GENOMIC ANALYSES OF FAMILIAL AND SPORADIC AUTOIMMUNE ADDISON'S DISEASE

Anna Louise Mitchell

MBBS (Hons), MRes, MRCP (UK)

Thesis submitted for the degree of Doctor of Philosophy

Institute of Genetic Medicine,

Faculty of Medical Sciences,

Newcastle University,

Newcastle upon Tyne, UK

May 2013

ABSTRACT

Autoimmune Addison's disease (AAD) is a rare and highly heritable endocrinopathy. It is a complex genetic disease, meaning that it is due to a combination of interacting environmental and genetic factors. To date, the majority of the substantial genetic component to AAD aetiology remains undefined. In this study, a combination of hypothesis-driven (candidate gene) and discovery-driven (genome-wide) approaches have been used to search for novel genetic determinants of AAD.

PCR-based approaches were undertaken to study the potential role of the *CYP21A1P* pseudogene in AAD. *CYP21A1P* is highly homologous to the *CYP21A2* gene which encodes 21-hydroxylase, the primary autoantigen in AAD. In individuals with AAD, *CYP21A1P* is more likely to be absent from the genomic DNA sequence than in controls. qPCR and *in situ* hybridisation have been successfully combined to identify *CYP21A1P* transcripts in thymic, and fetal adrenal, tissue. These data perhaps indicate a role for *CYP21A1P* in induction of immune tolerance, with its loss being associated with autoimmunity against the steroidogenic apparatus.

Taking a broader candidate gene approach, the largest association analysis in AAD to date, of twenty candidate genes in six European AAD cohorts, suggests a role for *NF-κB1*, *IL23A* and *GATA3* variants in susceptibility to AAD in individual European cohorts, and a role for *STAT4* more universally in AAD.

SNP array technology has been used to conduct the first genome-wide linkage and association analysis in AAD. The linkage study, including 23 families, has linked regions on chromosomes 6, 7, 9 and 18 to disease. A genome-wide association analysis, comparing the 50 familial AAD cases to the Wellcome Trust 1958 UK Birth Cohort control group, revealed clusters of associated SNPs on chromosomes 2 and 6.

This body of work has illustrated some of the challenges in investigating a rare, complex genetic disorder, and how international collaboration can help to resolve some of these issues. In the course of this work, in addition to identifying a number of novel genetic determinants to AAD, exciting preliminary results have been generated which will need to be followed up. It is hoped that once these preliminary findings are replicated and further investigated, they will contribute significantly to an increase in our understanding of the pathogenesis of AAD, with the long-term aim of identifying novel means of treating the disease, altering its natural history or even preventing it.

DECLARATION

This thesis is based upon work undertaken during the course of my studies at the Institute of Genetic Medicine, Faculty of Medical Sciences, Newcastle University, between April 2010 and April 2013. The content of this thesis is original and no part has been submitted for the award of any other degree.

DEDICATION

This thesis is dedicated to my wonderful parents, Carol and John Mitchell.

FUNDING

Dr Anna Louise Mitchell is in grateful receipt of an MRC clinical research training fellowship (reference number G0900390). This project was also supported by a European Union Framework 7 grant (reference 201167) to the Euradrenal Consortium.

ACKNOWLEDGMENTS

I am grateful to many individuals who have assisted me in completing this work. In particular, I wish to thank my supervisors, Professor Simon Pearce and Professor Heather Cordell, for their support, patience, guidance and encouragement. I have received excellent research training and have developed many useful and transferable skills with their help. I also wish to thank my PhD course assessors, Dr John Sayer and Dr Tim Cheetham, for their constructive advice and useful suggestions at my yearly appraisals.

The *CYP21* locus study, which forms a chapter of this thesis, was initially started by Alekhya Narravula, an MSc student in our group. Alekhya developed and optimised the competitive PCR assay to determine the presence or absence of *CYP21A1P* in genomic DNA. HLA genotype data for Norwegian AAD patients with and without *CYP21A1P* was made available by Miss Ingeborg Brønstad in Eystein Husebye's group, Bergen, Norway to help in the interpretation of data generated using Alekhya's assay. Dr Steve Lisgo in the Newcastle Human Developmental Biology Resource performed the *CYP21A2/CYP21A1P* tissue *in situ* hybridisation work and I am grateful to him for his expertise, enthusiasm and interest in this project. Thymus tissue for *in situ* hybridisation was collected by Dr Parth Narendaren and deposited in the Birmingham Human Biomaterials Resource Centre. Thanks to Dr Jane Steele from the Birmingham HBRC for her help in obtaining this tissue.

The candidate gene study was made possible through collaboration within the Euradrenal Consortium. Thanks to all consortium members for "voting" for their favourite candidate genes and for providing DNA samples: Eystein Husebye (Bergen, Norway), Dag Undlien and Beatte Skinningsrud (Oslo, Norway); Klaus Badenhoop (Frankfurt, Germany); Corrado Betterle (Padua, Italy) and Alberto Falorni (Perugia, Italy); Sophie Bensing, Anna Lena Hulting (Stockholm, Sweden) and Olle Kampe (Uppsala, Sweden); Anna Kasperlik-Zaluska, Barbara Czarnocka (Warsaw, Poland) and Marta Fichna (Poznan, Poland). DNA samples from rheumatoid arthritis cases and controls were provided by Tony

Merriman (Otago, New Zealand) and type 1 diabetes samples were provided by Beatte Skinningsrud and Dag Undlien (Oslo, Norway). Mandy Phipps-Green (Otago, New Zealand) kindly provided WTCCC data for *GATA3* SNPs.

For the linkage and association study on multiplex AAD families, I wish to thank affected families who came forward to participate in the study through the Addison's Disease Self Help Group. I also wish to thank referring clinicians including Dr Jolanta Weaver (Queen Elizabeth Hospital, Gateshead) and Dr Bijay Vaidya (Royal Devon & Exeter Hospital, Exeter). This part of the study was made possible by collaboration with the Bergen group in Norway. I would like to thank Dr Martina Erichsen for collecting samples and Dr Anette Bøe Wolff in particular for sending DNA samples and for providing detailed phenotype information for all members of the Norwegian multiplex families. I also wish to thank Ms Elisabeth Halvorsen (Bergen, Norway) for performing the 21OH assay on serum samples from UK family members. For the linkage study analysis, Professor Cordell wrote Perl programs to extract relevant data from original data files. For the analysis, I received a huge amount of support and guidance from Professor Heather Cordell and I wish to thank her for her patience. For the genome-wide association analysis, I am grateful to Dr Rebecca Darley in Professor Cordell's group for her help with EMMAX and generating QQ and Manhattan plots in R.

I would also like to thank the other members of the lab group for their help and support over the past three years. Dr Earn Gan for her cheerfulness and for many useful conversations over coffee. I also wish to thank Katie MacArthur for all her help, from designing Sequenom plexes for round one of the candidate gene study to mucking in when there were three dozen 96 well plates to PCR and checking RFLP results "blind" to ensure their validity. I am grateful to her for teaching me some of the laboratory techniques used in this study, for her help on countless arctic expeditions to find samples in the -80ºC freezer and for her general support and encouragement. I would also like to thank other colleagues at the IGM for their help troubleshooting lab problems, for sharing reagents in times of need/desperation, for hints and tips on experimental design and

control, and for their general support and encouragement: Dr Lorraine Eley, Dr Rosalyn Simms, Miss Anne-Marie Hynes, Miss Rachel Challis, Dr Debbie Hicks, Dr Edwin Wong and Dr Holly Anderson.

Finally, I wish to thank my friends and family for their endless support and encouragement, in particular my Dad for proof-reading a long thesis on a subject that he knew nothing about!

LIST OF ABBREVIATIONS

AAD	autoimmune Addison's disease
ACTH	adrenocorticotrophic hormone
AIRE	autoimmune regulator
APECED	autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy syndrome
APS	autoimmune polyendocrinopathy syndromes
AUTS2	autism susceptibility candidate 2
bp	base pairs
C4	complement factor 4
CAH	congenital adrenal hyperplasia
CCDS	Consensus Coding Sequence project
CDCV	common disease-common variant
cDNA	Complementary DNA
CIITA	MHC class II transactivator
CLEC16A	C-type lectin domain family 16, member A
cM	centimorgans
cpm	counts per minute
CRT_H2	chemoattractant receptor-homologous molecule expressed on T_H2 cells receptor
C_T	cycle threshold
CTDP1	carboxy-terminal domain RNA polymerase II, polypeptide A phosphatase, subunit 1
CTLA4	cytotoxic T-lymphocyte antigen 4
CYP2R	vitamin D 25-hydroxylase
CYP21A1P	cytochrome P450, family 21, subfamily A, polypeptide 1
CYP21A2	cytochrome P450, family 21, subfamily A, polypeptide 2

CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1
CYP27B1	cytochrome P450, family 27, subfamily B, polypeptide 1
DAX1	dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, 1 gene
DEC2	basic helix-loop-helix family, member e41; BHLHE41
DEPC	Diethylpyrocarbonate
Dig	digoxigenin
DNA	deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FBXO11	F-box only protein 11
FCRL	Fc receptor-like
FCS	fetal calf serum
FDR	false discovery rate
FOXN2	forkhead box N2
FOXP3	forkhead box P3
GALR1	galanin receptor 1
GATA3	GATA-binding protein 3
GC	vitamin D binding protein
gDNA	genomic DNA
GITR	glucocorticoid-induced TNF receptor-related protein
HDBR	Human Developmental Biology Resource
HDR	hypoparathyroidism, sensorineural deafness, renal abnormalities syndrome
HLA	human leukocyte antigen
HLOD	heterogeneity LOD
HPLC	high performance liquid chromatography
HWE	Hardy-Weinberg equilibrium
iAAD	isolated AAD
IBD	identical by descent

ICOS	inducible T-cell co-stimulator
IFIH1	interferon induced with helicase C domain 1
IFN-γ	interferon gamma
IgG1	immunoglobulin G subclass 1
IκB	inhibitor κB
IL	interleukin
IL17A	interleukin 17A
IL17RA	interleukin 17 receptor A
IL21	interleukin 21
IL23A	interleukin 23 alpha subunit p19
IPEX	immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome
IRF4	interferon regulatory factor 4
kb	kilobase
KCNG2	potassium voltage gated channel subfamily G, member 2
LD	linkage disequilibrium
LNA	locked nucleic acid
LOD	logarithm of the odds
LYP	lymphoid tyrosine phosphatase
MAF	minor allele frequency
MALDI-TOF	matrix-assisted laser desorption ionization-time of flight
Mb	megabase
MHC	major histocompatibility complex
MICA	MHC class I polypeptide-related sequence A
mRNA	messenger RNA
MSH6	mutS homolog 6
NBT-BCIP	nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate
NCBI	National Center for Biotechnology Information

NFATC1	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1
NFATC2	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
NLR	NOD-like receptor
NLRP1	NLR family, pyrin domain containing 1
NMD	Nonsense-mediated mRNA decay
NOD	non-obese diabetic
OD	optical density
OR	odds ratio
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD-L1	programmed cell death 1 ligand 1
PFA	paraformaldehyde
POMC	pro-opiomelanocortin
PPP1R21	protein phosphatase 1, regulatory subunit 21
PTPN22	protein tyrosine phosphatase, non-receptor type 22
qPCR	Fluorescence-based quantitative real-time PCR
QQ	quantile-quantile
REL	v-rel reticuloendotheliosis viral oncogene homolog
RFLP	restriction fragment length polymorphism
RH	Rel homology
RNA	Ribonucleic acid
RORA	RAR-related orphan receptor A
RORC	RAR-related orphan receptor C
ROR-γt	retinoic acid-related orphan receptor γt
RP	serine/threonine nuclear protein kinase

RR	relative risk
RT-PCR	reverse transcription PCR
SAP	shrimp alkaline phosphatase
SALL3	Sal-like 3
SD	standard deviations
SH2	Src homology 2
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
SSC	saline-sodium citrate
STAT2	signal transducer and activator of transcription 2
STAT4	signal transducer and activator of transcription 4
STON1	stonin 1
TBX21	T-box 21
TCR	T cell receptor
TDT	Transmission disequilibrium testing
T_{FH}	follicular helper T cells
TGFβ	transforming growth factor beta
T_H	helper T lymphocyte
TISH	tissue in situ hybridisation
Tm	melting temperature
TNC	tenascin C
TNX	tenascin X
T_{Reg}	Regulatory T cells
TSH	thyroid stimulating hormone (thyrotropin)
UCSC	University of California Santa Cruz
UV	ultraviolet
VDR	vitamin D receptor
wpc	weeks post-conception
WTCCC	Welcome Trust Case-Control Consortium

21OH	21 hydroxylase
3’UTR	3’ untranslated region
8.1 AH	8.1 ancestral haplotype (*HLA-A1-B8-DR3-DQ2*)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

1.1 TOLERANCE AND THE PATHOGENESIS OF AUTOIMMUNITY

1.1.1 IMMUNITY, TOLERANCE AND AUTOIMMUNITY

The immune system has evolved to protect the host from a wide variety of pathogens: this is achieved through a combination of innate and adaptive mechanisms. The innate immune system is concerned with the immediate and rapid defence of the host in a generic and nonspecific manner; it is made up of a number of different cell types and mechanisms. Primarily, its functions include pathogen recognition, immune and inflammatory cell recruitment through the production of chemokines and cytokines, activation of the complement cascade and activation of the adaptive immune system through antigen presentation. The main functions of the adaptive immune system are to recognise foreign antigens presented in the context of major histocompatibility complex (MHC) molecules during the process of antigen presentation, to respond rapidly and specifically to pathogens or to cells harbouring pathogens and to establish immunological memory for future defence. The adaptive immune response has two arms: an afferent, antigen presenting element, consisting of tissue macrophages and dendritic cells, and an effector arm comprising both T and B lymphocytes[1]. The immune system must achieve a fine balance between the need to fight infection and the need to maintain tolerance to self-antigens.

Tolerance refers to the immune system's ability to distinguish between native or self-antigens, which should not induce an immune response, and antigens from pathogens, which should provoke a vigorous immune response. Tolerance can be divided into two distinct mechanisms: central and peripheral tolerance (Figure 1).

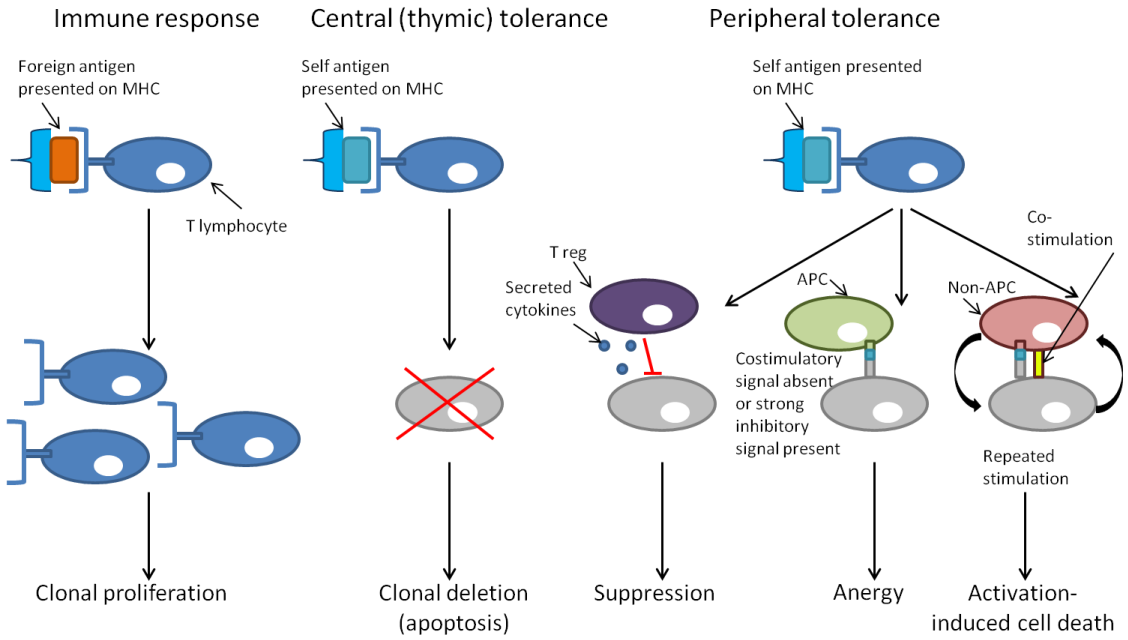


Figure 1: Mechanisms of T cell-mediated immunity and tolerance.

Foreign antigen presented on MHC molecules induces clonal proliferation and an immune response. In the thymus, self-antigen presented on MHC molecules induces clonal deletion (negative selection). Autoreactive T cells escaping to the periphery can be neutralised by one of three mechanisms. Firstly, regulatory T cells (TReg) may suppress their activity, directly through cell to cell contact, or through secreted cytokines. Secondly, encountering self-antigen presented by MHC molecules, but in the absence of a co-stimulatory molecule, induces anergy (hyporesponsiveness). Finally, repeated stimulation from a self-antigen, even in the presence of a co-stimulatory signal, induces activation-induced cell death.

1.1.1.1 CENTRAL TOLERANCE

Both T and B lymphocytes are generated in the bone marrow. B lymphocytes mature *in situ* while T cells must travel to the thymus to undergo maturation. In the thymus, T lymphocytes undergo induction of central tolerance. Initially, they must undergo a process of positive selection whereby T cells which do not recognise self-MHC molecules presented to them by thymic medullary epithelial cells are destroyed. T cells which "pass" the positive selection process then undergo negative selection in the thymic medulla. Here, any T cells with a T cell receptor which responds to self-antigen in the context of MHC molecules (autoreactive T cells) presented by dendritic cells are deleted[2, 3]. The autoimmune regulator (AIRE) protein, a transcription factor encoded by the *AIRE* gene, is thought to be crucial in establishing central tolerance. AIRE is expressed in medullary thymic epithelial cells and allows them to express a wide variety of peripheral self-antigens which would not otherwise be present in the thymus, allowing identification and destruction of autoreactive T lymphocytes[4]. In support of the crucial role of AIRE in induction of tolerance, *AIRE* deficient mice have defective T lymphocyte negative selection and develop organ-specific autoimmune conditions[5], as do humans with mutations in the *AIRE* gene[6]. Even with AIRE allowing expression of a diverse array of antigens, some autoreactive T cells inevitably escape deletion, in particular those with receptors which have a low affinity for self-antigens that are rare in the thymus.

B cells must also learn to be tolerant: this occurs in the bone marrow as they mature. Following the construction of the B cell receptor by heavy and light chain gene rearrangements, each B cell is "tested" to determine whether it is autoreactive. This is similar to the negative selection process which T lymphocytes must undergo. Autoreactive B cells are then thought to undergo receptor editing, whereby further light chain gene rearrangements may occur. Following receptor editing, any B lymphocyte that remains autoreactive is deleted. AIRE is not expressed in the bone marrow, and therefore B cells with

receptors that recognise antigens which are not expressed here may be released into the periphery.

1.1.1.2 PERIPHERAL TOLERANCE

Peripheral tolerance has evolved to control autoreactive lymphocytes which escape deletion centrally and involves a number of mechanisms. Newly released T and B lymphocytes are restricted to circulating through the secondary lymphoid organs, such as the lymph nodes and spleen, where they are most likely to encounter antigen presenting cells and therefore be of use. The self-antigens which are present in abundance in the secondary lymphoid organs are the same as those present in the bone marrow and thymus, therefore the cells likely to react to these antigens will already have been eliminated. If autoreactive T or B cells do leave the secondary lymphoid organs, other peripheral mechanisms are in place to prevent autoimmunity.

Regulatory T cells (T_{Reg}) are a subset of $CD4^+$ T lymphocytes which play an important role in maintaining tolerance in the periphery[7]. Natural and inducible T_{Reg} cells express CD25, CD4, cytotoxic T-lymphocyte antigen 4 (CTLA4) and forkhead box P3 (FOXP3)[8, 9]. Their primary function is to curb both physiological and pathological immune responses. However, the mechanism or mechanisms by which this suppression is achieved are not yet fully understood. T_{Reg} cells are thought to function in a contact-dependent manner, whereby they interact directly with T lymphocytes[10], perhaps allowing inhibitory co-stimulatory molecules such as CTLA4[11, 12] to be engaged. Cytokines such as interleukin-10 (IL10) and transforming growth factor beta (TGFβ) are also thought to play a role in the suppressive functions of the T_{Reg} population, although there is some debate as to whether this is a major or minor role. On the one hand, *in vitro* studies with T cells unable to produce these cytokines have demonstrated that T_{Reg} cells can still function in their absence[10, 13]; conversely, *in vivo* studies in mice suggest that cytokines may play a more important role in T_{Reg} function than originally thought, suppressing both innate and adaptive immunity through secreted factors[14, 15]. The critical role of T_{Reg} cells in maintaining tolerance in the periphery has been illustrated in a number

of animal models. If congenitally T cell deficient BALB/c athymic nude mice are inoculated with T cell suspensions free of $CD25^+$ cells from normal BALB/c mice, they develop autoimmune disease affecting multiple organs including the pancreas, gonads, thyroid and stomach[16]. If $CD25^+$ non-depleted T cell suspensions were inoculated at the same time, autoimmune disease did not ensue [16]. Mutations in the *FOXP3* gene in humans result in the immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome[17], a disease where autoimmunity predominates, further supporting the role of the T_{Reg} subpopulation in tolerance.

In the periphery, in addition to T_{Reg} cells exerting a suppressive effect on autoreactive T cells, they may also be rendered anergic or hyporesponsive if they encounter antigen in the absence of a co-stimulatory signal[18-20]. Classical antigen presenting cells, such as macrophages and activated B cells, present antigen on MHC molecules. A second co-stimulatory signal results in T lymphocyte activation, for example from a B7 molecule which interacts with CD28 on the T lymphocyte's cell surface. Non-antigen presenting cells, for example adrenal cortical cells, will display low levels of self-antigen on MHC but lack co-stimulatory molecules. T cells recognising self-antigen on a non-antigen presenting cell will therefore be rendered anergic by the lack of co-stimulation[21]. In addition, T cells may also be rendered anergic if they receive strongly inhibitory costimulatory signals[22]. B cells also require a second signal to be activated, either from a helper T lymphocyte (T_H), or via pattern recognition receptors which recognise bacterial or parasitic molecular patterns. These anergy-inducing mechanisms reduce the possibility of B cell mediated autoreactivity[23, 24].

Another important mechanism for preserving tolerance is activation-induced cell death. Autoreactive lymphocytes which escape deletion and encounter self-antigen in the periphery will be stimulated again and again, unlike antigens from infectious agents which are dealt with swiftly by the immune response and cleared. This repeated stimulation is recognised as abnormal and results in the death of the autoreactive lymphocyte, a process known as activation-induced

cell death[25, 26]. Finally, some autoreactive lymphocytes may be released but never encounter the native antigen that they recognise, because it is contained within an immunologically privileged site. Such sites are physically isolated from the immune system because lymphocytes are prevented from encountering self-antigens by a physical barrier such as the blood-brain barrier.

If the balance between immunity and tolerance is not achieved and the immune system responds inappropriately to self-antigens, tolerance is lost and autoimmunity arises. Autoimmunity can be subdivided into organ-restricted conditions where one organ, or part of an organ, is the target of the immune response, for example autoimmune Addison's disease (AAD), vitiligo or type 1 diabetes, and systemic diseases where multiple tissues and organs are affected, such as rheumatoid arthritis or systemic lupus erythematosus (SLE). The mechanisms through which tolerance is lost and autoimmunity arises are not yet fully understood. These mechanisms are likely to be complex and multiple, involving both intrinsic features of the affected individual and extrinsic factors, such as genetic factors and environmental influences respectively.

1.2 PRIMARY ADRENAL INSUFFICIENCY

Primary adrenal insufficiency arises if the adrenal glands are destroyed, absent or are unable to function appropriately. Some of the many diverse causes of this problem are outlined in Table 1[27]. Firstly, defects in steroidogenesis can result in primary adrenal insufficiency. The most common example is 21 hydroxylase (21OH) deficiency occurring as a result of mutations in the gene encoding the *enzyme cytochrome P450, family 21, subfamily A, polypeptide 2* (*CYP21A2*). These mutations cause congenital adrenal hyperplasia (CAH), one of the most common autosomal recessive conditions. Second, primary adrenal insufficiency may be caused by genetic defects that result in adrenal dysgenesis, for example mutations in the *dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, 1* (*DAX-1*) gene resulting in congenital adrenal hypoplasia. Finally, destruction of the adrenal glands, for example by infiltration from malignant metastases or amyloidosis, haemorrhage and infarction in disseminated meningococcal septicaemia (Waterhouse-

Friderichsen syndrome) or by autoimmune disease can result in primary adrenal insufficiency.

1.2.1 A BRIEF HISTORY OF ADDISON'S DISEASE

Dr Thomas Addison, a British physician born and raised in Newcastle upon Tyne, first described a condition with characteristic symptoms that include postural hypotension, lassitude, and increased pigmentation associated with pathological changes in the adrenal glands in his work 'On the Constitutional and Local Effects of Disease of the Suprarenal Capsules' in 1855[28]. This condition, now known eponymously as Addison's disease, is a chronic disease due to the destruction of adrenocortical steroidogenic cells which results in primary adrenal insufficiency characteristically with both glucocorticoid and mineralocorticoid hormone deficiencies.

Addison's original description referred to patients with primary adrenocortical failure due to tuberculous infiltration of the adrenal glands. In the developed world today, autoimmune destruction of the adrenals is now the most common cause of Addison's disease[29-32] .Regardless of aetiology, adrenal insufficiency was universally fatal until the 1940s, when the first cortisol precursors were synthesised and given to patients. These compounds revolutionised the management, turning it from a fatal condition into a chronic and manageable disease.

1.3 AUTOIMMUNE ADDISON'S DISEASE

Autoimmune Addison's disease (AAD) is a relatively rare endocrine condition with a prevalence in the Caucasian European adult population of 110–140 cases per million[31, 33, 34], making it 30- and 200-fold less prevalent than type 1 diabetes and autoimmune thyroid diseases respectively. AAD displays a predilection for females, affecting three times as many women as men. Most often, it presents in individuals between the ages of 30 and 50 years, although it can affect people at any age[31]. In AAD, the steroidogenic enzymes within the steroidogenic cells of the adrenal cortex, form the target of a misdirected

immunological attack which leads to adrenal cortex destruction and failure of glucocorticoid and mineralocorticoid hormone production. The primary autoantigen in AAD is the 21-hydroxylase (21OH) steroidogenic enzyme[35] and autoantibodies to 21OH can be detected in 85% of individuals presenting with primary adrenal failure, thus defining them as having AAD[35]. 21OH autoantibodies predominantly have an immunoglobulin G subclass 1 (IgG1) isotype and target the carboxy terminal of this enzyme[36]. Although *in vitro* these antibodies can inhibit the enzymatic activity of 21OH by preventing its interaction with cytochrome P450 oxidoreductase[37, 38] this finding has not been replicated *in vivo*[39].

Category	Examples
Impaired steroidogenesis	1) Defects in cholesterol biosynthesis - Smith-Lemli-Opitz syndrome - Abetalipoproteinaemia 2) Defects in steroid biosynthesis - StAR mutations - Mitochondrial mutations - Mutations in genes encoding steroidogenic enzymes resulting in congenital adrenal hyperplasia - Scavenger receptor BI mutations
Adrenal dysgenesis/ hypoplasia	- DAX1 mutations - SF1 mutations - ACTHR mutations - GPX1 mutations - NNT mutations
Adrenal destruction	- Autoimmune - Metastatic malignancy - Infectious - Amyloidosis - Haemochromatosis - Haemorrhagic - Adrenoleukodystrophy - Sarcoidosis

Table 1: Aetiologies of primary adrenal insufficiency.

StAR, steroidogenic acute regulatory protein; DAX-1, Dosage-sensitive sex reversal-adrenal hypoplasia gene 1; SF1, steroidogenic factor 1; ACTHR, ACTH receptor gene. GPX1, glutathione peroxidase 1; NNT, nicotinamide nucleotide transhydrogenase.

The 21OH enzyme is located intracellularly on the smooth endoplasmic reticulum of intact cells[40] and this location precludes its direct interaction with circulating autoantibodies. Despite their apparent lack of functional activity, the presence of circulating 21OH autoantibodies is a reliable predictor of the development of AAD, since they indicate an ongoing autoimmune process within the adrenal glands. The proportion of individuals with idiopathic Addison disease in whom 21OH autoantibodies cannot be detected but who have disease with an underlying autoimmune aetiology, for example individuals with an aberrant immune response to other adrenal antigens, remains unknown. However, in addition to 21OH, other autoantigens have been identified in patients with AAD, including steroid 17-α-hydroxylase[41] and the cholesterol side-chain cleavage enzyme[42]. It is possible that those individuals who appear to have an autoimmune aetiology underlying their Addison's disease have other autoantibodies not routinely measured. Biochemically, Addison's disease results in reduced or absent glucocorticoid (namely cortisol) production with a compensatory increase in pituitary adrenocorticotrophic hormone (ACTH) secretion, due to a disruption in the hypothalamic-pituitary-adrenal axis feedback loop (Figure 2). It also results in reduced or absent mineralocorticoid (namely aldosterone) production. Mineralocorticoid deficiency causes a fall in blood pressure and postural hypotension, which stimulate renal renin secretion.

1.3.1 CLINICAL PRESENTATION OF AAD

As would be predicted, the clinical signs and symptoms associated with AAD correlate closely with glucocorticoid and mineralocorticoid deficiency. Affected individuals often present following a protracted period of ill-health and complain of non-specific symptoms including weight loss, lethargy, postural dizziness and nausea. Individuals will often develop increasing skin pigmentation, often considered the clinical hallmark of Addison's disease. This is noted particularly in the skin flexures, scars and at areas of minor friction such as the knees and elbows. The increased pigmentation is widely thought to result from high levels of melanocyte-stimulating hormone, produced from the cleavage of ACTH from its precursor, pro-opiomelanocortin (POMC). However, anecdotally, we have noted that patients treated with synthetic ACTH also become pigmented, and

therefore this phenomenon may be directly due to the ACTH itself. Biochemically, affected individuals are often noted to be hyponatraemic and hyperkalaemic due to mineralocorticoid deficiency. A short synacthen test, whereby a bolus of synthetic ACTH analogue (synacthen) is given and the cortisol response measured at 30 and 60 minutes, reveals a subnormal response, while the ACTH level is often markedly elevated.

Addisonian crisis, characterised by severe hypotension, hypoglycaemia, electrolyte disturbances, coma and sometimes death, may occur, often during a relatively minor concurrent illness, if the insidious symptoms and signs go unrecognised.

Individuals diagnosed with AAD are dependent lifelong upon glucocorticoid and mineralocorticoid replacement and they face life-threatening consequences should these be omitted. Despite adequate replacement therapy, affected individuals have a poorer quality of life when compared to unaffected individuals[43], increased morbidity, resulting mainly from steroid-related osteoporosis and type 2 diabetes, and reduced life expectancy compared to their peers[44].

1.3.2 THE NATURAL HISTORY OF AAD

A series of phases have been proposed to occur in the natural history of 21OH autoantibody-positive AAD (Figure 3)[45]. During the earliest (potential) AAD phase, autoantibodies are present, but parameters of adrenal function remain entirely normal and no clinical features of disease are present. The subclinical phase follows, in which adrenal function gradually declines but clinical symptoms of AAD remain absent. Typically, during this phase, an initial rise in plasma renin and/or ACTH levels is evident before basal, and then stimulated, circulating cortisol concentrations become subnormal (<550nmol/l after synacthen stimulation) in a short synacthen test. Finally, clinical AAD develops. It is only at this stage that the affected individual becomes symptomatic, often with fatigue and pigmentation at first.

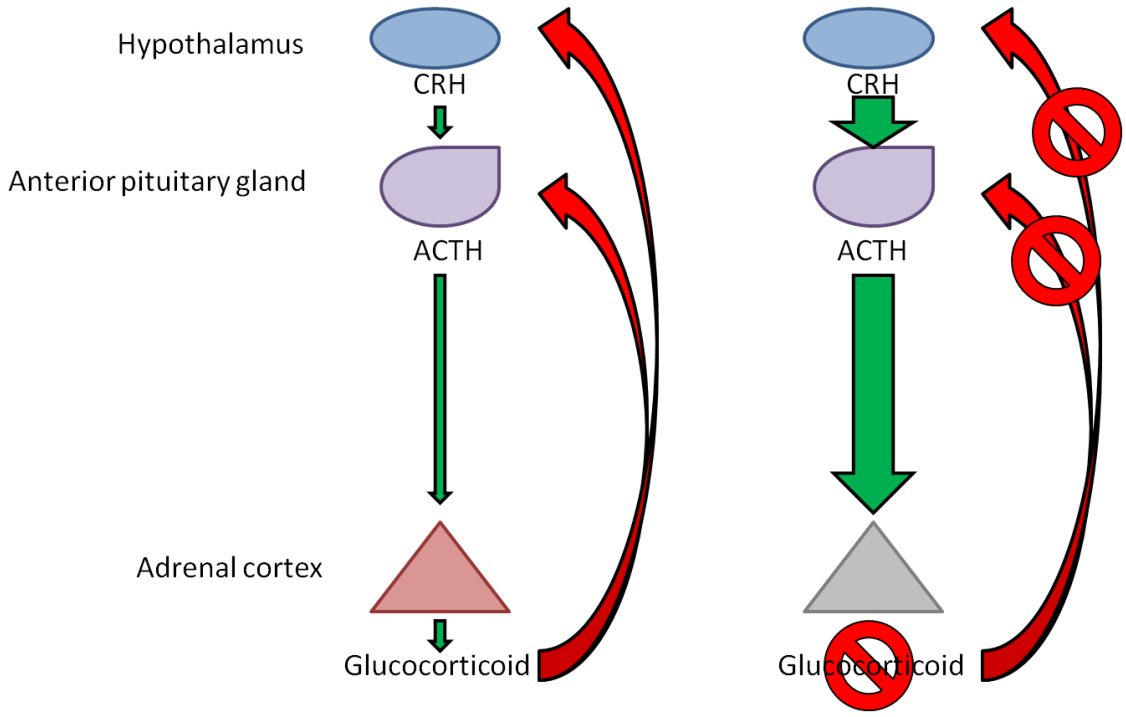


Figure 2: Simplified schematic of the hypothalamic-pituitary-adrenal axis.

Schematic simplification of the hypothalamic-pituitary-adrenal axis. Corticotropin-releasing hormone (CRH) is secreted from the hypothalamus and induces secretion of adrenocorticotrophic hormone (ACTH) from the anterior pituitary. This is turn results in glucocorticoid secretion from the adrenal cortex. Adequate glucocorticoid secretion prevents further ACTH and CRH secretion from the pituitary and hypothalamus respectively through a negative feedback loop (red arrows). If the adrenal cortex is destroyed, glucocorticoid secretion fails and the negative feedback is lost, resulting in increased CRH and ACTH secretion.

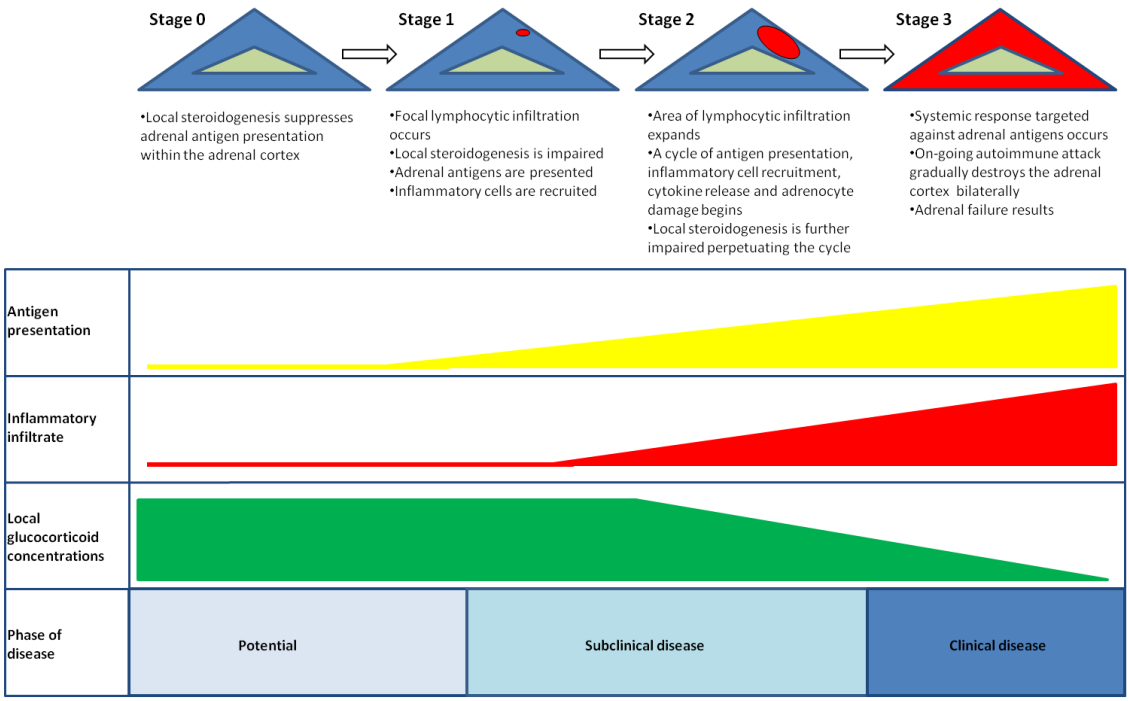


Figure 3: Glucocorticoid-induced immune privilege.

A hypothetical model of the pathogenesis of autoimmune Addison's disease (AAD) based on the breakdown of glucocorticoid-induced immune privilege. Gradual immune-mediated destruction of the adrenal gland occurs in patients with AAD. Stage 0 corresponds to the "potential" phase in the natural history of this disorder. Once tolerance to adrenal antigens is lost (stage 1), increased antigen presentation leads to increasing inflammatory infiltrates in the gland. Adrenocytes are damaged by this immune response and, in stage 2, local glucocorticoid production is impaired. Stages 1 and 2 correspond to the "subclinical" phase of AAD. Eventually, in stage 3, the adrenal cortex is destroyed and steroidogenesis ceases, which corresponds to the phase of "overt" clinical AAD[1].

In a clinical study, 31% of initially asymptomatic individuals with steroid 21OH autoantibodies developed AAD. There was marked variation in the time taken to progress from autoantibody positivity to overt adrenal failure, with the duration ranging from 3 months up to 11 years in different patients[46]. Moreover, not all autoantibody positive individuals went on to develop progressive disease: some remain in the potential phase, never developing AAD, and some revert to being antibody-negative[45, 46]. It is not currently clear what causes some people to progress to disease development while others clear their autoantibodies. However, it does appear that patients with high titers of 21OH autoantibodies tend to progress more rapidly than those with low titers, perhaps indicating a more aggressive immune response. A case describing a woman with a 9 year history of hyperpigmentation, raised circulating ACTH levels and high steroid 21OH autoantibody levels but a normal cortisol response to synacthen, highlights that individual responses to the presence of an ongoing autoimmune response are highly variable and rather unpredictable[47]. Similarly, case reports of two patients with remission of apparently established AAD[48, 49] and another with subclinical AAD and 21OH antibodies that became undetectable following glucocorticoid therapy[50, 51] suggest that even in patients with impaired steroidogenesis, adrenal autoimmunity might not always be permanent.

1.3.3 PATHOLOGICAL FINDINGS IN AAD

In AAD, the autoimmune destruction of the adrenal cortex is evident both macro- and microscopically, and is thought to be mediated by both the cellular and humoral arms of the immune system.

Macroscopically, individuals with AAD have small, atrophic adrenal glands identified by high-resolution imaging modalities or *post-mortem* examination[52]. Microscopically, histological examination of the adrenals of individuals with AAD reveals that the autoimmune process affects the entire cortex, from the outer zona glomerulosa, through to the inner zona reticularis. Cellular infiltrates are identifiable, consisting of lymphocytes, eosinophils, macrophages and plasma cells[53].

Cellular immunity is thought to be integral to the autoimmune destruction that occurs in AAD. In particular, T lymphocytes, identifiable in the cellular infiltrates seen histologically in the adrenals of individuals with AAD, are thought to play a crucial role in starting, and then perpetuating, the autoimmune process through their two key functions: cytokine release and cell lysis and destruction. As 21OH autoantibodies are common at the time of diagnosis of AAD[54, 55], humoral immunity is also thought to play a role in its pathogenesis. The enduring mystery, in AAD and other autoimmune conditions, is the immunological mechanism by which immune tolerance is lost in the first place.

1.3.4 ISOLATED AAD, APS1 AND APS2

AAD may occur in isolation or as part of one of the autoimmune polyendocrinopathy syndromes (APS).

APS type 1 (APS1, also known as autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy syndrome; or APECED) is a rare, monogenic, autosomal recessive syndrome resulting from mutations in the *AIRE* gene located on chromosome 21[6, 56, 57]. APS1 has a prevalence of only three per million in the UK population[58]. However, it occurs more frequently in certain groups, including Iranian Jewish (1:9,000)[59], Sardinian (1:14,000)[60], Finnish (1:25,000)[61] and Norwegian populations (1:90,000)[62]. Affected individuals usually present in childhood with chronic mucocutaneous candidiasis, adrenocortical failure and autoimmune hypoparathyroidism. They may also have associated dental enamel hypoplasia or nail dystrophy (ectodermal dystrophy)[63-65] and may go on to develop other autoimmune disorders later in life including type 1 diabetes, pernicious anaemia and primary gonadal failure[63].

APS type 2 (APS2) is much more common than APS1 and accounts for approximately half of cases of primary adrenal failure[66]. It is defined as AAD in conjunction with autoimmune thyroid disease and/or type 1 diabetes[63]. Other autoimmune conditions may also arise in individuals with APS2, including vitiligo, pernicious anaemia and primary gonadal failure[63]. APS2 is more

common in women than men and often presents in the fourth decade of life[55, 67] in marked contrast to APS1.

1.3.5 THE GENETIC BASIS OF ISOLATED AAD AND APS2

The strong genetic component to AAD susceptibility has long been recognised, with multiple reports of concordant monozygotic twins[68-71]. Familial clustering of cases of AAD also suggests a genetic basis[72, 73]. Pooled data from studies comprising more than 600 unrelated non-APS1 AAD probands from Norway and the UK have revealed only 14 families (2.3%) that contain affected AAD sibling pairs. From this data, a sibling recurrence risk ratio (that is, the risk of disease in a sibling of an affected individual, divided by the disease prevalence, also expressed as lambda-sibling, λs) can be estimated by using the formula[74]:

λs = sibling risk (%) / population prevalence (%)

Therefore, if the population prevalence of AAD is between 0.011% (110 per million)[31] and 0.014% (140 per million)[34], the λs for AAD is estimated to be between 160–210[1]. This is in contrast to type 1 diabetes, Graves' disease and Crohn's disease for example, which have a λs of 15, 10–15 and 17–35 respectively[75-78]. These epidemiological data therefore suggest that AAD is the complex autoimmune disorder with the highest genetic load. In addition, individuals with AAD will commonly report family members with other autoimmune conditions[79] and approximately 50% of individuals with AAD will have, or will go on to develop, other autoimmune conditions themselves, most commonly Hashimoto's thyroiditis and type 1 diabetes[32]. In a collection of 22 multiplex AAD pedigrees from the UK and Norway, 24 of the 48 individuals with AAD (50%) had an additional autoimmune condition. 14/48 (29%) had an autoimmune thyroid condition and 7/48 (15%) had type 1 diabetes. Within these pedigrees, 80 AAD-free relatives were also studied and 12/80 (15%) had an autoimmune condition, such as autoimmune thyroid disease, pernicious anaemia, vitiligo and type 1 diabetes. The clustering of multiple autoimmune conditions within individuals with AAD, and within families of those with AAD, suggests common susceptibility loci for these disorders.

