to bronchial hyperresponsiveness. Signals from damaged epithelium induce proliferation of fibroblasts through the action of profibrogenic growth factors such as TNF-α and TGF-β. TNF-α has indirect effects on the growth of smooth muscle and can stimulate collagen synthesis and fibroblast proliferation. TGF-β stimulates collagen production by fibroblasts and promotes differentiation of fibroblasts into myofibroblasts. ADAM33 is expressed in myofibroblast and fibroblast cell populations, so an increase in these cell populations could result in higher ADAM33 expression. Possible environmental and inflammatory modulators acting upon ADAM33 are as yet undetected.

This work has implicated Sp1 as a key regulator of *ADAM33* transcription. Sp1 is an ubiquitous transcription factor, implicated in many developmental pathways [295]. It may be that modification of *ADAM33* expression via Sp1, occurring in the developmental stages, affects smooth muscle proliferation in the developing lung [161], predisposing the individual to an asthmatic phenotype. This work suggests Sp1 expression has cell-specific effects on *ADAM33* expression. Aberrant Sp1 expression during fetal development, or infant lung maturation, could affect *ADAM33* expression and so modulate lung modelling and ultimately lung function. It could be that *in utero* modulation of Sp1 expression in epithelial cells, caused by unknown factors, leads to an up-regulation of *ADAM33* in cells that do not usually express it. An up-regulation of *ADAM33* in epithelial cells could be instrumental in epithelial-mesenchymal transition. More work needs to be done to define *ADAM33* activation more clearly.

New evidence suggests the catalytic domain may be important in lung function [173] and many transmembrane growth and differentiation factors require proteolytic cleavage to become active. It is possible that the metalloprotease domain of *ADAM33* cleaves growth factor(s) involved in fibrogenesis. The catalytic activity of *ADAM33* has not yet been tested on FGF and TGF-β [166]. Thus, modification of Sp1 in the fetal developing lung, or postnatally during lung maturation, could lead to increased

ADAM33 expression and proteolysis of its substrate. If the substrate were FGF, TGF-β or similar, cleavage by *ADAM33* could result in primitive mesenchymal cells converting to a myofibroblast phenotype. Myofibroblast over-proliferation results in excessive deposition of extra-cellular matrix, leading to thickening of the airway wall, and a reduction in lung function [318]. Myofibroblasts also secrete mitogens, important for smooth muscle cell proliferation, leading to an increase in smooth muscle mass and thickening of the airway. Cell adhesion and fusion, other putative functions of *ADAM33,* may also have a role here. If *ADAM33* were involved in proteolysis of such growth factors, the possibility of modulatory drug development targeting the metalloprotease domain exists, although much work would be needed to identify patients at risk, and to work out the best time for treatment. In addition, if disease modification by *ADAM33* occurs before the advent of symptoms, it will be difficult to persuade the population, even those with a family history of persistent asthma, to take part in treatments involving apparently healthy children.

The finding that *ADAM33* is not associated with reduced lung function in one month old infants, but is in three year old children [184] suggests a critical window may exist during which gene-environment interaction takes place, triggering asthma-like symptoms in a pre-disposed lung. If lung morphology is modulated during development, the effect may not be seen until the lung is challenged, by infection, for example. The infant lung continues to develop postnatally [248], so environmental modification of *ADAM33* expression may continue postnatally. Identification of environmental modifiers of *ADAM33* expression will enable at-risk patients to modify their environments to limit the scale of possible effects, an approach that would be publicly acceptable.

In summary, the true function of *ADAM33* remains unknown. However, evidence in this thesis supports association of *ADAM33* polymorphism with lung function. In addition it suggests that *ADAM33* may be regulated via Sp1-mediated mechanisms, supporting a role in lung morphogenesis. Modification of *ADAM33* regulation during development may result in a lung pre-disposed to asthma. Early life environmental interactions, possibly between birth and three years, may then lead to asthmatic symptoms in the predisposed individual.

Chapter 7

References

1. Masoli, M., et al., The global burden of asthma: executive summary of the GINA Dissemination Cornmittee report. Allergy, 2004. **59**(5): p. 469-78.

2. Matricardi, P.M. and S. Bonini, Why is the incidence of asthma increasing?, in Asthma Critical debates, S.L. Johnston and S.T. Holgate, Editors. 2002, Blackwell Science Ltd. p. 3-17.

3. European Respiratory Society and E.L. Foundation, Part 02 Major Respiratory Diseases, in European Lung White Book. 2003, European Respiratory Society Journals.

4. Soriano, J.B., et al., Increasing prevalence of asthma in UK primary care during the 1990s. Int J Tuberc Lung Dis, 2003. **7**(5): p. 415-21.

5. Fleming, D.M., et al., Declining incidence of episodes of asthma: a study of trends in new episodes presenting to general practitioners in the period 1989-98. Thorax, 2000. **55**(8): p. 657-61.

6. Holgate, S.T., The epidemic of allergy and asthma. Nature, 1999. **402**(6760 Suppl): p. B2-4.

7. Maddox, L. and D.A. Schwartz, The pathophysiology of asthma. Annu Rev Med, 2002. **53**: p. 477-98.

8. Roitt, I., J. Brostoff, and D. Male, Immunlogy. Five ed. 1998: Mosby.

9. Holgate, S.T., et al., The anti-inflammatory effects of omalizumab confirm the central role of IgE in allergic inflammation. J Allergy Clin Immunol, 2005. **115**(3): p. 459-465.

10. Howarth, P.H., et al., Tumour necrosis factor (TNFalpha) as a novel therapeutic target in symptomatic corticosteroid dependent asthma. Thorax, 2005. **60**(12): p. 1012-8.

11. Holgate, S.T., et al., Anti-immunoglobulin E treatment with omalizumab in allergic diseases: an update on anti-inflammatory activity and clinical efficacy. Clin Exp Allergy, 2005. **35**(4): p. 408-16.

12. Jeffery, P., Remodelling in asthma and chronic obstructive lung disease. Am J Respir Crit Care Med, 2001. **164**: p. S28-S38.

13. Carroll, N., et al., The structure of large and small airways in nonfatal and fatal asthma. Am Rev Respir Dis, 1993. **147**(2): p. 405-10.

14. Laitinen, L.A., et al., Damage of the airway epithelium and bronchial reactivity in patients with asthma. Am Rev Respir Dis, 1985. **131**(4): p. 599-606.

15. Puddicombe, S.M., et al., Involvement of the epidermal growth factor receptor in epithelial repair in asthma. Faseb J, 2000. **14**(10): p. 1362-74.

16. Bucchieri, F., et al., Asthmatic bronchial epithelium is more susceptible to oxidant-induced apoptosis. Am J Respir Cell Mol Biol, 2002. **27**(2): p. 179-85.

17. Warburton, D., et al., The molecular basis of lung morphogenesis. Mech Dev, 2000. **92**(1): p. 55-81.

18. Goldsby, R.A., T.J. Kindt, and B.A. Osborne, Kuby Immunology. Fourth ed. 2000: W.H.Freeman and Company.

19. Stewart, A.G., et al., Tumor necrosis factor alpha modulates mitogenic responses of human cultured airway smooth muscle. Am J Respir Cell Mol Biol, 1995. **12**(1): p. 110-9.

20. Vilcek, J., et al., Fibroblast growth enhancing activity of tumor necrosis factor and its relationship to other polypeptide growth factors. J Exp Med, 1986. **163**(3): p. 632-43.

21. Seals, D.F. and S.A. Courtneidge, The ADAMs family of metalloproteases: multidomain proteins with multiple functions. Genes Dev, 2003. **17**(1): p. 7-30.

22. McAnulty, R.J., et al., The effect of transforming growth factor beta on rates of procollagen synthesis and degradation in vitro. Biochim Biophys Acta, 1991. **1091**(2): p. 231-5.

23. Richter, A., et al., The contribution of interleukin (IL)-4 and IL-13 to the epithelial-mesenchymal trophic unit in asthma. Am J Respir Cell Mol Biol, 2001. **25**(3): p. 385-91.

24. Holgate, S.T., et al., Invited lecture: activation of the epithelial mesenchymal trophic unit in the pathogenesis of asthma. Int Arch Allergy Immunol, 2001. **124**(1-3): p. 253-8.

25. Fredberg, J.J., Bronchospasm and its biophysical basis in airway smooth muscle. 2004.

26. Ball, W., Interactive Respiratory Physiology. 1996, Office of Medical Informatics Education.

27. Lowe, L., et al., Specific airway resistance in 3-year-old children: a prospective cohort study. Lancet, 2002. **359**(9321): p. 1904-8.

28. Custovic, A., et al., The National Asthma Campaign Manchester Asthma and Allergy Study. Pediatr Allergy Immunol, 2002. **13 Suppl 15**: p. 32-7.

29. Turner, S.W., et al., Infants with flow limitation at 4 weeks: outcome at 6 and 11 years. Am J Respir Crit Care Med, 2002. **165**(9): p. 1294-8.

30. Ulrik, C.S., Outcome of asthma: longitudinal changes in lung function. Eur Respir J, 1999. **13**(4): p. 904-18.

31. Young, S., et al., The association between early life lung function and wheezing during the first 2 yrs of life. Eur Respir J, 2000. **15**(1): p. 151-7.

32. Cokugras, H., et al., Ultrastructural examination of bronchial biopsy specimens from children with moderate asthma. Thorax, 2001. **56**(1): p. 25-9.

33. Payne, D.N., et al., Early thickening of the reticular basement membrane in children with difficult asthma. Am J Respir Crit Care Med, 2003. **167**(1): p. 78-82.

34. Xuan, W., et al., Lung function growth and its relation to airway hyperresponsiveness and recent wheeze. Results from a longitudinal population study. Am J Respir Crit Care Med, 2000. **161**(6): p. 1820-4.

35. Koppelman, G.H., H. Los, and D.S. Postma, Genetic and environment in asthma: the answer of twin studies. Eur Respir J, 1999. **13**(1): p. 2-4.

36. Grol, M.H., et al., Risk factors for growth and decline of lung function in asthmatic individuals up to age 42 years. A 30-year follow-up study. Am J Respir Crit Care Med, 1999. **160**(6): p. 1830-7.

37. Jongepier, H., et al., Polymorphisms of the ADAM33 gene are associated with accelerated lung function decline in asthma. Clin Exp Allergy, 2004. **34**(5): p. 757-60.

38. van Diemen, C.C., et al., A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. Am J Respir Crit Care Med, 2005. **172**(3): p. 329-33.

39. Demeo, D.L. and E.K. Silverman, Genetics of chronic obstructive pulmonary disease. Semin Respir Crit Care Med, 2003. **24**(2): p. 151-60.

40. Sandford, A.J. and P.D. Pare, The genetics of asthma. The important questions. Am J Respir Crit Care Med, 2000. **161**(3 Pt 2): p. S202-6.

41. Kerem, B., et al., Identification of the cystic fibrosis gene: genetic analysis. Science, 1989. **245**(4922): p. 1073-80.

42. Group, T.H.s.D.C.R., The Huntington's Disease Collaborative Research Group, "A Novel Gene Containing a Trinucleotide Repeat That is Expanded and Unstable on Huntington's Disease Chromosomes. Cell, 1993. **72**(971).

43. Rommens, J.M., et al., Identification of the cystic fibrosis gene: chromosome walking and jumping. Science, 1989. **245**(4922): p. 1059-65.

44. Genes for asthma? An analysis of the European Community Respiratory Health Survey. Am J Respir Crit Care Med, 1997. **156**(6): p. 1773-80.

45. Holloway, J.W., et al., The Genetics of Asthma, in European Respiratory Monograph 23, Asthma, F. Chung and L.M. Fabbri, Editors. 2003, European Respiratory Society Journals.

46. Jenkins, M.A., J.L. Hopper, and G.G. Giles, Regressive logistic modeling of familial aggregation for asthma in 7,394 population-based nuclear families. Genet Epidemiol, 1997. **14**(3): p. 317-32.

47. Laitinen, T., et al., Importance of genetic factors in adolescent asthma: a population-based twin-family study. Am J Respir Crit Care Med, 1998. **157**(4 Pt 1): p. 1073-8.

48. Duffy, D.L., et al., Genetics of asthma and hay fever in Australian twins. Am Rev Respir Dis, 1990. **142**(6 Pt 1): p. 1351-8.

49. Hopp, R.J., et al., Genetic analysis of allergic disease in twins. J Allergy Clin Immunol, 1984. **73**(2): p. 265-70.

50. Hanson, B., et al., Atopic disease and immunoglobulin E in twins reared apart and together. Am J Hum Genet, 1991. **48**(5): p. 873-9.

51. Husby, S., et al., Cord blood immunoglobulin E in like-sexed monozygotic and dizygotic twins. Clin Genet, 1996. **50**(5): p. 332-8.

52. Wuthrich, B., et al., Total and specific IgE (RAST) in atopic twins. Clin Allergy, 1981. **11**(2): p. 147-54.

53. Proceedings of the ATS workshop on refractory asthma: current understanding, recommendations, and unanswered questions. American Thoracic Society. Am J Respir Crit Care Med, 2000. **162**(6): p. 2341-51.

54. health, N.I.o. and N.H.L.a.B. Institute, Global Initiative for asthma. Global strategy for asthma management and prevention. 1998-2005, Bethesda, MD.

55. Weiss, S.T., M.L. Van Natta, and R.S. Zeiger, Relationship between increased airway responsiveness and asthma severity in the childhood asthma management program. Am J Respir Crit Care Med, 2000. **162**(1): p. 50-6.

56. Martinez, F.D., Role of respiratory infection in onset of asthma and chronic obstructive pulmonary disease. Clin Exp Allergy, 1999. **29 Suppl 2**: p. 53-8.

57. Martinez, F.D., Development of wheezing disorders and asthma in preschool children. Pediatrics, 2002. **109**(2 Suppl): p. 362-7.

58. Borish, L., et al., Total serum IgE levels in a large cohort of patients with severe or difficult-to-treat asthma. Ann Allergy Asthma Immunol, 2005. **95**(3): p. 247-53.

59. Burrows, B., et al., Association of asthma with serum IgE levels and skin-test reactivity to allergens. N Engl J Med, 1989. **320**(5): p. 271-7.

60. Busse, W.W., S. Banks-Schlegel, and S.E. Wenzel, Pathophysiology of severe asthma. J Allergy Clin Immunol, 2000. **106**(6): p. 1033-42.