Isolated AAD, and AAD as a component of the APS2 syndrome, clearly have a strong genetic basis. However, these conditions are not inherited in a Mendelian fashion. Instead, they have a multifactorial or complex aetiology, in common with other autoimmune conditions such as type 1 diabetes and rheumatoid arthritis[80].

1.4 GENETIC ANALYSIS OF COMPLEX DISEASES

Like most autoimmune endocrinopathies[80], isolated AAD and APS2 have a complex genetic aetiology, which implies the interaction of genetic and environmental factors in a susceptible individual resulting in disease. These interactions are unpredictable, with genotype not always predicting phenotype, and *vice-versa*[81]. A number of factors may complicate the genetic analysis of complex conditions such as AAD. Incomplete penetrance means that an individual inheriting a known disease susceptibility allele may not manifest the disease, while an individual without any known disease susceptibility alleles may develop the disease for other reasons such as environmental factors or simply by chance, an occurrence called phenocopy. Mutations in any one of several genes may result in the same phenotype (genetic or locus heterogeneity), thus patients with the same disease may have entirely different genetic reasons for this. In addition, some conditions exhibit polygenic inheritance, where multiple mutations in different genes are required to produce the phenotype. Other conditions are difficult to map due to a high frequency of alleles linked to a condition in the general population. Finally, diseases may be transmitted by alternative mechanisms, for example through the mitochondrial genome, which might not be explored in a traditional genetic study[81].

Traditionally, there have been two approaches to identifying complex disease loci: linkage and association studies. Both approaches can be performed on a “hypothesis-driven” basis, using a candidate gene approach, or on a “discovery-driven” basis, using a genome-wide approach.

1.4.1 LINKAGE ANALYSIS

Linkage analysis aims to test for co-inheritance of chromosomal regions with a trait or disease within pedigrees[82]. Linkage analysis is based upon the principle that genes located on the same chromosome, in close proximity to one another, are likely to co-segregate, or be inherited together, because the likelihood of a recombination (crossover) event at meiosis is low. Specifically, two genetic loci are said to be linked if they are transmitted together from parent to offspring more often than expected under independent inheritance. Therefore, if a polymorphic marker is in close proximity to, and in linkage with, a disease-causing gene, the alleles of the polymorphic marker will co-segregate with the disease in affected families[83]. In the population as a whole, the non-random association of alleles at two or more loci, whereby they are found on the same haplotype more often than expected by chance, is termed linkage disequilibrium (LD).

Linkage studies were traditionally undertaken using microsatellites markers (sometimes referred to as short tandem repeats or simple sequence repeats), however single nucleotide polymorphisms (SNPs), polymorphic markers distributed throughout the genome, are now more commonly used for linkage studies[84]. SNPs account for the majority of genetic variation between individuals. A common SNP occurs at a frequency of greater than 1% in the general population, and each individual harbours approximately ten million common SNPs and numerous rarer variants, some of which may be “private” to them[85]. In linkage maps, distances between markers are expressed in centimorgans (cM), where one cM is defined as that distance between loci for which 1% of the products of meiosis are recombinant, crossover events. One cM corresponds to approximately one million base pairs (bp) in humans[86]. Linkage is reported as a logarithm of the odds (LOD) score. The LOD score is the logarithm of the likelihood of the odds that two loci are linked compared with the likelihood of the odds for independent assortment (i.e. that the two loci are unlinked). By convention, a LOD score >3, which indicates 1000 to 1 odds that the observed linkage is not a chance observation[87], in Mendelian disorders is considered strong evidence for linkage. A LOD score of -2 is used to reject

linkage[88]. For complex traits, more stringent significance levels have been suggested to reduce the type 1 (false positive) statistical error rate. For example, a LOD score of >3.3 might constitute definite evidence for linkage while a LOD score of 1.9 to 3.3 would be suggestive of linkage only[89]. In addition, if the heterogeneity LOD (HLOD) score is reported, which allows for locus heterogeneity in a linkage analysis, it has been suggested that a correction should be applied by deducting 0.3 from the HLOD score to give a more conservative estimate of linkage i.e. that a reported HLOD score of 3.0 be corrected to a score of 2.7[90]. Critics of this more stringent approach suggest that using such rigorous significance levels results in difficulty replicating identified loci[91, 92].

Linkage analysis can be performed using a parametric (model-based) approach, where the inheritance pattern of a disease is tested against a specific model, or a non-parametric (model-free) approach, where an inheritance pattern is not assumed.

1.4.1.1 PARAMETRIC LINKAGE ANALYSIS

Morton was the first to use parametric linkage analysis in his seminal work of 1956[93]. Parametric linkage analysis requires the mode of disease inheritance (disease model, for example recessive, dominant or co-dominant)[94], disease gene frequency and genotype penetrance for the disease of interest to be specified. This method has proved powerful in identifying susceptibility loci for monogenic disorders in large pedigrees, where the pattern of inheritance is clear, but is highly sensitive to model misspecification[94]. However, the very nature of complex genetic disorders means that the parameters that must be specified in order to perform parametric linkage analysis are often poorly defined. In addition, large pedigrees with multiple generations displaying a uniform mode of inheritance are not always available, particularly in late-onset diseases. In practice, to circumvent some of these issues, a number of analyses are often performed under different models. The researcher must then accept that the models represent an approximation of the real situation. Performing multiple analyses with different models introduces further potential for multiple

testing bias which should be considered in the final analysis. For these reasons, the parametric linkage method has proved less powerful for the investigation of complex genetic disorders to date. A non-parametric approach is often favoured, which circumvents the problem of selecting the correct model.

1.4.1.2 NON-PARAMETRIC LINKAGE ANALYSIS

Non-parametric linkage analysis does not require prior specification of a disease model. The rationale underlying non-parametric linkage analysis is that, within affected pedigrees, there will be excess sharing of haplotypes, greater than would be expected by chance, that are identical by descent (IBD), that is copies of the same ancestral allele, in the region of the gene causing disease. Non-parametric linkage is most commonly performed on affected sibling pairs[95, 96], but has also been adapted to other relationships within affected pedigrees[97-99]. It cannot be used for parent-offspring affected pairs as these always share one allele IBD.

1.4.1.3 LIMITATIONS OF LINKAGE STUDIES

Linkage studies do have a number of important limitations which should be considered when planning a study and interpreting the results. Linkage analysis has limited resolution, with disease causing alleles potentially a considerable physical distance from a linkage peak generated in a study. In addition, linkage peaks generated can be large, containing numerous genes. Linkage analyses should therefore be considered the first step in the genetic investigation of a disease, with further fine mapping studies often required following the initial linkage analysis. Denser linkage marker maps are now more widely available to increase resolution of studies. However, the use of more markers increases the time taken for computer analysis and results in the production of vast amounts of complex data. Genetic data for linkage analysis must be carefully quality controlled to minimise the potential for detecting a spurious finding. Using a stringent significance level goes some way to reducing false positive rates. To minimise these further, larger cohorts can be examined, or additional independent data sets can be tested as replication cohorts where these are

available. Results should be corrected for multiple testing which can further inflate false positive rates[100]. Unfortunately, in rarer complex diseases such as AAD, where affected pedigrees are uncommon, generating an independent data set can be challenging. In addition, AAD is a disease which most commonly affects Caucasian middle aged people, therefore the pedigrees are often relatively small, often with only two affected members, and parents may be missing from the data set because they are deceased. In addition, linkage cannot always successfully identify genes exerting modest effects[101] and this should be considered when investigating complex diseases where multiple loci may be contributing in modest ways to susceptibility.

1.4.2 ASSOCIATION STUDIES

Association studies have the aim of demonstrating a relationship between genetic variation at a locus and phenotypic variation. They may be population-based case-control studies or intra-familial, where affected family members constitute the cases and unaffected family members constitute the controls. In common with linkage studies, SNPs are the most commonly used markers. The development of large SNP databases such as HapMap (http://www.hapmap.org)[102], has greatly facilitated the design of association studies, allowing the most informative SNPs to be selected for analysis, and allowing marker data for other SNPs to be "imputed" according to LD patterns i.e. genotyping one SNP (a tag SNP) allows researchers to predict the genotypes of other SNPs in LD with the tag SNP.

1.4.2.1 POPULATION-BASED CASE-CONTROL ASSOCIATION STUDIES

A population-based case-control association study compares the frequency of an allele in affected people compared to its frequency in controls in the population. If the control cohort is unaffected by the disease and is unrelated, and matched ethnically, to cases, a significant difference in allele frequency between cases and controls suggests that the allele is influencing disease susceptibility, or is in LD with a susceptibility locus. The factor by which an individual's risk of disease must be multiplied for each copy of a particular risk

allele possessed is called the (allelic) relative risk (RR). The odds by which an individual's risk of disease must be multiplied for each copy of a particular risk allele possessed is reported as the (allelic) odds ratio (OR).

1.4.2.2 INTRA-FAMILIAL ASSOCIATION STUDIES

Intra-familial association studies are designed to look at association within families, using unaffected family members as an internal control group. This avoids population stratification, where cases and controls exhibit allele differences due to ancestry and ethnicity, rather than due to true disease association. Transmission disequilibrium testing (TDT)[103] is the most commonly used analysis method and requires trios (affected offspring and their parents). It relies on tracking the transmission of alleles from parents to offspring. At least one parent must be heterozygous for the allele of interest. If a specific allele is transmitted to affected offspring more often than expected, the allele may be causing the disease directly, or may be linked to a locus causing disease. The need for living parents in a pedigree is nonetheless problematic in later-onset diseases like AAD. Therefore, to circumvent this issue, a method using affected and unaffected sibling pairs has been developed[104].

1.4.2.3 LIMITATIONS OF ASSOCIATION STUDIES

In contrast to linkage studies, association studies are suitable for investigating genes with modest effects and for fine-mapping after an initial linkage analysis, however positive results have often proved difficult to replicate[105-109]. This may be because either the study is underpowered due to small cohort sizes, or because the selected cohort is not a well-defined group due to poor definition and characterisation of the disease or phenotype under investigation. Using large cohorts with strictly defined diagnostic criteria minimises these problems. Another reason why some studies have not been replicated is because the initial published report may have been a false positive result. Small studies should therefore be thought of as "hypothesis-generating" rather than definitive, and should be regarded as such until larger studies confirm or refute their findings[110]. In addition, genotyping errors may lead to problems with replicating

positive findings[111, 112]. These issues can be minimised by discarding genotyping data if the genotyping call rate (the percentage of samples studied yielding a genotype) is low, checking that control data sets are in Hardy-Weinberg equilibrium (HWE) and re-genotyping a proportion of samples to ensure genotyping accuracy. Finally population stratification may result in spurious association: in particular, if multiple ethnic groups are analysed as one in order to increase cohort size[113]. Controls should be carefully matched to cases to avoid this, and ethnicity considered when analysing different population groups together. Although challenging, replication of results and meta-analysis are essential to ensure generation of robust data sets[114].

1.4.3 HYPOTHESIS-DRIVEN VERSUS DISCOVERY-DRIVEN RESEARCH: CANDIDATE GENE VERSUS GENOME-WIDE STUDIES

Both linkage and association studies can be performed on a candidate gene basis, where a plausible candidate gene is chosen and markers in and around that gene are selected (the "hypothesis-driven" approach), or on a genome-wide basis, where no prior hypothesis is required and markers are distributed evenly throughout the genome (the "discovery-driven" approach).

1.4.3.1 CANDIDATE GENE STUDIES

A hypothesis-driven, candidate gene approach has been used to investigate a number of complex diseases to date. Candidate genes are selected based upon a hypothesis that the gene might be important in susceptibility to the disease of interest. There are approximately 23,000 genes in the human genome; therefore, candidates must be carefully and thoughtfully selected. A number of criteria have been suggested to aid the selection of a suitable candidate[115].

Firstly, candidates can be selected based upon biology, whereby a gene is chosen for investigation based on the knowledge that the protein that it encodes is implicated in some way in the biology of a disease. For example, the 21OH enzyme is a key enzyme in the steroidogenesis pathway and the major autoantigen in AAD, therefore the gene which encodes this enzyme,

cytochrome P450, family 21, subfamily A, polypeptide 2 (CYP21A2) would be an example of a plausible candidate gene in AAD based on biology.

Candidates may also be selected based upon studies of animal models of disease. The orthologue of a gene influencing a disease of interest in animals may be investigated in humans. For example, if a gene defect resulting in Addison's disease in another species was identified, the human orthologue could be investigated as a potential candidate in AAD.

Rare monogenic variants of a disease can also provide insights into disease aetiology and provide possible candidate genes for investigation. This approach has been used in AAD, where the role of *AIRE* gene mutations has been investigated in sporadic AAD after the discovery that these mutations result in the rare, autosomal recessive APS1[6]. Interestingly, *AIRE* mutations have not been found to contribute to AAD susceptibility[116], despite being causative in APS1.

Finally, data gathered from genome-wide or candidate gene studies into related diseases, in particular autoimmune diseases, can also be used to generate plausible candidates. In AAD, the *protein tyrosine phosphatase, non-receptor type 22* (*PTPN22*) gene was investigated after mutations at the locus were found to confer susceptibility to type 1 diabetes[117] and rheumatoid arthritis[118].

1.4.3.2 LIMITATIONS OF CANDIDATE GENE STUDIES

Candidate gene association studies are relatively cost-effective, as they focus on specific loci within the genome. In addition, when positive findings are reported, these should, in theory at least, have a meaningful interpretation given that the candidate approach is hypothesis based. Unfortunately, our understanding of the biology of many diseases is incomplete which makes selecting suitable candidates challenging. In addition, our understanding of genetics is also evolving. We now know that regions of deoxyribonucleic acid (DNA) outside of the exons, once considered "junk", actually contain important

regulatory elements, such as microRNAs. If the traditional approach to candidate gene studies is followed, whereby only protein-coding genes are chosen for investigation, important aetiological variants might be missed. Genome-wide studies have therefore been proposed as a promising alternative.

1.4.3.3 GENOME-WIDE STUDIES

High-throughput genotyping and sequencing technology is developing rapidly and allows the researcher to scrutinise the whole genome for susceptibility loci. It is a truly discovery-driven approach, needing no prior hypothesis other than that the disease or trait of interest has a genetic aetiology. Currently, there are three ways to execute this approach. The first is to genotype markers, for example SNPs, spread throughout the genome, often using an array-based platform. The second is to sequence the protein-coding DNA sequence, which many people consider to be the functionally important DNA, known as whole exome sequencing and the third is to sequence the whole genome. While the latter two genome-wide approaches circumvent the need to select suitable candidates for analysis, they do provide their own challenges.

1.4.3.4 LIMITATIONS OF GENOME-WIDE STUDIES

The technology for performing genome-wide studies, either array-based or sequencing, is developing at a rapid rate. Although costs are falling, it still depends on “high-tech” manufacturing processes and is therefore costly, sometimes making large-scale studies prohibitively expensive. Whole genome sequencing is currently the most expensive of the three options.

All three platforms generate vast quantities of data which must be manipulated, analysed and stored. The generation of such enormous quantities of data is a double-edged sword. While data can be analysed and meaningful results obtained, large amounts of data mean that false positive results may be gathered, or a biased analysis might result in incorrect conclusions being drawn from the data available. In addition, interrogating the entire genome at once raises the issue of how to deal with multiple testing bias.

Whole genome sequencing generates the most data and realistically, per individual sequenced, only a tiny proportion of the information gathered will be relevant to the disease under investigation. In addition, each Caucasian individual studied is estimated to harbour, on average, between 20,000 and 25,000 coding variants, of which up to 10,000 will be non-synonymous[119, 120]. Individuals from "older" populations, such as Africa, may harbour significantly more, up to 70,000[121]. Sorting through these data to determine which aspects are relevant to the disease of interest is time-consuming. There are methods and algorithms to help with the process of distinguishing the pathogenic from the non-pathogenic variants, such as ANNOVAR[122], which annotates sequencing data and guides the user as to which variants are predicted to be functional and which are not. However, many of these algorithms exclude common polymorphisms on the basis that, if they arise in the healthy population, they are not likely to be contributing to disease. However, in complex diseases, it is possible that relatively common polymorphisms can be contributing to the phenotype. This can also be the case in monogenic conditions, such as in the autosomal recessive condition cystic fibrosis, where between 2 and 3% of healthy Caucasians are heterozygous for the most common causative mutant allele *ΔF508*[123]. This would be considered a common variant in some algorithms and therefore excluded from further analysis despite being pathogenic when homozygous. The algorithms are adapting all the time to circumvent these issues, but are not yet perfect. In addition, the databases that contain the information about common variants, such as dbSNP (the single nucleotide polymorphism database - http://www.ncbi.nlm.nih.gov/projects/SNP[124]) and the 1000 Genomes Project (http://www.1000genomes.org[125]) are not infallible. These databases are constantly being updated and corrected. This means that the interpretation of genome-wide data, in particular whole genome sequencing data, may be different tomorrow than it is today. Whole exome sequencing has the advantage of being more focussed than whole genome sequencing and it generates less data in comparison. However, it also has the disadvantage of potentially missing an important variant in the 99% of DNA sequence that is not protein-coding. In addition, there are some technical difficulties with this approach that

current technologies are trying to rectify: for example, difficulty in sequencing CG-rich DNA sequence[126].

The analysis of genome-wide data is particularly challenging. Computer programs designed for this purpose have developed in parallel to high-throughput genotyping and sequencing technologies. However, it is difficult to select the appropriate program for your data set, format the data that you have in order to use your program of choice, and select the appropriate analyses to perform. In addition, if a large data set is analysed, the computer analyses can be very time-consuming. Genome-wide data analysis, in general, requires the input of a statistical geneticist to ensure that the results gained are meaningful and robust. The time and expertise taken to analyse genome-wide data, and in particular whole genome sequencing data, has led to the notion of the "$1,000 genome, $100,000 analysis"[127]. Following data analysis, the positive findings that are discovered in these studies must then be put into context and their significance in the disease of interest must be explored.

In addition to the logistical issues encountered when using these technologies, genome-wide genotyping and sequencing technologies raise a number of ethical and legal considerations. The interpretation of information gained from genome-wide studies, on an individual basis, is challenging. When used for diagnostic purposes, how do you gain informed consent from an individual for a complex test which might reveal numerous deleterious mutations, not all related to the condition for which they have had the test? For example, a healthy individual is estimated to be a heterozygous carrier of up to 100 highly penetrant deleterious variants[128]. Although this is likely to have no consequence for their own health, it may impact upon future children. Do all individuals who have whole genome sequencing therefore need genetic counselling regarding the possible impact of these variants on offspring? The possible use of personal genetic data must also be considered. Who should be allowed to access the information about an individual's genome? If it should not be publicly accessible, how can these data be stored securely?

Perhaps most importantly, we still know relatively little about what the protein-coding portions of our DNA do, how many proteins function and the impact that mutations, and even polymorphisms, have on the structure and function of these proteins. We know even less about what the non-protein coding regions of our DNA do. Cynics could say that genome-wide genotyping and sequencing technologies represent an example of advances in technology driving the research agenda, rather than *vice-versa*. However, the development of these technologies could also be viewed as an opportunity to fill in some of the gaps in our knowledge and to drive forward some of the basic science research needed to achieve a better understanding of human genetics and genomics. Regardless of the current challenges, genome-wide genotyping and sequencing provide extraordinary insight into our genetic makeup. The technical and logistical issues that arise will eventually be overcome by ever improving technologies and any disadvantages will hopefully be outweighed by the advantages gained from these exciting advances.

1.4.4 CONCEPTUALISING COMPLEX DISEASE GENETICS

In complex diseases to date, despite intensive research, a large proportion of the estimated heritability has not been accounted for, despite genome-wide association scans in large case-control cohorts. This has led to a change in the way that researchers conceptualise complex disease genetics[129], for both common and rare conditions, and new approaches are being developed as a consequence.

Originally, as genome-wide studies gained in popularity, the “common disease-common variant” (CDCV) hypothesis predominated[130-132]. This model states that common complex disease, for example hypertension, can be explained by a relatively small number of common variants, with each variant explaining a moderate proportion of disease risk (Figure 4). However, results from powerful genome-wide scans in common diseases have failed to demonstrate findings in line with this model and it has largely lost favour. When conceptualising complex disease genetics, three theories currently predominate, reviewed by Gibson[129]. The first is the infinitesimal model, which proposes that hundreds or

perhaps even thousands of common variants each contribute a tiny amount of risk to disease susceptibility or indeed to continuous traits such as height and weight[133, 134]. In this model, the associated loci observed in genome-wide studies are those with the largest effect sizes and they represent the tip of the iceberg. The second model is the rare allele model, which states that complex diseases are due to a large number of moderately penetrant rare variants[135, 136]. These alleles will also be present in the healthy population, but because they are not fully penetrant, and due to poorly understood modifying influences such as epigenetic and environmental factors, these individuals will not express the disease. The final model is the broad sense heritability, or “G x E”, model, which suggests that a combination of genomic, environmental, and epigenetic factors in a susceptible individual influence disease[137, 138]. Any of these three models could offer an insight into the genetic aetiology of AAD and need to be considered when planning genetic studies. It is also possible that reality lies somewhere between, or in a combination of, these models, with common and rare variants influencing AAD disease susceptibility in the context of environmental and epigenetic pressures[129].

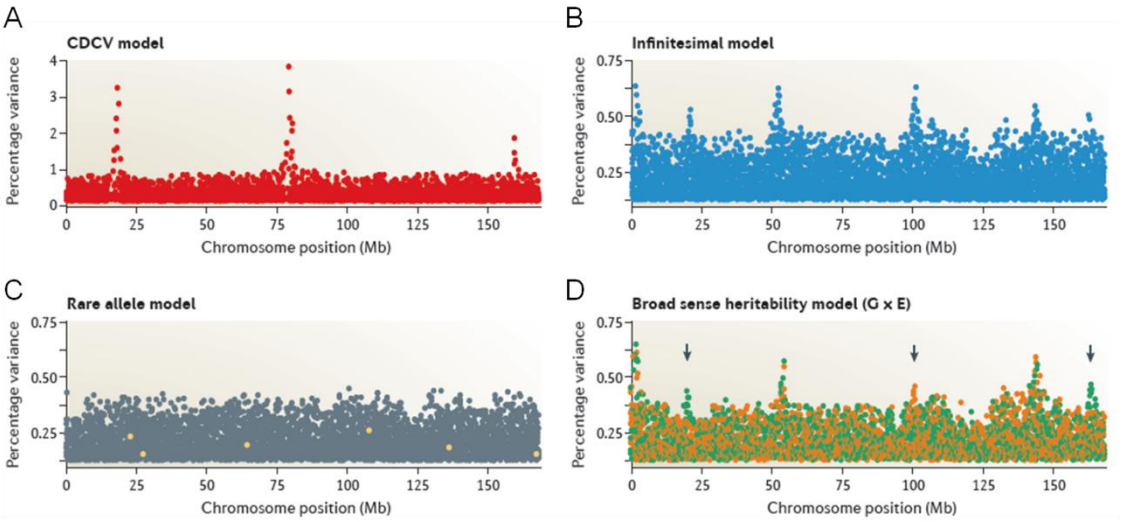


Figure 4: Different expected signatures from genome-wide association studies for four models of disease.

Each of the four plots shows the expected distribution of SNP effects, expressed as the percentage of variance for a disease or trait on the y axis and the position of SNPs on the x axis. In panel A, the common disease-common variant (CDCV) model, a small number of moderate effect loci would produce strong signals, each explaining several percent of the genetic variance. The infinitesimal model is shown in panel B. Many signals are seen, each explaining only a very small amount of the variance. In panel C, the rare allele model, causal variant effects are shown by yellow dots. These may be contributing a significant percentage of the variance in a few individuals, but are not common enough to result in genome-wide significance. In the broad sense heritability model shown in panel D, signals are seen in only certain environments, with different associations seen in the two environments (orange and green dots). When a mixed population, exposed to one or the other environments is studied, the overall effect at each locus is reduced (shown by the arrows) and fewer associations will be observed. Reproduced with permission of the authors from Nature Reviews Genetics “Rare and common variants: twenty arguments”[129].

1.5 GENETICS OF AUTOIMMUNE ADDISON'S DISEASE

1.5.1 ANIMAL MODELS IN AAD

Animal studies can provide useful genetic information because breeding and environment can be carefully controlled and manipulated[81]. However, to date, a faithful experimental animal model for AAD has been elusive. Researchers have successfully induced allergic adrenalitis in a range of laboratory animals, including mice and rabbits, and have observed lymphocytic infiltrates within the adrenal glands of affected animals and autoantibody production in some[139]. Unfortunately, these animals do not go on to develop adrenocortical dysfunction and adrenal failure, which limits the relevance of this model, and its application, to the human condition.

Dogs, however, may spontaneously develop Addison's disease with a clinical presentation very similar to that in humans[140, 141]. Canine Addison's disease predominantly affects middle-aged female dogs and affected dogs commonly present with lassitude, diarrhoea, vomiting, weight loss, poor appetite and weakness. Biochemically, electrolyte disturbances are frequently observed, in common with the human presentation. An ACTH stimulation test, similar to the short synacthen test in humans, is the gold standard diagnostic test in dogs with suspected Addison's. Canine Addison's may affect any breed of dog, however some breeds do seem to be particularly susceptible, including Portuguese Water dogs, Leonbergers, Bearded Collies, Standard Poodles and the Nova Scotia Duck-Tolling Retriever. The quoted incidence range is between 1.5–9.0% in these susceptible breeds[142-145]. The very high incidence of canine Addison's in certain breeds likely reflects founder effects due to significant in-breeding and the frequent use of popular sires[143]. These factors add weight to the hypothesis that, like human AAD, canine Addison's has a strong genetic component to susceptibility. Linkage studies have been used in large, in-bred dog pedigrees to identify susceptibility loci for canine Addison's. In the Portuguese Water dog, loci on chromosomes 12 and 37 have been linked with the disease, and these have subsequently been identified as orthologues of the

human leukocyte antigen (*HLA*) and the *CTLA4* gene regions, which are both associated with multiple human autoimmune conditions.

The observations that Canine Addison's tends to affect female, middle-aged dogs, and susceptibility loci found in regions orthologous to known autoimmune loci in humans, have given rise to the hypothesis that, like human AAD, canine Addison's might have an underlying autoimmune aetiology. However, this has not been definitively established. In dogs, lymphocytic infiltrates in the adrenal glands have been reported from *post-mortem* examinations of affected dogs and antibodies to adrenal cell antigens have been identified by indirect immunofluorescence[146]. However, autoantibodies to specific steroidogenic antigens have yet to be identified. It is therefore likely that canine Addison's disease, similar to human Addison's disease, is a heterogeneous condition: perhaps some affected dogs have an autoimmune aetiology while others have underlying defects in steroidogenesis or adrenocortical development. To date, candidate gene studies on affected dogs have yielded no useful insight into AAD in humans. Genome-wide studies are currently underway in small cohorts of susceptible breeds, and it will be interesting to see if the results of these studies parallel what is known about the genetic architecture of AAD in humans or whether they can shed new light on the human disease.

1.5.2 GENETIC STUDIES OF AAD IN HUMANS

Due to the rarity of AAD, large-scale genetic studies in humans have been impossible, and the studies carried out so far have been case-control candidate gene association studies conducted on relatively small patient cohorts. Despite the high genetic load of AAD, relatively few AAD loci have been identified to date and no single locus apart from MHC has thus far been shown to make a substantial contribution to genetic susceptibility. The majority of the substantial genetic component to AAD aetiology remains undefined. The genetics of AAD have been reviewed and the loci of interest are summarised below[1].

In patients with AAD, the majority of susceptibility loci identified to date exert their effects through the adaptive arm of the immune system; however, some

AAD susceptibility loci are now being discovered in genes of the innate immune system.

1.5.2.1 THE ADAPTIVE IMMUNE SYSTEM

Variants within the MHC class II genes are strongly associated with several autoimmune conditions. It is hypothesised that this is because some *HLA* alleles, in particular some *HLA-DR* or *HLA-DQ* polymorphisms, encode proteins which permit self-peptides to enter the antigen-binding pocket more readily than others. The association of multiple HLA alleles with AAD supports the hypothesis that AAD may result from autoreactivity against a number of different components of the steroidogenic machinery, rather than being triggered by an autoimmune response to a single peptide epitope of the steroid 21OH enzyme. The association between *HLA–DR3* alleles and AAD[147] has been confirmed a number of times[148-150]. An association with *HLA–DR4* has also been reported[147] and replicated. One study demonstrated an association between AAD and the *DRB1*04–DQA1*0301–DQB1*0302* haplotype[55, 151] and heterozygote carriers of the *DR3–DQ2* and *DR4–DQ8* haplotypes were particularly susceptible to AAD. In another study of Norwegian patients, the *DRB1*0404–DQ8* haplotype was strongly associated with AAD, while another haplotype, *DRB1*0401–DQ8,* had a protective effect[55]. A study from the USA then replicated the latter finding[152].

Other genes within the MHC region have also been associated with AAD. However, it has been difficult to establish definitely an independent effect of these alleles, above and beyond the association with HLA, due to the strong LD that exists in the region. In many cases, the genes close to MHC which have been associated with AAD probably do not confer an independent risk, but instead act as markers for the HLA risk alleles close by. One such example is the *CYP21A2* gene, located in the MHC class III region, which encodes the steroid 21OH enzyme and, as such, is a very plausible candidate for AAD. Variants in this gene have been studied in AAD, but any association has been widely attributed to LD with MHC class II alleles[150, 153]. One gene that does seem to be independently associated with AAD is the MHC class I polypeptide-

related sequence A (*MICA*) gene[154]. The strength of the association with the *MICA*5.1* polymorphism is greater than might be expected from linkage disequilibrium with the *HLA-DR3–HLA-DR4* haplotype alone[151, 155].

The *MHC class II transactivator* (*CIITA,* previously known as *MHC2TA*) gene (chromosome 16) encodes a crucial regulator of MHC class II expression, the CIITA protein. Loss of function mutations in *CIITA* result in the bare lymphocyte syndrome, a severe monogenic primary immunodeficiency disorder[156]. Polymorphisms in the *CIITA* gene have been linked to the systemic autoimmune conditions rheumatoid arthritis[157] and SLE[158]. In addition, variants in the promoter region and in intron 3 of *CIITA* have been associated with AAD in both Italian and Norwegian patient cohorts[159, 160], however it is unclear exactly how these polymorphisms exert an impact on AAD susceptibility.

Once a peptide has entered the MHC-binding groove, the MHC-peptide complex engages with the T cell receptor. This interaction in itself is not enough to stimulate an immune response. Co-stimulatory signals are needed and these may be either positive, promoting a vigorous immune response, or negative, which has the effect of downregulating, or dampening down the immune response. The *CTLA4* gene on chromosome 2 encodes a co-stimulatory molecule that has a role in downregulating T cell responses[161]. Polymorphisms at the *CTLA4* locus have been implicated in susceptibility to numerous organ-specific and systemic autoimmune conditions, including type 1 diabetes[162, 163], autoimmune thyroid disease[164-167] and rheumatoid arthritis[168, 169]. Variants in *CTLA4* have also been associated with AAD in a number of independent cohorts and these include an A/G SNP in exon 1[116], an AT repeat in the 3′ untranslated region (UTR) of exon 3[170] and a G/A SNP (*JO30*) downstream of the *CTLA4* gene[171]. The 3'UTR and downstream variant are thought to result in decreased levels of soluble CTLA4. Soluble CTLA4 can interact with CD80 and CD86 molecules on antigen-presenting cells, competing with CD28, an activating T cell surface molecule, to dampen down the T cell response. Thus, in decreasing soluble CTLA4 levels, these polymorphisms may act as a positive regulator of the adaptive immune response by allowing CD28 to access more of

its ligand. Plausible support for this mechanism comes from non-obese diabetic (NOD) mouse models[172, 173]. However, paradoxically increased levels of soluble CTLA4 have been reported in individuals with autoimmune conditions in some studies[174, 175]. This suggests that the CTLA4/CD28 interaction with cell-surface molecules, and subsequent downstream signalling, is incompletely understood.

The *PTPN22* gene on chromosome 1 encodes lymphoid tyrosine phosphatase (LYP) which, like CTLA4, is a negative regulator of T cell signalling. One particular *PTPN22* variant has been implicated in susceptibility to rheumatoid arthritis. This is the T allele of SNP *rs2476601*, which results in an Arg620Trp substitution at the amino acid sequence level[118]. This particular variant has also been associated with type 1 diabetes[176], Graves' Disease[177] and with AAD in several association studies in different populations[177-179]. Although the *PTPN22* protein product LYP has an established regulatory role in T cell activation and function, the mechanisms underlying the association of the Arg620Trp variant with autoimmunity were initially debated. Some studies demonstrated that B and T cell responses were increased in individuals homozygous for the variant, but other studies found exactly the opposite[180-182]. A study in 2011 added clarity, showing that, in both humans and mice, the Arg620Trp is a loss-of-function variant which causes an unstable LYP product that is more readily degraded than the wild-type protein. Levels of functional LYP are therefore reduced by the variant and, thus, its inhibitory effects on T cell signalling and activation are reduced. This creates an environment which favours the development of autoimmunity[183]. Two other co-stimulatory molecules have also been implicated in the pathogenesis of AAD. Allelic variants in *CD274* (which encodes programmed cell death 1 ligand 1, PD-L1) on chromosome 9 have been associated with autoimmune conditions, including AAD in the UK and Norwegian populations[184-186]. The *CD226* gene on chromosome 18 encodes the CD226 protein, also known as DNAX-accessory molecule-1 or DNAM-1[187]. The *Gly307Ser* variant, *rs763361*, has been associated with multiple autoimmune conditions in various populations[188, 189] and has been studied in AAD[190]. However, in AAD, the association was with APS2 and not with isolated AAD, suggesting that it is perhaps due to the

concomitant type 1 diabetes and autoimmune thyroid disease present in APS2 patients, rather than to AAD *per se*.

B lymphocytes are responsible for the generation of antibodies and the humoral immune response, but are also important and efficient antigen-presenting cells which express cell-surface markers that regulate T cell signalling and function. Fc receptor-like (FCRL) family members are B cell surface receptors of the immunoglobulin receptor superfamily encoded in the FcRL gene locus on chromosome 1. These genes have been implicated in susceptibility to several autoimmune diseases[191-194]. For example, FcRL3 is expressed predominantly on B lymphocytes, but is also expressed on T_{Reg} cells and natural killer cells. A promoter polymorphism in the *FCRL3* gene has been widely associated with rheumatoid arthritis and other systemic autoimmune disorders, particularly in Asian populations[192, 193]. A study of the *FCRL3* locus in a cohort of UK AAD patients found that the strongest association occurred with the T allele of the SNP *FCRL3_3*T* (*rs7528684* in the *FCRL3* promoter), which is predicted to have reduced promoter activity[195]. Intriguingly, a haplotype comprising seven *FCRL3* alleles including *FCRL3_3*T* was a susceptibility haplotype for AAD in white Europeans but appeared to be protective for multiple autoimmune conditions in a Japanese cohort. This contradiction could reflect population differences, or differences between the pathogenesis of AAD as an organ-specific autoimmune disorder versus the pathogenesis of other multisystem autoimmune conditions. However, this apparent inconsistency emphasises some of the difficulties encountered when studying the genetics of autoimmunity.

1.5.2.2 THE INNATE IMMUNE SYSTEM

Loci that confer susceptibility to AAD through variants in genetic components of innate immune pathways are now starting to emerge. Examples include the *NLR family, pyrin domain containing 1 (NLRP1)* gene, the *C-type lectin domain family 16, member A (CLEC16A)* gene, the *vitamin D receptor (VDR)* gene and the *cytochrome P450, family 27, subfamily B, polypeptide 1 (CYP27B1)* gene.

Cytoplasmic pattern-recognition receptors, such as the nucleotide-binding oligomerisation domain (NOD)-like receptors (NLRs), are an example of an innate mechanism which has evolved to detect foreign microbial products. The NLRs contribute to the formation of inflammasomes which are multiprotein complexes that activate proinflammatory cytokines and lead to a cascade which triggers the adaptive immune response. Defects in NLR function would be predicted to result in a reduced capacity to differentiate native from foreign antigens. The first report of an association between NLRs and autoimmunity came from research into the role of the *NOD2* (formerly *CARD15*) locus in Crohn's disease[196, 197]. Since then, polymorphisms in a number of NLR genes have also been associated with other organ-specific autoimmune diseases including vitiligo[198] and type 1 diabetes[199]. In Norwegian patients, a coding variant (*Leu155His*) of *NLRP1* (formerly known as *NALP1*) was associated with AAD[199] and this finding was replicated in a Polish cohort [200]. NLRP1 is thought to be important in the assembly of inflammasomes and is also a known activator of IL-1β, a proinflammatory cytokine. However, the effect of the *Leu155His* variant on these, and other functions, of NLRP1 remains unknown and requires further investigation.

The *CLEC16A* gene encodes a protein of unknown function and lies in close proximity to the *CIITA* gene on chromosome 16. The CLEC16A protein possesses a C-type lectin-binding domain, indicating that it might be a cell-surface receptor. It is known to be expressed on a number of professional antigen-presenting cells, including natural killer and dendritic cells[201]. Polymorphisms in *CLEC16A* have been implicated in a number of autoimmune conditions[202-205] including type 1 diabetes[117, 201]. In a study of AAD cases and controls from the UK and Norway, a *CLEC16A* intronic variant, *rs12917716,* was associated with AAD in the Norwegian cohort, but this finding was not replicated in the UK cohort[160].

Vitamin D is known to exert a suppressive effect on immunity and autoimmunity. Studies have demonstrated that vitamin D supplementation can prevent the onset of autoimmune diabetes in the NOD mouse which is prone to developing

diabetes[206]; in humans, vitamin D deficiency is associated with the development of some autoimmune conditions including type 1 diabetes[207]. Polymorphisms in two chromosome 12 genes, *VDR* and *CYP27B1* (which encodes a cytochrome P450 1α-hydroxylase enzyme that catalyses the conversion of 25-hydroxyvitamin D3 to the active 1,25-dihydroxyvitamin D3), have been associated with AAD. A small German study found evidence of association between *VDR* genotypes and AAD however, an association with alleles was not found[208]. To confirm the *VDR* gene as a susceptibility locus in AAD, further independent studies are required. In contrast, there is convincing evidence that a promoter polymorphism, *–1260C>A*, in the *CYP27B1* gene is associated with AAD, type 1 diabetes and autoimmune thyroid disease. This association was initially identified in a study from Germany[209] and subsequently confirmed in a UK study of AAD patients[210].

1.6 STUDY AIMS

The genetic determinants of AAD remain poorly defined despite on-going investigation. I plan to use both hypothesis-driven (candidate gene) approaches and a discovery-driven (genome-wide) approach to identify novel susceptibility alleles for AAD using unique sample resources.

1) A hypothesis-driven approach to the investigation of AAD – The *CYP21* locus.

To explore genomic variations in and around the *CYP21A2* gene, in unrelated individuals with AAD and controls, using polymerase chain reaction (PCR)-based approaches, tissue *in situ* hybridisation (TISH) and quantitative PCR (qPCR)

2) A hypothesis-driven approach to the investigation of AAD – A study of 20 candidate genes in six European AAD cohorts.

To perform a candidate gene study in unrelated individuals with AAD and controls from six European cohorts using the Sequenom iPlex genotyping platform.

3) A discovery-driven approach to the investigation of AAD – A genome-wide study of multiplex AAD families.

To perform a genome-wide linkage and association study on multiplex AAD families using the Affymetrix Genome-wide human SNP array 6.0.

CHAPTER 2 – STUDY SUBJECTS AND METHODS

2.1 SUBJECTS

Ethical approval for this study was granted by the Leeds (East) Research Ethics Committee in 2005 (REC reference number 05/Q1206/144).

2.1.1 TISSUE SAMPLES FOR FUNCTIONAL ASSAYS

Human tissue samples were used to investigate *CYP21A2* expression, and the expression of its closely related pseudogene, *cytochrome P450, family 21, subfamily A, polypeptide 1 pseudogene* (*CYP21A1P*), by TISH and qPCR. For TISH, skin and matched fixed and mounted adrenal samples were available for analysis from the Human Developmental Biology Resource (HDBR), Newcastle upon Tyne. In total, gDNA was extracted from 42 skin samples taken from fetal material ranging from eight to twelve weeks post-conception (wpc). These gDNA samples were used to establish the presence (*CYP21A1P*+) or absence (*CYP21A1P*–) of *CYP21A1P*. Mounted tissue sections corresponding to five of these samples (one *CYP21A1P*– sample, four *CYP21A1P*+ samples) were then used for TISH.

For qPCR, three of the above adrenal samples (one *CYP21A1P*–, two *CYP21A1P*+) with matched kidney tissue were used for assay optimisation and quality control, in addition to thymus samples from seven infants and children, taken during cardiothoracic surgery, from the Human Biomaterials Resource Centre (HBRC), University of Birmingham. These samples were taken from individuals age eight days to 16 months (median age ten weeks). Age information was not available for two of the seven samples and no other clinical data was available for these samples. gDNA was extracted from the seven thymus samples and used to establish the presence or absence of *CYP21A1P*. All were *CYP21A1P*+. RNA was extracted from the remainder of the sample and used to synthesise cDNA for qPCR to establish expression of *CYP21A2* and *CYP21A1P* transcripts in this organ.

2.1.2 UNRELATED AAD CASE-CONTROL COHORTS

In total, 2005 unrelated AAD DNA samples from six European countries were available for this study. To be included in the AAD cohorts, affected subjects had to have biochemical evidence of adrenal failure, with a maximum serum cortisol of less than 550nmol/L one hour following intravenous administration of synthetic ACTH analgoue (synacthen 250µg) and a raised ACTH level. Those with infective and infiltrative causes of Addison's, those with secondary adrenal failure and those with APS1 were excluded from this study. In addition, 1972 matched local healthy controls were available (Table 2). All controls were Caucasian and had no personal or family history of autoimmune disease. Clinically silent autoimmune disease was not excluded in these controls by checking autoantibody levels, adrenal or thyroid function, however this is likely to be very rare.

2.1.2.1 UK COHORT

For case-control and replication studies, 357 unrelated UK subjects with AAD were identified from local endocrine clinics within the north east of England and through the Addison's disease self-help group network. All AAD subjects were Caucasian.

- 273 were female (76.5%) and 84 were male (23.5%)
- The mean age of onset of AAD was 38.8 years (median 38 years, minimum 10 years, maximum 83 years)
- 203 (56.9%) had an additional autoimmune condition
    - 155 (43.4%) had autoimmune thyroid disease
    - 41 (15.0% of the female cohort) had premature ovarian failure
    - 21 (5.9%) had type 1 diabetes
    - 22 (6.2%) had pernicious anaemia
    - 11 (3.1%) had vitiligo
    - 8 (2.2%) had rheumatoid arthritis
    - 7 (2.0%) had coeliac disease
    - 3 (0.8%) had alopecia
    - 1 (0.3%) had Crohn's disease

- 1 (0.3%) had ankylosing spondylitis
- 1 (0.3%) had autoimmune hepatitis

Serum samples were not available for the majority of the AAD cases, however of the 61 that were tested for 21OH autoantibodies, 53 (86.9%) were positive. A maximum of 627 control samples were available from British Caucasian individuals for comparison.