61. Nieves, A., et al., Phenotypes of asthma revisited upon the presence of atopy. Respir Med, 2005. **99**(3): p. 347-54.

62. Martinez, F.J., C. Standiford, and S.E. Gay, Is it asthma or COPD? The answer determines proper therapy for chronic airflow obstruction. Postgrad Med, 2005. **117**(3): p. 19-26.

63. Louis, R., et al., The relationship between airways inflammation and asthma severity. Am J Respir Crit Care Med, 2000. **161**(1): p. 9-16.

64. Jatakanon, A., et al., Neutrophilic inflammation in severe persistent asthma. Am J Respir Crit Care Med, 1999. **160**(5 Pt 1): p. 1532-9.

65. Gauderman, W.J., Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol, 2002. **155**(5): p. 478-84.

66. Gauderman, W.J., Sample size requirements for matched case-control studies of gene-environment interaction. Stat Med, 2002. **21**(1): p. 35-50.

67. Purcell, S., S.S. Cherny, and P.C. Sham, Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics, 2003. **19**(1): p. 149-50.

68. Dib, C., et al., A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature, 1996. **380**(6570): p. 152-4.

69. The International HapMap Project. Nature, 2003. **426**(6968): p. 789-96.

70. Kruglyak, L. and D.A. Nickerson, Variation is the spice of life. Nat Genet, 2001. **27**(3): p. 234-6.

71. Palmer, L.J. and W.O. Cookson, Using single nucleotide polymorphisms as a means to understanding the pathophysiology of asthma. Respir Res, 2001. **2**(2): p. 102-12.

72. Spielman, R.S. and W.J. Ewens, The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet, 1996. **59**(5): p. 983-9.

73. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet, 1993. **52**(3): p. 506-16.

74. Ober, C., et al., Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. Hum Mol Genet, 1998. **7**(9): p. 1393-8.

75. Bland, J.M. and D.G. Altman, Multiple significance tests: the Bonferroni method. Bmj, 1995. **310**(6973): p. 170.

76. Rosenwasser, L.J. and L. Borish, Genetics of atopy and asthma: The rationale behind promoter based candidate gene studies (IL-4 and IL-10). Am J Respir Crit Care Med, 1997. **156**: p. S152-S155.

77. Thomas, N.S., J. Wilkinson, and S.T. Holgate, The candidate region approach to the genetics of asthma and allergy. Am J Respir Crit Care Med, 1997. **156**: p. S144-S151.

78. Laitinen, T., et al., A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. Nat Genet, 2001. **28**(1): p. 87-91.

79. Van Eerdewegh, P., et al., Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature, 2002. **418**(6896): p. 426-30.

80. Zhang, Y., et al., Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet, 2003. **34**(2): p. 181-6.

81. Allen, M., et al., Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nat Genet, 2003. **35**(3): p. 258-63.

82. Laitinen, T., et al., Characterization of a common susceptibility locus for asthma-related traits. Science, 2004. **304**(5668): p. 300-4.

83. Nicolae, D., et al., Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. Am. J. Hum. Genet, 2005. **76**: p. 349-357.

84. Noguchi, E., et al., Positional identification of an asthma susceptibility gene on human chromosome 5q33. Am J Respir Crit Care Med, 2005. **172**(2): p. 183-8.

85. Kimura, K., et al., Linkage and association of atopic asthma to markers on chromosome 13 in the Japanese population. Hum Mol Genet, 1999. **8**(8): p. 1487-90.

86. Daniels, S.E., et al., A genome-wide search for quantitative trait loci underlying asthma. Nature, 1996. **383**(6597): p. 247-50.

87. Hizawa, N., et al., Genetic regulation of Dermatophagoides pteronyssinus-specific IgE responsiveness: a genome-wide multipoint linkage analysis in families recruited through 2 asthmatic sibs. Collaborative Study on the Genetics of Asthma (CSGA). J Allergy Clin Immunol, 1998. **102**(3): p. 436-42.

88. Koppelman, G.H., et al., Genome-wide search for atopy susceptibility genes in Dutch families with asthma. J Allergy Clin Immunol, 2002. **109**(3): p. 498-506.

89. Wjst, M., et al., A genome-wide search for linkage to asthma. German Asthma Genetics Group. Genomics, 1999. **58**(1): p. 1-8.

90. Postma, D.S. and G.H. Koppelman, Confirmation of GPRA: a putative drug target for asthma. Am J Respir Crit Care Med, 2005. **171**(12): p. 1323-4.

91. Dizier, M.H., et al., Genome screen for asthma and related phenotypes in the French EGEA study. Am J Respir Crit Care Med, 2000. **162**(5): p. 1812-8.

92. Xu, J., et al., Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. Am J Hum Genet, 2001. **68**(6): p. 1437-46.

93. Cookson, W.O., et al., Genetic linkage of childhood atopic dermatitis to psoriasis susceptibility loci. Nat Genet, 2001. **27**(4): p. 372-3.

94. Xu, X., et al., A genomewide search for quantitative-trait loci underlying asthma. Am J Hum Genet, 2001. **69**(6): p. 1271-7.

95. A genome-wide search for asthma susceptibility loci in ethnically diverse populations. The Collaborative Study on the Genetics of Asthma (CSGA). Nat Genet, 1997. **15**(4): p. 389-92.

96. Lee, Y.A., et al., A major susceptibility locus for atopic dermatitis maps to chromosome 3q21. Nat Genet, 2000. **26**(4): p. 470-3.

97. Haagerup, A., et al., Allergic rhinitis--a total genome-scan for susceptibility genes suggests a locus on chromosome 4q24-q27. Eur J Hum Genet, 2001. **9**(12): p. 945-52.

98. Yokouchi, Y., et al., Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-q33 near the interleukin 12 B locus by a genome-wide search in Japanese families. Genomics, 2000. **66**(2): p. 152-60.

99. Xu, J., et al., Major genes regulating total serum immunoglobulin E levels in families with asthma. Am J Hum Genet, 2000. **67**(5): p. 1163-73.

100. Ober, C., et al., A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. Am J Hum Genet, 2000. **67**(5): p. 1154-62.

101. Mullis, K., et al., Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.

102. Syvanen, A.C., Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet, 2001. **2**(12): p. 930-42.

103. Kristensen, V.N., et al., High-throughput methods for detection of genetic variation. Biotechniques, 2001. **30**(2): p. 318-22, 324, 326 passim.

104. Lander, E.S. and D. Botstein, Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. Proc Natl Acad Sci U S A, 1986. **83**(19): p. 7353-7.

105. Day, I.N. and S.E. Humphries, Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. Anal Biochem, 1994. **222**(2): p. 389-95.

106. Holloway, J.W., S. Ye, and I.N. Day, Tools for molecular genetic epidemiology: a comparison of MADGE methodology with other systems. Biotechnol Genet Eng Rev, 2000. **17**: p. 71-88.

107. Sander, T., et al., Comparison of detection platforms and post-polymerase chain reaction DNA purification methods for use in conjunction with Cleavase fragment length polymorphism analysis. Electrophoresis, 1999. **20**(6): p. 1131-40.

108. Nataraj, A.J., et al., Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. Electrophoresis, 1999. **20**(6): p. 1177-85.