Centre	AAD	Controls	Total
UK	357	367	724
Norway	384	384	768
Germany	341	235	576
Italy - Perugia	100	188	288
Italy - Padua	180	134	314
Poland - Warsaw	159	50	209
Poland - Poznan	116	246	362
Sweden	368	368	736
Total	**2005**	**1972**	**3977**

Table 2: AAD DNA resources available for analysis.

2.1.2.2 NORWEGIAN COHORT

From Norway, 384 unrelated AAD cases were selected at random from a cohort of 426 individuals.

- 273 (64.0%) were females and 153 (36.0%) were males
- 359 (84.2%) were 21OH autoantibody positive
- The mean age was 53 years (minimum 18, maximum 95 years)
- The mean AAD disease duration was 14 years (minimum 0, maximum 56 years)
- Of the cohort as a whole, 66% reported autoimmune comorbidity
    - 198 (46.5%) had autoimmune thyroid disease
    - 49 (11.5%) had type 1 diabetes
    - 18 (6.6% of the female cohort) had premature ovarian failure
    - 42 (9.9%) had pernicious anaemia
    - 48 (11.3%) had vitiligo
    - 16 (3.8%) had alopecia
    - 3 (0.7%) had autoimmune hepatitis

Healthy blood donor controls from Norway were also available for comparison (maximum 1353 samples).

2.1.2.3 GERMAN COHORT

From Germany, DNA samples from 341 AAD cases and 235 healthy matched controls were available. These samples were selected from a cohort of 364 DNA samples. Of these, serum samples for 21OH autoantibody levels were available for 200: 170 (85%) were positive while 30 (15%) were negative.

- The mean age at collection of these samples was 50.7 years (median 49 years, minimum 22 years, maximum 88 years)
- 263 (72.2%) of this cohort were female
- 188 individuals (55.1%) had other autoimmune comorbidities in addition to AAD
    - 160 (46.9%) had autoimmune thyroid disease

- 24 (7.0%) had type 1 diabetes
- 3 (0.82%) had vitiligo
- 1 (0.27%) had Crohn's disease

2.1.2.4 SWEDISH COHORT

From Sweden, 368 21OH autoantibody positive AAD cases and 368 healthy matched controls were available.

- 143 (38.9%) were male, 225 (61.1%) were female
- The mean age of diagnosis was 34 years (median 32 years, minimum 0 years, maximum 71 years)
- 228 individuals (62.0%) had other autoimmune comorbidities in addition to AAD
  - 180 (48.9%) had autoimmune thyroid disease
  - 40 (10.9%) had type 1 diabetes

2.1.2.5 ITALIAN COHORT

From Italy, DNA samples were available from Padua (180 AAD cases, of which 166 (92.2%) were 21OH autoantibody positive and 14 (7.8%) were negative, and 134 matched healthy controls) and from Perugia (100 21OH autoantibody positive AAD cases, 188 healthy matched controls).

Of the 180 AAD cases from Padua, 117 (65%) were female and 63 (35%) were male.

- The mean age of onset of AAD was 34.7 years (median 34 years, minimum 6 years, maximum 84 years)
- Of those people, a total of 142 (78.9%) had another autoimmune comorbidity
  - 134 (74.4%) had autoimmune thyroid disease
  - 18 (10%) had type 1 diabetes

- 4 (3.4% of the females in the cohort) had premature ovarian failure
- 4 (2.2%) had vitiligo
- 1 had coeliac disease
- 1 had multiple sclerosis
- 1 had rheumatoid arthritis.

The AAD cohort from Perugia consisted of 67 (67%) females and 33 (33%) males.

- The mean age of onset of AAD was 40.6 years (median 38 years, minimum 10 years, maximum 78 years)
- 65 (65%) of those with AAD had an additional autoimmune condition
  - 49 (49%) had autoimmune thyroid disease
  - 15 (15%) had type 1 diabetes
  - 12 (17.9% of the female cohort) had premature ovarian failure
  - 9 (9%) had vitiligo
  - 3 (3%) had coeliac disease
  - 2 (2%) had rheumatoid arthritis
  - 1 had SLE
  - 1 had pernicious anaemia.

2.1.2.6 POLISH COHORT

The AAD cohort from Poznan consisted of 116 individuals

- 85 (73.3%) were females and 31 (26.7%) were males
- The mean age of onset of AAD in this cohort was 36.5 years (median 36 years, minimum 14 years, maximum 69 years)
- 98 individuals (84.5%) within the cohort had associated autoimmune comorbidity
  - 87 (75.0%) had autoimmune thyroid disease
  - 21 (18.1%) had chronic atrophic gastritis of which 8 (6.9%) had a diagnosis of pernicious anaemia
  - 10 (11.8% of the female cohort) had premature ovarian failure

    - 10 (8.6%) had type 1 diabetes
    - 7 (6.0%) had vitiligo
    - 3 (2.6%) had coeliac disease
    - 1 (0.86%) had alopecia

21OH autoantibody levels were not available for this cohort. 246 healthy controls from Poznan were available for comparison.

From Warsaw, 159 AAD cases were selected at random from a cohort of 229 AAD cases. Of the cohort of 229, 114 (49.8%) were 21OH autoantibody positive, 106 (46.3%) were 21OH autoantibody negative and results were not available for 9 (3.2%).

- 173 (75.5%) were females and 56 (24.5%) were males
- The mean age of onset of AAD in this cohort was 38.9 years (median 37 years, minimum 9 years, maximum 76 years). Age of onset was not available for 12 individuals
- Of the cohort as a whole, 165 (70%) reported autoimmune comorbidity
    - 138 (60.3%) reported autoimmune thyroid disease
    - 24 (10.5%) had vitiligo
    - 24 (10.5%) had type 1 diabetes
    - 15 (8.7% of the female cohort) had premature ovarian failure
    - 14 (6.1%) had pernicious anaemia
    - 5 (2.2%) had alopecia
    - 2 (0.9%) had coeliac disease
    - 2 (0.9%) had autoimmune hepatitis
    - 1 (0.4%) had SLE
    - 1 had rheumatoid arthritis
    - 1 had hypogonadism (1.8% of the male cohort).
- 42 individuals (18.3%) reported a family history of autoimmunity.

From Warsaw, 50 healthy matched blood donor controls were available for comparison.

2.1.3 OTHER AUTOIMMUNE COHORTS

2.1.3.1 UK GRAVES' DISEASE COHORT

A maximum of 447 DNA samples from individuals with Graves' disease (but without AAD) from the UK were available. Subjects had biochemical evidence of hyperthyroidism with confirmation by positive serum autoantibodies (anti-thyroid stimulating hormone (TSH) receptor and/or anti-thyroid peroxidase), diffuse uptake on $^{[99Tc]}$pertechnetate radionuclide scan, or presence of ophthalmopathy.

- This cohort comprised 352 (78.7%) females and 95 males
- At the time of the study, the mean age was 53 years (median 54 years, minimum 20 years, maximum 87 years). This information was missing for 18 individuals
    - 1 individual also had pernicious anaemia
    - 2 had type 1 diabetes
    - 2 females had premature ovarian failure.

2.1.3.2 NORWEGIAN TYPE 1 DIABETES COHORT

1195 DNA samples from individuals with type 1 diabetes from Norway were available for genotyping. These samples were selected at random from a previously described cohort[211] of 1331 Caucasian type 1 diabetes patients (51.9% boys and 48.1% girls) all diagnosed with type 1 diabetes before the age of 17 and meeting the EURODIAB[212] criteria. In addition, 1353 control DNA samples were available for comparison. These were all Caucasian, Norwegian blood donor controls.

2.1.3.3 NEW ZEALAND RHEUMATOID ARTHRITIS COHORT

650 DNA samples from individuals with a diagnosis of rheumatoid arthritis from New Zealand were available. These samples came from Caucasian individuals from Auckland, Bay of Plenty, Wellington, Otago and Southland regions of New Zealand. They all met the 1987 American College of Rheumatology criteria for disease (four or more of the following: morning stiffness for at least six weeks in

and around joints lasting at least one hour before maximal improvement, soft tissue swelling of three or more joint areas observed by a physician for at least six weeks, swelling of the proximal interphalangeal, metacarpophalangeal, or wrist joints for at least six weeks, symmetric swelling for at least six weeks, rheumatoid nodules, the presence of rheumatoid factor in the serum, radiographic erosions and/or periarticular osteopenia in hand and/or wrist joints). 452 matched healthy local controls were available for comparison.

2.1.4 MULTIPLEX AAD FAMILIES

Prior to inclusion in the study, multiplex AAD kindreds, that is kindreds containing two or more individuals with AAD, were carefully assessed for suitability for the study, with the aim of selecting a relatively homogeneous group. To achieve this, careful phenotyping of all of the families, face-to-face or by telephone interview, was undertaken with a view to excluding families with a non-autoimmune aetiology to the Addison's disease. A history of mucocutaneous candidiasis, hypoparathyroidism and dental problems was sought from all families to exclude those with APS1. In addition, all family members were asked about tuberculosis, to exclude this as a possible "familial" cause of Addison's. An underlying autoimmune aetiology to Addison's in families was accepted if the affected individuals had an additional autoimmune disease such as type 1 diabetes, autoimmune thyroid disease, pernicious anaemia or vitiligo. In the absence of a personal history of autoimmune disease, a family history of autoimmune disease in a first degree relative e.g. parent or sibling, in particular autoimmune thyroid disease or type 1 diabetes, lent support to an underlying autoimmune aetiology. Finally, 21OH autoantibody status, where serum was available, was considered.

2.1.4.1 UK FAMILIES

12 Caucasian UK AAD kindreds were identified from local endocrine clinics within the north east of England (n=2), from collaborators in Exeter (n=2) and from the Addison's disease self-help group network (n=8) (Figure 5).

The UK kindreds comprised 57 individuals in total (28 males, 29 females). 11 of the 12 UK kindreds had two members with AAD (affected sibling pairs n=8, affected parent-child pairs n=3) whilst one kindred comprised a trio of affected siblings.

Of those with AAD (n=25), 12 were female and 13 were male

- The mean age of onset of AAD was 39 years (median 36 years, minimum 18 years, maximum 67 years).
- 14 of the 25 individuals with AAD had one or more autoimmune comorbidities
    - 7 also had autoimmune thyroid disease
    - 6 had type 1 diabetes
    - 3 had seronegative arthritis
    - 2 had pernicious anaemia
    - 1 had vitiligo
    - 1 had rheumatoid arthritis
    - 1 had Sjögren's
    - 1 had SLE
    - 1 had hypogonadism
    - 1 had premature ovarian failure

Of those unaffected kindred members (n=31), 17 were female and 15 were male.

- 8 had autoimmune comorbidities
    - 4 had autoimmune thyroid disease
    - 1 had type 1 diabetes
    - 1 had pernicious anaemia
    - 1 had vitiligo
    - 1 had coeliac disease

One unaffected individual had had a subtotal thyroidectomy but this was for a large multinodular goitre and was not due to autoimmune thyroid disease. Ethylenediaminetetraacetic acid (EDTA) blood samples were available from all

57 individuals for DNA extraction. Serum samples for 21OH autoantibody status were available from 47 individuals in total. Of those 25 individuals with AAD, serum was available for 15 (9 were 21OH autoantibody positive, 6 were 21OH autoantibody negative). Of the unaffected family members, serum was available for 32 individuals. 3 were 21OH autoantibody positive and 29 were 21OH autoantibody negative.

2.1.4.2 NORWEGIAN FAMILIES

12 Norwegian kindreds were identified through the Norwegian registry of autoimmune diseases. These kindreds comprised 64 individuals in total (28 males, 36 females). 10 of the 12 Norwegian kindreds had 2 members with AAD (affected sibling pairs n=5, affected parent-child pairs n=1, other relationships e.g. avuncular, cousins *etc*. n=4). One kindred comprised a trio of affected siblings and one kindred comprised of two pairs of affected sibling pairs in two generations (Figure 6).

Of those with AAD (n=27), 18 were female and 9 were male

- The mean age of onset of AAD was available for 25 of the 27 individuals with AAD and was 30.4 years (median 27 years, minimum 10 years, maximum 67 years)
- Of those with AAD, 10 had autoimmune comorbidities
    - 9 also had autoimmune thyroid disease
    - 4 had type 1 diabetes
    - 2 had vitiligo
    - 2 had rheumatoid arthritis
    - 2 had coeliac disease
    - 1 had pernicious anaemia
    - 1 had gonadal failure
    - 1 had Sjögren's disease

Of those unaffected kindred members (n=37), 18 were female and 19 were male

- 4 had autoimmune comorbidities
    - 4 had autoimmune thyroid disease
    - 1 had diabetes

DNA samples were available from all 64 individuals and serum was available from 59 people. Of those 27 individuals with AAD, serum was available for 24 (22 were 21OH autoantibody positive, 2 were 21OH autoantibody negative). Of the unaffected family members, serum was available for 35 individuals. 2 were 21OH autoantibody positive and 33 were 21OH autoantibody negative.

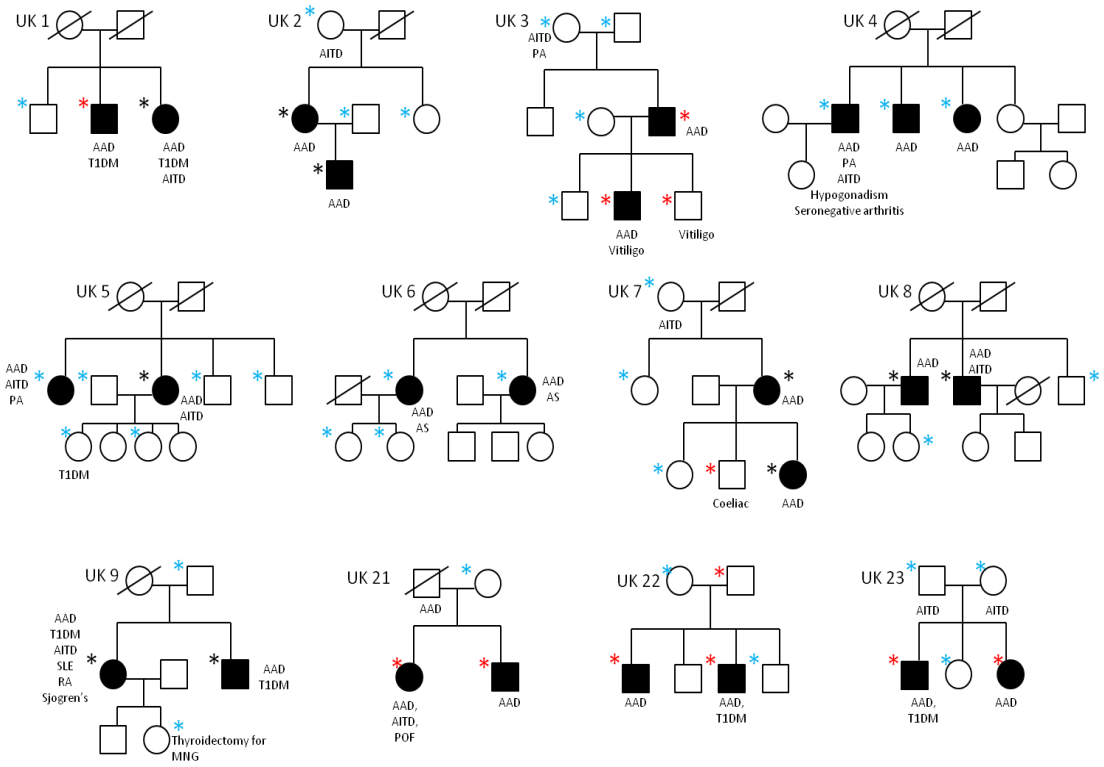


Figure 5: UK multiplex AAD family pedigree structures.

Circles represent females, squares represent males. Deceased individuals are represented by a score out circle or square. Filled circles/squares represent affected individuals and unfilled circles/squares represent unaffected individuals. Samples marked with an asterisk were available for 21OH autoantibody analysis: blue asterisk = 21OH autoantibody negative, red asterisk = 21OH autoantibody positive, black asterisk = 21OH autoantibody status unknown. AAD, autoimmune Addison's disease; T1DM, type 1 diabetes mellitus; AITD, autoimmune thyroid disease; PA, pernicious anaemia; AS, ankylosing spondylitis; SLE, systemic lupus erythematosus; RA, rheumatoid arthritis; MNG, multimodular goitre; POF, premature ovarian failure.

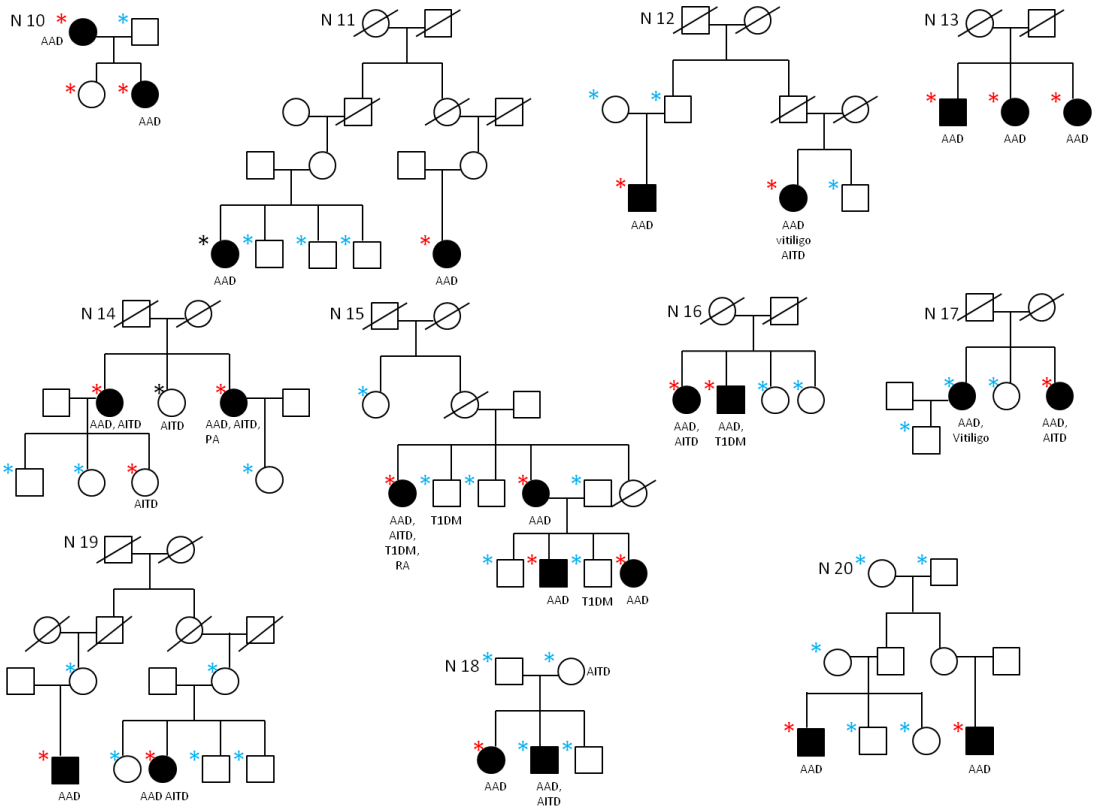


Figure 6: Norwegian multiplex AAD family pedigree structures.

Circles represent females, squares represent males. Deceased individuals are represented by a score out circle or square. Filled circles/squares represent affected individuals and unfilled circles/squares represent unaffected individuals. Samples marked with an asterisk were available for 21OH autoantibody analysis: blue asterisk = 21OH autoantibody negative, red asterisk = 21OH autoantibody positive, black asterisk = 21OH autoantibody status unknown. AAD, autoimmune Addison's disease; T1DM, type 1 diabetes mellitus; AITD, autoimmune thyroid disease; PA, pernicious anaemia; RA, rheumatoid arthritis.

2.2 GENERIC METHODS

2.2.1 EXTRACTION OF GENOMIC DNA FROM WHOLE VENOUS BLOOD

All centrifuge speeds are stated in g. This relates to revolutions per minute (rpm) as follows:

$g = (1.118 \times 10^{-5}) \times R \times S^2$

where R is the radius of the centrifuge rotor and S is the speed in rpm.

Genomic DNA (gDNA) was extracted from whole venous blood using the Nucleon BACC2 DNA extraction kit (GE Healthcare Life Sciences). 9ml of venous blood (anticoagulated with 0.5ml of 0.5M EDTA, pH 8), frozen at -80^{o}C for storage, was defrosted at room temperature for 1 hour. Once fully defrosted, the blood was transferred into a 50ml polypropylene tube, and 40ml of lysis buffer Reagent A (10mM Tris-HCl, 320mM sucrose, 5mM $MgCl_2$, 1%(v/v) Triton X-100, pH 8.0) was added. The tube was rotary mixed for 4 minutes at room temperature and then centrifuged at 1300g for 5 minutes. The supernatant was discarded, leaving a pellet which was resuspended in 2ml of Reagent B (400 mM Tris-HCL, 60 mM EDTA, 150 mM NaCl, 1% SDS, pH 8). The suspension was transferred into a clean 15ml polypropylene tube and 500µl of sodium perchlorate was added. The tube was then inverted 7 times by hand and 2ml of chloroform was added before inverting the tube again, by hand, 7 times. Without disturbing the phases, 300µl Nucleon™ resin was added to the tube and centrifuged at 1300g for 3 minutes.

The clear, colourless layer above the resin (approximately 3mls) was transferred to a clean 15ml polypropylene tube, without disturbing the resin layer. 2 volumes (approximately 6mls) of ice cold 100% ethanol were added to the solution, and the tube inverted, to precipitate gDNA. The opaque precipitated gDNA was extracted from the solution using a sterile plastic hook,

and suspended in 200µl of TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8) in a 1.5ml tube, and refrigerated at 4^{o}C overnight to allow the gDNA to dissolve fully.

2.2.2 EXTRACTION OF GENOMIC DNA FROM TISSUES

gDNA was extracted from tissues using the QIAamp DNA Mini Kit (Qiagen). Tissue samples stored in the -80^{o}C freezer were defrosted and weighed. A maximum of 20mg of tissue was finely chopped in a petri dish using sterile scalpels. The tissue was then placed in a 1.5ml microcentrifuge tube and 180µl of buffer ATL was added. 20µl of proteinase K was then added and the solution mixed by vortexing. This was then incubated overnight at 56°C in a shaker-incubator (Eppendorf Thermomixer comfort).

The following morning, the tube was briefly centrifuged and then 200µl of buffer AL was added to the sample. The solution was mixed by vortexing for at least 15 seconds and then incubated on a shaker-incubator for 10 minutes at 70°C. The tube was then briefly centrifuged and 200µl of 100% ethanol was added. This was then mixed by vortexing for a minimum of 15 seconds. The solution was then pipetted directly into a QIAamp Mini spin column placed into a 2ml collection tube and the cap closed. The column was centrifuged at 6000g for 1 minute. The QIAamp Mini spin column was then placed into a clean 2ml collection tube and the tube containing the filtrate was discarded. 500µl of buffer AW1 was then added to the spin column and the column centrifuged at 6000g for 1 minute. Again, the column was then placed into a new collecting tube and the tube containing the filtrate was discarded. 500µl of buffer AW2 was then placed into the column and this was centrifuged at full speed (maximum 20,000g) for 3 minutes. The tube containing the filtrate was discarded. The column was then placed in a new 2ml collection tube and centrifuged at full speed for 1 minute. Again, the tube containing the filtrate was discarded and the spin column was placed into a clean 1.5ml tube.

To elute the DNA captured on the spin column membrane, 200µl of buffer AE was applied to the column and this was incubated at room temperature for 1 minute before centrifuging at 6000g for 1 minute.

2.2.3 DNA QUANTIFICATION

After 24 hours, the concentration and integrity of gDNA in the solution was determined using ultraviolet (UV) spectrophotometry of the optical density (OD) at 260nm (NanoDrop ND-1000, NanoDrop Technologies Inc). An OD^{260} of 1.0 is taken as 50µg/ml for double-stranded DNA. For use with the Affymetrix Genome-Wide Human SNP Array 6.0 platform, a minimum of 10µl of gDNA of a minimum concentration of 100ng/µl with an $OD^{260/280}$ ratio between 1.7 and 2.0 was required. For other PCR applications, a minimum of 20ng/µl with an $OD^{260/280}$ ratio between 1.7 and 2.0 was required. DNA samples were also run out on an agarose gel to ensure sample integrity.

2.2.4 EXTRACTION OF RNA FROM WHOLE VENOUS BLOOD

Prior to each extraction, all equipment was cleaned with RNase ZAP (Sigma) to remove RNases. Ribonucleic acid (RNA) was extracted from whole venous blood using the PAXgene blood RNA kit (Qiagen), for extraction and purification of intracellular RNA.

2.5ml of whole venous blood was collected in PAXgene vacutainer tubes and frozen at -80ºC for storage. The PAXgene tubes were defrosted and allowed to stand at room temperature for 2 hours prior to the RNA extraction procedure. The PAXgene tube was then centrifuged at 3500g for 10 minutes and the supernatant discarded leaving a pellet, to which 4ml of RNase-free water was added. The PAXgene tube was then re-sealed using a fresh BD Hemogard closure lid and the pellet fully resuspended by vortexing. The tube was then centrifuged for a further 10 minutes at 3500g and the supernatant was again discarded, leaving a pellet.

350µl of Buffer BR1 was then added and the pellet resuspended by vortexing. This solution was then transferred to a fresh 1.5ml microcentrifuge tube. 300µl of Buffer BR2 and 40µl proteinase K were then added and the solution mixed by vortexing for 5 seconds. This solution was then incubated at 55ºC for 10 minutes in a shaker-incubator to allow protein digestion.

The lysate was pipetted carefully into a PAXgene shredder spin column placed in a 2ml processing tube, and centrifuged for 3 minutes at maximum speed (not exceeding 20,000g) to remove any residual debris and to homogenise the lysate. The entire supernatant (the flow-through fraction) was then transferred to a fresh 1.5ml microcentrifuge tube without disturbing the pellet in the processing tube.

350µl of 100% ethanol was then added and the solution mixed by vortexing before briefly centrifuging at 1000g for 2 seconds. 700µl of lysate was pipetted into the PAXgene RNA spin column placed in a 2ml processing tube, and centrifuged for 1 minute at 16,000g. During this step, RNA is selectively bound to the PAXgene silica membrane as contaminants pass through. The spin column was then removed and placed in a new 2ml processing tube, and the old processing tube, containing the flow-through, was discarded. This step was repeated with the remainder of the sample.

Remaining contaminants were removed in several wash steps. Initially, 350µl of Buffer BR3 was pipetted into the PAXgene RNA spin column and centrifuged for 1 minute at 16,000g. The spin column was then placed in a new 2ml processing tube and the old processing tube containing the flow-through was discarded. 10µl of DNase I stock solution was added to 70µl of Buffer RDD and mixed by hand. 80µl of the DNase I mix was applied directly onto the PAXgene RNA spin column membrane, and left at room temperature for 15 minutes to remove any traces of DNA.

350µl of Buffer BR3 was pipetted into the PAXgene RNA spin column, and centrifuged for 1 minute at 16,000g. The spin column was placed into a fresh 2ml processing tube, and the old processing tube and flow-through discarded. 500µl of Buffer BR4 was then pipetted into the PAXgene RNA spin column, and again centrifuged for 1 minute at 16,000g. Again, the spin column was placed in a new 2ml processing tube, and the old processing tube containing the flow-through was discarded.

A further 500µl of Buffer BR4 was then pipetted into the PAXgene RNA spin column and centrifuged for 3 minutes at 16,000g. The tube containing the flow-through was then discarded and the PAXgene RNA spin column was placed into a fresh 2ml tube. This was again centrifuged for 1 minute at 16,000g and the tube containing the flow through was discarded while a fresh 1.5ml microcentrifuge tube was used for the spin column.

To elute RNA, 40µl of the Buffer BR5 was applied directly onto spin column membrane and centrifuged for 1 minute at 16,000g. This step was then repeated for maximum RNA yield. The eluate was then heat-denatured by incubating for 5 minutes at 65°C and then chilled immediately on ice following incubation. RNA samples were then stored at -80ºC.

2.2.5 EXTRACTION OF RNA FROM TISSUE

Tissue samples, stored at -80ºC, were defrosted on ice in a 1.5ml tube. 1ml of TRIzol (Invitrogen) was added. The tube was incubated at room temperature for 5 minutes. 0.2ml of chloroform was then added and the mixture vortexed for 15 seconds. This mixture was then incubated at room temperature for a further 3 minutes. The tube was then centrifuged at 4ºC for 15 minutes, at 12,000g. The upper, aqueous layer was then removed by pipette and put into a fresh tube. 0.5ml of 100% isopropanol was added and the solution incubated at room temperature for 10 minutes. The tube was then centrifuged at 4ºC for 10 minutes, at 12,000g. The layer of isopropanol was then removed to leave a pellet. This was resuspended in 1ml of 70% ethanol (1ml of ethanol per 1ml of TRIzol). The tube was then centrifuged at 4ºC for 5 minutes, at 7500g. The ethanol was then removed and the pellet allowed to air-dry for 2 minutes. The residual pellet was resuspended in 50µl Diethylpyrocarbonate (DEPC) water.

To remove traces of gDNA contamination, RNA samples were treated with DNase, using the Primerdesign Precision DNase kit (Primerdesign, Southampton). In a PCR tube, for every 50µl of RNA, 5µl of 10X Precision

DNase reaction buffer was added, making the final reaction concentration 1X. 1µl of Precision DNase enzyme was added for every 100µl of RNA solution. This solution was then incubated at 30ºC for 10 minutes and then at 55ºC for 5 minutes to inactivate the DNase.

Integrity and concentration of RNA samples was determined by UV spectrophotometry of the OD at 260nm using the NanoDrop ND-1000 spectrophotometer (an OD^{260} of 1.0 being taken as 40µg/ml for RNA). Samples with an $OD^{260/280}$ ratio of less than 1.80 or greater than 2.20 were rejected. RNA samples were also run out on an agarose gel to ensure sample integrity.

2.2.6 FIRST STRAND cDNA SYNTHESIS

5µg of RNA was pipetted into a PCR tube with 50ng of random hexamers and 40u RNaseOut (Invitrogen). The volume was made up to 12µl with nuclease-free filtered, sterile distilled water if necessary. This was incubated at 70ºC for 10 minutes in a PCR machine, briefly centrifuged and then placed on ice. To each tube, 8µl of master mix (0.5mM dNTPs, 1X buffer, 0.01M DTT and 200u Superscript III RT (Invitrogen) enzyme) was added to make a total reaction volume of 20µl. This was incubated in a PCR machine at 50ºC for 1 hour and then 95ºC for 5 minutes. After this, samples were briefly centrifuged and placed on ice. Complementary DNA (cDNA) was then stored at -20ºC for later use.

2.2.7 EXTRACTION AND PURIFICATION OF PCR PRODUCTS FROM AGAROSE GEL

The QIAquick Gel Extraction Kit (Qiagen) was used to extract and purify PCR products from agarose gel for sequencing.

PCR products were cut out from the agarose gel under direct UV visualisation, placed in a 1.5ml tube and weighed. 3 volumes of buffer QG were added and then each tube was incubated at 50^0C on a heat block for 10 minutes, vortexing

regularly until the agarose gel had dissolved. 1 volume of 100% isopropanol was then added to the tube and mixed by hand. This solution was then applied to a QIAquick column placed into a 2ml collecting tube and centrifuged at 17,900g for 1 minute. The flow-through was then discarded. 500µl of buffer QG was added to the column which was centrifuged for 1 minute at 17,900g. Once again, the flow-through was discarded. 750µl of buffer PE was then applied to the column and centrifuged for 1 minute at 17,900g. The QIAquick column was then placed in a fresh tube. To elute the PCR product, 50µl of buffer EB (10mM Tris-Cl, pH 8.5) was added and the column left to stand for 1 minute before centrifuging at 17,900g for 1 minute. The extracted product was sent to Eurofins MWG Operon for dideoxynucleotide (Sanger) sequencing.

2.2.8 21OH AUTOANTIBODY ASSAY

Available serum samples were shipped to Bergen, Norway on dry ice, for analysis of 21OH autoantibody levels. A fluid phase radioimmuno-precipitation assay was used as previously described[213]. In brief, each patient or control serum sample, diluted 1 in 10 in assay buffer (150 mM NaCl, 20 mM Tris–HCl, 0.15 % (v/v) Tween–20, pH 8.0), was incubated overnight at 4°C with 15,000–20,000 counts per minute (cpm) of radiolabeled 21OH antigen (in triplicate). Immune complexes were isolated by precipitation with protein A Sepharose (GE Healthcare Life Sciences). The amount of radiolabeled antigen bound by serum samples was analysed in filter-bottomed microtiter plates (Millipore) by a TopCount liquid scintillation counter (PerkinElmer). The results were expressed as an autoantibody index, AI:

[(cpm sample – cpm negative standard)/(cpm positive standard – cpm negative standard) x 1000]

For each assay plate, serum from an AAD patient with previously confirmed autoantibodies targeting the respective antigens was used as a positive control, while pooled human sera from healthy volunteers was used as a negative control. The upper normal limit (positive cut-off) for each antigen was set as the

autoantibody index calculated from the mean cpm of triplicates of 100 Norwegian blood donors + 3 standard deviations (SD).

2.2.9 GENOTYPING ON THE SEQUENOM PLATFORM

The Sequenom MassARRAY platform is a high-throughput SNP genotyping platform which allows up to 40 SNPs to be genotyped in each reaction, through multiplexing of compatible assays. Initially, primers pairs specific to each SNP of interest are used in parallel in a PCR reaction. This is followed by an extension reaction, again specific to each SNP of interest, whereby an extension primer anneals immediately upstream of its target SNP. The reaction is then incubated with mass-modified di-deoxynucleotide terminators and each extension primer is extended by a single mass-modified base. The base extension is determined by the underlying sequence. The mass of the extended primer can then be determined using matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry and the alleles present can be derived from the primer's final mass[214].

2.2.9.1 SNP SELECTION FOR SEQUENOM

To select SNPs for genotyping, the Ensembl database (http://www.ensembl.org)[215] was used to search for coding variants in genes or regions of interest and Haploview[216] was used to analyse LD patterns and haplotypes. The HapMap database (http://hapmap.ncbi.nlm.nih.gov)[102] was then used to select SNPs covering common haplotypes.

2.2.9.2 SEQUENOM ASSAY DESIGN

Following SNP selection, the Ensembl database was then used to identify the DNA sequence flanking each SNP. This sequence (approximately 300bp in length) was uploaded into the MySequenom website (https://www.mysequenom.com). The MassARRAY® Designer software was then used to design compatible, sequence-specific forward, reverse and iPLEX single base extension primers. All forward and reverse primers were ordered unmodified (Metabion) and the extension primers were high performance liquid

chromatography (HPLC) purified and MALDI-TOF checked (Metabion). Primer sequences for the twenty candidate gene study and the linkage validation association study are shown in appendices A and C respectively.

2.2.9.3 SEQUENOM PCR METHOD

Two Sequenom providers were used in the course of this work. Either PCR products were provided to NewGene (Newcastle, UK) for a post-PCR Sequenom genotyping service or gDNA samples were shipped to CIGMR (Manchester University, Manchester, UK) for a complete Sequenom genotyping service.

To provide PCR products for genotyping, multiplex PCR was carried out in 96 well plates and at least five negative controls and a number of duplicates were included, per plate, to ensure that the PCR was not contaminated and to ensure genotyping fidelity respectively. The PCR reaction volume was 10µl and contained 15ng of template gDNA, 1.25X PCR buffer (Qiagen), 1.63mM $MgCl_2$ (Qiagen), 0.5mM dNTPs (New England Biosciences) and 1u of HotStar Taq DNA polymerase (Qiagen). A pool of primers (Metabion) was made to include 0.5 µM of each forward and reverse primer. This pool was added to the PCR reaction to give a final concentration of each primer of 0.1µM. The thermal cycling conditions for the reaction included an initial denaturation step at 95°C for 15 minutes, followed by: 45 cycles of 95°C for 20 seconds, 56°C for 30 seconds and 72°C for 1 minute, followed by a final extension step of 72°C for 3 minutes. A number of PCR products were then selected at random, including at least one negative control, and were visualised on a 1.8% agarose gel stained with SafeView Nucleic Acid Stain (5%, NBS Biologicals) to ensure that the PCR had been successful and was not contaminated. Following PCR, 5µl of each product was transferred into a 384 well plate and this plate was given to the Sequenom service provider for further processing.

In brief, post-PCR, shrimp alkaline phosphatase (SAP) is transferred into each well of the 384 well plate. This removes non-incorporated dNTPs from

amplification products when incubated at 37°C for 50 minutes. The SAP enzyme is then inactivated by incubating at 85°C for 20 minutes.

An extension reaction mix is then prepared, which contains an extension primer for each SNP of interest, di-deoxynucleotide terminators and polymerase enzyme. The post-PCR primer extension reaction generates allele-specific DNA products that, based on their unique mass values, allow two alleles at a site of interest to be discriminated.

The extension reaction mix is transferred into the SAP-treated PCR plates. A SpectroCLEAN cationic resin is used to remove salt particles from the extension reaction. 15nl of the extension product from each well of the sample plate is then spotted onto a 384 format SpectroCHIP. The SpectroCHIPs are loaded into the MALDI-TOF mass spectrometer. Each spot is then excited with a laser under vacuum conditions and sample molecules are vaporised and ionised. They are then transferred electrostatically into a time-of-flight mass spectrometer (TOF-MS), where they are separated from the matrix ions, individually detected based on their mass-to-charge ratios, and analysed. Detection of an ion at the end of the tube is based on its flight time, which is proportional to the square root of its mass-to-charge ratio. This process generates data for each sample, and from this, a genotype can be derived[214].

2.2.9.4 SEQUENOM DATA MANAGEMENT AND QUALITY CONTROL

Case-control association study genotype data from Sequenom providers were returned in a Microsoft Excel spreadsheet. Failed genotypes were recorded as 0, while successful genotypes were shown as A (indicating a homozygous genotype of AA), B (indicating a homozygous genotype of BB) and AB indicating a heterozygote.

Initially, data quality control checks were carried out. All non-template control (water blank) samples were assessed to ensure that these did not produce genotype results suggesting a contaminated PCR reaction. The results of

duplicate samples were then checked to ensure that the genotype results per SNP replicated, thus ensuring assay fidelity. Replicate samples were then removed from the data set. Sequenom summary plots were reviewed to ensure that distinct clusters of genotypes were seen (Figure 7).

Data were then sorted and analysed using the filter function to display genotyping results for each SNP individually. For each SNP, a count was made of each genotype (AA, AB, BB) in cases and then in controls. Genotype frequencies were then used to calculate percentage call rates for each SNP:

Genotyping call rate = AA+AB+BB/sample size x 100

SNPs with a call rate of less than 95% were excluded from further analysis. Allele frequencies were then calculated from the genotype frequencies

Allele A frequency = 2(AA) + AB

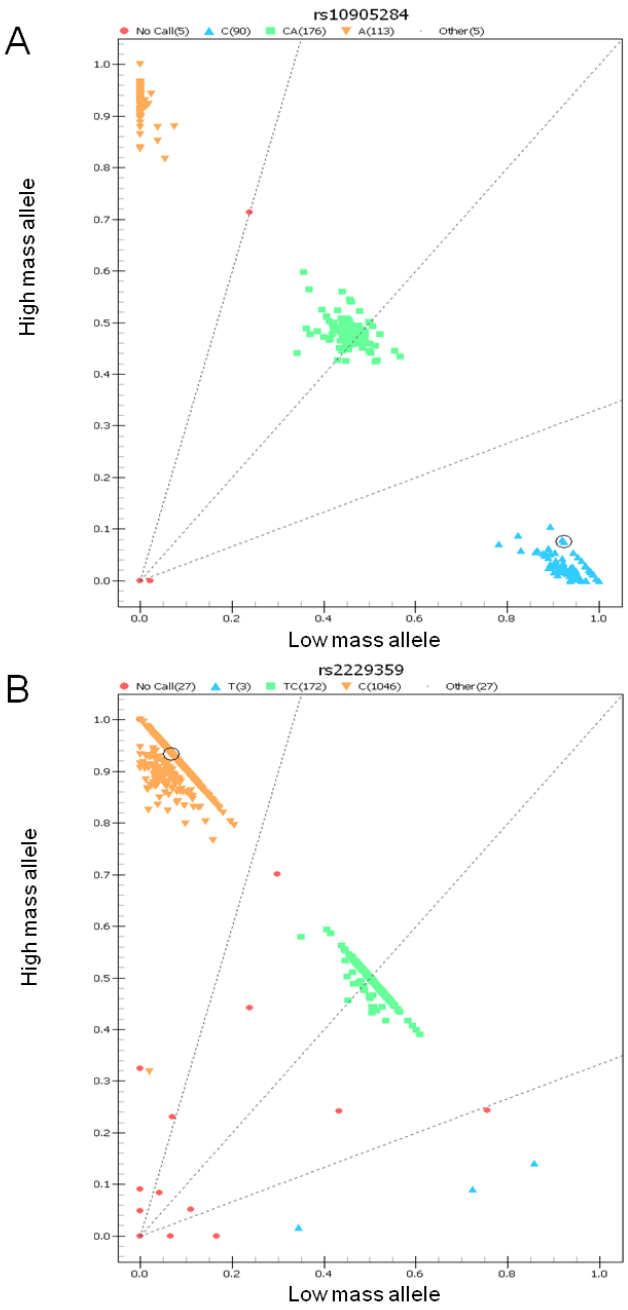


Figure 7: Sequenom cluster genotype plots.

Cluster plots for two GATA3 SNPs genotyped in 335 UK AAD patients. The plot for SNP rs10905284 is shown in panel A, the plot for rs2229359 in panel B. Each genotype is represented by a coloured dot. Panel A shows an example of a reliable Sequenom assay, with three distinct clusters of genotype calls (orange = AA, blue = CC, green = CA) and few samples falling between the groups (shown as failed genotypes in red). In contrast, panel B shows an assay which has not worked as well on this occasion, with more failed calls and less distinct clusters, particularly for the less common TT genotype (shown in blue). The results for this assay were excluded from analysis.

Genotype data from the control cohorts at each SNP were assessed for deviation from HWE using a free access computer program available at http://www.tufts.edu. The HWE principle states that genotype and allele frequencies in a population remain constant from generation to generation provided no disrupting factors are introduced, for example non-random matings or mutations. Therefore, if a single locus has three possible genotypes, AA, AB and BB, and the frequencies of A and B are expressed as *p* and *q* respectively, genotype data in HWE should obey the formula:

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

The above computer program performs a chi squared (χ^2) test of independence (as genotype data is nominal), with 1 degree of freedom, comparing observed genotype data in controls with expected genotypes under the HWE principle. A P value is then derived from the χ^2 distribution, with the P value representing the probability that the deviation of the observed from that expected is due to chance alone. Any SNP in any control cohort with a HWE P value of <0.01 was excluded. In genetic studies, significant deviation from HWE often indicates a small sample size or a genotyping problem.

2.2.9.5 STATISTICAL ANALYSIS OF SEQUENOM GENOTYPE DATA

Following all quality control checks, frequencies of genotypes and alleles in cases and controls at each SNP were compared to look for differences. This was done in one of two ways depending on the structure and format of the raw data. For some analyses, contingency tables were constructed. For genotypes, 3 x 2 contingency tables were used (2 degrees of freedom) and for alleles, 2 x 2 contingency tables were used (1 degree of freedom). χ^2 statistics and P values, for both genotypes ($P_{genotype}$) and alleles (P_{allele}), were then calculated using an online calculator (http://www.obg.cuhk.edu.hk). Alternatively, PLINK was used to perform association tests[217]. PLINK generates a genotypic P value from a 3 x 2 χ^2 test and a Cochran-Armitage allelic test P value. Results were considered indicative of association if the P value was <0.05. All P values are reported

uncorrected for multiple testing. The two analysis methods were compared and were found to give comparable results.

Odds ratios (OR) and 95% confidence intervals were calculated using an online calculator (http://www.hutchon.net/ConfidOR.htm), which uses the formula:

OR = (*a* x *d*)/(*b* x *c*)

where *a* and *c* are the numbers of patients and controls with the risk allele A respectively, and *b* and *d* are the numbers of patients and controls with the alternative B allele[218, 219]. All odds ratios are stated in respect to the minor allele. Where the appropriate demographic data were available, AAD cohorts were subdivided into individuals with AAD alone (isolated AAD; iAAD) and those with AAD in conjunction with another autoimmune condition (APS2). Additionally, subgroup analyses were performed using 21OH autoantibody status where this information was available.

In some analyses, to determine whether there was significant genetic heterogeneity between the cohorts, control cohort allele frequency data between countries were compared using a 2 x 2 χ^2 test. The allele frequencies between cohorts were considered significantly different, suggesting genetic heterogeneity, if the P value was <0.05.

Where appropriate, haplotype analysis was performed using the UNPHASED software package[220]. For this, rare haplotypes (<5%) were excluded, as these can result in spurious results. An initial basic association analysis is performed and the most associated marker identified. The analysis is then conditioned upon this marker and the next most associated marker selected in sequential order, until all association has been accounted for. A haplotype analysis is then performed, using the most common haplotype as the “reference” and an overall estimate of association calculated.

2.2.9.6 META-ANALYSIS

The RevMan 5 software package (The Nordic Cochrane Centre[221]) was used to perform meta-analysis, using a conservative random effects model to allow for heterogeneity. Should a correction for multiple testing be required, an adapted method of the Bonferroni correction was applied using the formula:

α/n

where α is the desired significance level (e.g. 0.05) and n is the number of independent tests carried out on a data set (the number of hypotheses tested)[222, 223].

2.3 *CYP21* LOCUS METHODS

2.3.1 *CYP21A1P* GENOMIC DNA DELETION

A competitive PCR assay to differentiate the *CYP21A2* gene from the *CYP21A1P* pseudogene (Figure 8) had already been established by a previous member of the laboratory group, Alekhya Narravula. To establish this assay, primers on either side of the 8bp deletion that differentiates *CYP21A1P* from *CYP21A2* were designed in order to amplify both sequences from gDNA in the same reaction. The forward primer sequence (5'–3') was TCCTCCTGCAGACAAGCTG and the reverse primer sequence (5'–3') was CTTCTTGTGGGCTTTCCAGA (Tm 60ºC).

Primer pairs were checked for specificity using both the University of California Santa Cruz (UCSC) *in silico* PCR design tool (http://www.genome.ucsc.edu) and the National Center for Biotechnology Information (NCBI) nucleotide BLAST tool (http://www.ncbi.nlm.nih.gov). Primers were ordered from Eurofins MWG Operon.

The PCR reaction volume was 25µl and contained 1X PCR buffer (Qiagen), 0.15mM $MgCl_2$ (Qiagen), 0.1mM dNTPs (New England Biosciences), 1.6µM each primer (Integrated DNA technologies) and 0.75u HotStar Taq DNA polymerase (Qiagen). The thermal cycling conditions included an initial denaturation step at 94°C for 15 minutes, followed by 35 cycles of 94°C for 30 seconds, 54°C for 30 seconds and 72°C for 30 seconds, followed by a final extension step of 72°C for 10 minutes. The PCR products were run out, next to a 100bp ladder, on a 3.8% agarose gel at 100V for 90 minutes and visualised using a UV light. The 85bp *CYP21A2* product could be distinguished by size from the 77bp *CYP21A1P* product.

This assay had been previously used to genotype 295 AAD patients and 299 healthy controls for the presence or absence of the *CYP21A1P* pseudogene. I

used this assay to study an additional 20 individuals with AAD, 447 people with Graves' disease and 328 controls. 2 x 2 contingency tables were created in order to analyse the results. A P value and odds ratio were generated using a χ^2 test.

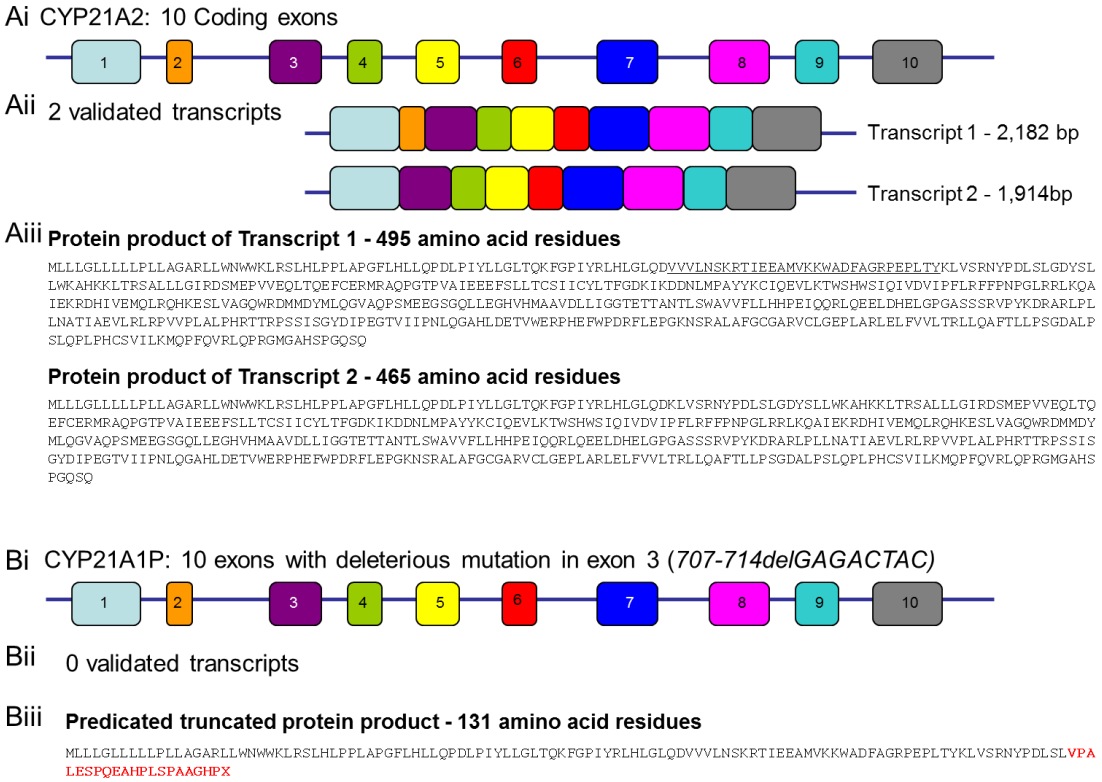


Figure 8: Schematic of the CYP21A2 and CYP21A1P genes, their validated transcripts and their known and predicted protein products.

The CYP21A2 gene is shown in panel Ai, with exons indicated by the coloured boxes marked 1 to 10 inclusive. The two validated transcripts are shown in panel Aii and the protein products derived from translation of these transcripts is shown in panel Aiii. The underlined amino acid residues in panel Aiii indicate the difference between the amino acid sequence for the protein product derived from transcript 1 compared to that derived from transcript 2. A schematic representation of the CYP21A1P pseudogene is shown in panel Bi. There is a deleterious mutation in exon 3 labelled 707-714delGAGACTAC which differentiates it from the gene. There are no validated transcripts for the pseudogene however the predicted amino acid sequence for a truncated protein product produced from the frameshift mutation and premature stop codon introduced by the deletion in exon 3 is shown in panel Biii. Those amino acids in Biii which differ from those in panel Aiii due to the frameshift mutation are shown in red, bold text.

2.3.2 TISSUE IN SITU HYBRIDISATION TO ESTABLISH *CYP21A2* AND *CYP21A1P* EXPRESSION

2.3.2.1 OVERVIEW

Tissue *in situ* hybridisation (TISH) is a technique used to study gene expression in fixed tissues. Tissues must first be prepared by removing paraffin, retrieving the antigens and permeabilising tissue. Tissues can then be exposed to labelled probes designed to be specific and complementary to a messenger RNA (mRNA) or a microRNA sequence of interest. In this study, digoxigenin (DIG) labelled probes were used. A full length RNA probe, synthesised from a *CYP21A2* cDNA template, was designed to hybridise to both *CYP21A2* and *CYP21A1P* mRNA transcripts, in order to demonstrate their presence or absence in the tissues of interest. In addition, two short locked nucleic acid (LNA) probes were designed, one to hybridise specifically to *CYP21A2* and another to hybridise to *CYP21A1P,* allowing us to determine whether both transcripts were expressed in tissues of interest, and to study their specific distributions. LNAs are chemically modified nucleotides which contain an additional 2'–O,4'–C–methylene bridge within the ribose ring[224, 225]. LNA probes are oligonucleotides in which LNAs replace ribonucleic acids at specific intervals along their length. The introduction of LNAs with their modified, rigid ribose ring changes the properties of the probe, conferring high binding affinity to complementary sequence and reduced degradation by nucleases, resulting in increased stability[226]. Despite being very short compared to RNA probes, LNA probes hybridise very specifically. This means that they can be used to distinguish between highly similar transcripts, with only a single base pair difference[227, 228] and can be used to detect short sequences, such as microRNAs in tissue sections[229-231]. This makes them ideal to study the *CYP21A2* and *CYP21A1P* transcripts which are highly homologous. Following hybridisation of the probes, Fab fragments conjugated to alkaline phosphatase (AP) were used to bind to the DIG labels which were conjugated to the 5’ and 3’ bases of the LNA probe when it was synthesised. A nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate (NBT-BCIP) system was then used for

detection. BCIP is an alkaline phosphatase substrate which reacts after dephosphorylation to give a dark blue dye as an oxidation product. NBT acts as an oxidant and also gives a dark blue dye. This intensifies the colour, making detection more sensitive. In this study, fetal tissues were studied for *CYP21A2* and *CYP21A1P* expression.

2.3.2.2 TISSUE RESOURCES

For this study, adrenal tissue from subjects with and without the *CYP21A1P* pseudogene were required. To identify appropriate tissues, the Newcastle HDBR database was used to search for any specimen between 8 and 12 wpc where both a frozen skin sample, for gDNA extraction, and fixed adrenal tissue, for TISH, were available. In total, 42 potential specimens were identified. gDNA was extracted from the 42 fetal skin samples and this was used to genotype each specimen for *CYP21A1P*. 40 specimens (95.2%) were found to have the pseudogene (*CYP21A1P+)* and two (4.8%) were found to have no *CYP21A1P* (*CYP21A1P*–). Fixed and mounted adrenal and kidney sections from two *CYP21A1P*+ samples (N1822, 10 wpc; N1593, 12 wpc) and from one *CYP21A1P*– sample (N1643, 11 wpc) were therefore selected for the TISH experiment. In addition, central nervous system sections from two samples (N1481, 9 wpc; N646, 8 wpc) were available for use with a positive control LNA probe, which hybridises strongly and specifically to an abundant microRNA (hsa–miR–124) expressed exclusively in neuronal tissue. The use of this probe on these neuronal sections was to act as a positive control for the LNA probe technique optimisation. In addition, other specific controls were used: to determine if the CYP21A2 LNA probe was hybridising specifically, limb and cardiac tissue sections were available from two 12wpc fetuses (N1593 and N1782 respectively) for use as control tissues, as these are not expected to express *CYP21A2*. A sonic hedgehog mRNA riboprobe, which hybridises specifically to bone, was used as a positive control for this experiment.

2.3.2.3 TISH PROTOCOL

TISH was performed through the “In House Gene Expression” service available from the Human Developmental Biology Resource (http://www.HDBR.org), Newcastle. Initially, I designed and ordered primer pairs (Integrated DNA Technologies) (Table 3), with SP6 and T7 extensions, to anneal to cDNA from both *CYP21A2* and *CYP21A1P.* These were supplied to HDBR.

A PCR was performed, using 100ng of whole cDNA from a Carnegie Stage 22 embryo as a template. The PCR reaction volume was 50µl and contained 1X GoTaq PCR buffer (Promega), 0.5µM dNTPs (New England Biosciences), 0.5µM each primer (Integrated DNA technologies) and 1.25u GoTaq DNA polymerase (Promega). The thermal cycling conditions included an initial denaturation step at 94°C for 5 minutes, followed by 30 cycles of 95°C for 1 minute, 55°C for 1 minute and 74°C for 1 minute, followed by a final extension step of 74°C for 10 minutes. PCR products were run out on a 1% agarose gel for approximately 60 minutes at 80V. Amplicons were produced for primer pairs 2 and 3 only. The product for primer pair 2 was strongest and therefore this amplicon was cut out and extracted from the gel using a QIAquick Gel Extraction Kit (Qiagen). This purified PCR product was then diluted 1/10, 1/100 and 1/1000. 1µl of each dilution was then used as the template for a second round of PCR, using the above protocol. The PCR products were again run out on a 1% agarose gel.

The strongest band was cut out and extracted using the QIAquick Gel Extraction Kit (Qiagen) and quantified using the Nanodrop. This PCR DNA was then used to set up two *in vitro* transcription reactions using an RNA DIG Labelling Kit (SP6/T7) (Roche) to produce an SP6 (antisense) riboprobe and a T7 (sense) riboprobe.

Primer pair name	Sense probe (T7) (5'-3')	Antisense probe (SP6) (5'-3')	T_m (°C)	Product size (bp)
Pair 1	TAAGTTAATACGACTCACTATAGGGCGACAGGCTCCACCTTGGGCTGC	AATACGATTTAGGTGACACTATAGAATACTCTTCCATGCTCGGCTGCGC	60	636
Pair 2	TAAGTTAATACGACTCACTATAGGGCGAAAGCTCACCCGCTCAGCCCT	AATACGATTTAGGTGACACTATAGAATACACCAGGGCCCAGTTCGTGGT	60	624
Pair 3	TAAGTTAATACGACTCACTATAGGGCGAGAGCGCATGAGAGCCCAGCC	AATACGATTTAGGTGACACTATAGAATACGCGGTGGGGCAAGGCTAAGG	60	657

Table 3: Primer pairs for TISH.

Three primer pair combinations for TISH. The T7 sequence is shown in blue text and the SP6 sequence in red text.

In a 20µl reaction volume, 75ng of PCR DNA was added to 1X Transcription buffer and 1X DIG labelling mix. 20u of RNase inhibitor was then added, with 40u of either SP6 or T7 RNA polymerase. This was incubated at 37°C for 2 hours. 20u of DNase I was then added and the solution mixed thoroughly before incubating at 37°C for 15 minutes. ProbeQuant G-50 Micro Columns (GE Healthcare) were then placed in a centrifuge and spun at 735g for 1 minute.

Following incubation with DNase I, the *in vitro* transcription reaction solutions were made up to 50µl with nuclease-free DEPC water and added to the ProbeQuant G-50 columns. These were centrifuged at 735g for 2 minutes to remove unincorporated NTPs. The Nanodrop was then used to quantify RNA produced. A 2% RNA formamide gel was then made by adding 36mls of DEPC sterile distilled water to 5mls of 10 x MOPS (3-(N-morpholino)propanesulfonic acid)-EDTA-sodium acetate buffer (400mM MOPS, 100mM sodium acetate, 10mM EDTA, pH 8.3. Sigma). 1g of RNAse free agarose (Invitrogen) was added and the solution weighed. This was then heated in a microwave for 90 seconds to dissolve the agarose. DEPC sterile distilled water was then added to make the solution up to the pre-heated weight (41 mls total volume). In a fume hood, 9mls of 37% formamide (Sigma) was added to make the total volume to 50mls. This was then poured into a gel tray, with a comb *in situ* to form wells, and left in a hood to set.

1µl of each product was then added to loading dye (1X) and heated at 70ºC for 10 minutes prior to loading. Samples were run on the gel at 50V for 2 hours in 1X MOPS-EDTA-Sodium acetate buffer and visualised by UV to determine that they were the appropriate size. The first reaction, with SP6, produced an “antisense” probe complementary to *CYP21A2* and *CYP21A1P* mRNA for use in the TISH protocol. The second reaction was set up with T7 in place of SP6, producing a “sense” probe, for use as a negative control in the TISH protocol. The SP6 and T7 riboprobes were then carefully labelled and stored at -80ºC for future use.

Prior to starting the TISH protocol, all solutions were made using DEPC-treated water or phosphate buffered saline (PBS). All glass/steelware were baked at 180ºC for at least 4 hours before use. For each reaction, a T7 probe slide was included as a negative control.

Slides were placed in metal racks and paraffin removed by soaking in Xylene for 5 minutes x 3, Xylene/Ethanol 1:1 for 3 minutes, 100% ethanol for 3 minutes x 2, 90% ethanol for 3 minutes, 70% ethanol for 3 minutes and 50% ethanol for 3 minutes. Slides were then washed in PBS for 2 minutes, the PBS discarded and fresh PBS added for a further 2 minute wash.

To permeabilise tissues, the slides were left in a PBS solution containing Proteinase K (20µg/ml) at 37ºC for 8 minutes. They were then rinsed for 30 seconds in PBS and the PBS discarded. Slides were then fixed for 20 minutes in 4% paraformaldehyde (PFA) in PBS. They were then washed for 2 minutes in PBS twice and the PBS discarded at the end of each wash. The sections were then left in 0.1M Triethanolamine pH 8.0, 0.25% Acetic anhydride DEPC PBS for 10 minutes before washing twice, for 2 minutes in PBS. Sections were then dehydrated using a series of ethanol washes (50%, 70%, 90%, 100% and 100% for 2 minutes each). At the end of the final ethanol wash, slides were placed in fresh DEPC 100% ethanol for 2 minutes. Slides were then air-dried by placing them in a rack covered in foil (cleaned in DEPC ethanol) in a filtered air stream for 1 to 3 hours.

To hybridise the RNA probe to mRNA, the slides were placed in a plastic slide tray which had been cleaned in DEPC ethanol. The solution containing the labelled probe was made up, using DIG Easy Hyb Mix (Roche), 3ng/µl, allowing 100µl of hybridisation solution per slide (300ng of mRNA probe was added to 100µl of Hyb mix per slide). This solution was then pipetted onto each slide and a cover slip gently applied to protect the section. A hybridisation chamber was prepared, with paper towel soaked in 2X saline-sodium citrate (SSC) (to make a 20X solution, 3M NaCl, 0.3M Na-citrate, pH 7.2) to keep the slides moist. The slides were then placed into the chamber and left at 64ºC overnight.

Post-hybridisation, the cover slips were removed by rinsing the slides in pre-warmed 5X SSC (60ºC). Slides were then placed in a black (to avoid gradient staining) plastic slide rack and washed in 5X SSC for 10 minutes at 60ºC, 5X SSC for 10 minutes at 60ºC, 2X SSC for 10 minutes at 60ºC, 2X SSC for 10 minutes at 60ºC. Slides were then allowed to cool to room temperature. For antibody detection, slides were washed in buffer 1 (0.1M Tris (pH 7.6), 0.15M NaCl, made up to 1L with distilled water) on a rocker for 10 minutes at room temperature x 3. Each slide was then covered in 10% fetal calf serum (FCS, previously heat inactivated at 58ºC for 30 minutes in buffer 1) for 1 hour. 150µl of anti-DIG AP mix (diluted 1:1000 – 2% FCS/buffer 1 - Roche) was then pipetted onto each slide. Slides were then covered with a fresh parafilm coverslip. The slide tray was moistened with paper towels soaked in buffer 1 and left overnight at 4ºC.

The next day, slides were washed in buffer 1 on a rocker for 10 minutes at room temperature x 3. They were then equilibrated in buffer 2 (0.1M Tris (pH 9.5), 0.1M NaCl, make up to 1L with distilled water) on a rocker for 5 minutes at room temperature x 3 and placed in a slide tray. An NBT-BCIP (20µl/ml - Roche) solution in buffer 2 was made up in a foil covered falcon tube (light sensitive). Each slide was then flooded with this solution and the slide tray covered in foil to protect from the light. Slides were then left to develop in a dark room and checked regularly for staining under the microscope. Fresh solution was added to avoid the slides drying out if needed. Development typically took from 30 minutes to several days. Once staining is detected, to stop the reaction, slides were rinsed in buffer 2 and then in several changes of distilled water. Sections were then mounted in Aquamount (Fisher Scientific) and the edges of the coverslips painted in clear varnish to seal.

The above protocol was also adapted for use with *CYP21A2* and *CYP21A1P* LNA probes, designed and made by Exiqon. The probe sequences are shown in Table 4.

Probe name	Probe sequence
CYP21A2 LNA probe	/5DIG_N/TCCAGAGCAGGGAGTAGTCT/3DIG_N/
CYP21A1P LNA probe	/5DIG_N/AGCAGAGACCAACGACAG/3DIG_N/
hsa-miR-124 positive control LNA probe	/5DIG_N/GGCATTCACCGCGTGCCTTA/3DIG_N/
Scramble-ISH negative control LNA probe	/5DIG_N/GTGTAACACGTCTATACGCCCA/3DIG_N/

Table 4: LNA probe sequences.

Sequence information for the four end-labelled LNA probes used for TISH.

For each experiment, a scrambled probe, not known to hybridise to any known human RNA or micro-RNA sequence (Scramble–ISH, Exiqon) was used to act as a negative control. A positive control probe known to specifically hybridise to neuronal tissue (has-miR-124, Exiqon) was also used.

The has–miR–124 control probe was used at 20nM concentration as per instructions. The Scramble-ISH probe was used at equivalent concentrations to the *CYP21A2* and *CYP21A1P* probes. For the *CYP21A2* and *CYP21A1P* LNA probes, a titration series was first used to determine the optimum concentration for each of the two probes in the hybridisation solution. Probes were tried at a concentration of 10nM, 20nM, 40nM, 100nM and 250nM.

LNA probes were left to hybridise overnight at 53ºC as per the manufacturer's instructions. Various combinations of hybridisation temperature (50ºC to 56ºC) were used to optimise conditions. The stringency of the post-hybridisation washes was also adjusted. In addition to the medium stringency washes (as above), a high stringency wash protocol of 1X SSC followed by two washes of 0.1X SSC was used to minimise non-specific hybridisation signals. The remainder of the TISH protocol was unchanged.

2.3.3 NORTHERN BLOTTING

2.3.3.1 OVERVIEW

Northern blotting was developed to study expression of transcripts. Initially, RNA samples in a tissue of interest are separated by size using gel electrophoresis. The RNA may then be transferred onto a blotting membrane for detection by probes specific and complementary to the sequence within the transcript of interest. In this study, Northern blotting was performed with the aim of confirming that the SP6 antisense mRNA riboprobe and CYP21A2 and CYP21A1P LNA probes were hybridising to an appropriate sized transcript.

2.3.3.2 NORTHERN BLOTTING PROTOCOL

Prior to starting the Northern blotting protocol, all solutions were made using DEPC treated water and all glass/steelware was baked at 180ºC for 4 hours before use. A 2% RNA formamide gel was then made (see TISH protocol). 400ng adrenal RNA samples were prepared with 1X RNA loading dye (Fermentas) and each sample made up to 12µl with DEPC water. In addition, a ladder was prepared allowing 0.5µl of ladder for every mm lane width e.g. 4µl for an 8mm wide lane. Two ladders were tried: the unlabelled Riboruler High Range ready-to-use RNA ladder (Fermentas) and the DIG labelled RNA molecular weight marker (Roche). Each sample and ladder were heated at 70ºC for 10 minutes and then put immediately onto ice for 2 minutes prior to loading.

Electrophoresis was carried out in 1X MOPS-EDTA-Sodium acetate buffer at 50v for 2 hours. Following electrophoresis, the ladder lane was cut out and soaked in a 3X Gel Red nucleic acid stain solution (Biotium, made up to 3X from a 10000X stock with sterile, distilled water) in the dark for 15 minutes and then visualised using UV to check that the bands of the ladder had adequately separated prior to transfer. In the case of the unlabelled ladder, bands were marked on the gel using a scalpel.

A transfer stack was constructed (Figure 9) and transfer was performed overnight. 30 minutes prior to transfer, Hybond N nylon membrane (GE Healthcare Life Sciences) was soaked in 20X SSC (3M NaCl, 0.3M sodium citrate). 20X SSC solution was poured into a dish and an inverted gel tray was placed into this. A sheet of filter paper (Whatman, GE Healthcare) 20cm x 10cm was folded around the gel tray so that the ends extended into the SSC solution. 2 sheets of filter paper were then placed onto this and the gel placed on top of the filter paper. The nylon membrane was placed carefully onto the gel and any bubbles removed by gently rolling a glass pipette over the surface. At this point, the gel with ladder lanes marked was lined up with the unmarked gel and the ladder lanes marked on the membrane with a scalpel to allow sizing of any band seen.

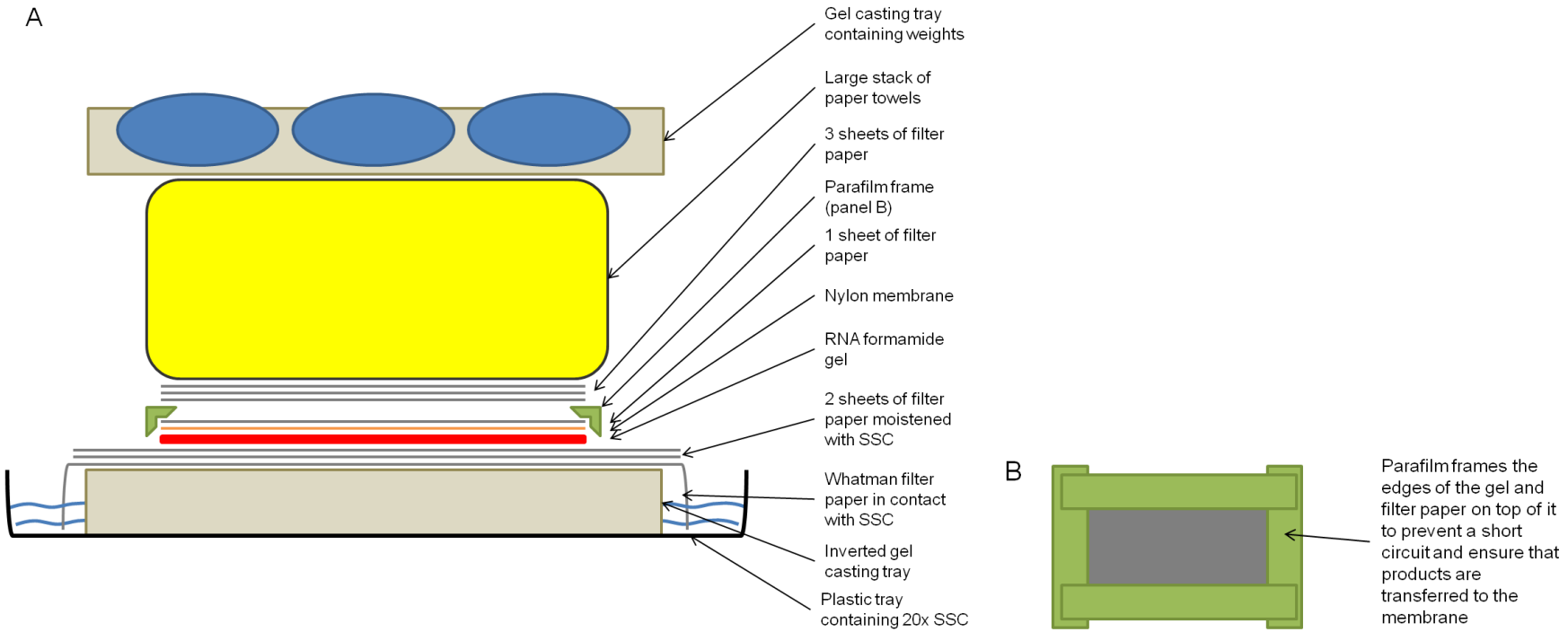


Figure 9: The components of the Northern blotting stack.

Panel A shows the stack components for overnight transfer. Panel B illustrates the parafilm frame that forms part of the stack.

One sheet of filter paper, cut to the same size as the membrane, was placed on top of the membrane and then a frame of parafilm was placed around the edges of this. Three further sheets of filter paper were placed on top of this and then a stack of absorbent paper towels were stacked on top. Finally, a weighted plate was placed on top to add pressure. The stack was covered in clingfilm and left overnight.

Following transfer, the membrane was cut into separate lanes so that each RNA sample was on a separate piece of membrane. Each lane had ladder bands marked by scalpel for product sizing. The top of the membrane was marked with a scalpel to facilitate orientation. While the membrane was still damp, the RNA was fixed to it by UV cross-linking at 120mJ for 1 minute (BioRad Genelinker). Specimen bags and weigh boats were then labelled to identify each sample. Dig Easy Hyb Mix (Roche), 3ng/µl, was made up, allowing 10ml of hybridisation solution per 1 x 10cm blot. This hybridisation buffer was warmed to hybridisation temperature (68ºC for RNA probes, 53ºC for LNA probes). Each piece of membrane was placed into a labelled sealable specimen bag and 10ml of Dig Easy Hyb Mix was added per 1 x 10cm strip of membrane. This was left at hybridisation temperature for 30 minutes and then poured off.

Probes were made up in 10ml of the hybridisation solution. 150ng of the T7 and SP6 probes were used while a range of LNA probe concentrations were tried, from 20 and 100nM. All probes were heated to 95ºC for 5 minutes and then put onto ice. Each probe was then added to 10ml of Dig Easy Hyb mix and this solution was then added to the appropriate specimen bag containing the membrane. The LNA probes were placed in a shaking water bath at 53ºC overnight while the mRNA probes were incubated at 68ºC overnight.

Following overnight hybridisation, the DIG Easy Hyb mix was poured off and the membranes washed. All wash steps were carried out at 53ºC for LNA probes and 68ºC for the RNA probe. Membranes were washed in 5X warmed SSC for 10 minutes, 5X warmed SSC for 10 minutes, 2X warmed SSC for 10 minutes. The final wash was carried out at room temperature with 2X warmed SSC for 10

minutes. The membranes were then equilibrated with buffer 1 by carrying out three 5 minute washes on a shaker.

1X Blocking solution (Life Technologies, Invitrogen), made up from 10X concentrate with buffer 1, was then used to block the membrane, by soaking at room temperature for 30 minutes. Membranes were then washed for 5 minutes in buffer 1. The Anti-digoxigenin-AP antibody (Roche) was then diluted 1 in 5000 in buffer 1 (1µl in 5 ml). This was added to the washed membranes in bags, allowing 10 ml per 1x10cm blot and incubated at room temperature for 30 minutes.

Following incubation with the antibody, the blots were washed on a shaker 3 x for 5 minutes each in buffer 1. The blots were then washed in buffer 2 on a shaker for 5 minutes. For detection, NBT/BCIP (Roche) 20µl/ml solution in buffer 2 was made up in a foil covered falcon tube (light sensitive). This solution was then added to the blots, allowing 10ml per 1x10cm blot and these were then left in the dark for 30 minutes at room temperature. To stop staining, membranes were rinsed in buffer 2 and then in several changes of distilled water.

2.3.4 FLUORESCENCE-BASED QUANTITATIVE REAL-TIME PCR TO DETERMINE *CYP21A2* AND *CYP21A1P* EXPRESSION

2.3.4.1 OVERVIEW

Fluorescence-based quantitative real-time PCR (qPCR) can be used to monitor the progress of the PCR reaction in “real time” by measuring fluorescence produced during DNA amplification. The basis for this technique is that, early in the PCR reaction, when little amplification has occurred, there is little fluorescent signal released. The baseline for the amplification plot is therefore defined early in the reaction. As amplified PCR products accumulate, the fluorescent signal increases and can be detected above the baseline. A fixed

fluorescence threshold can be set arbitrarily above the baseline and the cycle number at which the fluorescent signal crosses this threshold determines the threshold cycle (C_T). The point at which the target is first detected during the PCR cycle can therefore be monitored. The higher the starting copy number of the nucleic acid target, the earlier an increase in the fluorescent signal is detected and the lower the C_T value will be. Conversely, if only tiny amounts of target nucleic acid are present in the tested sample, a high C_T will be observed. In these reactions, fluorescence is provided by fluorogenic probes and there are many of these available, both specific and non-specific. In this project, qPCR assays specific to *CYP21A2* and *CYP21A1P* were designed by Primerdesign, Southampton and included specific primers (Table 5, Figure 10) and double-dye hydrolysis (Taqman-style) probes. The aim of qPCR in this study was to determine whether *CYP21A2* and *CYP21A1P* were expressed in human thymus samples through the use of two FAM-labelled probes specific to *CYP21A2* and *CYP21A1P* respectively, using cDNA as the PCR template, an absolute quantification method and human adrenal tissue to optimise the protocol.

2.3.4.2 TISSUE RESOURCES

Following genotyping of fetal skin samples for the presence or absence of *CYP21A1P*, RNA was extracted from three fetal adrenal glands available from Newcastle HDBR (one *CYP21A1P*– specimen, N1733, 11wpc; two *CYP21A1P*+ specimens, N1736, 9wpc and N1812, 12wpc) and cDNA was synthesised. In addition, seven thymus glands from infants and children (*CYP21A1P*+) were available for analysis from Birmingham. Finally, matched kidney samples from N1812, N1736 and N1733 were available for analysis.

2.3.4.3 METHOD

2.3.4.3.1 STANDARD CURVE PRODUCTION

A standard curve was produced using a positive control template for *CYP21A2* and *CYP21A1P* respectively. These control templates were produced by Primerdesign. The undiluted *CYP21A2* control (positive control tube 1) contained 200,000 copies and was used to produce a series of dilutions (Table 6). A series of dilutions was then made using the *CYP21A1P* positive control in the same manner.

Assay	Sense primer	Anti-sense primer	Tm (ºC)	Product length (bp)
CYP21A2	CAAGAGGACCATT GAGGAAGC	TCCAGAGCAGGG ACTAGTCTC	57	131
CYP21A1P	CGGACCTGTCGTT GGTCTC	CTCACAGAACTCC TGGGTCA	57	123

Table 5: Primer sequences for the CYP21A2 and CYP21A1P qPCR assays.

qPCR assays were designed and made by Primerdesign, Southampton.

CATCTACAGGCTCCACCTTGGGCTGCAAGATGTGGTGGTGCTGAACTC***CA***
AGAGGACCATTGAGGAAGCCATGGTCAAAAAGTGGGCAGACTTTGCTGG
CAGACCTGAGCCACTTACCTACAAGCTGGTGTCTAGGAACTACCCGGACC
TGTCNTTGG***GAGACTACTCNCTGCTCTGGA***AAGCCCACAAGAAGCTCACC
CGCTCAGCCCTGCTGCTGGGCATCCGTGACTCCATGGAGCCAGTGGTGG
AGCAGCTGACCCAGGAGTTCTGTGAGCGCATGAGAGCCCAGCCCGGCAC
CCCTGTGGCCATTGAGGAGGAATTCTCTCTCCTCACCTGCAGCATCATCTG
TTACCTCACCTTCGGAGACAAGATCAAGGACGACAACTTAATGCCTGCCTA
TTACAAATGTATCCAGGAGGTGTTAAAAACCTGGAGCCACTGGTCCATCCA
AATTGTGGACGTGATTCCCTTTCTCAGGTTCTTCCCCAATCCAGGTCTCCG
GAGGCTGAAGCAGGCCATAGAGAAGAGGGATCACATCGTGGAGATGCAG
CTGAGGCAGC

Figure 10: cDNA sequence of CYP21A2 illustrating qPCR primer positions.

Exon-exon boundaries are shown underlined. The 8bps that are deleted in CYP21A1P are shown highlighted in pink. The CYP21A2 assay primer pair is shown in bold italic green text. The CYP21A1P assay primer pair is shown in red text. The forward primer for the CYP21A1P assay overlaps with the reverse primer of the CYP21A2 assay (shown in red, bold text). The CYP21A2 assay spans the exon-exon boundary however the CYP21A1P assay does not.

Standard curve	Absolute copy number
Positive control tube 1	200 000
Positive control tube 2	20 000
Positive control tube 3	2 000
Positive control tube 4	200
Positive control tube 5	20
Positive control tube 6	2

Table 6: Standard curve production for qPCR absolute quantification.

2.3.4.3.2 qPCR PROTOCOL

cDNA samples were prepared from DNase-treated RNA samples, using 5µg of RNA per reverse transcription reaction. Prior to use, each cDNA sample was diluted 1 in 10 with nuclease-free water. Each real-time PCR reaction was set up in a 384-well optical plate (Starlab) on ice. Standards and samples for each gene were run in triplicate. Water blanks were also run for each gene as controls. The *CYP21A2* assay spans the exon-exon boundary, however the *CYP21A1P* assay could not be designed to be intron-spanning in the same way. Therefore RNA samples were run for each sample and each gene to ensure that starting material was not contaminated with gDNA. 20µl reaction volumes were used and each reaction contained 1X Primerdesign Precision qPCR Master Mix (0.25u Taq polymerase, 0.03mM MgCl, dNTP mix: 1.25µM each), 1µl of the primer/probe mix of interest (*CYP21A2* or *CYP21A1P*, containing 300nM of each primer) and 5µl of template. Reaction volumes were made up with sterile, distilled, nuclease-free water. The PCR plate was then sealed using optical PCR film (Starlab). The qPCR reactions were carried out using a 7900HT real-time PCR system (Applied Biosystems) and cycled at 95ºC for 10 minutes, then 50 cycles at 95ºC for 15 seconds and 60ºC for 1 minute. Data was collected during each cycle during the 60ºC step through the FAM channel. Results were then analysed using the 7900HT Sequence Detection System software (Applied Biosystems). An absolute quantification method was used to determine *CYP21A2* and *CYP21A1P* copy numbers/µl of sample. Each standard and sample were run in triplicate producing 3 C_T measurements for each. The coefficient of variation was calculated for each set of replicates:

$$CV = SD/\text{mean } C_T \times 100$$

If this was less than 5%, the results were accepted. If the CV was greater than 5%, the sample was rejected. Samples where a single replicate failed, giving an “undetermined” result, were accepted if the CV of the two remaining samples was less than 5%. Samples where two of the three replicates failed were rejected.

An initial quality control experiment was conducted. For this, a standard curve for both *CYP21A2* and *CYP21A1P* was produced by plotting the triplicate C_T values on the Y axis against the $\log_{10}$ of the starting copy number on the X axis. To determine primer efficiencies, the C_T values for each standard sample, for each assay, were plotted and the gradient of the slope for each assay was calculated by linear regression in Prism 5 (GraphPad). The slope gradient should be approximately -3.32, indicating a priming efficiency of close to 100%. The primer efficiency (E) is calculated from the gradient using the following formula:

$$E = 10^{(-1/slope)} -1 \times 100$$

In general, slopes in the range of -3.60 to -3.10 are generally considered acceptable for qPCR, corresponding to primer amplification efficiencies of between 90 and 110%. Primer efficiencies were calculated from triplicate C_T values generated from standard samples and these were replicated on three separate occasions. The standard curve for each plate was then used to determine the linear dynamic range of the assay. This range determines the limits of assay detection and sets a minimum and maximum C_T for each assay. C_T values falling outside this range are discarded.

A second quality control experiment was then conducted to test the specificity of the *CYP21A1P* assay by using it, alongside the *CYP21A2* assay, on *CYP21A1P*- and CYP21A1P+ fetal adrenal samples. This experiment aimed to ensure that the *CYP21A1P* assay was specific for pseudogene transcripts and was not erroneously amplifying the highly homologous *CYP21A2* transcripts. To this end, cDNA synthesised from a *CYP21A1P*- fetal adrenal sample (N1733, 11wpc) and from two *CYP21A1P*+ fetal adrenal samples (N1736, 9wpc and N1812, 12wpc) were used in the qPCR assay. If C_T values were generated for the *CYP21A1P*- sample using the *CYP21A1P* assay, this would suggest that the assay was non-specific. RNA samples from each specimen were also included to ensure that the RNA samples were not contaminated by gDNA.

Following quality control measures, the qPCR assays were used on samples of interest. Using the standard curve, Prism 5 software was used to “interpolate” $\log_{10}$ copy numbers for each sample, based on its C_T measurements. These were converted to copy numbers:

Copy number in assay sample = 10^x (where x is the interpolated value)

As 5µl of cDNA diluted 1 in 10 was used in the qPCR protocol, the copy number value calculated related to the number of copies in 0.5µl of undiluted sample. To allow comparison of samples and replicate results, all values were therefore converted to an absolute copy number/µl by multiplying by two.

As each sample was run in triplicate, and on at least two days, multiple absolute copy number values were generated per sample. This allowed a mean and standard error of the mean to be plotted per sample for *CYP21A2* and *CYP21A1P* expression. The mean absolute copy number results for each sample for *CYP21A1P* and *CYP21A2* could then be calculated. The mean of *CYP21A1P* copies could then be normalised to the mean of *CYP21A2* copies. The mean was used for this as the *CYP21A1P* and *CYP21A2* assays for each sample were run in separate wells, and therefore the results of individual components of the triplicates for *CYP21A1P* and *CYP21A2* were not directly comparable. The assays were used to determine *CYP21A2* and *CYP21A1P* expression in human thymus obtained from infants and children during cardiothoracic surgery.

2.4 20 CANDIDATE GENE ASSOCIATION STUDY METHODS

2.4.1 *NF-κB1* GENOTYPING BY RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP) GENOTYPING

In this method, a restriction enzyme specific to a single allele of the SNP of interest is identified using a restriction site mapping tool, for example Webcutter version 2.0 (http://rna.lundberg.gu.se/cutter2). The RFLP method was used to genotype the *rs4698861* SNP in the *NFKB1* gene in a cohort of UK Graves' disease patients (392 cases).

The genomic sequence containing the *rs4698861* SNP was found on the Ensembl database. The Webcutter 2.0 program was then used to identify an NlaIV restriction site introduced by the single nucleotide A to G substitution at this locus. Two primer pairs to amplify the SNP by PCR were designed in the Primer3 program[232]. Pair 1 gave an expected product of 390bp and pair 2 gave an expected product of 517 bp (Table 7). Each PCR was performed in a 10µl volume. Each reaction contained 200ng DNA template, 1X GoTaq buffer (Promega), 0.2mM each dNTP (New England Biolabs), 0.25µM each primer (MWG) and 0.25u GoTaq DNA polymerase (Promega). The thermal cycling conditions included an initial denaturation step at 94°C for 2 minutes, followed by 31 cycles of 94°C for 30 seconds, 60°C for 1 minute and 72°C for 1 minute, followed by a final extension step of 72°C for 10 minutes. 5µl of the PCR products were run out on a 2% agarose gel next to a 100bp ladder at 90v for 45 minutes and visualised under UV light to check that the PCR had been successful and that the correct size product was visible (Figure 11).