109. Keen, J., et al., Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. Trends Genet, 1991. **7**(1): p. 5.

110. Strachan, T. and A.P. Read, Human Molecular Genetics 2. Second edition ed. 1999: BIOS Scientific Publishers Ltd.

111. Livak, K.J., SNP genotyping by the 5'-Nuclease Reaction, in Methods in Molecular Biology. 2002, Humana press inc.: Totowa, New Jersey. p. 129-147.

112. Livak, K.J., Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal, 1999. **14**(5-6): p. 143-9.

113. Livak, K.J., et al., Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl, 1995. **4**(6): p. 357-62.

114. Tyagi, S., D.P. Bratu, and F.R. Kramer, Multicolor molecular beacons for allele discrimination. Nat Biotechnol, 1998. **16**(1): p. 49-53.

115. Tyagi, S. and F.R. Kramer, Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol, 1996. **14**(3): p. 303-8.

116. Landegren, U., et al., A ligase-mediated gene detection technique. Science, 1988. **241**(4869): p. 1077-80.

117. Nickerson, D.A., et al., Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. Proc Natl Acad Sci U S A, 1990. **87**(22): p. 8923-7.

118. Newton, C.R., et al., Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res, 1989. **17**(7): p. 2503-16.

119. Nazarenko, I.A., S.K. Bhatnagar, and R.J. Hohman, A closed tube format for amplification and detection of DNA based on energy transfer. Nucleic Acids Res, 1997. **25**(12): p. 2516-21.

120. Ye, S., et al., An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res, 2001. **29**(17): p. E88-8.

121. Ronaghi, M., et al., Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem, 1996. **242**(1): p. 84-9.

122. Nyren, P., B. Pettersson, and M. Uhlen, Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem, 1993. **208**(1): p. 171-5.

123. Alderborn, A., A. Kristofferson, and U. Hammerling, Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res, 2000. **10**(8): p. 1249-58.

124. Stocker, W., et al., The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. Protein Sci, 1995. **4**(5): p. 823-40.

125. Chang, C. and Z. Werb, The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. Trends Cell Biol, 2001. **11**(11): p. S37-43.

126. Izumi, Y., et al., A metalloprotease-disintegrin, MDC9/meltrin-gamma/ADAM9 and PKCdelta are involved in TPA-induced ectodomain shedding of membrane-anchored heparin-binding EGF-like growth factor. Embo J, 1998. **17**(24): p. 7260-72.

127. Nath, D., et al., Meltrin gamma(ADAM-9) mediates cellular adhesion through alpha(6)beta(1)integrin, leading to a marked induction of fibroblast cell motility. J Cell Sci, 2000. **113 (Pt 12)**: p. 2319-28.

128. Hooper, N.M., E.H. Karran, and A.J. Turner, Membrane protein secretases. Biochem J, 1997. **321 (Pt 2)**: p. 265-79.

129. Primakoff, P. and D.G. Myles, The ADAM gene family: surface proteins with adhesion and protease activity. Trends Genet, 2000. **16**(2): p. 83-7.

130. Peschon, J.J., et al., An essential role for ectodomain shedding in mammalian development. Science, 1998. **282**(5392): p. 1281-4.

131. Black, R.A., et al., A metalloproteinase disintegrin that releases tumour-necrosis factor-alpha from cells. Nature, 1997. **385**(6618): p. 729-33.

132. Black, R.A., Tumor necrosis factor-alpha converting enzyme. Int J Biochem Cell Biol, 2002. **34**(1): p. 1-5.

133. Mohan, M.J., et al., The tumor necrosis factor-alpha converting enzyme (TACE): a unique metalloproteinase with highly defined substrate selectivity. Biochemistry, 2002. **41**(30): p. 9462-9.

134. Reddy, P., et al., Functional analysis of the domain structure of tumor necrosis factor-alpha converting enzyme. J Biol Chem, 2000. **275**(19): p. 14608-14.

135. Asakura, M., et al., Cardiac hypertrophy is inhibited by antagonism of ADAM12 processing of HB-EGF: metalloproteinase inhibitors as a new therapy. Nat Med, 2002. **8**(1): p. 35-40.

136. Schwettmann, L. and H. Tschesche, Cloning and expression in Pichia pastoris of metalloprotease domain of ADAM 9 catalytically active against fibronectin. Protein Expr Purif, 2001. **21**(1): p. 65-70.

137. Yan, Y., K. Shirakabe, and Z. Werb, The metalloprotease Kuzbanian (ADAM10) mediates the transactivation of EGF receptor by G protein-coupled receptors. J Cell Biol, 2002. **158**(2): p. 221-6.

138. Loechel, F., et al., Human ADAM 12 (meltrin alpha) is an active metalloprotease. J Biol Chem, 1998. **273**(27): p. 16993-7.

139. Iba, K., et al., The cysteine-rich domain of human ADAM 12 supports cell adhesion through syndecans and triggers signaling events that lead to beta1 integrin-dependent cell spreading. J Cell Biol, 2000. **149**(5): p. 1143-56.

140. Krymskaya, V.P. and J.M. Shipley, Lymphangioleiomyomatosis: a complex tale of serum response factor-mediated tissue inhibitor of metalloproteinase-3 regulation. Am J Respir Cell Mol Biol, 2003. **28**(5): p. 546-50.

141. Sagane, K., et al., Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23. Gene, 1999. **236**(1): p. 79-86.

142. Bigler, D., et al., Sequence-specific interaction between the disintegrin domain of mouse ADAM 2 (fertilin beta) and murine eggs. Role of the alpha(6) integrin subunit. J Biol Chem, 2000. **275**(16): p. 11576-84.

143. Almeida, E.A., et al., Mouse egg integrin alpha 6 beta 1 functions as a sperm receptor. Cell, 1995. **81**(7): p. 1095-104.

144. Cho, C., et al., Fertilization defects in sperm from mice lacking fertilin beta. Science, 1998. **281**(5384): p. 1857-9.

145. Hougaard, S., et al., Trafficking of human ADAM 12-L: retention in the trans-Golgi network. Biochem Biophys Res Commun, 2000. **275**(2): p. 261-7.

146. Loechel, F., et al., Regulation of human ADAM 12 protease by the prodomain. Evidence for a functional cysteine switch. J Biol Chem, 1999. **274**(19): p. 13427-33.

147. Gilpin, B.J., et al., A novel, secreted form of human ADAM 12 (meltrin alpha) provokes myogenesis in vivo. J Biol Chem, 1998. **273**(1): p. 157-66.

148. Shi, Z., et al., ADAM 12, a disintegrin metalloprotease, interacts with insulin-like growth factor-binding protein-3. J Biol Chem, 2000. **275**(24): p. 18574-80.

149. Howard, L., R.A. Maciewicz, and C.P. Blobel, Cloning and characterization of ADAM28: evidence for autocatalytic pro-domain removal and for cell surface localization of mature ADAM28. Biochem J, 2000. **348 Pt 1**: p. 21-7.

150. Roberts, C.M., et al., MDC-L, a novel metalloprotease disintegrin cysteine-rich protein family member expressed by human lymphocytes. J Biol Chem, 1999. **274**(41): p. 29251-9.

151. Hotoda, N., et al., A secreted form of human ADAM9 has an alpha-secretase activity for APP. Biochem Biophys Res Commun, 2002. **293**(2): p. 800-5.

152. Yavari, R., et al., Human metalloprotease-disintegrin Kuzbanian regulates sympathoadrenal cell fate in development and neoplasia. Hum Mol Genet, 1998. **7**(7): p. 1161-7.