Primer pair 2, which gave a larger product pre-digest, was selected for further work. Following PCR, a digest was set up. The digest volume was 10µl and contained 2µl of PCR product, 1u NlaIV enzyme (New England Biolabs), 1X Buffer 4 (New England Biolabs) and 10% bovine serum albumin (New England Biolabs). The mixture was incubated at 37^0C for 1 hour. The NlaIV enzyme cuts if the G allele is present at the SNP site. The digest products were run out on a

1.0% agarose gel, next to a 100bp ladder, at 90v for 1 hour and visualised using UV light. If the AA genotype is present, 2 products of approximately 400 and 60bp are seen. If the genotype is GG, 2 products of 200 and 60bp are seen. If the AG genotype is present, 3 products of 400, 200 and 60bp are seen (Figure 11). To increase genotyping call accuracy, the post-digest PCR products were independently assessed by two individuals within the lab (Katie MacArthur and I). Any genotypes that were not agreed upon were repeated or discarded to reduce the chance of erroneous calls.

Primer name	Pair number	Primer Sequence (5'-3')	Tm (ºC)	Product size (bp)
rs4698861 forward 1	1	GCCAAAGGGATCAGAAATGA	60	390bp
rs4698861 reverse 1		CACGCTGTGTGCATATGTTG	60	
rs4698861 forward 2	2	CTTCTTTCTGCCACTTCTTTGTGT	61	517bp
rs4698861 reverse 2		GCATATGTTGCTCATCATTCAAGA	61	

Table 7: Primer sequences for genotyping *rs4698861* in the *NF-κB1* gene by RFLP.

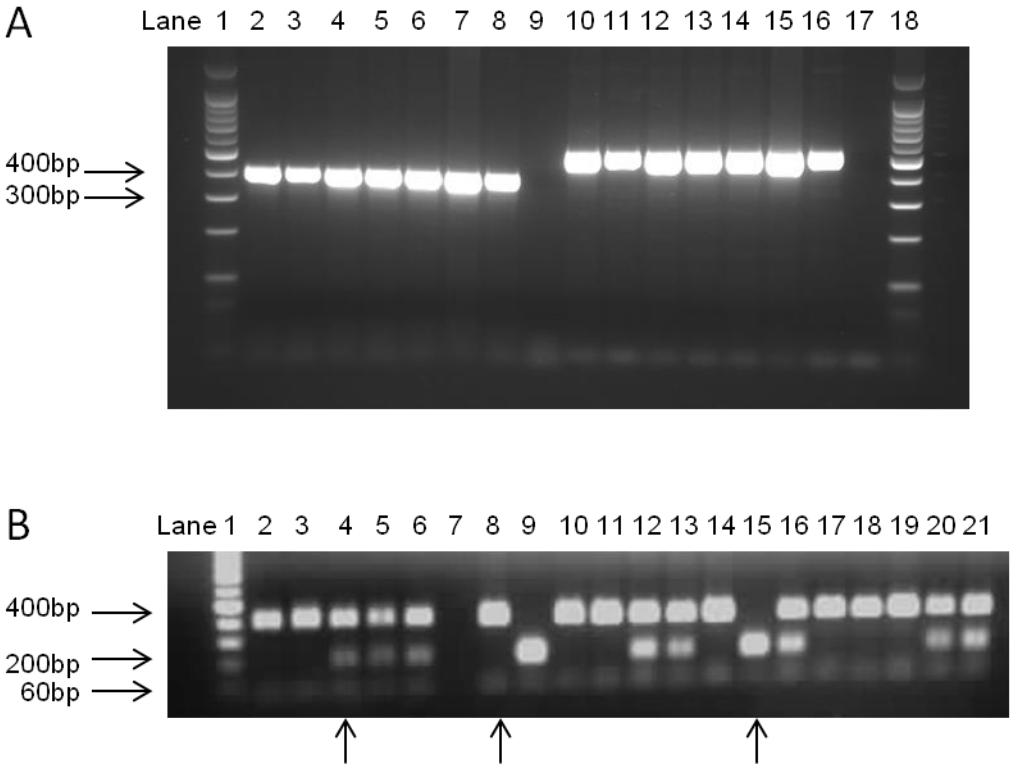


Figure 11: rs4698861 RFLP gel images.

2% agarose gel electrophoresis of PCR products pre-digest (panel A) and 1% agarose gel electrophoresis of digested PCR products (panel B). In panel A, products for NF-κB1 SNP rs4698861 primer pairs 1 and 2, designed to amplify a portion of DNA flanking rs4698861, are shown. In this image, a 100bp ladder is shown in lanes 1 and 18 and arrows indicate a product of 400 and 300bp. PCR products from primer pair 1 of 390bp are shown in lanes 2 to 8. PCR products from primer pair 2 of 517bp are shown in lanes 10 to 16. Lanes 9 and 17 contain a no template control. In panel B, post-digest products are shown for primer pair 2. A 100bp ladder is shown in lane 1 and arrows indicate a product of 400, 100 and 60bp in size. An example of a heterozygote (AG) at SNP rs4698861 is seen in lane 4. In this case, three products are seen: 400, 200 and 60bp in size. An example of a wild type homozygote (AA) is seen in lane 8 where two products, 400 and 60bp in size, are seen. Lane 15 is an example of a mutant homozygote (GG) where two products of 200 and 60bp are seen (all indicated with arrows). The sample in lane 7 is a no template control.

2.5 GENOME-WIDE LINKAGE AND ASSOCIATION STUDY METHODS

2.5.1 LINKAGE STUDY POWER CALCULATION

The available pedigrees from the UK and Norway were used to perform power calculations. To do this, the pedigree was converted to linkage format using the MAKEPED program[233], and the SLINK program[234, 235] was used to simulate data at a polymorphic marker in the members of the 100 pedigree replicates. A rare, dominant model was used, assuming a disease allele frequency in the population of 1 in 10,000 and assuming a disease penetrance of 0.1% if 0 risk alleles are present and 99.9% if 1 or 2 risk alleles are present. Study power was calculated for four levels of heterogeneity between the families: no heterogeneity, 25% heterogeneity, 50% heterogeneity and 75% heterogeneity. Merlin[236] was used to run a linkage analysis on the simulated data and data were analysed in the statistics package Stata[237].

2.5.2 GENOTYPING ON THE AFFYMETRIX GENOME-WIDE HUMAN SNP ARRAY 6.0

Following extraction and quality control checks, gDNA samples were shipped, with ice packs, to genotyping companies. All samples were genotyped on the Affymetrix genome-wide human SNP array 6.0. Genotyping was undertaken in three phases. The first kindreds (43 samples in total from the UK) were genotyped by Almac Diagnostics in June 2010, while the second (69 samples from Norway and the UK) and third (ten samples from the UK, including one repeat sample from batch two) batches were genotyped by AROS Applied Biotechnology in April 2011 and April 2012 respectively. In total, 121 individuals from 24 families were genotyped (12 families from the UK, 12 from Norway) (Figure 12).

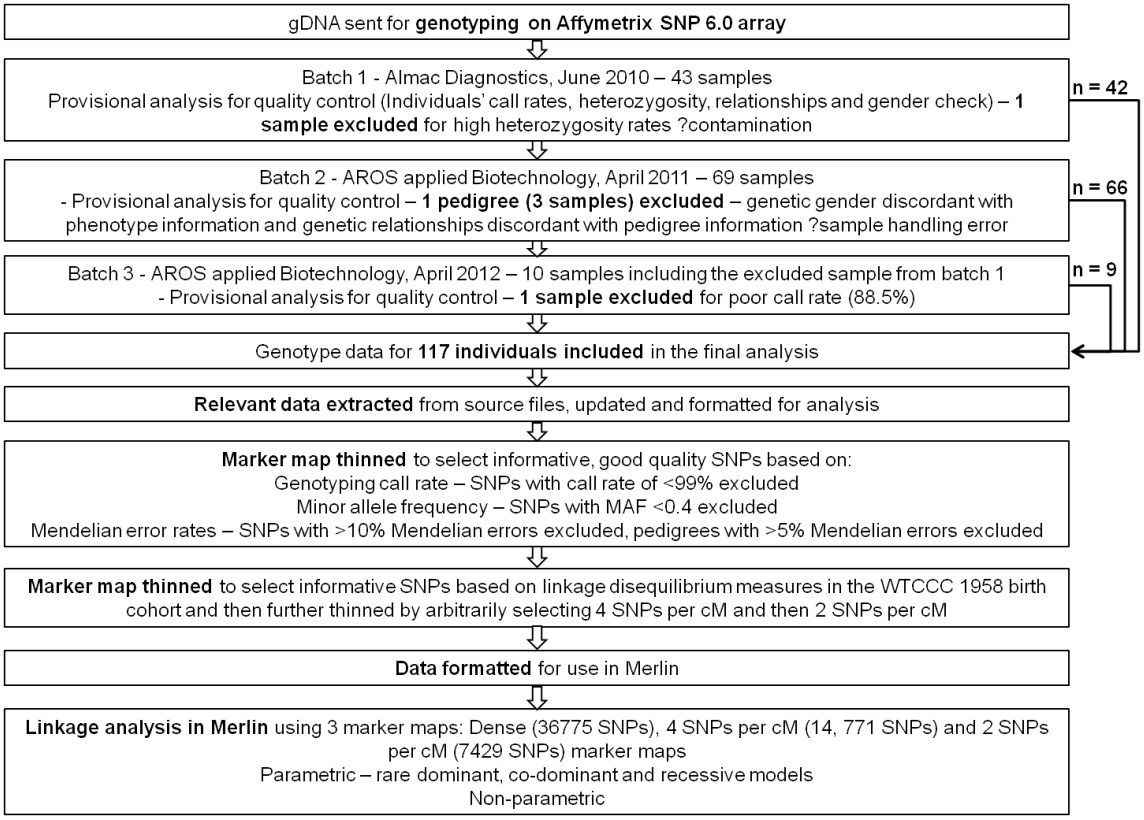


Figure 12: Flowchart of method used for the linkage analysis in multiplex AAD families.

2.5.2.1 THE AFFYMETRIX GENOME-WIDE HUMAN SNP ARRAY 6.0

SNP microarrays (also referred to as SNP arrays or SNP chips) have been developed as a genotyping platform to exploit the uniform distribution of SNPs throughout the human genome[238]. SNP arrays allow multiple SNPs to be genotyped simultaneously[239], and work on the principle that complementary nucleotide sequences hybridise to one another. Each microarray is made up of numerous short sequences of single-stranded DNA fragments (oligonucleotide probes, corresponding to SNPs) which are synthesised *in situ* and arranged in a grid pattern on a solid surface, usually glass or silica. DNA is digested with Nsp I and Sty I restriction enzymes and ligated to adaptors that recognise the cohesive 4bp overhangs. A generic primer that recognises the adaptor sequence is used to amplify DNA fragments which have been ligated to adaptors in the 200 to 1100bp size range. PCR amplification products for each restriction enzyme digest are combined and purified using activated beads and then the amplified DNA is fragmented, labelled and hybridised to the array. Any sequences in the sample that find a "match" on the array will bind to that complementary sequence at a specific spot and then a computer program can be used to determine the amount of sample bound to each spot on the microarray[240]. A number of SNP microarray platforms are now commercially available. The Affymetrix Genome-Wide Human SNP Array 6.0 claims to "represent the most genetic variation on a single array"[240], featuring 1.8 million genetic markers, including more than 900,000 SNPs and more than 900,000 probes for the detection of copy number variation.

Samples were genotyped in batches at either Almac diagnostics, UK or at AROS Applied Biotechnology, Denmark. Raw data files in birdseed-v2 format were returned for analysis.

2.5.2.2 RAW DATA MANAGEMENT AND FORMATTING FOR LINKAGE

Initially, the raw data files received were formatted to create manageable files that could be used for an initial linkage analysis. To extract the relevant data from the source data files, Professor Cordell wrote short programs using Perl, a computer programming language.

A Perl program was initially used to extract fields of interest (the probe SNP ID, the dbSNP RS number, chromosome, physical position, strand, allele A, allele B, allele frequencies, minor allele, minor allele frequency) from the Affymetrix SNP 6.0 annotation file. These fields were used to create a map file describing the position of each genotyped SNP.

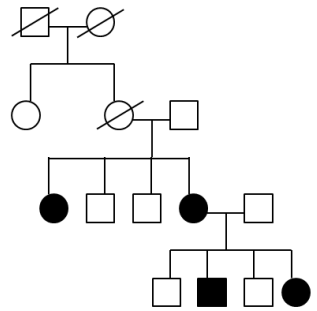
Genotype data was returned for each individual sample analysed in a separate file. Relevant data was extracted from these source files using a Perl program, including SNP unique identifier code, call (genotype), signal A and signal B. If there was no call for a SNP, this was coded as 00.

The data for each individual were then combined, again using a Perl script, to produce a single genotype data file (lgen file). Initially, each individual genotyped was allocated a unique “person” number (1–121 inclusive). Samples were genotyped in three separate batches. Individuals genotyped in the first batch were coded as belonging to family “0” while those added in batch two were coded as belonging to family “1”. Finally, individuals genotyped in the final batch were coded as belonging to family “new”. A pedigree file, reflecting all participants in the study, was then created in Microsoft Excel. This file identified relatives and defined their relationships to each other. Each family was allocated a family identifier code and each individual within a family was allocated a unique person identifier code (Figure 13).

PLINK (version 1.07, Shaun Purcell, http://pngu.mgh.harvard.edu/purcell/plink/)[217] was then used to update the lgen file with the correct information from the pedigree file, thus defining individuals

belonging to each of the 24 families studied, each individual's gender, their parents (defined by their father's and mother's person identifier code) and phenotype (affected or unaffected). PLINK is a free access whole genome association study analysis engine and is relatively user-friendly, can be used to generate summary statistics and is particularly useful for formatting and managing large data sets, including recoding and reordering data files, merging files and extracting subsets of data such as individual families or sets of SNPs.

A



B

Family ID	Individual ID	Father ID	Mother ID	Sex	Phenotype
15	1	0	0	1	1
15	2	0	0	2	1
15	3	1	2	2	1
15	4	1	2	2	1
15	5	0	0	1	1
15	6	5	4	2	2
15	7	5	4	1	1
15	8	5	4	1	1
15	9	5	4	2	2
15	10	0	0	1	1
15	11	10	9	1	1
15	12	10	9	1	2
15	13	10	9	1	1
15	14	10	9	2	2

Figure 13: Pedigree file structure for linkage analysis.

Representation of pedigree information in diagram (panel A) and pedigree file format (panel B). Column 1 of the pedigree file gives the family ID number, in this case family 15 is used as an example. Each family member is given an individual ID number, shown in column 2. Family 15 comprises 14 individuals. The gender of each individual is defined in column 5, with males coded as 1 and females coded as 2. Phenotype is defined in column 6 with unaffected individuals coded as 1 and affected individuals coded as 2. Intrafamilial relationships are defined by the father and mother ID. For example, individuals 3 and 4 are unaffected female siblings, as both have person 1 as their father and person 2 as their mother. Individuals who are founders, that is they have no relatives in generations above them in the family, have a father and mother code of 0 and 0. Individuals 1, 2, 5 and 10 are examples of founders in this family.

PLINK is also useful for data quality control, allowing the user to calculate missing genotype rates, determining the genetic relationships of individuals genotyped, checking for errors in family data and performing gender checks using X chromosome genotype data[217].

A preliminary map file was then made. For ease, each SNP was initially labelled as being on chromosome 1 and arbitrary cM and kilobase (kb) positions were allocated. This file was then updated to include the correct chromosome information for each SNP (1-22, X=23, Y=24), with an accurate cM and kb position defined by the Affymetrix SNP 6.0 annotation file.

2.5.2.3 DATA QUALITY CONTROL AND MARKER DENSITY REDUCTION FOR LINKAGE ANALYSIS

I undertook a number of checks to ensure that only high quality genotyping data were used for the linkage analysis. The first step was to check for missing data at each SNP and for each individual using PLINK. High levels of missing data for a SNP might indicate a genotyping problem at that probe, while high levels of missing data for an individual might indicate a problem with the DNA sample. Individuals and SNPs with poor calls should be excluded for these reasons. The data were then checked to ensure that the allocated gender for each individual in the pedigree file matched the genotyping data, using PLINK to check that X chromosome heterozygosity data related appropriately to each individual's allocated gender. Any discrepancies were resolved using an amelogenin PCR assay.

The data were then checked for an excess of Mendelian errors both at SNPs and within individuals. This was to ensure that the samples had not been mislabelled or become muddled up in the lab. This check would also identify mis-paternity should there be any such cases in our families. SNPs with a Mendelian error rate of 10% or more, and families with a Mendelian error rate of 5% or more, would be excluded from further analysis.

Individuals born as a result of random matings within a population will have genotypes that are in HWE. Calculated inbreeding co-efficients (F) represent a measure of deviation from HWE averaged over all SNPs in an array. Deviation from HWE resulting in an outlying F value can indicate a number of issues. A low F (high heterozygosity) may indicate sample contamination. A high F value (low heterozygosity) may indicate either true inbreeding or suggests that the individual in question originates from a different population i.e. a different ethnic group. The latter is known as the Wahlund effect[241]. Heterozygosity rates were calculated for each individual genotyped as part of the quality control measures employed in this study. To do this, PLINK was used to calculate the observed number of homozygous genotypes and the total number of genotypes. Heterozygosity rates could then be calculated as follows:

number of non-missing genotypes-number of homozygous genotypes/number of non-missing genotypes

PLINK was used to calculate average rates of IBD allele sharing across the genome in pedigrees. R[242], a statistical program, was then used to generate graphs of these data. I checked that there was no excess IBD allele sharing in unrelated individuals, and that related individuals shared an appropriate proportion of alleles IBD (e.g. parent-offspring pairs should share 1 allele IBD while full siblings share 0 alleles IBD 25% of the time, 1 allele IBD 50% of the time and 2 alleles IBD 25% of the time, resulting in an average of 1 allele shared IBD).

The Affymetrix SNP 6.0 array contains more than 900,000 SNP probes and therefore produces very dense SNP genotyping data. The linkage analysis was to be performed using Merlin (version 1.1.2, Goncalo Abecasis, http://www.sph.umich.edu/csg/abecasis/Merlin)[236], a free computer program that has been designed to perform linkage analyses on pedigree data. Merlin is an acronym for “multi-point engine for rapid likelihood inference” and uses sparse binary trees to track gene flow through pedigrees, allowing the detection of alleles co-segregating with a phenotype of interest[236]. Merlin can handle

denser marker maps than other linkage programs however, to make the analyses computationally efficient and to reduce the chances of detecting false positive linkage due to SNPs in LD, the SNP map was thinned for linkage analysis. Information gathered during the quality control steps and from SNP LD patterns was first used to remove poor quality genotype data and to select informative and independent SNPs, thus reducing the chances of a false positive finding.

SNPs were first thinned based on genotyping call rate. Any SNP with a genotyping call rate of less than 99% was excluded (80,150 SNPs excluded). Heterozygous SNPs are most informative for linkage analysis, allowing allele flow through pedigrees to be followed, therefore the dataset was thinned further based on heterozygosity. Initially, SNPs with a minor allele frequency of less than 0.3 were excluded (773,909 SNPs excluded). SNPs were then thinned further on the basis of Mendelian error rates, where families with greater than 5% Mendelian error rates were excluded (0 families excluded) and SNPs with greater than 10% Mendelian error rate were excluded (2 SNPs excluded). SNPs were thinned on the basis of LD, with SNPs not in LD with each other being selected to avoid false positive results. To do this, we used the Welcome Trust Case-Control Consortium (WTCCC) 1958 birth cohort genome-wide data to calculate LD and then select SNPs on the basis of those LD calculations. At this point, more than 50,000 SNPs remained which was too dense a map, therefore a minor allele frequency cut off of 0.4 was applied. This then left 36,775 SNPs. This map was used for an initial linkage analysis, however the files generated were difficult to manage as they were so large, therefore the SNPs were thinned further using MapThin[243], a computer program, to arbitrarily select 4 and then 2 SNPs per cM, leaving 14,771 and 7429 SNPs respectively in total for the final linkage analyses.

2.5.2.4 DATA FORMATTING FOR MERLIN

The pedigree, map and genotype data files that had been created from the raw data files were appropriately formatted for use with Merlin. In addition, an allele frequency file was created from the WTCCC 1958 birth cohort data. This file

contained details of the major and minor allele at each SNP and the frequencies of these. PLINK was used to update the allele annotation within the genotype data file, changing alleles originally coded as A or B to the correct nucleotide (i.e. A, C, G or T). The nucleotides were then recoded numerically from A, C, G and T to 1, 2, 3, 4 respectively.

2.5.2.5 LINKAGE ANALYSIS USING MERLIN

Merlin was used to perform the linkage analysis. The linkage analysis was performed under the supervision of Professor Cordell. Initially, a parametric linkage analysis was performed on the autosomal chromosomes using a number of models: rare dominant, rare recessive and rare co-dominant (Table 8). Both LOD and HLOD scores were calculated. The HLOD score is generated based on heterogeneity likelihood. That is to say, the linkage analysis can allow for the possibility that clinically indistinguishable forms of a disease may be seen in different kindreds, caused by genes at different loci. When Merlin is used to perform this method, it generates a parameter called α which indicates the proportion of kindreds under investigation whose disease is due to a gene linked to the marker being studied, where 1 is equal to 100%. The HLOD is then calculated. Allowing for genetic heterogeneity between families in this way is often considered more powerful than non-parametric linkage analysis[244, 245]. A non-parametric linkage analysis was then also performed. The X chromosome was analysed separately using MINX, a version of Merlin designed for markers on the X chromosome. Again, a parametric analysis, using the same models, and a non-parametric analysis were performed.

A second analysis was then performed, taking 21OH autoantibody status as the trait. 21OH autoantibody positive individuals were re-coded in PLINK as cases and 21OH autoantibody negative individuals were re-coded as controls. The linkage analysis was then repeated.

Model name	Disease allele frequency	Penetrance if 0 disease alleles are present	Penetrance if 1 disease allele is present	Penetrance if 2 disease alleles are present
Rare dominant	0.0001	0.001	0.999	0.999
Rare recessive	0.0001	0.001	0.001	0.999
Rare co-dominant	0.0001	0.001	0.75	0.999

Table 8: Models used for parametric linkage analysis in Merlin.

2.5.2.6 ASSOCIATION ANALYSIS IN MULTIPLEX AAD FAMILIES USING EMMAX

Association analyses were performed using the EMMAX (Efficient mixed-model association expedited) program[246]. EMMAX is another free program which is based upon EMMA (efficient mixed-model association), a variance component approach, which allows for sample structure by explicitly accounting for pairwise relatedness between individuals. A useful feature of EMMAX is that it accepts transposed files generated by PLINK. Two association analyses were performed. The first used the affected AAD family members as cases and unaffected family members as controls. The second analysis included the affected AAD family members as cases and compared their genotyping results to those of 2706 1958 UK birth cohort controls, available through the WTCCC. Control data from the WTCCC had already undergone strict quality control filtering on the basis of call rate (SNPs excluded if call rate was <95% or <99% if the minor allele frequency was less than 0.05) and HWE ($<5.7 \times 10^{-7}$). In the multiplex AAD family cases and controls, SNPs with a call rate of less than 99% were excluded as were SNPs with a Mendelian error rate of 10% or more. In the multiplex AAD family controls used in the first analysis, in addition to the above measures, SNPs which were out of HWE were also excluded. To do this, two thresholds were applied, a stringent threshold of $P < 0.01$ and a less stringent threshold of $P < 1.0 \times 10^{-8}$, and the results compared. Finally, any SNP with a minor allele frequency of less than 5% was excluded as these rare variants can be difficult to genotype.

For the first analysis, 595,118 SNPs met the quality control criteria if a HWE threshold of $P < 1.0 \times 10^{-8}$ was applied. If a HWE threshold of $P < 0.01$ was applied, a further 2275 SNPs were excluded (592,843 SNPs in total). For the second association analysis, 551,634 SNPs met the quality control criteria. For each analysis, R[242], a statistical package that can be used to plot graphs of complex data, was used to generate quantile-quantile (QQ) and Manhattan plots from the results. EMMAX and R were used with guidance from Dr Rebecca Darley in the statistical genetics group, Institute of Genetic Medicine, Newcastle University. The LocusZoom program[247] was then used to visualise

data on a regional basis by selecting the most associated SNP and plotting SNPs, both in LD and not in LD, around it.

2.5.3 AMELOGENIN SEX IDENTIFICATION ASSAY

The amelogenin gene, found on both the X and Y chromosomes, can be used in sex determination of samples by PCR, since the X chromosome amelogenin allele (*AMELX*) contains a 6bp deletion in intron 1 which is not present in the Y chromosome allele (*AMELY*). For this assay, primers specific for intron 1, which span the deletion, are designed: forward primer sequence (5'-3') CCCTGGGCTCTGTAAAGAATAGTG, reverse primer sequence (5'-3') ATCAGAGCTTAAACTGGGAAGCTG, Tm 61ºC. A female DNA sample (XX) will produce a 106bp PCR product and a sample from a male (XY) will result in two PCR products of 106 and 112bp. These products can be visualised by agarose gel electrophoresis, thus allowing differentiation of sex from unknown samples.

DNA samples were diluted to a concentration of 20ng/µl. PCR was carried out in PCR tubes with a reaction volume of 25µl. The PCR reaction contained 1X PCR buffer (Qiagen), 0.15mM $MgCl_2$ (Qiagen), 0.1mM dNTPs (New England Biosciences),1.6µM of each primer (Eurofins MWG Operon), 0.75u HotStar Taq DNA polymerase (Qiagen) and 40ng of DNA. The thermal cycling conditions included an initial denaturation step at 94°C for 15 minutes, followed by 35 cycles of 94°C for 30 seconds, 57°C for 30 seconds and 72°C for 30 seconds, followed by a final extension step of 72°C for 10 minutes. The PCR products were run out, next to a 100bp, on a 4% agarose gel at 90V for 1 hour and visualised using a UV light. Samples where 2 bands were present at 106 and 112bp were designated male, while samples where a single 106bp product was apparent were designated female.

CHAPTER 3 – A HYPOTHESIS-DRIVEN APPROACH TO THE INVESTIGATION OF AAD – THE *CYP21* LOCUS

3.1 BACKGROUND

The central dogma of molecular biology was originally that the only function of RNA was to act as an intermediate, allowing information contained within the gDNA to be transferred into an amino acid sequence and therefore a protein. Non-coding sequence was thought to be junk DNA awaiting evolutionary elimination. However, the discovery that non-coding gDNA sequence is often transcribed into non-coding RNA generated the hypothesis that these non-coding transcripts could have a function, perhaps influencing and regulating coding genes and therefore health and disease[248]. This has since led to extensive research into non-coding RNAs and their functions. One potential source of non-coding RNA is pseudogenes. Pseudogenes are sequences of DNA which have derived from functional genes, and often closely resemble these ancestor genes, but are rendered non-functional by mutations. Pseudogenes were first described in the *Xenopus laevis* (African clawed frog) species by Jacq *et al*[249] and have since been found in the sequence of numerous other species including humans. Pseudogenes arise in the gDNA sequence by two main mechanisms: retrotransposition events (resulting in processed or retro-pseudogenes) and duplication events (resulting in non-processed or duplicated pseudogenes).

Processed pseudogenes arise due to retrotransposition events. This is when a single-stranded transcript or part of a transcript is spontaneously reverse-transcribed back into the gDNA sequence as double-stranded sequence by an RNA polymerase. The re-inserted sequence is often referred to as a retrotransposon. The presence of a poly(A) tail and the lack of intronic sequence in retrotransposons reveals that they have derived from an mRNA source[250, 251]. Although a retrotransposed gene can be functional[252], the process of retrotranscription frequently results in mutation and inactivation, in which case the inserted sequence is known as a processed pseudogene[250].

Non-processed pseudogenes arise as a result of duplication events. These occur commonly in the genome and form the basis of gene families which arise

as a result of duplication of a single common ancestor: for example, the human globin genes which encode haemoglobin subunits[253]. In some cases, mutations render these duplicated sequences non-functional[253], in which case they are referred to as non-processed pseudogenes.

The RCCX module on chromosome 6 is an example of a portion of DNA sequence which has undergone duplication resulting in non-processed pseudogenes. The RCCX module is approximately 30kb and is found in the MHC class III region. The module contains four genes arranged in tandem: the *serine/threonine nuclear protein kinase* (*RP*) gene; the *complement component C4* gene (*C4*, present as a *long C4A* or a *short C4A* and a *long C4B* or a *short C4B*); the *CYP21A2* gene (also known as *CYP21B*); and the *extracellular matrix protein tenascin* (*TNX*) gene[254]. The RCCX module may exist in three forms: as a monomodular, bimodular or trimodular structure containing zero, one or two duplications of its constituent genes. If the RCCX is monomodular, as found in 17% of the Caucasian population, it contains only one functional copy of each of the four genes (*RP1, C4, CYP21A2* and *TNXB*). If the RCCX is bimodular (69%), it contains a functional copy of each of the four constituent genes and a duplication containing three non-processed pseudogenes (*CYP21A1P, TNXA*, *RP2)* and an additional functional *C4* gene. A trimodular RCCX (14%) contains two duplications (Figure 14)[254]. In addition to functional C4 genes, non-functional copies may be present, termed “null” alleles. The RCCX module is of interest in AAD as it contains the *CYP21A2* gene. This gene encodes the steroid 21OH enzyme which is a key enzyme in the steroidogenesis pathway[255] (Figure 15). The functional *CYP21A2* gene is approximately 3.4kb in length and is composed of ten exons[256]. It is expressed from as early as 50 to 52 days post-conception in the adrenal gland, gonad, liver, thymus and brain. Expression at much lower levels has also been reported in the heart, lung, kidney, pancreas, prostate and stomach[257, 258].

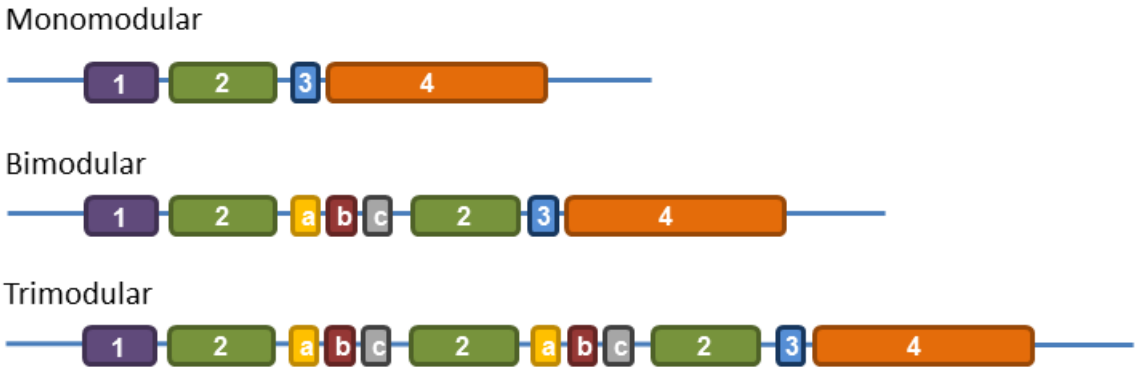


Figure 14: Schematic of RCCX haplotypes in Caucasians.

The monomodular RCCX comprises a single functional copy of the serine/threonine nuclear protein kinase (RP) gene (purple rectangle, 1), the complement component *C4* gene (green rectangle, 2), the *CYP21A2* (cytochrome P450, family 21, subfamily A, polypeptide 2, also known as *CYP21B*) gene (blue rectangle, 3) and the extracellular matrix protein *tenascin* (*TNX*) gene (orange rectangle, 4). If the RCCX is bimodular, it contains a functional copy of each of the four constituent genes and a duplication containing three non-processed pseudogenes (*RP2* – yellow rectangle a, *CYP21A1P* – red rectangle b, *TNXA* – grey rectangle c) and an additional functional *C4* gene. A trimodular RCCX contains two duplications.

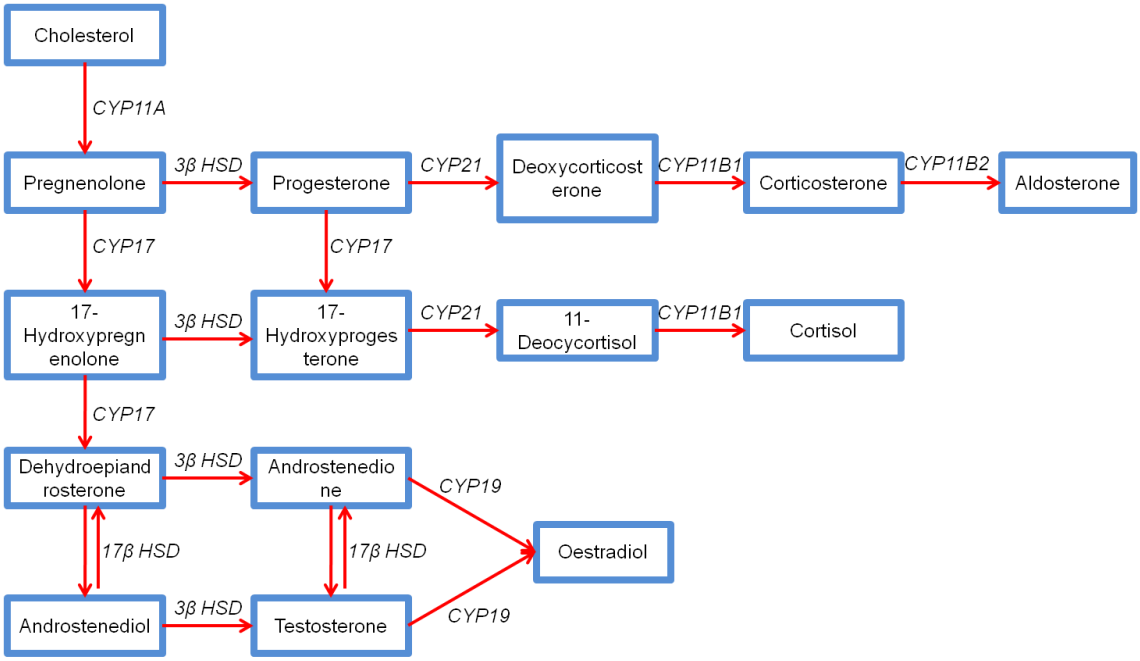


Figure 15: The steroid biosynthesis pathway.

CYP11A, Cholesterol side-chain cleavage enzyme; CYP11B1, 11 beta hydroxylase; CYP11B2, aldosterone synthase; CYP17, 17 alpha hydroxylase; CYP19, oestrogen synthase; 17β HSD, 17 beta hydroxysteroid dehydrogenase; CYP21, 21-hydroxylase; 3β HSD, 3 beta-hydroxysteroid dehydrogenase.

The *CYP21A1P* non-processed pseudogene is highly homologous to *CYP21A2*[259]*.* It contains exons, introns and promoter regions, however it has been rendered non-functional by a deleterious 8bp deletion in exon three which results in a frameshift and premature stop codon[259]. *CYP21A2* has six predicted transcripts but only two, 1914 and 2182bp in size, are listed in the Consensus Coding Sequence Project (CCDS) database[260], demonstrating that they have been validated by multiple sources. *CYP21A1P* has two predicted transcripts, 1481 and 1971bp in length, however these are not validated and therefore not listed in CCDS. The predicted CYP21A1P protein product is truncated due to the deleterious mutations in *CYP21A1P* and is therefore not predicted to have any enzymatic activity. The deleterious nature of the pseudogene's mutations is demonstrated in congenital adrenal hyperplasia (CAH), which is an inherited condition where 21OH deficiency results in an inability to make cortisol. CAH is most commonly due to recombination events between the gene and the pseudogene: either deleterious mutations in the pseudogene are transferred into the functional *CYP21A2* gene (gene-conversion events) rendering the gene non-functional[261], or the functional gene is deleted and replaced by the pseudogene[262, 263].

Like many pseudogenes, *CYP21A1P* is evolutionarily conserved in humans and its sequence exhibits reduced nucleotide variability and an excess of synonymous SNPs compared to non-synonymous SNPs; all of which are hallmarks of DNA sequence with a functional role[264]. Indeed, *CYP21A1P* transcripts have been isolated from cultured adrenal cells, demonstrating that non-coding RNA is derived from the pseudogene sequence[265]. Furthermore, a previous member of the laboratory group, Alekhya Narravula, had previously determined that almost 1 in 5 individuals with AAD had zero copies of *CYP21A1P* compared to just 1 in 27 controls ($P < 0.0001$). At a molecular level, having no *CYP21A1P* copies is synonymous with being homozygous for the monomodular RCCX. With this in mind, we hypothesised that *CYP21A1P* transcripts and/or CYP21A1P protein products are expressed in human thymus to promote and induce tolerance to components of the steroidogenic enzymes. While *CYP21A2* gene transcripts or the CYP21A2 protein product could serve this purpose, we hypothesise that the high concentration of immunosuppressive

steroids produced by the functionally active 21OH enzyme could inhibit lymphocyte activity and therefore perhaps the tolerance-inducing process. Having a non-functional but highly homologous pseudogene might allow transcripts or a defective product to be expressed in the absence of steroids, perhaps in the thymus, allowing tolerance to the steroidogenic apparatus to be developed. Thus, it follows that an individual without any copies of the *CYP21A1P* pseudogene could be more susceptible to developing AAD compared to individuals with copies of the pseudogene who have been rendered tolerant to the steroidogenic apparatus through *CYP21A1P* pseudogene expression.

3.2 AIM

I aimed to determine whether individuals with another autoimmune condition, Graves' disease, were more likely to have no copies of *CYP21A1P* compared to controls or whether this finding is restricted to AAD. My second aim was to determine whether transcripts of *CYP21A1P* could be detected in fetal tissues. The presence of *CYP21A1P* transcripts in fetal tissue might imply a role for it in the development of tolerance in early life.

3.3 SUMMARY OF STUDY DESIGN

A PCR method was used to genotype additional individuals with AAD and healthy controls (in addition to those already studied by Alekhya Narravula) to determine the proportion of AAD cases and controls with no *CYP21A1P* copies (*CYP21A1P*–). In addition, 447 individuals from the UK with Graves' disease were also genotyped to determine whether any differences were restricted to AAD or whether they might also be observed in another autoimmune condition.

To study *CYP21A2* and *CYP21A1P* expression, a series of tissue *in situ* hybridisation (TISH) experiments were planned. TISH was initially attempted on 10 to 12 week post-conception (wpc) fetal adrenal tissue. Adrenal tissue was chosen as a positive control tissue for initial protocol optimisation because *CYP21A2* is known to be expressed here in fetal life, from as early as 50 days post-conception[257]. In addition, as *CYP21A1P* transcripts have previously been detected in adrenal cell lines[265], adrenal tissue would seem to be the most likely site of expression of *CYP21A1P,* if indeed it is expressed. A pair of traditional mRNA riboprobes were designed (an antisense probe and a sense negative control probe) which would hybridise to both *CYP21A2* and *CYP21A1P* transcripts in fixed, mounted adrenal tissue. As mRNA riboprobes are long, ideally between 250 and 1500bp in length, it is not possible to design two probes which could differentiate between two highly homologous transcripts such as *CYP21A2* and *CYP21A1P*. The mRNA probe was therefore used as a

positive control in this experiment, giving an indication of the expected pattern of distribution of *CYP21A2* and/or *CYP21A1P* transcripts in the tissues studied.

To differentiate *CYP21A2* from *CYP21A1P* expression specifically, a pair of locked nucleic acid (LNA) probes (Exiqon) were also designed: one specific to *CYP21A2* and the other specific to *CYP21A1P.* The published literature suggests that LNA probes can be used to distinguish transcripts differing at only one or two bases[266], due to their short nature and the increased hybridisation specificity conferred by the LNAs introduced. This, in theory, makes them ideal for this project: two LNA probes could be used to test for the presence or absence of each transcript individually, even when the transcripts differ by only an 8bp deletion. For this study, to ascertain the specificity of the CYP21A1P probe, both *CYP21A1P*– and *CYP21A1P*+ samples would be used, thus ensuring that the CYP21A1P probe is not hybridising to *CYP21A2* transcripts. In conjunction with TISH, Northern blotting, using fetal adrenal RNA, was performed to determine the size of transcripts to which each probe was hybridising.

Following TISH, a qPCR expression experiment was designed, with double-dye hydrolysis probes, to detect *CYP21A2* and *CYP21A1P* transcripts in fetal adrenal gland and thymus obtained from children.

3.4 RESULTS

3.4.1 ABSENCE OF *CYP21A1P* IN INDIVIDUALS WITH AUTOIMMUNE DISEASE

A previous MSc student in the laboratory group, Alekhya Narravula, used a competitive PCR assay to test for the presence of *CYP21A1P* in gDNA samples from 295 individuals with AAD and 299 controls. The results of this work are summarised in Table 9 (panel A). I used the same assay to assess an additional 20 individuals with AAD, 447 samples from individuals with Graves' disease and 328 additional healthy controls. The results of this work are summarised in Table 9 (panel B). All results from this assay are shown in Table 9 (panel C). An image generated from UV visualisation of the PCR products from this assay is shown in Figure 16.

55 of the 315 AAD samples tested (17.5%) were *CYP21A1P*– compared to just 19 of 627 (3.0%) of controls (P <0.00001, OR 6.77 [95% CI 3.94 – 11.63]). In addition, 35 of 447 (7.8%) people with Graves' disease were *CYP21A1P*– compared to 3.0% of controls (P 0.0007, OR 2.72 [95% CI 1.53 – 4.82]).

A	AAD patients (n=295)	Controls (n=299)
CYP21A1P absent	50 (16.9%)	11 (3.7%)*
CYP21A1P present	245 (83.1%)	288 (96.3%)

B	AAD patients (n=20)	Graves' disease patients (n=447)	Controls (n=328)
CYP21A1P absent	5 (25.0%)	35 (7.8%)	8 (2.4%)
CYP21A1P present	15 (75.0%)	412 (92.2%)	320 (97.6%)

C	AAD patients (n=315) *	Graves' disease patients (n=447) $	Controls (n=627)
CYP21A1P absent	55 (17.5%)	35 (7.8%)	19 (3.0%)
CYP21A1P present	260 (82.5%)	412 (92.2%)	608 (97.0%)

*χ^2 P <0.00001 (8.1×10^{-15}), OR 6.77 [95% CI 3.94 – 11.63]

$ χ^2 P 0.0007, OR 2.72 [95% CI 1.53-4.82]

Table 9: Competitive PCR assay results demonstrating frequency of CYP21A1P absence in AAD, Graves' disease and controls.

Panel A: Rates of *CYP21A1P* absence in AAD cases and controls (work done by A. Narravula). Panel B: Rates of *CYP21A1P* absence in AAD cases, Graves' disease cases and healthy controls (work done by A. Mitchell). Panel C: Summary results table of *CYP21A1P* competitive PCR assay.

Figure 16: Competitive CYP21A2/CYP21A1P PCR gel image.

3.8% agarose gel electrophoresis of PCR products obtained using the CYP21A2/CYP21A1P competitive PCR assay. The 85bp and the 77bp bands are the CYP21A2 and CYP21A1P amplicons respectively. A 100bp ladder is shown in lane 1 and lane 2 is a no template (water) control. The sample in lane 14 has failed to produce any PCR products. Samples in lanes 3 to 12 inclusive and the sample in lane 15 have all produced two bands consistent with amplicons derived from the CYP21A2 gene and the CYP21A1P pseudogene. The sample in lane 13 does not have a 77bp band and is therefore CYP21A1P deleted.