153. Katagiri, T., et al., Human metalloprotease/disintegrin-like (MDC) gene: exon-intron organization and alternative splicing. Cytogenet Cell Genet, 1995. **68**(1-2): p. 39-44.

154. Powell, R.M., et al., The Splicing and Fate of ADAM33 Transcripts in Primary Human Airways Fibroblasts. Am J Respir Cell Mol Biol, 2004. **31**(1): p. 13-21.

155. Wen, C., M.M. Metzstein, and I. Greenwald, SUP-17, a Caenorhabditis elegans ADAM protein related to Drosophila KUZBANIAN, and its role in LIN-12/NOTCH signalling. Development, 1997. **124**(23): p. 4759-67.

156. Kang, Q., Y. Cao, and A. Zolkiewska, Metalloprotease-disintegrin ADAM 12 binds to the SH3 domain of Src and activates Src tyrosine kinase in C2C12 cells. Biochem J, 2000. **352 Pt 3**: p. 883-92.

157. Huang, L., et al., Screen and identification of proteins interacting with ADAM19 cytoplasmic tail. Mol Biol Rep, 2002. **29**(3): p. 317-23.

158. Costa, F.F., et al., Epigenetic silencing of the adhesion molecule ADAM23 is highly frequent in breast tumors. Oncogene, 2004. **23**(7): p. 1481-8.

159. Hagihara, A., et al., Identification of 27 5' CpG islands aberrantly methylated and 13 genes silenced in human pancreatic cancers. Oncogene, 2004. **23**(53): p. 8705-10.

160. Takada, H., et al., ADAM23, a possible tumor suppressor gene, is frequently silenced in gastric cancers by homozygous deletion or aberrant promoter hypermethylation. Oncogene, 2005.

161. Haitchi, H.M., et al., ADAM33 expression in asthmatic airways and human embryonic lungs. Am J Respir Crit Care Med, 2005. **171**(9): p. 958-65.

162. Sagane, K., et al., Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2 and MDC3: novel human cellular disintegrins highly expressed in the brain. Biochem J, 1998. **334 (Pt 1)**: p. 93-8.

163. Yoshinaka, T., et al., Identification and characterization of novel mouse and human ADAM33s with potential metalloprotease activity. Gene, 2002. **282**(1-2): p. 227-36.

164. Van Wart, H.E. and H. Birkedal-Hansen, The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. Proc Natl Acad Sci U S A, 1990. **87**(14): p. 5578-82.

165. Garlisi, C.G., et al., Human ADAM33: protein maturation and localization. Biochem Biophys Res Commun, 2003. **301**(1): p. 35-43.

166. Zou, J., et al., Catalytic activity of human ADAM33. J Biol Chem, 2004. **279**(11): p. 9818-30.

167. Davies, D.E., et al., Airway Remodelling in Asthma-New Insights. J Allergy Clin Immunol, 2003. **111**: p. 215-225.

168. Bridges, L.C., D. Sheppard, and R.D. Bowditch, ADAM disintegrin-like domain recognition by the lymphocyte integrins alpha4beta1 and alpha4beta7. Biochem J, 2005. **387**(Pt 1): p. 101-8.

169. Keith, T., Novel Human Gene Relating to Respiratory Diseases, obesity, and Inflammatory Bowel Disease. 2001, Genome Therapeutics Corporation.

170. Cao, Y., et al., Intracellular processing of metalloprotease disintegrin ADAM12. J Biol Chem, 2002. **277**(29): p. 26403-11.

171. Umland, S.P., et al., Human ADAM33 mRNA expression profile and post-transcriptional regulation. Am J Respir Cell Mol Biol, 2003.

172. Haitchi, H.M., R. Powell, and D.I. Wilson, ADAM33 expression in embryonic mouse lung. Am J Respir Cell Mol Biol, 2003. **167**: p. A377.

173. Lee, J.Y., et al., A Disintegrin and Metalloproteinase 33 Protein in Asthmatics : Relevance to Airflow Limitation. Am J Respir Crit Care Med, 2005.

174. Vignola, A.M., et al., Airway inflammation in mild intermittent and in persistent asthma. Am J Respir Crit Care Med, 1998. **157**(2): p. 403-9.

175. Larsen, K., et al., Functional and phenotypical comparison of myofibroblasts derived from biopsies and bronchoalveolar lavage in mild asthma and scleroderma. Respir Res, 2006. **7**: p. 11.

176. Laporte, J.D., et al., Expression of ADAM-33 in cultured human airway smooth muscle cells. Am J Respir Crit Care Med, 2003. **167**(7): p. A329.

177. Beelman, C.A. and R. Parker, Degradation of mRNA in eukaryotes. Cell, 1995. **81**(2): p. 179-83.

178. Dalgleish, G.D., et al., Localisation of a reporter transcript by the c-myc 3'-UTR is linked to translation. Nucleic Acids Res, 1999. **27**(22): p. 4363-8.

179. De Sanctis, G.T., et al., Quantitative locus analysis of airway hyperresponsiveness in A/J and C57BL/6J mice. Nat Genet, 1995. **11**(2): p. 150-4.

180. Gunn, T.M., et al., Identification and preliminary characterization of mouse Adam33. BMC Genet, 2002. **3**(1): p. 2.

181. Howard, T.D., et al., Association of a disintegrin and metalloprotease 33 (ADAM33) gene with asthma in ethnically diverse populations. J Allergy Clin Immunol, 2003. **112**(4): p. 717-22.

182. Werner, M., et al., Asthma is associated with single-nucleotide polymorphisms in ADAM33. Clin Exp Allergy, 2004. **34**(1): p. 26-31.

183. Lee, J.H., et al., ADAM33 polymorphism: association with bronchial hyper-responsiveness in Korean asthmatics. Clin Exp Allergy, 2004. **34**(6): p. 860-5.

184. Simpson, A., et al., Polymorphisms in a disintegrin and metalloprotease 33 (ADAM33) predict impaired early-life lung function. Am J Respir Crit Care Med, 2005. **172**(1): p. 55-60.

185. Blakey, J., et al., Contribution of ADAM33 polymorphisms to the population risk of asthma. Thorax, 2005. **60**(4): p. 274-6.

186. Cheng, L., et al., Polymorphisms in ADAM33 are associated with allergic rhinitis due to Japanese cedar pollen. Clin Exp Allergy, 2004. **34**(8): p. 1192-201.

187. Raby, B.A., et al., ADAM33 polymorphisms and phenotype associations in childhood asthma. J Allergy Clin Immunol, 2004. **113**(6): p. 1071-8.

188. Lind, D.L., et al., ADAM33 is not associated with asthma in Puerto Rican or Mexican populations. Am J Respir Crit Care Med, 2003. **168**(11): p. 1312-6.

189. Gern, J.E., R.F. Lemanske, Jr., and W.W. Busse, Early life origins of asthma. J Clin Invest, 1999. **104**(7): p. 837-43.

190. Martinez, F.D., et al., Asthma and wheezing in the first six years of life. The Group Health Medical Associates. N Engl J Med, 1995. **332**(3): p. 133-8.

191. Sears, M.R., et al., A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. N Engl J Med, 2003. **349**(15): p. 1414-22.

192. Turner, S.W., et al., The Relationship Between Infant Airway Function, Childhood Airway Responsiveness and Asthma. Am J Respir Crit Care Med, 2004.