3.4.2 *CYP21A1P* EXPRESSION IN FETAL TISSUE

3.4.2.1 TISSUE IN SITU HYBRIDISATION TO DETERMINE CYP21A2 AND CYP21A1P EXPRESSION

In the TISH experiment, the SP6 and T7 mRNA riboprobes were hybridised to adrenal and renal tissue obtained from 10 to 12 wpc human foetuses. Marked, specific hybridisation throughout the fetal adrenal gland, most marked at the periphery, was seen with the SP6 probe in both of the *CYP21A1P*+ samples (N1822, N1593) and in the *CYP21A1P*– sample (N1643). No hybridisation was seen with the T7 (sense) negative control probe, confirming that the hybridisation to adrenal with the SP6 probe was specific (Figure 17). Using Northern blotting with RNA derived from fetal adrenal gland, a band at approximately 2000bp was seen with the SP6 probe, while no band was seen with the T7 probe (Figure 18). This is the appropriate size for the validated CYP21A2 protein coding transcripts (1914 and 2182bp).

The detection of a strong signal in the *CYP21A1P*– sample confirms the presence of *CYP21A2* transcripts in fetal adrenal gland. The signal detected in the two *CYP21A1P*+ samples could be due to the presence of *CYP21A2* transcripts with or without *CYP21A1P* transcripts (Figure 17). No hybridisation was seen in the renal tissue adjacent to the adrenal gland, suggesting that the transcripts are either not expressed here or not abundant enough to be detected by this method. mRNA probes which hybridise specifically to kidney, such as FRMD7 and collagen 6 riboprobes, have been applied successfully to all three samples used in this experiment for other projects at the HDBR, Newcastle, suggesting that the absence of signal in kidney was not due to localised RNA degradation in this tissue. An example, showing positive signal from kidney with the FRMD7 riboprobe for sample N1593 is shown in Figure 17. This finding therefore confirmed *CYP21A2* expression in fetal adrenal gland.

The TISH protocol was then adapted for use with the *CYP21A2* and *CYP21A1P* LNA probes, with the aim of establishing whether *CYP21A1P* transcripts are expressed in fetal tissues. To optimise the protocol, tissue sections from the *CYP21A1P*+ fetuses (N1822, N1593) were used, in order to give optimal hybridisation signals.

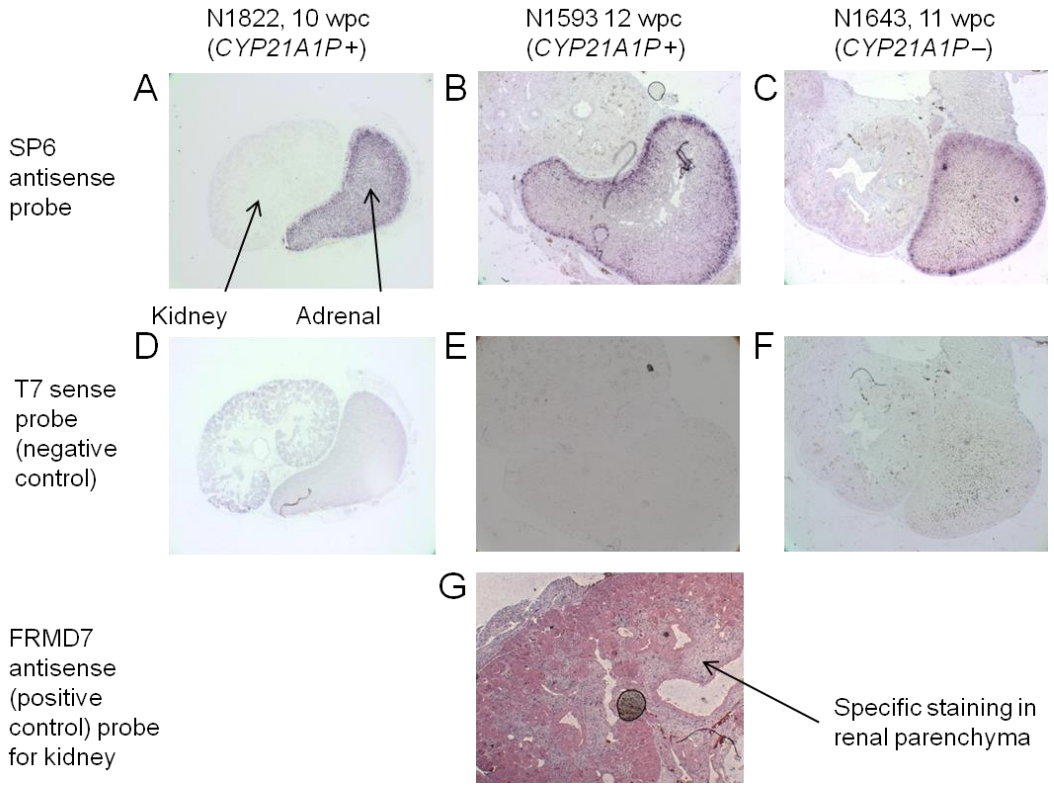


Figure 17: CYP21A2 and/or CYP21A1P expression in fetal adrenal demonstrated with CYP21-specific mRNA riboprobes.

Tissue in situ hybridisation with the SP6 (upper panel) and T7 negative control (middle panel) mRNA riboprobes under light-field conditions at 2.5x magnification (300ng of probe per slide). Samples N1822 (panels A, D) and N1593 (panels B, E) are known to have the CYP21A1P sequence in their gDNA (CYP21A1P+). N1643 (panels C, F) does not (CYP21A1P–). There is marked, specific hybridisation throughout the fetal adrenal gland in all three samples with the SP6 probe (panels A, B, C), suggesting the presence of CYP21A2 and/or CYP21A1P transcripts. There is no hybridisation to fetal renal in these sections. To establish the presence of intact RNA in fetal kidney, an FRMD7 mRNA probe, which hybridises specifically to kidney, is included as a positive control (lower panel, G) and was used on a renal tissue section from 12 wpc fetus N1593. There is no non-specific hybridisation with the T7 sense probe to either the fetal adrenal or kidney (panels D, E, F). wpc, weeks post-conception.

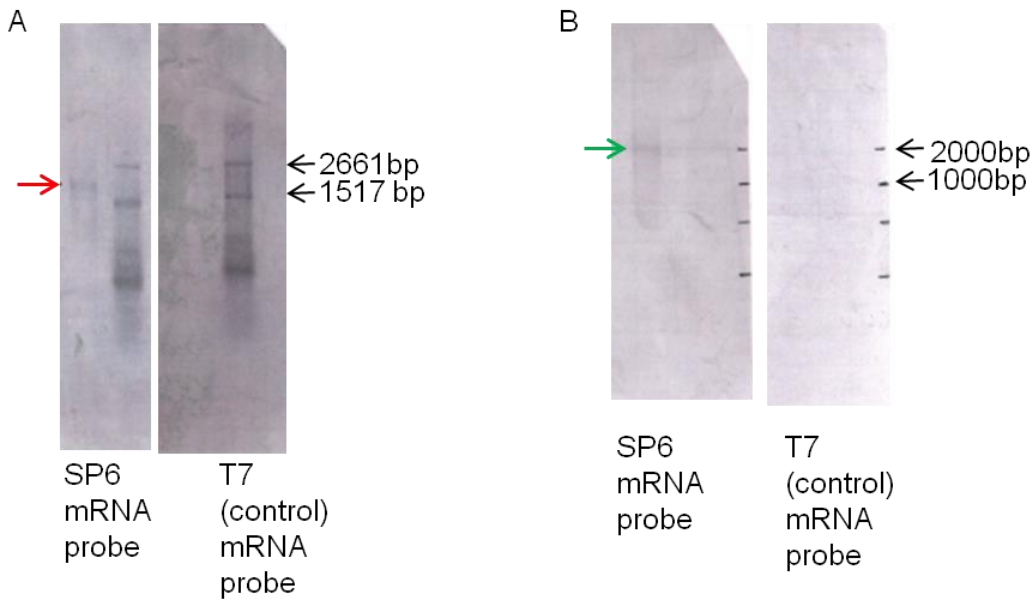


Figure 18: Northern blotting of fetal adrenal RNA, with the SP6 and T7 mRNA riboprobes.

Panel A shows the results with the SP6 and T7 mRNA probes compared to a DIG-labelled RNA ladder (Roche), on the right of each blot. A single band (red arrow) between the 2661 and 1517bp markers was seen with the SP6 probe, while no band is seen on the blot hybridised to the T7 probe. Panel B shows the results with the SP6 and T7 mRNA probes compared to a non-labelled ladder (Riboruler, high range ladder, Fermentas) marked using indelible pen on the right of each blot. The ladder marker positions are marked onto the membrane prior to transfer. A single band (green arrow) at 2000bp was seen with the SP6 probe, while no band is seen on the blot hybridised to the negative control T7 probe.

Initially, the LNA probes were used as per the manufacturer's (Exiqon) instructions, at concentrations of 20 to 40nM. When both of the LNA probes were used at this concentration, a strong signal was seen in both the fetal adrenal and kidney with the CYP21A2 probe, but no signal was seen with the CYP21A1P probe. The manufacturer's recommendation in this instance is to perform a titration gradient experiment and then to proceed to use the probe at the lowest concentration at which a positive hybridisation signal is seen.

The titration gradient experiment, conducted on sample N1593 which had been used for other studies with good results, using 10, 20, 40, 100 and 250nm concentrations of each LNA probe for hybridisation, demonstrated that a concentration of 250nM of CYP21A1P probe was needed to generate a signal (Figure 19). At this concentration, a strong signal was seen in both the fetal adrenal and kidney, in the same distribution as with the CYP21A2 LNA probe. There are two predicted *CYP21A1P* transcripts of 1481 and 1971bp and two validated *CYP21A2* protein coding transcripts. However, when the CYP21A2 and CYP21A1P LNA probes were used at varying concentrations, from 40 to 250nM, in the Northern blotting protocol, using fetal adrenal RNA, no bands were seen. Therefore the size of the transcript that the LNA probes were hybridising to could not be determined.

At such a high probe concentration there was concern that the signal seen with the CYP21A1P LNA probe might be non-specific. To investigate this, the CYP21A1P LNA probe was used at the lowest concentration required to give a signal (200nM) on *CYP21A1P*+ sample N1593 and on *CYP21A1P*– sample N1643. A strong hybridisation signal was seen from the fetal adrenal and kidney in all sections in this experiment, including from the *CYP21A1P*– fetus, confirming non-specific binding at this high concentration (Figure 20). To further confirm this, a similar pattern of staining was seen with the scrambled negative control probe when used at 250nM (Figure 19). When the CYP21A1P probe was used at the recommended concentration of 40nM on the *CYP21A1P*– tissue, no signal was seen.

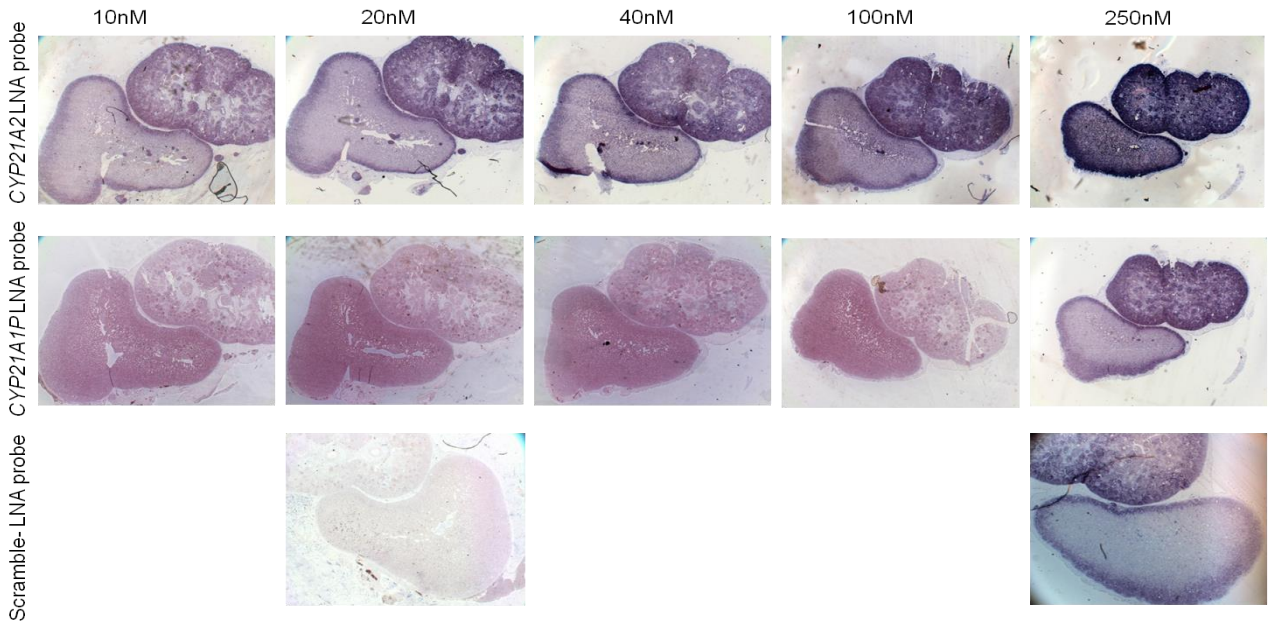


Figure 19: LNA probe in situ hybridisation titration gradient experiment.

12 weeks post-conception CYP21A1P+ fetal adrenal and renal tissue sections used in an LNA probe concentration titration gradient in situ hybridisation experiment. Results are shown from the CYP21A2 (upper panel), the CYP21A1P (middle panel) and the Scramble–ISH negative control (lower panel) LNA probes at increasing concentrations under light-field conditions, at 2.5x magnification. There is marked hybridisation with the CYP21A2 LNA probe throughout the fetal adrenal gland and kidney, but most marked at the periphery of each tissue at all probe concentrations, with signal intensifying with increased probe concentration. With the CYP21A1P LNA probe, hybridisation is seen at 250nM concentration of probe only. Using the scrambled probe at the same concentration gives a similar pattern of hybridisation which suggests non-specific hybridisation at high concentrations.

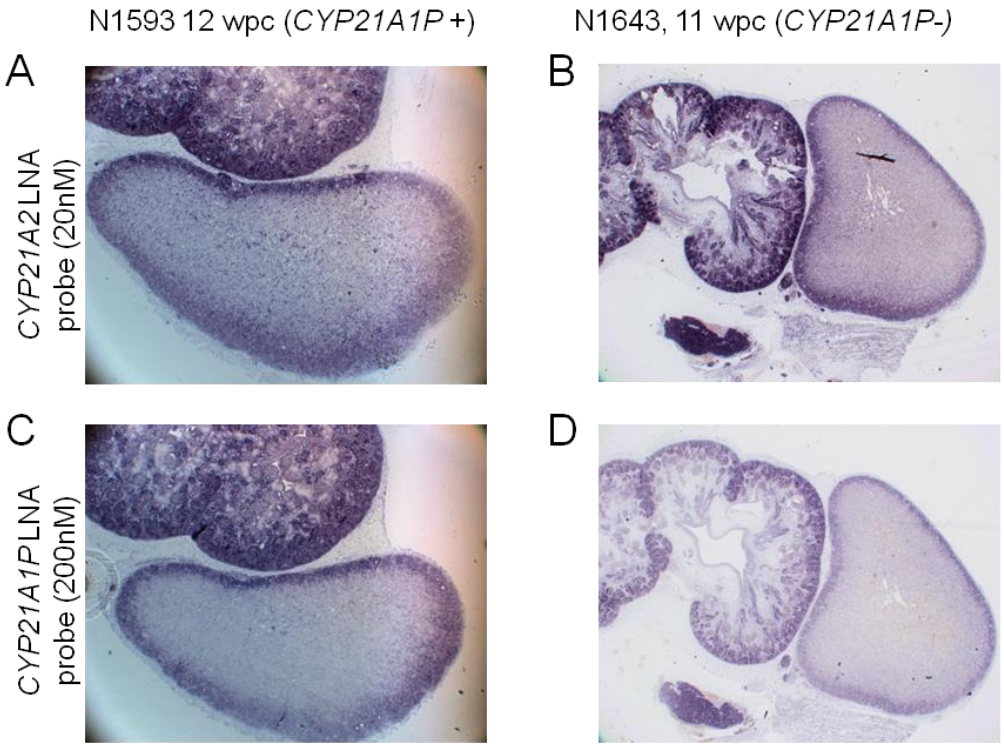


Figure 20: CYP21A1P LNA probe in situ hybridisation specificity experiment.

In situ hybridisation results with the CYP21A2 (upper panel) and the CYP21A1P (lower panel) LNA probes under light-field conditions at 2.5x magnification. 12 weeks post-conception (wpc) fetus N1593 (panels A, C) is CYP21A1P+ while 11 wpc fetus N1643 (panels B, D) is CYP21A1P–. There is hybridisation throughout the fetal adrenal gland and kidney in all tissue sections, most marked at the periphery. The strong hybridisation signal in the CYP21A1P– sections with the CYP21A1P LNA probe (panel D) suggests non-specific hybridisation of the CYP21A1P LNA probe.

In summary, if the CYP21A1P LNA probe is used at a concentration of 20–40nm, as recommended by the manufacturer, no hybridisation is seen. Used at a sufficient concentration to produce a hybridisation signal, the hybridisation is non-specific. One interpretation of these results is that the LNA probe, used at the manufacturer's recommended concentration, can be used to demonstrate that *CYP21A1P* transcripts are not expressed in fetal adrenal or kidney, or that transcripts are expressed at such a low level that they could not be detected. However, the lack of a positive control tissue for this series of experiments, where *CYP21A1P* is definitely expressed, meant that this could not be confirmed satisfactorily. Therefore the results from this LNA probe were discounted and an alternative method, using qPCR, was developed to investigate *CYP21A1P* expression in these tissues.

The *CYP21A2* LNA probe gave a strong hybridisation signal in the fetal adrenal gland from 20nM concentration in both the *CYP21A1P*+ and *CYP21A1P*– tissues, with marked staining at the periphery of the gland. This finding was expected and in keeping with published findings[257, 258]. Moreover, an even stronger signal was observed in fetal kidney and again, the signal was strongest at the periphery of the kidney. Although these results appeared to suggest marked *CYP21A2* expression in both fetal adrenal and kidney at 20–40nM concentration, the strong signal seen in the kidney was viewed with suspicion for four reasons. Firstly, the traditional SP6 mRNA riboprobe results showed no signal in the kidney, indicating no, or only very low, expression of *CYP21A2* transcripts in the kidney. Secondly, the distribution of staining in the kidney and adrenal with the CYP21A2 LNA probe was identical to the non-specific hybridisation signal seen with the CYP21A1P and scrambled LNA probes when used at high concentrations. Thirdly, the stronger staining in kidney compared to adrenal is contrary to previous published findings[258]. Where *CYP21A2* transcripts have previously been detected in the fetal kidney by qPCR, the levels have been significantly lower than levels in the fetal adrenal[258]. Finally, when used in the Northern blotting protocol, no bands were seen when either of the LNA probes were hybridised to adrenal fetal RNA. A series of experiments were conducted altering the stringency of the TISH conditions, including hybridising at higher temperatures (up to 56ºC) and increasing the length and

stringency of the post-hybridisation washes to reduce non-specific hybridisation. However, this made no difference and the same pattern of hybridisation was still observed. The possible explanations for this observation are that either *CYP21A2* is expressed in both fetal adrenal and kidney at equal levels, or that the CYP21A2 LNA probe binds either entirely non-specifically to fixed tissue sections, or binds specifically but to a homologous transcript present in abundance in fetal adrenal and kidney.

To determine whether the CYP21A2 probe was binding in a totally non-specific manner to fixed tissue, the probe was hybridised at 40nM to cardiac and limb tissue. These tissues were selected as they are not expected to express high levels of *CYP21A2.* Sections of cardiac tissue and limb tissue, both from 12wpc fetuses (N1782 and N1593 respectively) were used for this experiment. A sonic hedgehog mRNA probe, expressed in bone, was used on the limb tissue as a positive control. No signal was seen in limb or heart with either the CYP21A2 or the scrambled (control) LNA probe (Figure 21). A positive signal was seen in bone with the sonic hedgehog mRNA control probe. This suggests that the CYP21A2 LNA probe does not bind to fixed tissue in a completely non-specific matter.

The probe designs for the CYP21A2 and CYP21A1P LNA probes were reassessed by the designers at Exiqon, however due to the high homology between gene and pseudogene, the original designs could not be improved.

In summary, the SP6 mRNA probe demonstrated *CYP21A2* expression in fetal adrenal only and a 2000bp band was seen by Northern blotting, appropriate to the *CYP21A2* transcript. The CYP21A1P LNA probe did not produce a signal in fetal adrenal or kidney at the recommended concentration; at higher concentrations, it was found to hybridise non-specifically. In contrast, the CYP21A2 LNA probe showed marked and equal expression in fetal adrenal and kidney when used at the recommended concentration. However, hybridisation to an appropriate sized transcript could not be determined by Northern blotting. As we lacked a positive control tissue for the CYP21A1P probe, it was

impossible to determine whether *CYP21A1P* is just not expressed in adrenal, or whether the CYP21A1P LNA probe was simply not working. In addition, the discrepancy between the mRNA riboprobe and CYP21A2 LNA probe TISH results could not be satisfactorily explained. Due to these technical challenges, the use of further fetal tissue, a rare resource, for continuing TISH experiments was not felt to be justified, as the results would be difficult to interpret. Therefore, the TISH investigation was abandoned and an alternative line of investigation, using qPCR, was planned. The aim of qPCR was twofold: firstly we aimed to determine if *CYP21A1P* is expressed in fetal adrenal and thymus; secondly, we aimed to use qPCR to investigate *CYP21A2* expression in kidney, thus clarifying the contradictory results generated by TISH.

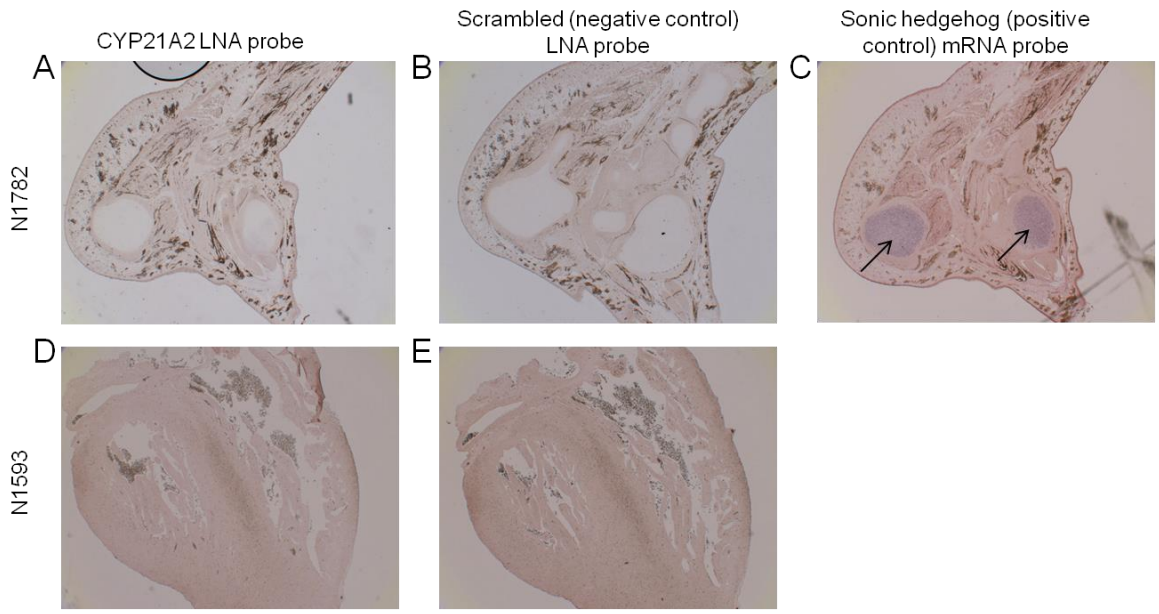


Figure 21: LNA probe in situ hybridisation control tissue experiment.

In situ hybridisation using the CYP21A2 and scrambled LNA probes at 40nM on 12 weeks post-conception fetal limb (N1782, upper panel) and cardiac tissue (N1593, lower panel) under light-field conditions at 2.5x magnification. A sonic hedgehog mRNA probe, which hybridises to bone, was used at a concentration of 150ng per slide as a positive control for the experiment (panel C, hybridisation to bone indicated by arrows). The CYP21A2 and scrambled LNA probes do not hybridise to limb tissue (panels A and B respectively) or cardiac tissue (panel D and E respectively). The specific hybridisation to bone seen with the sonic hedgehog positive control riboprobe (panel C) demonstrates that the *in situ* hybridisation technique has worked and that lack of staining with the other two probes is not due to experimental failure.

3.4.2.2 FLUORESCENCE-BASED QUANTITATIVE REAL-TIME PCR TO DETERMINE CYP21A2 AND CYP21A1P EXPRESSION

3.4.2.2.1 qPCR PRIMER EFFICIENCY

For quality control, the primer efficiencies were first determined by running the standard curve samples in triplicate and plotting the results (Figure 22). The gradient of the slope for both the *CYP21A2* and the *CYP21A1P* assays was then determined by linear regression and PCR efficiency calculated. Excellent primer efficiencies were confirmed as the gradients for the slope plotted for the standard curve for the *CYP21A2* and *CYP21A1P* assays were -3.22 [95% CI -3.07 to -3.37] and -3.26 [95% CI -3.18 to -3.35] corresponding to primer efficiencies of 104.5% and 102.6% respectively. These results replicated well when run a further two times on different days (Table 10).

3.4.2.2.2 qPCR LINEAR DYNAMIC RANGE

Using the standard curve samples run on each plate, the linear dynamic ranges of the assay were determined (Figure 22). The linear dynamic range of each assay varied slightly from plate to plate. For the *CYP21A2* assay, generally C_T values of less than 16.0 and greater than 37.0 were considered beyond the limits of assay detection for the *CYP21A2* assay and C_T values of less than 14.0 and greater than 37.0 were considered beyond the limits of assay detection for the *CYP21A1P* assay. Samples with low CT values could be diluted for analysis (however this was not necessary for any sample analysed).

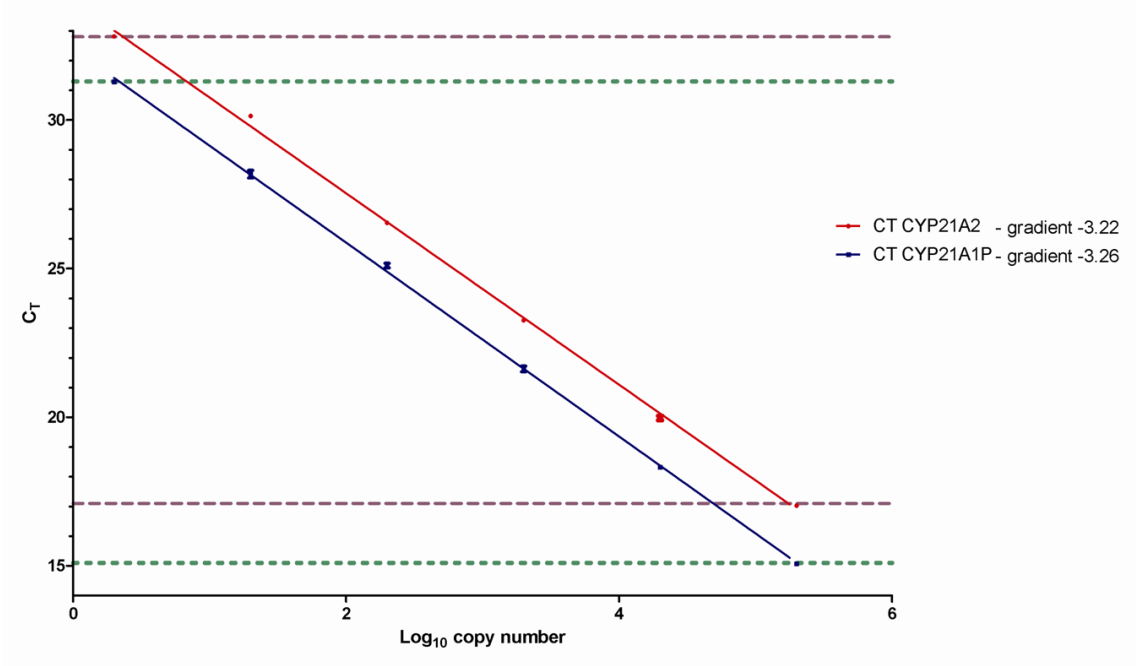


Figure 22: qPCR primer efficiencies and linear dynamic range for the CYP21A2 and CYP21A1P assays.

Standard curves plotted for the CYP21A2 (red) and CYP21A1P (blue) qPCR assays. The mean C_T values are plotted with the standard error of the mean. The Log_{10} copy number is shown on the x axis and the C_T value on the y axis. The gradients of each slope were calculated by linear regression and confirmed excellent primer efficiency for both assays. The linear dynamic range of each assay is shown by the dashed purple line for the CYP21A2 assay and the dashed green line for the CYP21A1P assay.

Assay	Gradient of slope replicate 1 [95% CI]	Gradient of slope replicate 2 [95% CI]	Gradient of slope replicate 3 [95% CI]
CYP21A2	-3.22 [-3.07 to -3.37]	-3.13 [-3.00 to -3.26]	-3.15 [-2.97 to -3.33]
CYP21A1P	-3.26 [-3.18 to -3.35]	-3.14 [-3.05 to -3.23]	-3.34 [-3.30 to -3.37]

Table 10: CYP21A2 and CYP21A1P qPCR primer efficiency replicates.

Standard curve gradients, with 95% confidence intervals, generated from standard samples run in triplicate on three separate occasions.

3.4.2.2.3 CYP21A1P ASSAY SPECIFICITY

The qPCR assays were used on cDNA synthesised from two *CYP21A1P*+ fetal adrenal samples (N1736, 9wpc and N1812, 12wpc) and from one *CYP21A1P*– fetal adrenal sample (N1733, 11wpc). The aim was to ensure that the *CYP21A1P* assay was specific for pseudogene transcripts and was not erroneously amplifying the highly homologous *CYP21A2* transcripts. RNA samples from each specimen were also included to ensure that the RNA samples were not contaminated by gDNA, which could also give rise to spurious results. This experiment was repeated on two separate days to ensure the validity of the results and to allow direct comparison of results. An absolute copy number/µl was calculated for each sample (up to 6 replicates obtained from samples run in triplicate on two occasions).

No amplification was detected with the non-template controls or the RNA samples included in this experiment on either occasion, which confirmed that the qPCR plates were not contaminated and that the RNA samples were free from gDNA contamination. *CYP21A2* transcripts were detected in all three fetal adrenal samples as would be expected (mean 447,057 copies/µl). Sample N1812 had a significantly lower number of copies/µl compared to samples N1736 and N1733 (mean 55,750, 726,170 and 559,251 copies/µl respectively). *CYP21A1P* transcripts were detected in the adrenal samples from N1812 and N1736 (mean 1374 and 8858 copies/µl respectively) but were not detected in any of the replicates from *CYP21A1P*- sample N1733 demonstrating assay specificity. In samples N1812 and N1736, where *CYP21A2* and *CYP21A1P* transcripts were both expressed, *CYP21A2* transcripts were 40 to 80 fold more abundant than *CYP21A1P* transcripts. These data confirm that the CYP21A1P assay is specific to *CYP21A1P* transcripts. In addition, it confirms that both *CYP21A2* and *CYP21A1P* transcripts are expressed in the fetal adrenal gland, with *CYP21A1P* transcripts expressed at 40 to 80 fold lower levels than the more abundant *CYP21A2* transcripts (Figure 23).

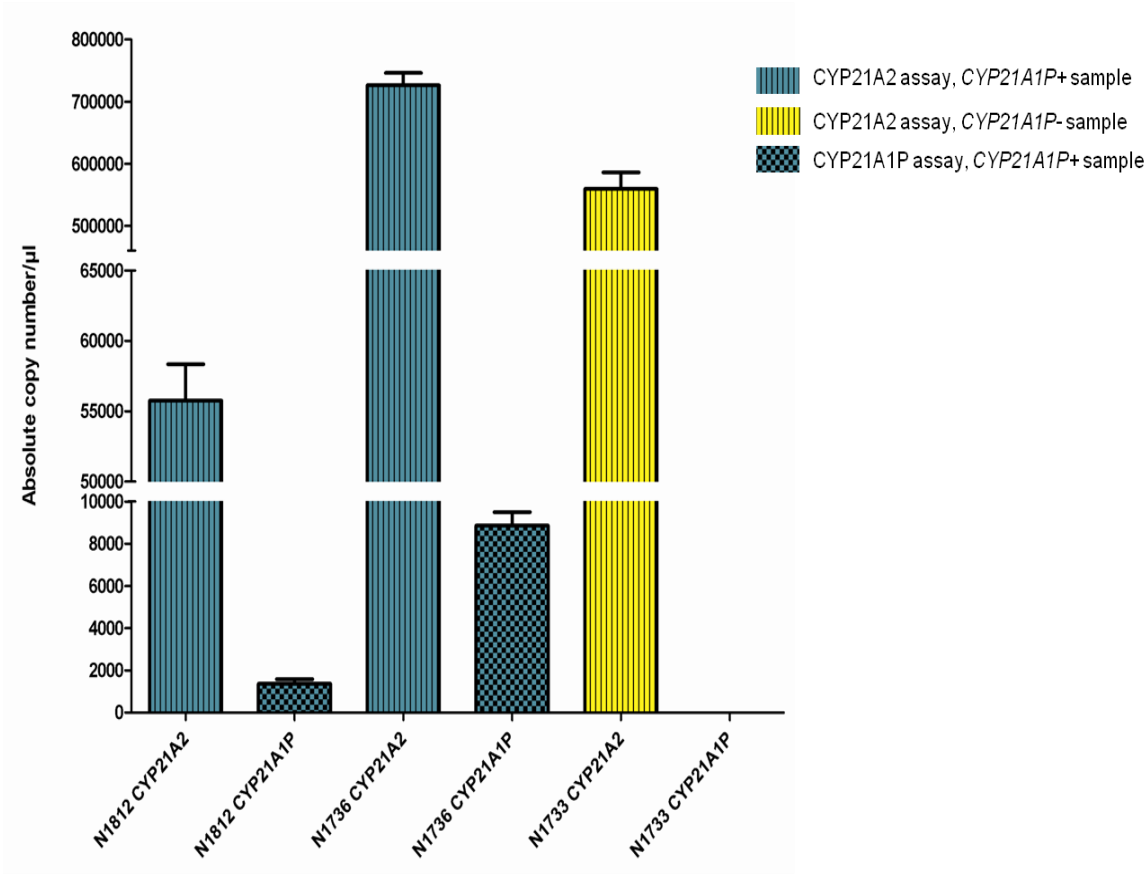


Figure 23: CYP21A1P qPCR assay specificity.

Mean absolute copy number/µl (and standard error of the mean) are shown for three fetal adrenal samples, analysed in triplicate and on two different days. Samples N1812 and N1736 are gDNA *CYP21A1P*+ (green) while N1733 is *CYP21A1P*– (yellow). *CYP21A2* transcripts (striped columns) were detected in all three samples, with N1812 having lower copies/µl (mean 55,750 copies/µl) than samples N1736 and N1733 (mean 726,170 and 559,251 copies/µl respectively). *CYP21A1P* transcripts (chequered columns) were detected in the adrenal samples from N1812 and N1736 (1374 and 8858 copies/µl respectively) but were not detected in any of the triplicates from *CYP21A1P*– sample (N1733 shown in yellow), which demonstrated assay specificity.

3.4.2.2.4 CYP21A2 AND CYP21A1P EXPRESSION IN THYMUS

Of the seven cDNA samples prepared from thymus tissue, two were rejected (thymus samples 4 and 5) due to gDNA contamination, despite careful preparation and DNase treatment, detected by running RNA samples with the cDNA samples in the qPCR assay. Of the five remaining samples, *CYP21A2* was expressed at very low levels in three of the five samples (thymus samples 1, 3 and 7: 2.5, 1.7 and 6.5 copies/µl respectively) (Figure 24). These levels were at the very limit of assay detection. No *CYP21A2* expression was detected in thymus samples 2 and 6. In contrast, *CYP21A1P* was expressed at low levels in all five thymus samples. The mean number of copies/µl of *CYP21A1P* was 8.16 (median 4.7, minimum 3.4, maximum 23.5). *CYP21A1P* was 1.5 to 5 fold more abundant in thymus than *CYP21A2*. A Mann-Whitney U test was performed to determine if the copy number/µl of *CYP21A2* was significantly different to the copy number/µl of *CYP21A1P* per sample, but the result did not reach statistical significance (P 0.075).

When the mean *CYP21A1P* copy number was normalised to the mean *CYP21A2* copy number in each sample for the three fetal adrenal samples and the five thymus samples, we observed that *CYP21A1P* appeared to be more abundant in thymus compared to *CYP21A2* (which was indeed absent in two thymus samples in multiple replicates). Conversely, in adrenal, *CYP21A1P* is significantly less abundant in adrenal when compared to *CYP21A2* transcript levels (Figure 25).

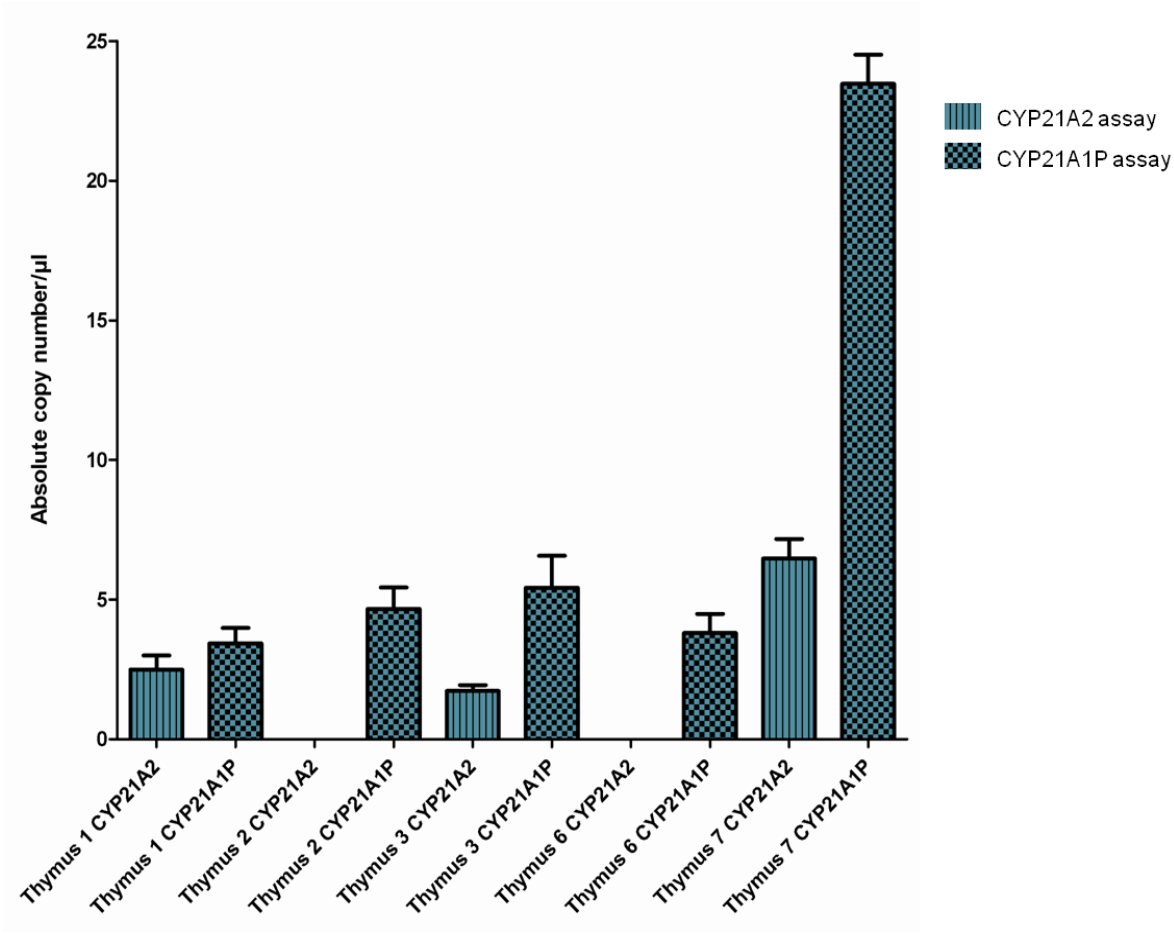


Figure 24: CYP21A2 and CYP21A1P expression in thymus.

Mean absolute copy number/µl (and standard error of the mean) are shown for five thymus samples (all *CYP21A1P*+). *CYP21A2* transcripts (striped columns) were detected at low levels in three of the five samples, however no copies were detected in thymus samples 2 and 6. *CYP21A1P* transcripts (chequered columns) were detected in all five samples and were 1.5 to 5 fold more abundant than *CYP21A2* transcripts, although this difference was not statistically significant.

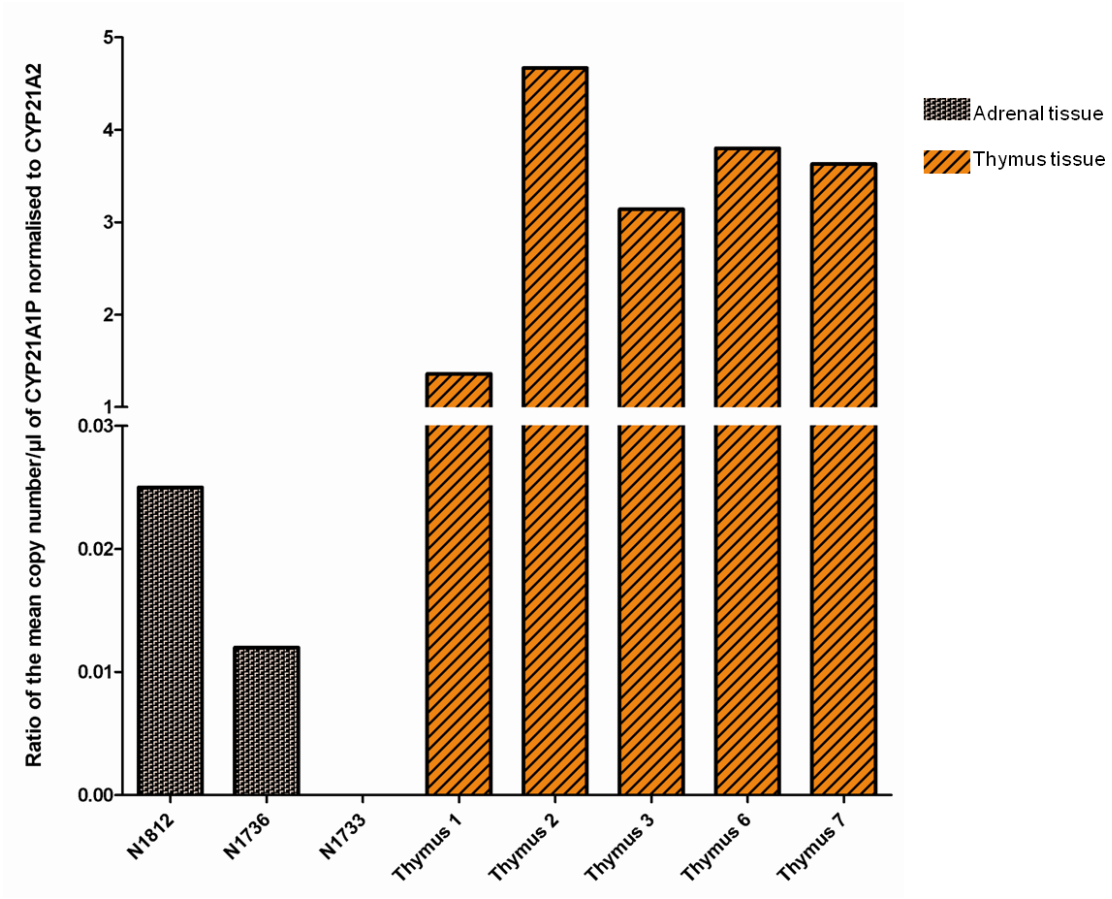


Figure 25: Mean copy number of CYP21A1P normalised to the mean copy number of CYP21A2 for three fetal adrenal samples (N1812, N1736 and N1733) and five thymus samples (thymus 1, 2, 3, 6 and 7).