193. Custovic, A., et al., Effect of environmental manipulation in pregnancy and early life on respiratory symptoms and atopy during first year of life: a randomised trial. Lancet, 2001. **358**(9277): p. 188-93.

194. Stick, S.M., et al., Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants. Lancet, 1996. **348**(9034): p. 1060-4.

195. Le Souef, P.N., Pediatric origins of adult lung diseases. 4. Tobacco related lung diseases begin in childhood. Thorax, 2000. **55**(12): p. 1063-7.

196. Woodcock, A., et al., Early life environmental control: effect on symptoms, sensitization, and lung function at age 3 years. Am J Respir Crit Care Med, 2004. **170**(4): p. 433-9.

197. Le Souef, P., Prediction of asthma in children at 6 and 12 years of age: Perth infant asthma follow-up study. Pediatr Allergy Immunol, 2002. **13 Suppl 15**: p. 44-6.

198. Schaid, D.J., et al., Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet, 2002. **70**(2): p. 425-34.

199. Zaykin, D.V., et al., Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered, 2002. **53**(2): p. 79-91.

200. Clayton, D.G., SNPHAP. 2001, Cambridge Institute for Medical Research: Cambridge.

201. Maniatis, N., et al., Positional cloning by linkage disequilibrium. Am J Hum Genet, 2004. **74**(5): p. 846-55.

202. Maniatis, N., et al., The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2228-33.

203. Collins, A., Mapping disease genes using the Malecot model for allelic association and the beta model for linkage. Clin Exp Allergy, 1999. **29 Suppl 4**: p. 53-6.

204. Maniatis, N., et al., The optimal measure of linkage disequilibrium reduces error in association mapping of affection status. Hum Mol Genet, 2005. **14**(1): p. 145-53.

205. Devlin, B., N. Risch, and K. Roeder, Disequilibrium mapping: composite likelihood for pairwise disequilibrium. Genomics, 1996. **36**(1): p. 1-16.

206. Sachidanandam, R., et al., A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 2001. **409**(6822): p. 928-33.

207. Griffiths, A.J.F., et al., eds. An introduction to genetic analysis. 7th edition ed. 2000, W.H.Freeman: New York.

208. Kong, A., et al., A high-resolution recombination map of the human genome. Nat Genet, 2002. **31**(3): p. 241-7.

209. AppliedBiosystems, ABI prism 7900HT sequence detection system-User guide.

210. Holland, P.M., et al., Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A, 1991. **88**(16): p. 7276-80.

211. Kutyavin, I.V.A., I. A. Mills, A.Gorn, V.V. Lukhtanov, E. A. Belousov, E. S. Singer, M. J. Walburger, D. K. Lokhov, S. G. Gall, A. A. Dempcy, R. Reed, M. W. Meyer, R. B. Hedgpeth, J., 3'-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res, 2000. **28**(2): p. 655-661.

212. Jeffreys, A.J., J. Murray, and R. Neumann, High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell, 1998. **2**(2): p. 267-73.

213. Jeffreys, A.J., D.L. Neil, and R. Neumann, Repeat instability at human minisatellites arising from meiotic recombination. Embo J, 1998. **17**(14): p. 4147-57.

214. Ke, X. and L.R. Cardon, Efficient selective screening of haplotype tag SNPs. Bioinformatics, 2003. **19**(2): p. 287-8.

215. Pritchard, J.K. and M. Przeworski, Linkage disequilibrium in humans: models and data. Am J Hum Genet, 2001. **69**(1): p. 1-14.

216. von Mutius, E. and M.R. Sears, Risk factors for development of asthma, in European respiratory monograph Asthma, F. Chung and L.M. Fabbri, Editors. 2003, European respiratory society journals Ltd.

217. von Mutius, E., et al., Prevalence of asthma and allergic disorders among children in united Germany: a descriptive comparison. Bmj, 1992. **305**(6866): p. 1395-9.

218. von Mutius, E., et al., Increasing prevalence of hay fever and atopy among children in Leipzig, East Germany. Lancet, 1998. **351**(9106): p. 862-6.

219. Wennergren, G. and S. Kristjansson, Wheezing in infancy and its long term consequences, in European Respiratory Monograph. 2002, European Respiratory Society Journals. p. 116-130.

220. Martinez, F.D., et al., Diminished lung function as a predisposing factor for wheezing respiratory illness in infants. N Engl J Med, 1988. **319**(17): p. 1112-7.

221. Martinez, F.D., et al., Initial airway function is a risk factor for recurrent wheezing respiratory illnesses during the first three years of life. Group Health Medical Associates. Am Rev Respir Dis, 1991. **143**(2): p. 312-6.

222. Murray, C.S., et al., Lung function at one month of age as a risk factor for infant respiratory symptoms in a high risk population. Thorax, 2002. **57**(5): p. 388-92.

223. Dezateux, C., et al., Impaired airway function and wheezing in infancy: the influence of maternal smoking and a genetic predisposition to asthma. Am J Respir Crit Care Med, 1999. **159**(2): p. 403-10.

224. Dezateux, C., et al., Airway function at one year: association with premorbid airway function, wheezing, and maternal smoking. Thorax, 2001. **56**(9): p. 680-6.

225. Taussig, L.M., et al., Tucson Children's Respiratory Study: 1980 to present. J Allergy Clin Immunol, 2003. **111**(4): p. 661-75; quiz 676.

226. Stein, R.T. and F.D. Martinez, Asthma phenotypes in childhood: lessons from an epidemiological approach. Paediatr Respir Rev, 2004. **5**(2): p. 155-61.

227. Stein, R.T., et al., Peak flow variability, methacholine responsiveness and atopy as markers for detecting different wheezing phenotypes in childhood. Thorax, 1997. **52**(11): p. 946-52.

228. Hoo, A.F., et al., Development of Lung Function in Early Life: Influence of Birthweight in Infants of Non-smokers. Am J Respir Crit Care Med, 2004.

229. Lucas, J.S., et al., Small Size at Birth and Greater Postnatal Weight Gain: Relations to Diminished Infant Lung Function. Am J Respir Crit Care Med, 2004.

230. Singh, A.M., et al., Bronchiolitis to Asthma:A Review and Call for Studies of Gene-Viral Interactions in Asthma Causation. Am J Respir Crit Care Med, 2006.

231. Papi, A., et al., Respiratory viruses and asthma, in European Respiratory Monograph 23, Asthma, F. Chung and L.M. Fabbri, Editors. 2003, European Respitratory Society Journals.

232. Stokes, G.M., et al., Lung function abnormalities after acute bronchiolitis. J Pediatr, 1981. **98**(6): p. 871-4.

233. Kattan, M., et al., Pulmonary function abnormalities in symptom-free children after bronchiolitis. Pediatrics, 1977. **59**(5): p. 683-8.

234. Gurwitz, D., C. Mindorff, and H. Levison, Increased incidence of bronchial reactivity in children with a history of bronchiolitis. J Pediatr, 1981. **98**(4): p. 551-5.

235. Phelan, P.D., C.F. Robertson, and A. Olinsky, The Melbourne Asthma Study: 1964-1999. J Allergy Clin Immunol, 2002. **109**(2): p. 189-94.