There is no *CYP21A1P* expression in N1733 as it is gDNA *CYP21A1P*–. In fetal adrenal *CYP21A2* is much more abundant than *CYP21A1P*. In contrast, in thymus, *CYP21A1P* is expressed at low levels and is slightly more abundant than *CYP21A2.* However, the difference was not statistically significantly. *CYP21A2* is not always expressed in thymus (thymus 2 and 6 both express *CYP21A1P* but not *CYP21A2*).

3.4.2.2.5 CYP21A2 AND CYP21A1P EXPRESSION IN KIDNEY

In order to investigate the TISH results further, in particular whether *CYP21A2* is expressed in renal tissue as indicated by the CYP21A2 LNA probe, we used the qPCR assays on kidney tissue obtained from samples N1812, N1733 and N1736. This was carried out in order to compare copy numbers of *CYP21A2* transcript in kidney and adrenal.

CYP21A2 was expressed in kidney tissue in all three samples. Sample N1812 had the least copies of *CYP21A2* in adrenal (mean 55,750 copies/µl) but the most copies in kidney (mean 3266 copies/µl), with *CYP21A2* being 17 fold more abundant in adrenal compared to kidney. In contrast, samples N1733 and N1736 had greater copy numbers in adrenal (mean of 726,170 and 559,251 copies/µl respectively) but much lower expression levels in kidney (56 and 22 copies/µl respectively). In sample N1733, *CYP21A2* was expressed 9900 fold more abundantly in adrenal than kidney and in sample N1736, it was expressed 33,000 fold more abundantly in adrenal than kidney (Figure 26). These data are contrary to the TISH results with the CYP21A2 LNA probe, which suggested that *CYP21A2* is expressed at an equal or greater level in kidney than adrenal.

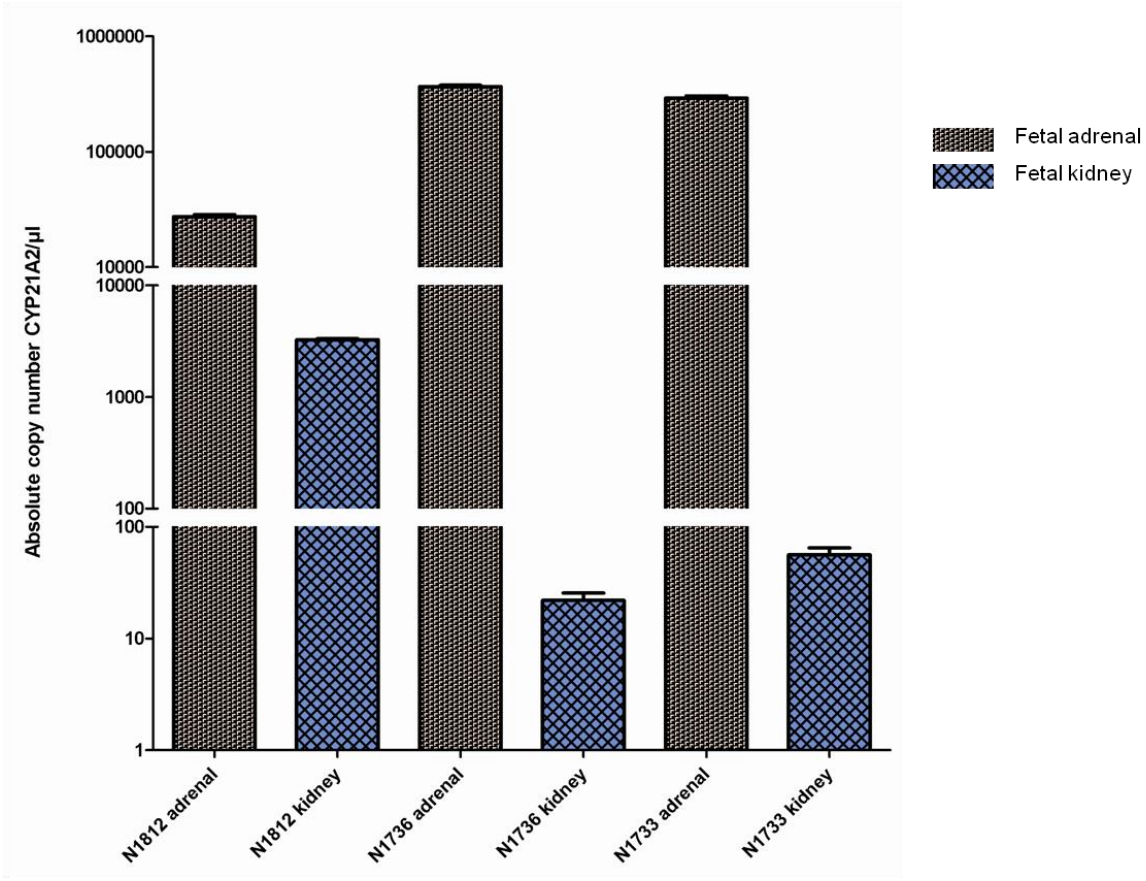


Figure 26: CYP21A2 expression by qPCR in three matched fetal adrenal and kidney samples.

Mean absolute copy number/µl (and standard error of the mean) are shown for three fetal adrenal (grey spotted bars) and kidney (blue chequered bars) samples. These samples have been analysed in triplicate twice. *CYP21A2* transcripts were detected in both adrenal and kidney tissue in all three samples. Sample N1812 had the least copies of *CYP21A2* in adrenal (mean 55,750 copies/µl) but the most copies in kidney (mean 3266 copies/µl). N1736 had a mean of 726,170 copies/µl in adrenal but only 22 copies/µl in kidney. Sample N1733 had a mean of 559,251 copies/µl in adrenal and 56 copies/µl in kidney. In all samples, there was at least a 17 fold increase in copies of *CYP21A2* in adrenal when compared to kidney.

3.5 DISCUSSION

The *CYP21* locus on chromosome 6 is perhaps the most obvious candidate for investigation in AAD, as it encodes the 21OH enzyme which is the primary autoantigen in this condition. While this locus has been extensively studied in individuals with CAH, relatively little is known of its role, if any, in susceptibility to AAD. Two previous studies[153, 267], both conducted in the 1990s, have investigated polymorphisms in the *CYP21A2* gene. However, due to the strong LD which extends some distance from the HLA region, they were both unable to identify an independent effect of *CYP21A2* polymorphisms, above and beyond that conferred by the MHC class II alleles. Therefore, the *CYP21* locus remains the "elephant in the room" for AAD geneticists: it is the most obvious and biologically plausible candidate locus for AAD and as such, warrants further investigation, but is frequently overlooked due to the difficulties of dissecting its role from that of the HLA alleles nearby.

The preliminary work in this study has established that individuals with AAD are more likely to have no copies of the *CYP21A1P* pseudogene (synonymous with being homozygous for the monomodular RCCX), when compared with controls: 17.5% of AAD individuals have zero copies of *CYP21A1P* compared to just 3.0% of controls (P <0.00001, OR 6.77 [95% CI 3.94 – 11.63]). On review of the available clinical information for all AAD individuals with and without *CYP21A1P* copies, there were no differences in clinical characteristics between the two groups. This finding led to the hypothesis that *CYP21A1P* might be important in inducing tolerance to the steroidogenic machinery and that its absence might therefore predispose to AAD. However, following genotyping of a cohort of individuals with Graves' disease without concomitant AAD, where 7.8% had no *CYP21A1P* gene compared to 3.0% of controls (P 0.0007, OR 2.72 [95% CI 1.53 – 4.82]), it appears that, although the association is much stronger with AAD than with Graves' disease, this finding is not restricted to AAD alone.

Previous studies have demonstrated that individuals who are homozygous for the monomodular *RCCX*, in particular those with a null *C4A* gene, are predisposed to autoimmune conditions including SLE[268] and type 1 diabetes[269, 270]. However, the monomodular RCCX associated with these conditions is, in European Caucasians, in strong LD with the *HLA–A1–B8–DR3–DQ2* haplotype, also known as the 8.1 ancestral haplotype (8.1 AH). The 8.1 AH is strongly linked to autoimmunity, with the *HLA–DRB1*0301* allele appearing to have a particularly strong influence[271]. The 8.1 AH has been associated with multiple autoimmune conditions including type 1 diabetes[272], SLE[273], coeliac disease[274, 275] and also AAD[45, 276]. There is therefore some debate as to whether components of the monomodular RCCX (or, more likely, absence of components from the duplicated RCCX, such as the *CYP21A1P* pseudogene[277] or additional copies of the functional *C4* gene[278]), are conferring independent risk to autoimmunity[279-281] or simply "tagging" an HLA haplotype which is influencing disease susceptibility.

The UK AAD, Graves' disease and control cohorts have not been HLA-genotyped as this is an expensive undertaking. However, HLA data has been gathered for a significant proportion of the Norwegian AAD cohort and this data has been kindly shared with us by our collaborators in Bergen. Using our competitive PCR assay, of 319 AAD cases genotyped for *CYP21A1P* presence (*CYP21A1P*+) or absence (*CYP21A1P*–), 42 (13.2%) were *CYP21A1P*–. This compared to just 19/413 (4.6%) of controls (P 3.2×10^{-5}, OR 3.14 [95% CI 1.79 – 5.52]). This result was similar to our findings in the UK AAD case and control cohorts.

HLA data were available for 35 of the 42 *CYP21A1P*– Norwegian AAD cases. 22 of 35 (62.9%) *CYP21A1P*– individuals were homozygous for both *HLA–DRB1*0301* and *HLA–DQB1*02 (HLA-DR3)*, compared with only 3 of 230 (1.3%) *CYP21A1P*+ individuals. In addition, 24 of 35 (68.6%) *CYP21A1P*– individuals were homozygous for the *HLA-B*08* allele compared to just 3 of 228 (1.3%) *CYP21A1P*+ people. 13 of 34 (38.2%) of *CYP21A1P*– people were found to be homozygous for *HLA-A*01* compared to 10 of 228 (4.4%) of

CYP21A1P+ individuals. Finally, 28 of 34 (82.4%) *CYP21A1P*– individuals with AAD were homozygous for the *MICA*5.1* allele compared to 108 of 229 (47.2%) of *CYP21A1P*+ AAD patients. In addition, HLA data were available for 323 healthy controls (303 *CYP21A1P*+, 20 *CYP21A1P*–). 9/20 (45.0%) *CYP21A1P*– controls were homozygous for both *HLA–DRB1*0301* and *HLA–DQB1*02 (HLA-DR3)*, compared with only 6 of 303 (2.0%) *CYP21A1P*+ controls. Furthermore, 9 of 20 (45.0%) *CYP21A1P*– controls were homozygous for the *HLA-B*08* allele compared to just 1 of 303 (0.3%) *CYP21A1P*+ controls. 6 of 20 (30.0%) of *CYP21A1P*– controls were found to be homozygous for *HLA-A*01* compared to 3 of 303 (1.0%) of *CYP21A1P*+ controls. Finally, 10 of 20 (50.0%) *CYP21A1P*– controls were homozygous for the *MICA*5.1* allele compared to 91 of 302 (30.1%) of *CYP21A1P*+ controls (Husebye *et al*, unpublished data).

HLA haplotype reconstruction using PHASE software, performed by Ingeborg Bronstad (Husebye group, Bergen, Norway), was used to compare the frequency of the HLA haplotype of interest, the 8.1 AH (*HLA–A*01–B*08–DRB1*0301–DQB1*02*) in *CYP21A1P*+ and CYP21A1P– AAD cases and then in controls, before comparing cases with controls. This analysis was performed using χ^2 tests. This analysis showed that the 8.1 AH is associated with *CYP21A1P*– in both AAD cases and in controls (P 9.6×10^{-7}, OR 3.62 [95% CI 2.11-6.21]; P 1.2×10^{-7}, OR 5.81 [95% CI 2.86-11.82] respectively). There was no significant difference in the frequency of the 8.1 AH haplotype between *CYP21A1P*+ AAD cases and controls (P 0.076) and between *CYP21A1P*– AAD cases and controls (P 0.81).

These data suggest that absence of *CYP21A1P* is commonly associated with the 8.1 AH in both cases and controls and can be viewed as tagging this autoimmune-associated haplotype. This association explains why we have observed an excess of *CYP21A1P*– individuals in our cohort of Graves' disease patients in addition to those with AAD, as Graves' disease is also associated with the 8.1 AH (reviewed in [76]).

The association between the absence of *CYP21A1P* (indicating a homozygously monomodular RCCX) and AAD, noted in both the UK and Norwegian AAD cohorts, was more marked than that seen with Graves' disease (OR 6.77 and 2.72 respectively). Although available HLA data and HLA haplotype analysis suggest that at least some of this association is due to LD with the 8.1 AH, this observation begs the question of whether this explains the entire association or whether other factors are at play. In Caucasian individuals, the presence of the *C4A null* allele, which is often seen in conjunction with a monomodular RCCX, has been shown to increase risk of SLE[282]. Although this was originally attributed to the influence of HLA on susceptibility to SLE, the increased susceptibility was noted both in those with, and without, the 8.1 AH[273], suggesting independent effects. In support of the *C4 null* allele being an independent risk factor for SLE, it is also associated with SLE in populations where the 8.1 AH is rare, for example in Chinese and Korean SLE cohorts[283-285]. The *C4 null* allele is thought to result in partial C4 deficiency. Complete C4 deficiency is rare and is associated with homozygosity of the *C4A* and *C4B* null alleles. It has been shown to be associated with impaired antibody responses to antigens presented by T cells in both humans and guinea pigs[286] and with an immune-complex mediated lupus-like disease in humans[287]. In contrast, partial C4 deficiency results in reduced clearance of immune complexes and autoimmune disease[288]. In AAD, it is possible that those people who are homozygous for the monomodular RCCX, and therefore *CYP21A1P*–, might also carry the *C4 null* allele, which would result in partial C4 deficiency. This could lead to subtle perturbations of the complement cascade which may be contributing to the autoimmune process in AAD, independent of the association with HLA. This hypothesis has not been tested in this study and requires investigation.

The finding of *CYP21A1P* absence in a significantly greater proportion of AAD patients compared to controls provoked interest in the expression of *CYP21A1P* transcripts and led to the hypothesis that a steroid-rich thymic environment might be toxic for emerging T lymphocytes and impede the positive/negative selection process, and that pseudogene transcripts could therefore be expressed in thymus as a mechanism for inducing tolerance in a steroid-neutral

environment. In a *CYP21A1P*– individual, failure to express pseudogene transcripts in thymus during development might mean that tolerance to key steroidogenic components is not established, ultimately leading to an aberrant immune response to those elements later in life, which could cause AAD. Therefore, the possibility of pseudogene transcripts being expressed in the thymus was of particular interest. As the thymus gland involutes in adults, thymus samples for this project were obtained from infants and children undergoing cardiothoracic surgery for congenital heart defects from the Birmingham University HBRC. In order to establish and optimise assays of *CYP21A1P* expression, a positive control tissue was required. Adrenal tissue was chosen as a positive control tissue because *CYP21A2* is known to be expressed here in fetal life[257] and levels of *CYP21A1P* would be compared, per sample, to levels of *CYP21A2*. A known positive control tissue for *CYP21A1P* expression has not previously been established. However, *CYP21A1P* transcripts have previously been detected in adrenal cell lines[265], and therefore adrenal tissue would seem to be the most likely site of expression of *CYP21A1P* if indeed it is expressed.

Using TISH and an antisense SP6 mRNA riboprobe on *CYP21A1P*– fetal adrenal tissue, we established expression of *CYP21A2* transcripts in fetal adrenal tissue obtained from 10 to 12 wpc fetuses. A strong positive signal, most marked at the periphery of the fetal adrenal gland, was seen in the three samples tested (Figure 17). We confirmed that the mRNA SP6 antisense riboprobe was hybridising to a transcript approximately 2000bp in length by Northern blotting (Figure 18). The 2000bp single band seen in Northern blotting corresponds to two known protein-coding *CYP21A2* transcripts: 1914bp and 2182bp in length. The finding of abundant CYP21A2 expression in fetal adrenal is in keeping with reports in published literature[257, 258]. No signal was seen in the adjacent fetal kidney, suggesting that *CYP21A2* transcripts are either not expressed here, or are expressed at very low levels which could not be detected with the mRNA probe. qPCR has previously been used to demonstrate abundant *CYP21A2* expression in fetal adrenal. *CYP21A2* transcripts were also detected in fetal kidney, but at much lower levels, between 4 to 70 fold less than in the adrenal[258]. As qPCR is an amplification method, it is likely to be more

sensitive than TISH for targets of low abundance, explaining the discrepancy between our negative TISH kidney results and previous qPCR data.

When the SP6 mRNA probe was applied to *CYP21A1P*+ tissue, a positive signal was also seen, in the same distribution (Figure 17) and there was little discernible difference in hybridisation when compared with the results from *CYP21A1P*– tissue. The signal seen in this tissue could be due to the presence of *CYP21A2* and/or *CYP21A1P* transcripts, as the mRNA probe could not be designed to differentiate between the two transcripts. The LNA probes were therefore applied to both *CYP21A1P*+ and *CYP21A1P*– tissue to try to establish whether *CYP21A1P* is expressed in fetal adrenal tissue prior to investigating thymus.

The CYP21A1P LNA probe did not hybridise to fetal adrenal or kidney when used at the manufacturer's recommended concentration (20–40nM). At a concentration sufficient to generate a signal in *CYP21A1P*+ tissue (approximately 200nM), non-specific hybridisation occurred (Figure 19). At 200nM, a signal was seen in both fetal adrenal and kidney in *CYP21A1P*– tissue. Indeed, when the scrambled probe was used at this high concentration, the same pattern of non-specific hybridisation was seen. As we did not have a positive control tissue, we were unable to ascribe the lack of signal when used at 20–40nM concentration to *CYP21A1P* not being expressed and therefore the results of the CYP21A1P LNA probe experiments had to be disregarded.

In contrast, the CYP21A2 LNA probe gave a strong signal in both fetal adrenal and kidney when used at 20nM (Figure 19). However, when this probe was used in conjunction with Northern blotting at a range of concentrations, no band was seen and therefore we could not confirm the size of the transcript to which the probe was hybridising. We successfully demonstrated that the CYP21A2 LNA probe did not hybridise to limb or cardiac tissue sections (Figure 21), and therefore was not binding in a completely non-specific manner to fixed tissue. The observation of equal or stronger signal from the kidney than the adrenal was in stark contrast to the mRNA riboprobe results, which suggested no, or

only very low, expression of *CYP21A2* in the fetal kidney. While LNA probes are short and end-labelled with DIG molecules for detection, the mRNA riboprobes used in this project have multiple DIG molecules incorporated throughout the length of the longer probe. These are introduced by labelling 1 in 3 dUTP molecules with DIG, which are incorporated into the probe during the *in vitro* transcription reaction. As a consequence, one would anticipate that the signal gained from a DIG-labelled mRNA riboprobe would be significantly greater than that gained from an end-labelled LNA probe targeting the same transcript.

Following these initial TISH results, and in light of the discrepancy between expression patterns with the different probes, both the mRNA riboprobe and CYP21A2 LNA probe results were scrutinised carefully. The mRNA riboprobe results were felt to be accurate for a number of reasons, while the CYP21A2 LNA probe results were eventually viewed with suspicion.

The mRNA riboprobe results were in keeping with previously published findings[258] and Northern blotting was used successfully to establish that the SP6 probe was hybridising to an appropriate sized transcript. The lack of signal from kidney with the mRNA riboprobe was not felt to be due to protocol failure, as strong and specific signal was seen in the adrenal gland. In addition, good signal from the adrenal eliminates the possibility that RNases were introduced to the tissue sections during the TISH protocol. Had this been the case, RNA degradation would be seen across the whole tissue section, leading to lack of signal in both adrenal and kidney. Furthermore, tissue sections from the three samples used for this TISH project had previously been used in unrelated projects with riboprobes specific to renal tissue, for example FRMD7 and collagen 6, with excellent results. This suggests that significant and localised RNA degradation in the renal tissue adjacent to the adrenal gland prior to tissue fixation and storage was not the cause for the absence in signal from the kidney. In contrast, the expression pattern seen with the CYP21A2 LNA probe, with strong expression more marked in kidney than adrenal, is entirely contrary to previous findings. Even using the CYP21A2 LNA probe at high concentration in Northern blotting, the size of transcript to which the probe was binding could

not be identified. As entirely non-specific hybridisation had been eliminated by using the CYP21A2 LNA probe on limb and cardiac tissue with negative results, this raised the possibility that the probe might be hybridising to a small, abundant, homologous target transcript: for example, an uncharacterised microRNA, small enough in size to have completely “run off” the RNA formamide gel during electrophoresis.

As the CYP21A1P LNA probe results had been unsatisfactory, an alternative line of investigation was sought to establish *CYP21A1P* expression. Two qPCR assays were therefore designed, one specific to *CYP21A2* and one to *CYP21A1P*, and an absolute copy number method was used to analyse results. qPCR would also be used to investigate *CYP21A2* and *CYP21A1P* expression in fetal kidney, thus investigating further the discrepancy between the TISH mRNA riboprobe and CYP21A2 LNA probe results.

Using qPCR, initial quality control experiments on three fetal adrenal tissue samples (two gDNA *CYP21A1P*+, one gDNA *CYP21A1P*-) demonstrated that both the CYP21A2 and CYP21A1P assays were sensitive, being able to detect as few as 1 copy/µl, and specific. Using these assays, we demonstrated, for the first time, that in individuals who are gDNA *CYP21A1P*+, *CYP21A1P* transcripts are expressed in fetal adrenal gland between 10 and 12 wpc. In this tissue, *CYP21A1P* transcripts are 40 to 80 fold less abundant than *CYP21A2* transcripts, which are expressed at high levels (mean 5116 copies/µl *CYP21A1P;* 447,057 copies/µl *CYP21A2*) (Figure 23).

Following assay optimisation in fetal adrenal tissue, thymus samples were assessed to determine *CYP21A2* and *CYP21A1P* expression. Using qPCR, of five thymus samples that met with quality control criteria, very low levels of *CYP21A2* were expressed in three thymus samples (1.7, 2.5 and 6.5 copies/µl) and in two samples, *CYP21A2* was not expressed. This finding reflects previous published results, where qPCR has been used to demonstrate *CYP21A2* expression in thymus[258]. In the same five samples, we established, for the first time, that *CYP21A1P* is expressed in thymus, again at low levels. Expression

was seen in all five samples, with a mean copy number/µl of 8.16 (median 4.7, minimum 3.4, maximum 23.5). In contrast to fetal adrenal tissue, in thymus, *CYP21A1P* was expressed at 1.5 to 5 fold higher levels than *CYP21A2* (mean 2.14 copies/µl) (Figure 24), although this difference was not statistically significant (P 0.075). In comparison to fetal adrenal tissue, levels of both *CYP21A2* and *CYP21A1P* transcripts in thymus were considerably lower.

In order to further investigate the discrepancy between TISH results with the mRNA riboprobe and the CYP21A2 LNA probe, RNA was extracted from kidney samples matched to adrenal samples already studied (N1812, N1733 and N1736). Matched samples would allow comparison of *CYP21A2* copy number/µl in adrenal and kidney. If the copy number in kidney was significantly lower than the copy number in adrenal, this would support that the results gained from the TISH experiment with the mRNA riboprobe and suggest that the results gained with the CYP21A2 LNA probe were spurious. However, if copy number was approximately equal in the two tissues, or greater in kidney than adrenal, this would be in keeping with the findings with the CYP21A2 LNA probe.

CYP21A2 transcripts were detected in all three of the matched kidney samples. There was significant variation in copy number between the three samples. *CYP21A2* transcripts were expressed at 17 to 33,000 fold lower levels in kidney compared to adrenal (Figure 26). The mean number of copies/µl in kidney was 1115 (minimum 22, maximum 3266 copies/µl) compared to 447,057 copies/µl in adrenal (minimum 55,750, maximum 726,170 copies/µl). This result suggests that the TISH results gained from the CYP21A2 LNA probe, which showed an equal or stronger signal from kidney than adrenal, were not reliable.

There was considerable variability between samples in the number of copies of each transcript present/µl. This variability was apparent even between samples of the same developmental age. This variability could reflect true differences in transcript copy number between samples but could also reflect technical factors, for example, conditions affecting RNA quality and levels degradation, such as

the length of time taken from sample procurement to storage, sample transport conditions prior to freezing and the length of time the sample had been frozen prior to use. In addition, some variation is inevitably introduced during the RNA extraction, cDNA preparation and qPCR procedure. Finally, adrenal and thymus tissue samples were obtained from Newcastle and Birmingham biobank facilities. As only very small samples were available, we were unable to spare tissue for histological examination in order to determine the proportions of medulla and cortex in each sample. Differing proportions of each tissue may account for some variability between the samples.

In summary, these results indicate that *CYP21A1P* transcripts are expressed in relatively low numbers, both in the adrenal gland in fetal life, and in the thymus in infants and children. This observation is in keeping with a previous study that demonstrated *CYP21A1P* transcripts in an adrenal cell line[265]. This observation is intriguing as these transcripts would be predicted to undergo nonsense-mediated mRNA decay (NMD). NMD is an important cellular “quality control” mechanism whereby mRNAs containing premature translation termination codons are rapidly degraded. This pathway is in place to prevent accumulation of truncated and abnormal proteins in cells[289]. The observation of *CYP21A1P* transcripts in both fetal adrenal tissue and in thymus tissue can perhaps be explained in one of two ways: first, that the *CYP21A1P* transcripts are subject to NMD and that we are simply detecting a small number which are yet to be removed by this regulatory system. In thymus, where copy numbers were very low, usually in single figures, this explanation seems likely. The other possibility is that the *CYP21A1P* transcripts are NMD-insensitive. This phenomenon has previously been reported in both humans and in other species[290-292], however the mechanism or mechanisms by which some transcripts are able to evade NMD are not yet fully understood and are subject to continuing research. The key question now is whether a CYP21A1P protein product can be detected. If a protein product cannot be detected, this supports the idea that the transcripts observed in adrenal and thymus tissue are transient and subject to NMD, however if a CYP21A1P protein product can be detected, this supports the latter suggestion that the transcripts are protected in some way from the NMD mechanism.

This study has also demonstrated that LNA probes do not always hybridise specifically, and cannot always be used to differentiate between highly homologous transcripts. In this study, comparison of LNA probe TISH results with those gained using a traditional mRNA riboprobe highlighted significant discrepancies and these were resolved using qPCR in favour of the mRNA riboprobe results. Comparison of LNA probe data with riboprobe data is therefore a useful tool in differentiating specific from non-specific hybridisation.

3.6 CONCLUSIONS AND FUTURE DIRECTIONS

While this small study has provided new evidence that *CYP21A1P* transcripts are expressed in both fetal adrenal and thymus, with such small sample sizes it would be inappropriate to draw wider conclusions at this stage. The sample size is limited by availability of fetal tissue and tissue from infants and children, which is an extremely rare resource. qPCR could be used to analyse a greater sample set and to look at different developmental stages and a range of different tissues. This would give a more accurate overview of variability of inter-individual transcript copy numbers and variation in transcript levels at different developmental stages in various tissues.

The next step in this study is to determine whether a CYP21A1P protein product is expressed. This could be done using Western blotting and an antibody to the N-terminus, which is predicted to be shared by the CYP21A2 protein and the putative CYP21A1P protein product. While the protein product of the *CYP21A2* gene is between 465 and 495 amino acids in length, the predicted pseudogene protein product, if one does exist, would be considerably shorter, approximately 131 amino acids in length. This difference would be due to the deletion in exon three, which introduces a premature stop codon and frameshift, resulting in truncation of the predicted pseudogene protein. These different sized proteins could be easily differentiated by size by Western blotting in tissues of interest. Expression of a pseudogene protein product would confirm that *CYP21A1P* transcripts in some way evade the NMD mechanism and would also lend support to our hypothesis that it plays a role in establishing tolerance to essential components of the steroidogenic machinery.

CHAPTER 4 – A HYPOTHESIS-DRIVEN APPROACH TO THE INVESTIGATION OF AAD – A STUDY OF TWENTY CANDIDATE GENES IN EUROPEAN AAD COHORTS

4.1 BACKGROUND

The hypothesis-driven candidate gene approach has seen a number of successes in the investigation of AAD. Most notably, the finding of association of *PTPN22*[176, 177] and *CTLA4*[116, 171] polymorphisms with AAD which have been widely replicated. In order to attain adequate power in these studies, large case-control cohorts are needed. In AAD, previous candidate gene studies have been relatively small, including at most 300 affected individuals, and some findings have proved difficult to replicate. The Euradrenal consortium has recently provided a platform for collaboration between researchers interested in AAD in Europe and has, for the first time, allowed almost 2000 AAD samples to be collected together for genetic analysis.

While this size of cohort would provide power for an initial genome wide study, the lack of a replication cohort has prohibited this approach to date. A large candidate gene study provides an alternative, hypothesis-driven approach to learning more about the underlying genetic aetiology of AAD and makes good use of this valuable and unique DNA resource.

Candidate genes for investigation in this study were selected from four broad biological pathways known to be important in the pathogenesis of autoimmune conditions and the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) was used to review recent literature to aid selection of plausible candidate genes within these pathways. The selected candidates included genes influencing $CD4^+$ lymphocyte fate (*GATA3, GATA binding protein 3; IL17A, interleukin 17A; IL17RA, interleukin 17 receptor A; IL21, interleukin 21; IL23A, interleukin 23 alpha subunit p19; RORA, RAR-related orphan receptor A; RORC, RAR-related orphan receptor C; STAT2, signal transducer and activator of transcription 2; STAT4, signal transducer and activator of transcription 4* and *TBX21, T-box 21*), T cell signalling (the *CD28-CTLA4-ICOS (inducible T-cell co-stimulator*) cluster), transcription factors which alter the immune response (*NFATC2, nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2, NFKB1, nuclear factor of kappa light polypeptide gene enhancer*

in B-cells 1 and *REL, v-rel reticuloendotheliosis viral oncogene homolog*) and genes important for innate immune mechanisms (*CYP2R1, vitamin D 25-hydroxylase; CYP24A1, CYP27B1, 25-hydroxyvitamin D-1 alpha hydroxylase; GC, vitamin D binding protein; IFIH1, interferon induced with helicase C domain 1* and *VDR*).

4.1.1 GENES INFLUENCING CD4$^+$ CELL FATE DETERMINATION

$CD4^+$ T lymphocytes play a critical role in the adaptive immune response and are therefore thought to play a central part in loss of tolerance and the development and perpetuation of autoimmune responses. Our understanding of these cells is constantly evolving. Historically, they were thought of as a uniform population, defined purely on the basis of having both CD3 and CD4 surface molecules, and were referred to as "helper T cells". Now $CD3^+CD4^+$ T lymphocytes are being increasingly subdivided and re-classified as our understanding of T cell biology, and its complexity, increases. Two main subdivisions now exist: T_H and T_{Reg} cells. These can then be further subdivided. The T_H cells are commonly subdivided into T_H1, T_H2, T_H9[293], T_H17, T_H22[294] and follicular helper T cells (T_{FH})[295, 296]. T_{Reg} cells all have a CD25 cell surface molecule, in addition to CD3 and CD4[13, 297-299]. They are currently subdivided into T_R1, natural T_{Reg} and inducible T_{Reg} cells[300], although other subdivisions have also been proposed. These subdivisions are made according to their function, regulating transcription factors, surface marker phenotype (in addition to the T cell receptor (TCR), CD3 and CD4 molecules) and their cytokine secretion profile when stimulated. It is likely that the number of subdivisions will increase in the future.

As our understanding of T lymphocyte biology increases, some long-held paradigms are being challenged. One such paradigm is the "master regulator" theory of $CD4^+$ differentiation which defined a single transcription factor regulating differentiation and cell fate of each subtype. TBX21 was thought to be the master regulator for T_H1 cells, GATA3 for T_H2 cells, retinoic acid-related orphan receptor γt (ROR-γt) for T_H17 cells and FOXP3 for T_{Reg} cells. However, it is increasingly recognised that T cell lineage fate is far more complex and

dynamic than previously thought: these factors can be co-expressed and can interact with other factors, all influenced by the cytokine environment, to alter T_H cell balance *in vivo*[301, 302]. In addition, the model that T_H1 dominated responses result in autoimmune disease and that T_H2 dominated responses result in allergic disease[303] is being challenged. We now recognise that, in autoimmune disease, a number of interacting subsets contribute towards pathophysiology, in particular T_H1, T_H2, T_H17 and T_{Reg} cells. This makes genes involved in these lineages plausible candidates for AAD. In this study, ten genes contributing to $CD4^+$ cell fate were selected for investigation: three involved in T_H1 cell lineage specification (*TBX21, STAT4* and *STAT2*), one in T_H2 specification (*GATA3,* which also has a role in T_{Reg} function) and six in T_H17 specification (*IL17A, IL17RA, IL21, IL23A, RORA* and *RORC*).

4.1.1.1 THE T_H1 CELL LINEAGE

The development of T_H1 cells is regulated by IL12[304, 305]. Their surface marker molecules include the IL12 receptor, the interferon gamma (IFN-γ) receptor[306] and the CXCR3 chemokine receptor[307] and, when stimulated, they secrete IFN-γ, IL2[306] and lymphotoxin alpha[300, 308]. T_H1 cells present antigens in the context of MHC class II and activate macrophages. They have a crucial role in immunity against intracellular pathogens. The key transcription factors for this subset are TBX21[309], STAT4[305] and STAT1[310]. 15 SNPs in three T_H1 genes were selected for investigation: two independent SNPs in *TBX21*, 11 SNPs in *STAT4*, including a cluster of seven in moderate LD, and two SNPs in complete LD in *STAT2*.

The *TBX21* gene on chromosome 17 contains six exons and is a member of a family of genes which contain a T-box binding domain. It encodes the TBX21 transcription factor which is an important regulator of T_H1 lymphocyte development, with a critical role in inducing T_H1 cell specification from T_H cells[309]. TBX21 is also thought to modulate T_H17 cell responses[311]. Polymorphisms in TBX21 have been associated with autoimmune disease in various populations, with most studies coming from Asia. Associations include type 1 diabetes in a cohort from Japan[312], SLE in a cohort of Chinese

patients[313], Graves' disease in a smaller cohort from Japan[314] and rheumatoid arthritis in Korean patients[315].

STATs are cytosolic proteins which all have an Src homology 2 (SH2) domain and a carboxyl terminal tyrosine phosphorylation site. STAT proteins all contain a DNA binding site[316] or multiple DNA binding sites[317]. STATs are activated in response to cytokines and growth factors and, when activated, they dimerise. These dimers translocate to the nucleus and exert effects on signal transduction and activation of transcription by binding to the promoters of specific target genes[318]. The *STAT4* gene is found on chromosome 2 and is composed of 24 exons. It encodes STAT4, which is expressed in myeloid cells, T lymphocytes, the thymus and testes. It is phosphorylated specifically in response to IL12, type 1 interferons (e.g. interferon alpha and beta) and IL23[319] and is essential for the T_H1 response and cellular immunity. In support of this, *STAT4* knock-out mice are viable and fertile but their IL12-mediated functions are impaired, including the induction of the primary T_H1 cytokine IFN-γ and T_H1 differentiation[320]. As the balance between T_H1 and T_H2 responses is thought to be critical in immune tolerance and susceptibility to autoimmunity, *STAT4* is an excellent candidate gene for all autoimmune conditions. Indeed, polymorphisms in *STAT4* have been associated with both rheumatoid arthritis and SLE in a large American study[321].

The *STAT2* gene is found on chromosome 12 and consists of 24 exons. It encodes the STAT2 protein which is widely expressed and, like STAT4, is phosphorylated in response to type 1 interferons[320, 322, 323]. However, STAT2 lacks the ability to bind DNA directly and therefore forms a transcriptional complex with STAT1 and p48 to exert its actions[323, 324]. *STAT2* polymorphisms have been associated with psoriasis in a European genome-wide study[325].

4.1.1.2 THE T_H2 CELL LINEAGE

T_H2 cells activate B cells and therefore promote humoral immunity. They are important in the immune response to extracellular pathogens, but can also mediate allergic disease if not tightly regulated[326]. Their development and

maintenance are regulated by IL4[327], IL25[328, 329] and IL33[330]. T_H2 cells express cell surface IL4, IL17 receptor B and IL33 receptors[331], the chemokine receptor CCR4[307], and the chemoattractant receptor-homologous molecule expressed on T_H2 cells (CRT_H2) receptor[307]. When stimulated, they secrete IL4[327], IL5[327], IL10[332] and IL13[308], and their regulating transcription factors include GATA3[333], STAT6[334], interferon regulatory factor 4 (IRF4)[335, 336], basic helix-loop-helix family, member e41; BHLHE41 (DEC2)[337, 338] and v-maf musculoaponeurotic fibrosarcoma oncogene homolog[339]. In this study, four SNPs in the *GATA3* gene, including a pair in moderate LD (*rs3802604* and *rs570613*, r^2 0.67), were chosen to investigate the influence of *GATA3* polymorphisms in AAD susceptibility.

GATA3 is found on chromosome 10 and encodes the GATA3 transcription factor. GATA3 is a member of the GATA-binding protein family which has six members in mammals (GATA1 to GATA6), each containing two highly conserved zinc finger domains: one at the C-terminal which allows DNA binding; one at the N-terminal which stabilises the protein-DNA interaction[340]. The GATA family members all interact with 5'-(A/T)GATA(A/G)-3' sequence on DNA, altering transcription and regulating gene expression[341-343]. It is from this interaction that they derive their name. The GATA3 protein is widely expressed both in embryonic development and in adult life[344], is an important regulator of T cell development and plays a crucial role in determining T cell fate, in particular the differentiation of naïve $CD4^+$ T cells into T_H2 cells[345-347]. Until very recently, GATA3 was thought to be the master regulator of T_H2 differentiation. However, it is now recognised that GATA3 interacts with a number of other factors, such as STAT4 and TBX21, to perform this, and other, functions[302]. GATA3 is also an important regulator of T_{Reg} cells[348] which have potent immunosuppressive properties and therefore promote the dampening down of immune responses[7, 10, 349]. T_{Reg} cells are thus critical for maintaining immune homeostasis and tolerance to self[350]. One group of T_{Reg} cells are generated from naïve T cells centrally, in the thymus, when they are termed natural T_{Reg} cells[3, 351], or peripherally, when they are called inducible T_{Reg} cells[352]. They are recognised by their $CD25^+$ cell surface phenotype[13, 297-299], and may also bear CTLA4[351] and glucocorticoid-induced TNF receptor-related protein (GITR)[353]

cell surface markers. FOXP3 was thought to be the master regulator of this subset of cells[8, 9], but other transcription factors are also important in this lineage, including STAT5[354], FOXO1[355], FOXO3[355], SMAD2[356] and SMAD3[356]. T_{Reg} lymphocytes secrete IL10 and TGF-β[357].

Loss of function mutations in the *GATA3* gene result in the hypoparathyroidism, sensorineural deafness and renal abnormalities (HDR) syndrome[358] while *GATA3* polymorphisms have been associated with allergic rhinitis[359], breast cancer[360] and Hodgkin's lymphoma[361]. To date, no *GATA3* polymorphisms have yet been associated with autoimmune disease, although altered GATA3 expression has been linked to both systemic sclerosis[362] and flares of SLE[363].

4.1.1.3 THE T_H17 CELL LINEAGE

T_H17 cells were originally named because they produce IL17[364], although we now know that their cytokine secretion profile is not limited to IL17 alone: they also secrete IL17A, IL17F, IL21[365], IL22 and IL26[366]. The function of T_H17 cells is to provide additional immunological defence against microbes, particularly at epithelial surfaces such as in the gut, and to secrete cytokines which are chemoattractant to neutrophils. T_H17 cells are extremely proinflammatory[367]. Their surface marker profile includes the IL23 receptor, the IL1 receptor, the chemokine receptor CCR6[368] and CD161 (killer cell lectin-like receptor subfamily B, member 1 also known as KLRB1)[369]. They are produced in response to TGF-β[370], IL6[371] and IL21[372] and are maintained by IL23[373] and IL1[367]. Their regulating transcription factors include RORγt[374], RORα[375] and STAT3[376]. Three SNPs in *IL17A* (*rs3819024* and *rs16882180* in moderate LD (r^2 0.55) and one independent SNP), two independent SNPs in *IL17RA*, two independent SNPs in *IL21*, one SNP in *IL23A*, four independent SNPs in *RORA* and five largely independent SNPs in *RORC* (maximum r^2 0.40) were selected for genotyping.

The *IL17A* gene comprises three exons and is located on chromosome 6. It encodes the IL17A cytokine. Polymorphisms in the *IL17A* gene have previously been associated with both ulcerative colitis[377] and Crohn's disease[378]. The *IL17RA* gene, found on chromosome 22, encodes the IL17A receptor which is expressed mainly on immune cells. *IL17RA* polymorphisms have been associated with alopecia areata[379] and Crohn's disease[378]. The *IL21* gene on chromosome 4 encodes the interleukin 21 cytokine which plays a role in both innate and adaptive immunity by inducing the differentiation, proliferation and activation of many cell types including macrophages, natural killer cells, B lymphocytes and T lymphocytes[365]. IL21 levels have been correlated with disease severity in psoriasis[380], with alterations in lymphocyte subsets in patients with SLE[381] and with immune reconstitution autoimmune conditions following alemtuzumab (Campath) therapy[382]. Polymorphisms in the *IL21* gene have been associated with Graves' disease[383] and with SLE[384, 385].

The *IL23A* (interleukin 23, alpha subunit p19) gene comprises four exons and is found on chromosome 12. It is widely expressed and encodes a subunit of the IL23 cytokine, a dimeric cytokine made up of IL23A and IL12B, which can promote differentiation of naïve T_H cells to T_H17 cells[386]. The *RORC* gene is found on chromosome 1 and two transcripts are expressed: RORγ and RORγt, both of which are orphan nuclear receptors which can function as transcription factors. RORγ is expressed in a number of tissues, including the liver and thymus, however its function is poorly defined. The expression of RORγt is limited to the thymus where it plays an important role in inhibiting apoptosis of undifferentiated T cells and enhances their differentiation into T_H17 cells[375]. The final member of the T_H17 pathway that we selected for analysis was the *RORA* gene on chromosome 15, which encodes the RORα protein. This protein, like RORγt, is a nuclear receptor and transcription factor important for T_H17 lineage specification[375].