236. Stein, R.T., et al., Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. Lancet, 1999. **354**(9178): p. 541-5.

237. Henderson, J., et al., Hospitalization for RSV bronchiolitis before 12 months of age and subsequent asthma, atopy and wheeze: a longitudinal birth cohort study. Pediatr Allergy Immunol, 2005. **16**(5): p. 386-92.

238. Sigurs, N., et al., Asthma and immunoglobulin E antibodies after respiratory syncytial virus bronchiolitis: a prospective cohort study with matched controls. Pediatrics, 1995. **95**(4): p. 500-5.

239. Lemanske, R.F., Jr., et al., Rhinovirus illnesses during infancy predict subsequent childhood wheezing. J Allergy Clin Immunol, 2005. **116**(3): p. 571-7.

240. Rakes, G.P., et al., Rhinovirus and respiratory syncytial virus in wheezing children requiring emergency care. IgE and eosinophil analyses. Am J Respir Crit Care Med, 1999. **159**(3): p. 785-90.

241. Young, S., et al., Parental factors affecting respiratory function during the first year of life. Pediatr Pulmonol, 2000. **29**(5): p. 331-40.

242. Clarke, J.R., et al., Evidence for genetic associations between asthma, atopy, and bronchial hyperresponsiveness: a study of 8- to 18-yr-old twins. Am J Respir Crit Care Med, 2000. **162**(6): p. 2188-93.

243. Jenkins, M.A., et al., Factors in childhood as predictors of asthma in adult life. Bmj, 1994. **309**(6947): p. 90-3.

244. Raby, B.A., et al., Paternal history of asthma and airway responsiveness in children with asthma. Am J Respir Crit Care Med, 2005. **172**(5): p. 552-8.

245. Friedman, N.J. and R.S. Zeiger, The role of breast-feeding in the development of allergies and asthma. J Allergy Clin Immunol, 2005. **115**(6): p. 1238-48.

246. Remes, S.T., et al., Dog exposure in infancy decreases the subsequent risk of frequent wheeze but not of atopy. J Allergy Clin Immunol, 2001. **108**(4): p. 509-15.

247. Rosenthal, M. and A. Bush, The Growing Lung: normal development, and the long-term effects of pre- and postnatal insults., in European respiratory monograph. 2002, ERS Journals Ltd. p. 1-24.

248. Jones, C.A. and J.O. Warner, Fetal origins of adult disease, in Fetal origins of cardiovascular and lung disease, D. Barker, Editor. 2000.

249. Hislop, A.A. and H. Pandya, Structural development, in Childhood asthma and other wheezing disorders, M. Silverman, Editor. 2002, Arnold.

250. Martin, T.R., et al., Airway size is related to sex but not lung size in normal adults. J Appl Physiol, 1987. **63**(5): p. 2042-7.

251. Sly, P.D. and F.S. Flack, Lung Function, in Childhood asthma and other wheezing disorders, M. Silverman, Editor. 2002, Arnold.

252. Merkus, P., J. Jongste, and J. Stocks, Respiratory function measurments in infants and children, in European Respiratory Monograph, R. Gosselink and H. Stam, Editors. 2005, European Respiratory Society.

253. Klug, B. and H. Bisgaard, Measurement of lung function in awake 2-4-year-old asthmatic children during methacholine challenge and acute asthma: a comparison of the impulse oscillation technique, the interrupter technique, and transcutaneous measurement of oxygen versus whole-body plethysmography. Pediatr Pulmonol, 1996. **21**(5): p. 290-300.

254. Klug, B. and H. Bisgaard, Measurement of the specific airway resistance by plethysmography in young children accompanied by an adult. Eur Respir J, 1997. **10**(7): p. 1599-605.

255. Nielsen, K.G. and H. Bisgaard, Discriminative capacity of bronchodilator response measured with three different lung function techniques in asthmatic and healthy children aged 2 to 5 years. Am J Respir Crit Care Med, 2001. **164**(4): p. 554-9.

256. Feher, A., et al., Flow limitation in normal infants: a new method for forced expiratory maneuvers from raised lung volumes. J Appl Physiol, 1996. **80**(6): p. 2019-25.

257. The raised volume rapid thoracoabdominal compression technique. The Joint American Thoracic Society/European Respiratory Society Working Group on Infant Lung Function. Am J Respir Crit Care Med, 2000. **161**(5): p. 1760-2.

258. Kurzius-Spencer, M., et al., Familial correlation in the decline of forced expiratory volume in one second. Am J Respir Crit Care Med, 2001. **164**(7): p. 1261-5.

259. Morgan, W.J. and F.D. Martinez, Maternal Smoking and Infant lung function. Further evidence for An in Utero Effect. Am J Respir Crit Care Med, 1998. **158**: p. 689-690.

260. Stein, R.T., et al., Influence of parental smoking on respiratory symptoms during the first decade of life: the Tucson Children's Respiratory Study. Am J Epidemiol, 1999. **149**(11): p. 1030-7.

261. Palmer, L.J., et al., Airway responsiveness in early infancy predicts asthma, lung function, and respiratory symptoms by school age. Am J Respir Crit Care Med, 2001. **163**(1): p. 37-42.

262. Simpson, B.M., et al., NAC Manchester Asthma and Allergy Study (NACMAAS): risk factors for asthma and allergic disorders in adults. Clin Exp Allergy, 2001. **31**(3): p. 391-9.

263. Bisgaard, H., The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. Ann Allergy Asthma Immunol, 2004. **93**(4): p. 381-9.

264. Joseph-Bowen, J., et al., Lung function, bronchial responsiveness, and asthma in a community cohort of 6-year-old children. Am J Respir Crit Care Med, 2004. **169**(7): p. 850-4.

265. Stovin, P.G., Early lung development. Thorax, 1985. **40**(6): p. 401-4.

266. Nielsen, K.G., et al., Serial lung function and responsiveness in cystic fibrosis during early childhood. Am J Respir Crit Care Med, 2004. **169**(11): p. 1209-16.

267. Ranganathan, S.C., et al., Exploring the relationship between forced maximal flow at functional residual capacity and parameters of forced expiration from raised lung volume in healthy infants. Pediatr Pulmonol, 2002. **33**(6): p. 419-28.

268. Loland, L., et al., Sensitivity of bronchial responsiveness measurements in young infants. Chest, 2006. **129**(3): p. 669-75.

269. Bentur, L., et al., Methacholine bronchial provocation measured by spirometry versus wheeze detection in preschool children. BMC Pediatr, 2005. **5**: p. 19.

270. Yoo, Y., et al., Percentage fall in FVC at the provocative concentration of methacholine causing a 20% fall in FEV1 in symptomatic asthma and clinical remission during adolescence. Chest, 2006. **129**(2): p. 272-7.

271. Jeffery, P.K., Pathology of asthma, in European Respiratory Monograph 23 Asthma, F. Chung and L.M. Fabbri, Editors. 2003, European Respiratory Society.

272. Griffiths, A.M., et al., Introduction to Genetic Analysis. 8th edition ed. 2005: W.H. Freeman and Company.

273. de Wet, J.R., et al., Firefly luciferase gene: structure and expression in mammalian cells. Mol Cell Biol, 1987. **7**(2): p. 725-37.

274. Strachan, D. and A.P. Read, Human Molecular Genetics 3. 3rd ed. 2004, London and New York: Garland Science.

275. Ikawa, M., et al., Green fluorescent protein (GFP) as a vital marker in mammals. Curr Top Dev Biol, 1999. **44**: p. 1-20.

276. corporation, P., pGL3 luciferase reporter vectors. 2002.

277. Kadonaga, J.T., et al., Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science, 1988. **242**(4885): p. 1566-70.

278. Dynan, W.S. and R. Tjian, Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell, 1983. **32**(3): p. 669-80.

279. Dynan, W.S. and R. Tjian, The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell, 1983. **35**(1): p. 79-87.

280. Mayor, C., et al., VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics, 2000. **16**(11): p. 1046-7.

281. Frazer, K.A., et al., VISTA: computational tools for comparative genomics. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W273-9.

282. Quandt, K., et al., MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res, 1995. **23**(23): p. 4878-84.

283. Chae, S.C., K.H. Yoon, and H.T. Chung, Identification of novel polymorphisms in the Adam33 gene. J Hum Genet, 2003. **48**(5): p. 278-81.

284. Wills-Karp, M., et al., Interleukin-13: central mediator of allergic asthma. Science, 1998. **282**(5397): p. 2258-61.

285. Wills-Karp, M., Interleukin-13 in asthma pathogenesis. Immunol Rev, 2004. **202**: p. 175-90.

286. Ito, I., et al., Effect of cytokines/chemokines on ADAM33 expression in human airway smooth muscle cells. 2005: ATS A237.

287. Hewson, C.A., M.R. Edbrooke, and S.L. Johnston, PMA induces the MUC5AC respiratory mucin in human bronchial epithelial cells, via PKC, EGF/TGF-alpha, Ras/Raf, MEK, ERK and Sp1-dependent mechanisms. J Mol Biol, 2004. **344**(3): p. 683-95.

288. Perry, R.L., M.H. Parker, and M.A. Rudnicki, Activated MEK1 binds the nuclear MyoD transcriptional complex to repress transactivation. Mol Cell, 2001. **8**(2): p. 291-301.

289. Trouche, D., et al., Repression of c-fos promoter by MyoD on muscle cell differentiation. Nature, 1993. **363**(6424): p. 79-82.

290. Vinals, F., et al., Myogenesis and MyoD Downregulate Sp1. The Journal of Biological Chemistry, 1997. **272**(20): p. 12913-12921.

291. Mayer, D.C. and L.A. Leinwand, Sarcomeric gene expression and contractility in myofibroblasts. J Cell Biol, 1997. **139**(6): p. 1477-84.

292. Roche, W.R., et al., Subepithelial fibrosis in the bronchi of asthmatics. Lancet, 1989. **1**: p. 520-4.

293. Brewster, C.E., et al., Myofibroblasts and subepithelial fibrosis in bronchial asthma. Am J Respir Cell Mol Biol, 1990. **3**(5): p. 507-11.

294. Kadonaga, J.T., et al., Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell, 1987. **51**(6): p. 1079-90.

295. Zhao, C. and A. Meng, Sp1-like transcription factors are regulators of embryonic development in vertebrates. Develop. Growth Differ., 2005. **47**: p. 201-211.

296. Lomberk, G. and R. Urrutia, The family feud: turning off Sp1 by Sp1-like KLF proteins. Biochem J, 2005. **392**(Pt 1): p. 1-11.

297. Courey, A.J., et al., Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell, 1989. **59**(5): p. 827-36.

298. Mastrangelo, I.A., et al., DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5670-4.

299. Li, R., et al., Sp1 activates and inhibits transcription from separate elements in the proximal promoter of the human adenine nucleotide translocase 2 (ANT2) gene. J Biol Chem, 1996. **271**(31): p. 18925-30.

300. Hagen, G., et al., Sp1-mediated transcriptional activation is repressed by Sp3. Embo J, 1994. **13**(16): p. 3843-51.

301. Freely associating. Nat Genet, 1999. **22**(1): p. 1-2.

302. Weiss, S.T., Association studies in asthma genetics. Am J Respir Crit Care Med, 2001. **164**(11): p. 2014-5.

303. Gold, D.R., et al., Gender- and race-specific effects of asthma and wheeze on level and growth of lung function in children in six U.S. cities. Am J Respir Crit Care Med, 1994. **149**(5): p. 1198-208.

304. O'Donnell, A.R., et al., Age-specific relationship between CD14 and atopy in a cohort assessed from age 8 to 25 years. Am J Respir Crit Care Med, 2004. **169**(5): p. 615-22.

305. Takemoto, N., et al., Cutting edge: chromatin remodeling at the IL-4/IL-13 intergenic regulatory region for Th2-specific cytokine gene cluster. J Immunol, 2000. **165**(12): p. 6687-91.

306. Goll, M.G., et al., Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science, 2006. **311**(5759): p. 395-8.

307. Holgate, S.T., et al., Bronchial epithelium as a key regulator of airway allergen sensitisation and remodeling in asthma. Am J Respir Crit Care Med, 2000. **162**: p. S113-S117.

308. Galliano, M.F., et al., Binding of ADAM12, a marker of skeletal muscle regeneration, to the muscle-specific actin-binding protein, alpha -actinin-2, is required for myoblast fusion. J Biol Chem, 2000. **275**(18): p. 13933-9.

309. Brown, T.A., Gene Cloning and DNA analysis. Fourth ed. 2001: Blackwell. 363.

310. Peebles, R.S., Jr., Viral infections, atopy, and asthma: is there a causal relationship? J Allergy Clin Immunol, 2004. **113**(1 Suppl): p. S15-8.

311. Holgate, S.T., et al., The bronchial epithelium as a key regulator of airway inflammation and remodelling in asthma. Clin Exp Allergy, 1999. **29 Suppl 2**: p. 90-5.

312. Beum, P.V., et al., Mucin biosynthesis: upregulation of core 2 beta 1,6 N-acetylglucosaminyltransferase by retinoic acid and Th2 cytokines in a human airway epithelial cell line. Am J Physiol Lung Cell Mol Physiol, 2005. **288**(1): p. L116-24.

313. Chaves, A.C., et al., IL-4 and IL-13 regulate the induction of indoleamine 2,3-dioxygenase activity and the control of Toxoplasma gondii replication in human fibroblasts activated with IFN-gamma. Eur J Immunol, 2001. **31**(2): p. 333-44.

314. Darley, R., et al., Interactions between interferon gamma and retinoic acid with transforming growth factor beta in the induction of immune recogrition molecules. Cancer Immunol Immunother, 1993. **37**(2): p. 112-8.

315. Djavaheri-Mergny, M., et al., Ultraviolet A decreases epidermal growth factor (EGF) processing in cultured human fibroblasts and keratinocytes: inhibition of EGF-induced diacylglycerol formation. J Invest Dermatol, 1994. **102**(2): p. 192-6.

316. Baginski, T.K., et al., Cigarette smoke synergistically enhances respiratory mucin induction by proinflammatory stimuli. Am J Respir Cell Mol Biol, 2006. **35**(2): p. 165-74.

317. Shao, M.X., I.F. Ueki, and J.A. Nadel, Tumor necrosis factor alpha-converting enzyme mediates MUC5AC mucin expression in cultured human airway epithelial cells. Proc Natl Acad Sci U S A, 2003. **100**(20): p. 11618-23.

318. Shaw, T.J., et al., Endobronchial ultrasound to assess airway wall thickening: validation in vitro and in vivo. Eur Respir J, 2004. **23**(6): p. 813-7.