4.1.2 GENES INFLUENCING T CELL SIGNALLING

T cell activity must be carefully regulated in order to achieve a fine balance of effective protection against pathogens and maintenance of tolerance to self-

antigens. Tolerance in T lymphocytes is maintained by a number of mechanisms, both centrally, in the thymus, where autoreactive T lymphocytes are silenced or destroyed, and peripherally, outside of the thymus. Costimulatory molecules, present on antigen presenting cells, are important mediators of tolerance and have important roles in T lymphocyte regulation and function. Costimulatory molecules contribute one half of the classic two signal model for T cell activation. In the two signal model, following presentation of an antigen in the context of MHC, a costimulatory signal is required for T cell activation[387-389]. Many costimulatory molecules have now been described and these can provide positive, activating signals or negative signals which downgrade the T cell response.

The *CD28/CTLA4/ICOS* gene cluster is found on chromosome 2. The three genes which make up the cluster are relatively close together and significant LD exists in the region, particularly between *CTLA4* and *ICOS*. These three genes encode cell surface molecules that act as costimulatory molecules for T cell responses. While the CTLA4 molecule has a role in downregulating T cell responses[161], the CD28 and ICOS molecules have the opposite effect, providing positive signals which upregulate T cell activity[390]. Genomic variants at the *CTLA4* locus have been implicated in the aetiology of numerous auto-immune conditions, including autoimmune thyroid diseases[164-167], type 1 diabetes[162, 163] and rheumatoid arthritis[168, 169]. *CD28* polymorphisms have previously been associated with rheumatoid arthritis[391] and an increase in the circulating levels of soluble CD28 have also been reported both in this condition[392] and in other autoimmune diseases including SLE and Sjögren's syndrome[393, 394]. *ICOS* polymorphisms have been associated with rheumatoid arthritis[395] and SLE[396]. Due to the strong LD in the region, it can be difficult to ascertain whether *CD28* and *ICOS* are conferring risk of autoimmune disease independent of *CTLA4*. In this study, four SNPs in *CD28*, including a pair in moderate LD (*DIL107* and *rs1181389*, r^2 0.71), were chosen for genotyping. In addition, two independent SNPs situated between *CD28* and *CTLA4* were selected and seven SNPs in or close to the *CTLA4* gene, including a trio of SNPs in significant LD (*rs231775, rs231726* and *rs231727*). Finally, one independent SNP between *CTLA4* and *ICOS*, and two independent SNPs in or

upstream of the *ICOS* gene, were also genotyped, to capture as much variation in this region as possible.

4.1.3 GENES ENCODING TRANSCRIPTION FACTORS INVOLVED IN DYNAMIC IMMUNE RESPONSES

The NF-κB proteins comprise a family of transcription factors which influence a wide variety of biological processes including innate and adaptive immune responses, inflammation, cell growth, tissue differentiation and apoptosis[397-399]. Members of this family share a highly conserved DNA-binding domain called the Rel homology (RH) domain, nuclear localisation sequences and dimerisation sequences[400].

The family members can be subdivided into two groups depending on the structure of the c-terminal portion of the protein. The class 1, or NF-κB, proteins include NF-κB1 (sometimes called p50/p105) and NF-κB2 (p52/p100) and have ankyrin repeats in this section, while the class 2, or Rel, proteins, including RelA, RelB and c-Rel, have transcriptional activation domains here[399]. The class 1 proteins must form dimers with the class 2 proteins in order to act as transcription factors.

In most cells, NF-κB proteins are present in the cytoplasm in a complex with an inhibitor protein which is a member of the inhibitor κB (IκB) family[401]. In response to a stimulus, for example cytokines or bacterial products, NF-κB proteins are activated by degradation of the IκB protein by an IκB kinase. They then form dimers and translocate to the nucleus and alter gene expression[402]. The wide-ranging effects on immunity and autoimmunity make components of this pathway excellent candidates for AAD susceptibility. Therefore, for this study six SNPs in the *NF-κB1* (*nuclear factor of kappa light polypeptide gene enhancer in B-cells 1*) gene and two in *REL* were selected for investigation. The *NF-κB1* gene is comprised of 24 exons and is found on chromosome 4. *NF-κB1* encodes two isoforms: p105 is a non-DNA binding protein and p50 is capable of binding to DNA[403]. When activated, NF-κB1 forms biologically active

heterodimers or dimers with Rel proteins, which translocate to the nucleus to alter transcription. NF-κB1 is widely expressed at low levels in many tissues, and is expressed at higher levels in lymphocytes.

Polymorphisms in the *NF-κB1* gene have been implicated in a number of autoimmune and inflammatory conditions such as inflammatory bowel disease[404], type I diabetes[405] and Graves' disease[406]. The REL gene, on chromosome 2, consisting of 11 exons, encodes the proto-oncogene c-Rel which is widely expressed at low levels and highly expressed in B and T lymphocytes. c-REL is thought to play a critical role in T and B cell differentiation, proliferation and survival[407-411], in particular in T_{Reg} development[412]. It is also thought to have an overlapping role with T_H17 cell function[413, 414]. *REL* polymorphisms have been associated with a number of autoimmune conditions including rheumatoid arthritis[415], coeliac disease[416] and psoriasis[417, 418]. In this study, six SNPs in, and around, the *NF-κB1* gene were selected for investigation: there was moderate LD between four of these and two were independent. Two SNPs in complete LD in the *REL* gene were also selected for genotyping.

The NFAT proteins fulfil a similar role to NF-κB proteins, allowing rapid initiation of gene expression during immune responses. They are expressed on cells of the immune system such as lymphocytes, mast cells and macrophages, and play a critical role in the transcription of genes required for a vigorous immune response[419-421]. NFAT proteins have a REL-homology region and an NFAT-homology region and they can bind similar DNA sequences as the NF-κB dimers, but are not known to form dimers with NF-κB proteins[399, 422]. Like NF-κB proteins, NFAT proteins are present in the cytoplasm but rapidly translocate to the cell nucleus when surface T cell receptors, or other receptors coupled to calcium mobilisation, are stimulated. The translocation process is controlled by the protein phosphatase calcineurin, which interacts with the NFAT domain. In the nucleus, NFAT proteins form transcription complexes with other NFAT proteins and with molecules such as GATA4, and regulate gene expression in response to T cell activation during immune and inflammatory responses[422].

NFATC2 encodes the nuclear factor of activated T-cells, cytoplasmic calcineurin-dependent 2 protein. It is located on chromosome 20 and comprises 11 exons. Polymorphisms in *NFATC2* have been associated with narcolepsy[423] and altered *NFATC2* expression has been linked to complicated sarcoidosis in a recent, small study[424]. However, polymorphisms in this gene have not been associated with autoimmune disease to date, despite it being a very plausible candidate. 14 SNPs in *NFATC2* were selected for genotyping, including two pairs of SNPs in moderate LD (*rs959996* and *rs2024582*, r^2 0.72; *rs3787189* and *rs2273642*, r^2 0.61) and a cluster of eight SNPs in moderate LD.

4.1.4 GENES INFLUENCING INNATE IMMUNE MECHANISMS

The innate immune system is concerned with defending the host from infection by pathogens in a rapid, generic and nonspecific manner. Its primary functions include pathogen recognition, immune-cell recruitment through the production of chemokines and cytokines, activation of the complement cascade and activation of the adaptive immune system through antigen presentation. The innate immune system is complex: it involves numerous cell types, mechanisms and biomolecular pathways.

Vitamin D has long been recognised to have significant effects on the immune system and on autoimmunity. In the late 1800s, sunlight, one of the main sources of vitamin D, was a favoured treatment for a number of infectious and autoimmune diseases, including tuberculosis and lupus. This was long before the discovery of antibiotics and vitamin D itself, which was discovered in 1922 by Mellanby, in the course of his research into rickets[425]. Since the discovery of vitamin D, much research has focussed on the effects of its metabolites, vitamin D pathway components, and the vitamin D receptor, on immunity and in autoimmune disease. Alterations in the function of components of the vitamin D pathway, in addition to vitamin D levels, have been associated with autoimmune conditions, as have polymorphisms in the genes encoding these components, making them good candidates in AAD. Five genes from the vitamin D pathway were selected for this study. Three SNPs in moderate to high LD were selected in *CYP27B1,* 9 SNPs, including *rs4809959* and *rs2296241* which are in

moderate LD (r^2 0.78), were selected from *CYP24A1,* three independent SNPs in *CYP2R1,* three largely independent SNPs in *GC* (maximum r^2 0.42) and six SNPs in *VDR* (including a pair in moderate LD, *rs2189480* and *rs3819545*, r^2 0.57) were selected for genotyping.

CYP27B1 is a member of the cytochrome P450 superfamily of enzymes. These catalyse a large number of reactions involved in the synthesis of lipids, cholesterol and steroids, and in drug metabolism. The primary function of CYP27B1 is to hydroxylate 25-hydroxyvitamin D3 to 1,25-dihydroxyvitamin D3, an active metabolite which can bind to the vitamin D receptor and regulate calcium homeostasis and exert a suppressive effect on immunity and autoimmunity[206]. CYP27B1 is encoded by the *CYP27B1* gene, found on chromosome 12. The CYP27B1 protein is located on the inner mitochondrial membrane[426] and it is widely expressed. *CYP27B1* polymorphisms have been associated with type 1 diabetes[427], autoimmune thyroid disease[209] and AAD already, but in a relatively small study that has yet to be replicated[209].

The *CYP24A1* gene is found on chromosome 20 and encodes another cytochrome P450, the cytochrome P450 family 24, subfamily A, polypeptide 1 (CYP24A1, previously known as vitamin D 24-hydroxylase). This hydroxylase enzyme is found in mitochondria and its primary function is to hydroxylate 25-hydroxyvitamin D3 and 1,25-dihydroxyvitamin D3, resulting in their degradation[428, 429]. Polymorphisms in *CYP24A1* have been linked to reduced vitamin D levels in individuals with type 1 diabetes, although no significant association was found with type 1 diabetes and these polymorphisms *per se*[430].

The *CYP2R1* gene is found on chromosome 11 and encodes the cytochrome P450, family 2, subfamily R, polypeptide 1 (CYP2R1, previously known as vitamin D 25-hydroxylase). This enzyme is found in the cytoplasm and hydroxylates vitamin D to form the active 25-hydroxyvitamin D. Polymorphisms in *CYP2R1* have previously been associated with type 1 diabetes[430].

The *GC* gene on chromosome 4 encodes the vitamin D binding protein, which is a plasma protein belonging to the same family as albumin. The vitamin D binding protein transports vitamin D and its metabolites to target tissues. *GC* polymorphisms have been associated with the development of arthritis and uveitis in Korean patients with ankylosing spondylitis[431] and with reduced levels of vitamin D in individuals with type 1 diabetes, although this study failed to demonstrate a direct association between *GC* polymorphisms and type 1 diabetes[430].

The *VDR* gene on chromosome 12 encodes the vitamin D receptor transcription factor which belongs to the steroid receptor family. On activation, VDR forms heterodimers; these bind to vitamin D responsive elements in the promoters of target genes and activate their expression[432]. VDR can be expressed on any dividing cells and it is known to be expressed in at least 30 tissues, including on many immune cells[433-435]. Polymorphisms in *VDR* have previously been associated with autoimmune thyroid disease, type 1 diabetes[436], SLE and rheumatoid arthritis[437].

One additional gene involved in innate immunity was selected for inclusion in this study. The *IFIH1* gene on chromosome 2 encodes the IFIH1 protein which is a member of the DEAD box protein family. Members of this family are putative RNA helicases and are implicated in a number of cellular processes involving RNA metabolism and alteration of RNA secondary structure[438]. IFIH1 is thought to promote mRNA degradation by specific RNases and is involved in defending the host against viruses[439]. *IFIH1* polymorphisms have been associated with a number of common autoimmune conditions including Graves' disease[440] and type 1 diabetes[441], making it a plausible candidate in AAD, in particular in APS2. Three independent SNPs were selected for genotyping in the *IFIH1* gene.

4.2 AIM

My aim was to investigate the role of twenty candidate genes in the pathogenesis of AAD in six European case-control cohorts, using the Sequenom iPlex genotyping platform.

4.3 SUMMARY OF STUDY DESIGN

A list of 36 candidate genes of interest was constructed and members of the Euradrenal consortium were asked to vote for the 20 genes that they would like to be included in the study. 20 candidate genes were thus selected for analysis in this study. Sequenom iPlex assays were designed for multiple SNPs within these genes.

In round one, 101 SNPs in these 20 genes were chosen for genotyping in UK (309 AAD, 335 controls) and Norwegian (382 AAD, 380 controls) AAD case-control cohorts. Primer sequences are detailed in electronic appendix A. These SNPs were genotyped in four separate Sequenom plexes by CIGMR, Manchester. Data were analysed for each cohort as a whole, and then the cohorts were subdivided into individuals with isolated AAD (iAAD) and into those with AAD in addition to another autoimmune condition (APS2) and the analysis repeated. Genotype data are shown in full in electronic appendix B.

Following analysis of round 1 data, a second round of genotyping was undertaken to determine whether any of the findings in round 1 could be replicated in other European cohorts. In round two, 21 SNPs in 11 genes, associated in either the UK or Norwegian cohort in round 1, were genotyped by CIGMR, Manchester, in a single Sequenom plex, in case and control cohorts from Germany (341 AAD, 235 controls), Poland (275 AAD, 296 controls), Italy (280 AAD, 322 controls) and Sweden (368 AAD, 368 controls). Data from each cohort were analysed individually. Genotype data are shown in full in electronic appendix B. A subgroup analysis was not performed as clinical information to divide the cohorts into iAAD and APS2 were not available for all cohorts. A

meta-analysis was then performed, using the RevMan 5 software, across the six cohorts as a whole, initially including all individuals, and then excluding any individual without positive serum 21OH autoantibodies.

Following round 1 and round 2 data analyses, some candidates were taken forward to further investigation in a series of replication studies. A single SNP in *NF-κB1,* found to be associated with AAD in the UK cohort, was investigated by RFLP in a cohort of individuals from the UK with Graves's disease, to see if the round 1 findings could be replicated in another autoimmune disease. In addition, 15 SNPs in the *GATA3* gene were selected for further investigation by Sequenom genotyping in UK and Norwegian AAD cohorts (335 UK AAD, 302 UK controls, 352 Norwegian AAD, 1353 Norwegian controls) to narrow down the association detected in rounds 1 and 2. The primer sequences for these *GATA3* SNPs can be found in electronic appendix A. These SNPs were also genotyped in 283 UK Graves' disease patients, 1195 type 1 diabetes patients from Norway and in a cohort of 650 rheumatoid arthritis patients from New Zealand and 452 matched controls, to see if the findings could be replicated in other autoimmune cohorts. For the *GATA3* replication study, PCR was carried out in house for the AAD, Graves' disease and rheumatoid arthritis cohorts and PCR products given to Newgene, Newcastle for further processing. For the type 1 diabetes cohort, PCR and Sequenom genotyping were undertaken by CIGMR, Manchester. *GATA3* genotype data can be found in electronic appendix B.

4.4 RESULTS

4.4.1 STUDY POWER

Power calculations were performed using QUANTO[442]. The power of the study as a whole, and per cohort, was calculated before genotyping was undertaken. Including all 1955 AAD cases and 1936 controls, and assuming a minor allele frequency (MAF) of 0.3 and a type I error rate, or α, of 0.00078 (allowing for testing 64 independent markers in round 1 of genotyping: 0.05/64) the study has 100% power to detect a locus with an odds ratio of 1.4 and 65% power to detect a locus with an odds ratio of 1.2. Assuming a lower MAF reduces the power significantly: if a MAF of 0.2 is assumed, the power falls to 93% to detect a locus with an odds ratio of 1.4 and 48% to detect a locus with an odds ratio of 1.2.

Including only samples known to be 21OH autoantibody positive (1204 cases), assuming a MAF of 0.3, the study has 47% power to detect a locus with an odds ratio of 1.2 and 100% power to detect a locus with an odds ratio of 1.4. Again, if a MAF of 0.2 is assumed, the power is reduced (32% to detect a locus with an odds ratio of 1.2 and 80% to detect a locus with an odds ratio of 1.4).

The study power per individual European cohort was calculated under the same assumptions. Assuming a MAF of 0.3, each individual cohort would have more than 60% power to detect a locus with an odds ratio of 1.6 and more than 90% power to detect a locus with an odds ratio of 1.8. If a MAF of 0.2 or 0.1 is assumed, the power drops significantly (Figure 27).

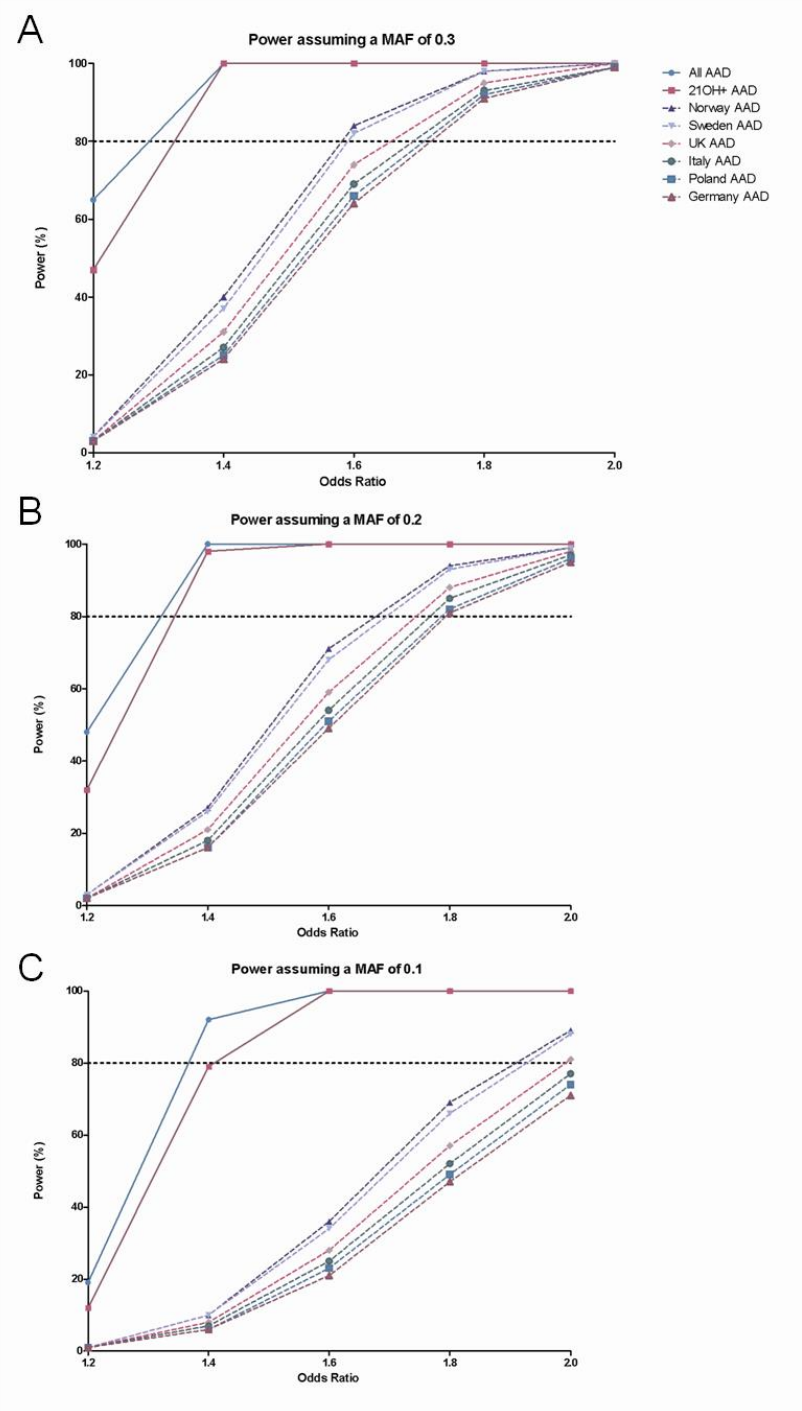


Figure 27: Power estimates for cohorts in the 20 candidate gene study.

Line graphs demonstrating the power of the candidate gene cohorts to detect loci with differing odds ratios. Each assumes a type I error rate (α) of 0.00078. Panel A shows the power assuming a minor allele frequency (MAF) of 0.3, panel B 0.2 and panel C 0.1. While the cohort as a whole, and the 21OH autoantibody positive cohort, both shown with solid lines, are well powered to detect a locus with an odds ratio of 1.4 or above, the individual cohorts, shown with dashed lines, have significantly less power.

4.4.2 ROUND 1 RESULTS

4.4.2.1 ROUND 1 DATA QUALITY CONTROL

SNP genotyping call rates for the UK and Norwegian AAD and control cohorts are shown in Table 11. In total, 32 of the 101 SNPs (31.7%) had a call rate of less than 95% in either the UK case or control cohorts and 27 of the 101 SNPs (26.7%) had a call rate of less than 95% in either the Norwegian case or control cohorts. These SNPs were therefore excluded from further analysis. The "drop-out" rate appears to be high. However, of the excluded SNPs, the mean genotyping call rate was 90.1% (minimum 27%, maximum 94%, median 93%), therefore most of these were close to meeting the study quality control inclusion criteria. Had a less stringent genotyping call rate threshold been arbitrarily applied, the drop-out rate would have been reduced considerably. However, the genotyping call rate threshold was set at 95% and the relatively high drop-out rate tolerated in order to ensure that only high quality genotyping data was taken forward for analysis. In the UK control cohort, no SNP deviated significantly from HWE (P<0.01) however, one of the remaining SNPs in the Norwegian control cohort, *rs10735810* was out of HWE (P 0.0087) and was consequently also excluded from the final analysis. Therefore, in total, 69 and 73 SNPs were included in the final analysis in the UK and Norwegian cohorts respectively.

Following quality control checks, the allele frequencies in control cohorts were compared by χ^2 testing to determine whether there was significant heterogeneity between the UK and Norwegian cohorts. Significant heterogeneity was noted (P <0.05) at 1 in 5 SNPs (20%) tested in both cohorts, therefore the UK and Norwegian data were analysed separately.

Cohort	SNP genotyping call rate (%)				Number of SNPs excluded on the basis of call rate <95%
	Mean	Minimum	Median	Maximum	
UK AAD cases	98	89	99	100	32/101
UK controls	96	88	97	99	
Norwegian AAD cases	95	27	97	99	27/101
Norwegian controls	96	71	97	99	

Table 11: SNP genotyping call rates for the UK and Norwegian cohorts in round 1 of the 20 candidate gene study.

4.4.2.2 ROUND 1 UK COHORT RESULTS

In total, 13 SNPs in nine genes were associated with AAD in the UK cohort and all significant results ($P_{genotype}$ or P_{allele} <0.05) are summarised in Table 12. Full genotype and allele data are listed in electronic appendix B.

Maximal association was seen with the *NF-κB1* gene. Six SNPs were genotyped in, and close to, this gene (Figure 28). Alleles at markers B, C and E, all in moderate LD (*rs10026278*, *rs230532* and *rs4698861,* r^2 0.39 – 0.68)*,* were associated with AAD in the UK cases compared to controls. The maximum evidence for association was found at the intergenic marker E where the GG genotype was found in 22 of 308 cases (7.1%) compared with 45 of 321 controls (14.0%) (P 0.00084). A similar decrease in the frequency of the minor G allele was observed at this SNP in AAD cases compared with controls (27.4% and 37.4% respectively) [odds ratio (OR) 0.63, 95% confidence interval (CI) 0.50 – 0.80; P 0.00017)]. When the AAD case cohort was subdivided into iAAD (n=135) and into those with APS2 (n=174), the association was with APS2 ($P_{genotype}$ 0.00004, P_{allele} 0.0000076, OR 0.51 [95% CI 0.38 – 0.69]). Haplotype analysis in UNPHASED[220] revealed that this marker accounts for all of the association with disease and that the other markers are simply associated because they are in LD with this SNP.

Gene	SNPs typed	SNPs excluded	rs ID	Minor allele	MAF cases/controls	Pgenotype/Pallele	OR [95% CI]	Contribution	LD between associated markers*
			rs10026278	T	0.27 / 0.35	0.012 / 0.0034	0.69 [0.54-0.88]		
			rs230532	T	0.30 / 0.40	0.0016 / 0.00041	0.65 [0.52-0.82]		
NFKB1	6	0	*rs4698861*	G	0.27 / 0.37	0.00084 / 0.00017	0.63 [0.50-0.80]	APS2	Moderate
			rs4646536	G	0.26 / 0.33	0.012 / 0.0091	0.72 [0.56-0.92]		
CYP27B1	3	1	*rs703842*	G	0.27 / 0.33	0.027 / 0.014	0.74 [0.58-0.94]	APS2	Significant
IL23A	1	0	*rs11170816*	A	0.05 / 0.09	N/A / 0.0047	0.53 [0.34-0.84]	iAAD only	
REL	2	1	*rs13017599*	A	0.41 / 0.33	0.0099 / 0.0028	1.40 [1.12-1.76]	iAAD>APS2	
			rs569421	C	0.26 / 0.19	0.0092 / 0.003	1.50 [1.15-1.96]		
GATA3	4	0	*rs444929*	C	0.21 / 0.28	0.012 / 0.0053	0.69 [0.54-0.90]	iAAD	Low
IL21	2	1	*rs907715*	T	0.32 / 0.39	0.015 / 0.012	0.74 [0.59-0.93]	APS2	
STAT2	2	1	*rs2066808*	G	0.05 / 0.09	N/A / 0.012	0.57 [0.36-0.90]	iAAD	
CYP24A1	9	3	*rs4809959*	G	0.48 / 0.53	0.012 / 0.046	0.80 [0.64-0.99]	APS2	
IL17A	3	0	*rs16882180*	T	0.32 / 0.38	0.13 / 0.043	0.79 [0.63-1.00]	APS2	

Table 12: Summary of significant associations in the UK AAD cohort in round 1 of genotyping.

Only significant results (P < 0.05) are shown. Associated SNPs, the minor allele and its frequency (MAF) in cases and controls, and the P value and odds ratio (OR) generated by comparing both genotypes and allele frequencies between cases and controls are shown. The "contribution" column shows whether the association was with isolated AAD (iAAD) or AAD with other autoimmune conditions (APS2) when the cohort was subdivided. If the minor genotype was not present, a genotyping P value was not generated (marked N/A). *Low LD = r2 <0.40, moderate LD = r2 0.40-0.79, significant LD = r2 >0.79.

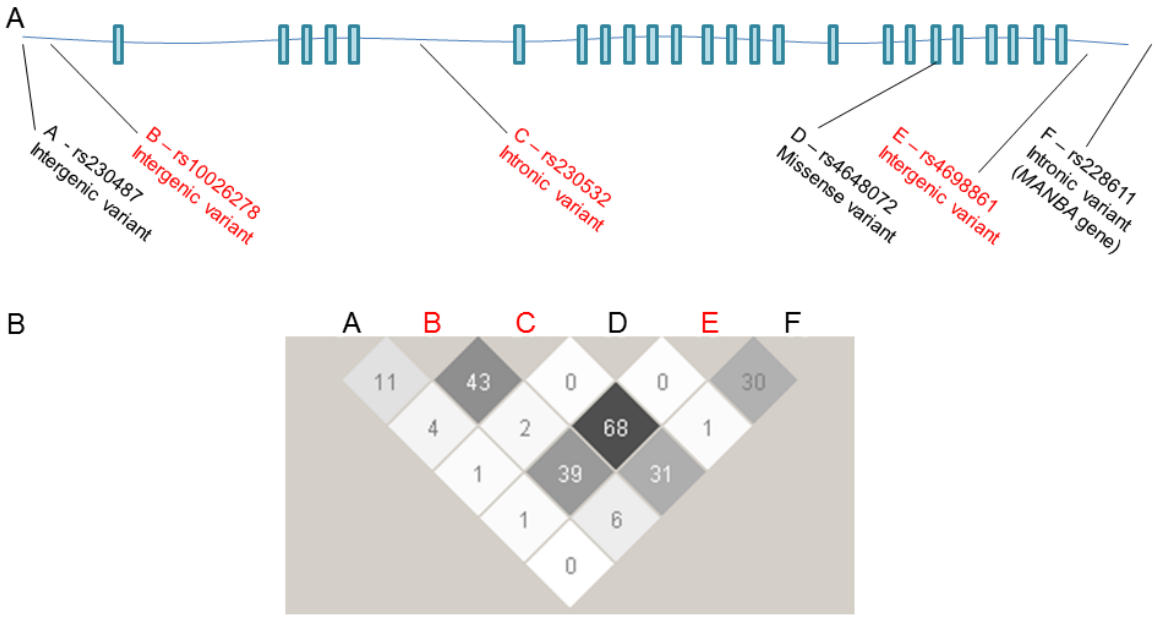


Figure 28: Schematic representation of the NF-κB1 locus (panel A) and pairwise linkage disequilibrium (r2) measures between 6 SNPs genotyped in and around the NF-κB1 gene (panel B).

In panel A, exons are depicted by the blue boxes and intronic sequence is illustrated by the line. The six genotyped variants are shown with their approximate locations. SNPs associated with AAD in the UK cohort are shown in red (diagram not to scale). In panel B, LD measures between six SNPs in and around the *NF-κB1* gene are shown in a Haploview LD plot generated using genotype data derived from Caucasian individuals from HapMap. White boxes represent the lowest r^2 values and black boxes represent the highest r^2 values. Two of the six genotyped SNPs, C and E are in moderate LD (r^2 0.68) and there is also some LD between these two SNPs and B and F. Markers A and D are independent.

4.4.2.3 ROUND 1 NORWEGIAN COHORT RESULTS

In total, 11 SNPs in seven genes were associated with AAD in the Norwegian cohort and all significant results ($P_{genotype}$ or P_{allele} <0.05) are summarised in Table 13. Genotype and allele data can be found in electronic appendix B.

Maximal association was seen at intronic marker G (*rs4274624*) in the *STAT4* gene (Figure 29). In total, nine SNPs were genotyped in this gene but, in this cohort, two SNPs, F and I (*rs10931481* and *rs4853543)* were excluded due to poor call rates in the case cohort. Of the remaining seven, only alleles at marker G were associated with AAD. At this SNP, the CC genotype was observed in 25 of 375 cases (6.7%) compared to 9 of 375 controls (2.4%) (P 0.0013). The C allele appeared to be conferring disease risk as it was observed in 26.9% of cases and in only 19.5% of controls (P 0.00045, OR 1.52, [95% CI 1.20 – 1.9]). The cohort was then subdivided into individuals with iAAD (n=162) and into those with APS2 (n=220). At this SNP, the association was with both iAAD and APS2 (iAAD $P_{genotype}$ 0.0065, P_{allele} 0.0027, OR 1.58 [95% CI 1.16 – 2.14]; APS2 $P_{genotype}$ 0.0087, P_{allele} 0.0038, OR 1.50 [95% CI 1.13 – 1.98]). This associated SNP is in moderate LD with two other neighbouring intronic SNPs (marker F, *rs10931481*, r^2 0.59 and marker H, *rs16833260* r^2 0.56). Marker F was excluded from the analysis, however no association was observed with alleles at marker H and AAD.

Gene	SNPs typed	SNPs excluded	rs ID	Minor allele	MAF cases/controls	$P_{genotype}/P_{allele}$	OR [95% CI]	Contribution	LD between associated markers*
STAT4	11	2	*rs4274624*	C	0.27 / 0.19	**0.0013 / 0.00045**	1.52 [1.20-1.94]	iAAD>APS2	
RORA	4	1	*rs1234805*	T	0.37 / 0.30	**0.0068 / 0.0018**	1.41 [1.13-1.75]	APS2>iAAD	
CD28/ CTLA4/ ICOS	16	4	*rs3181096*	T	0.32 / 0.38	**0.038 / 0.012**	0.76 [0.61-0.94]	APS2	Low
			rs231775	G	0.49 / 0.42	**0.016 / 0.0042**	1.35 [1.10-1.66]		Significant
			rs231726	T	0.44 / 0.37	**0.0093 / 0.0026**	1.37 [1.11-1.69]		
			rs231727	A	0.44 / 0.37	**0.025 / 0.0070**	1.33 [1.08-1.63]		
			rs2882973	C	0.40 / 0.35	0.087 / **0.035**	1.25 [1.01-1.54]		Low
GATA3	4	0	*rs3802604*	G	0.32 / 0.37	**0.04 / 0.038**	0.79 [0.64-0.98]	iAAD	
NFKB1	6	2	*rs228611*	A	0.43 / 0.49	0.063 / **0.024**	0.79 [0.65-0.97]	iAAD	
CYP24A1	9	1	*rs2209314*	C	0.25 / 0.25	**0.031 / 0.91**	0.99 [0.78-1.25]	N/A	
IL17A	3	1	*rs4711998*	A	0.22 / 0.27	0.060 / **0.039**	0.77 [0.61-0.98]	APS2	

Table 13: Summary of significant associations in the Norwegian AAD cohort in round 1 of genotyping.

Only significant results ($P < 0.05$) are shown. Associated SNPs, the minor allele and its frequency (MAF) in cases and controls, and the P value and odds ratio (OR) generated by comparing both genotypes and allele frequencies between cases and controls are shown. The "contribution" column shows whether the association was with isolated AAD (iAAD) or AAD with other autoimmune conditions (APS2) when the cohort was subdivided. *Low LD = r^2 <0.40, moderate LD = r^2 0.40-0.79, significant LD = r^2 >0.79.

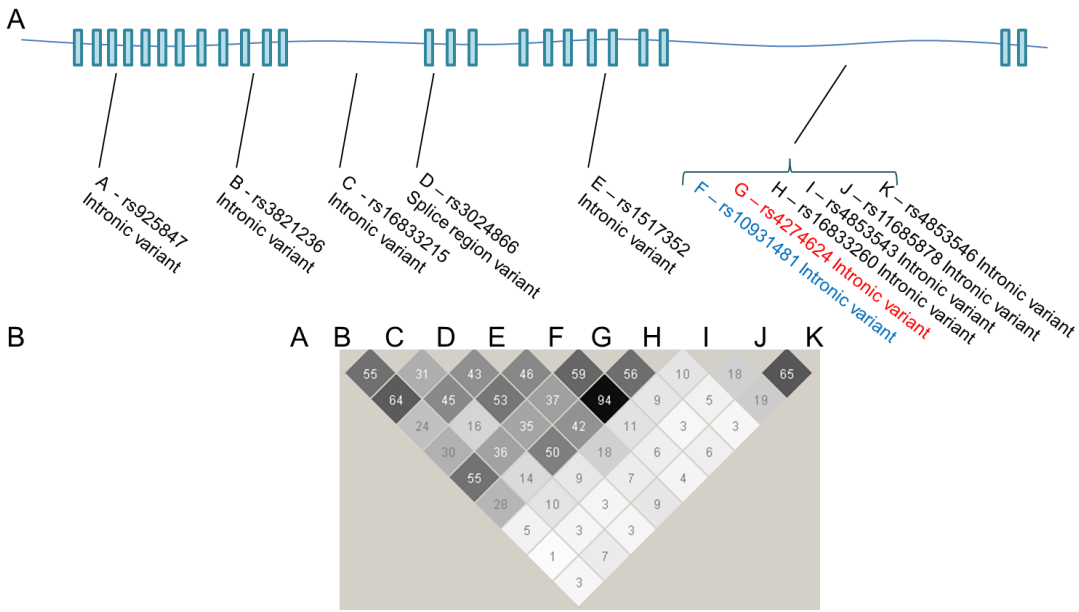


Figure 29: Schematic representation of the STAT4 locus (panel A) and pairwise linkage disequilibrium (r2) measures between 11 SNPs genotyped in and around this locus (panel B).

In panel A, exons are depicted by the blue boxes and intronic sequence is illustrated by the line. The 11 genotyped variants are shown with their approximate locations. The associated SNP, marker G, in the Norwegian AAD cohort is shown in red. This SNP, and marker F (in blue) were also associated in the six cohort meta-analysis. In panel B, LD measures between the SNPs are shown in a Haploview LD plot generated using genotype data derived from Caucasian individuals from HapMap. White boxes represent the lowest r^2 values and black boxes represent the highest r^2 values. There is significant LD between markers F and H (*rs10931481* and *rs16833260*; r^2 0.94). There is also moderate LD between a number of other SNPs. No LD information for marker A was available.

4.4.3 ROUND 2 RESULTS

4.4.3.1 ROUND 2 DATA QUALITY CONTROL

SNP genotyping call rates in round 2 for the German, Swedish, Italian and Polish AAD and control cohorts are shown in Table 14. Genotyping call rates at each SNP were greater than 95% in all cohorts and therefore no SNP was excluded from further analysis on this basis. In the German control cohort, SNP *rs3819024* was out of HWE (P 0.005); in the Polish control cohort, SNP *rs13017599* was out of HWE (P 0.0091). These two SNPs were therefore excluded. Therefore, in total, 21 SNPs were included in the final analysis in the Italian and Swedish cohorts, while 20 SNPs were analysed in the German and Polish cohort groups.

Full genotype data for all cohorts studied in round 2 can be found in electronic appendix B.

4.4.3.2 ROUND 2 GERMAN COHORT RESULTS

Of the 20 SNPs analysed, alleles at one independent SNP in the *IL21* gene (*rs907715*) were associated with AAD, with the T allele having a protective effect ($P_{genotype}$ 0.0078 and P_{allele} 0.018, OR for T allele 0.73 [95% CI 0.56 – 0.95]) (Table 15).

4.4.3.3 ROUND 2 SWEDISH COHORT RESULTS

Of the 21 SNPs analysed, the C allele at *rs4274624* in the *STAT4* gene appeared to be conferring a modest risk of AAD. The association was seen with alleles only and did not reach statistical significance with genotypes ($P_{genotype}$ 0.056, P_{allele} 0.017, OR for C allele 1.33 [95% CI 1.05 – 1.68]). This SNP is in moderate LD (r^2 0.59) with a neighbouring marker, *rs10931481,* which was also genotyped in this round. However, no association with AAD was noted at this SNP in this cohort (Table 15).

4.4.3.4 ROUND 2 ITALIAN COHORT RESULTS

The Italian cohort was divided into samples from two geographically separate regions of Italy; Padua (180 AAD, 134 controls) in northern Italy and Perugia (100 AAD, 188 controls) in central Italy. Following genotyping, the genotype and allele frequencies were compared between the two control cohorts. This was done using a 3x2 and 2x2 χ^2 test respectively, to determine whether the two control cohorts were significantly different. The control groups differed significantly at only 1 of the 21 SNPs genotyped, *rs907715* in *IL21* ($P_{genotype}$ 0.04, P_{allele} 0.02). The cohorts were initially analysed separately and then combined for a final analysis, given that there was little heterogeneity between the control cohorts.

Of the 21 SNPs genotyped in the Padua cohort, seven were associated with AAD. In the smaller Perugia cohort, only one SNP was associated with AAD (Italian subgroup genotype data can be found in appendix B).

In the Italian cohort as a whole, six SNPs were associated with AAD and these significant results ($P_{genotype}$ or P_{allele} <0.05) are summarised in Table 15. Maximal association was seen with alleles at *rs11171806* in *IL23A.* Here, the T allele was conferring disease risk ($P_{genotype}$ 0.012, P_{allele} 0.0028, OR for T allele 2.37 [95% CI 1.32 – 4.23]).

4.4.3.5 ROUND 2 POLISH COHORT RESULTS

The Polish cohort was divided into samples from Warsaw in the east of Poland (159 AAD, 50 controls) and Poznan in the west (116 AAD, 246 controls). The genotype and allele frequencies were compared between the two control cohorts and the control groups differed significantly at two of the 21 SNPs genotyped (*rs3802604* in *GATA3:* $P_{genotype}$ 0.02, P_{allele} 0.016; *rs1234805* in *RORA:* $P_{genotype}$ 0.036, P_{allele} 0.047). The cohorts were initially analysed

separately and then combined for a final analysis, given that there was little heterogeneity between the control cohorts.

Two SNPs were associated with AAD in the Warsaw cohort while two different SNPs were associated with AAD in the Poznan cohort (Polish subgroup genotype data can be found in appendix B).

In the Polish cohort as a whole, of the 20 SNPs which met the quality control inclusion criteria, three were associated with AAD. These significant results ($P_{genotype}$ or P_{allele} <0.05) are summarised in Table 15. Maximal association was seen with alleles at the independent *rs2221903* SNP in the *IL21* gene. Here, the C allele appeared to be conferring modest disease protection ($P_{genotype}$ 0.04, P_{allele} 0.02, OR for C allele 0.75 [95% CI 0.58 – 0.96]).

Following analysis of genotype data for each individual cohort, allele frequencies in control cohorts were compared by χ^2 testing between the six different European cohorts, to determine whether there was significant heterogeneity between the populations studied (Table 16). The highest levels of heterogeneity between control cohorts were seen between the Italian and UK cohorts. Of the 15 SNPs that passed the quality control criteria in these two cohorts, allele frequencies at 11 (73.3%) were significantly different (P <0.05), suggesting significant genetic differences between these two populations. The least heterogeneity was seen between the German and Swedish cohorts. Of the 20 SNPs that passed the quality control criteria in these two cohorts, allele frequencies did not differ significantly at any marker.

Cohort	SNP genotyping call rate (%)				Number of SNPs excluded on the basis of call rate <95%
	Mean	Minimum	Median	Maximum	
German AAD cases	97	96	98	98	0/21
German controls	98	97	98	99	
Swedish AAD cases	100	100	99	100	0/21
Swedish controls	99	98	100	100	
Italian AAD cases	99	97	99	99	0/21
Italian controls	100	98	100	100	
Polish AAD cases	98	96	98	98	0/21
Polish controls	100	98	100	100	

Table 14: SNP genotyping call rates for the German, Italian, Polish and Swedish cohorts in round 2 of the 20 candidate gene study.

Gene	SNPs typed	SNPs excluded	rs ID	Minor allele	MAF cases/controls	$P_{genotype}/P_{allele}$	OR [95% CI]	LD between associated markers*
Round 2 data – Germany								
IL21	2	0	*rs907715*	T	0.24 / 0.31	**0.0078 / 0.018**	0.73 [0.56-0.95]	
Round 2 data – Sweden								
STAT4	3	0	*rs4274624*	C	0.29 / 0.24	0.056 / **0.017**	1.33 [1.10-1.68]	
Round 2 data – Italy								
STAT4	3	0	*rs10931481*	G	0.36 / 0.29	**0.016 / 0.0056**	1.41 [1.11-1.80]	moderate
			rs4274624	C	0.28 / 0.22	0.059 / **0.02**	1.37 [1.10-1.78]	
IL23A	1	0	*rs11171806*	A	0.06 / 0.03	**0.012 / 0.0028**	2.37 [1.32-4.23]	
NFKB1	3	0	*rs10026278*	T	0.27 / 0.23	**0.049** / 0.078	1.27 [0.97-1.65]	
STAT2	2	0	*rs2066808*	G	0.07 / 0.04	**0.037 / 0.014**	1.93 [1.13-3.28]	significant
			rs2066807	G	0.06 / 0.03	**0.023 / 0.0063**	2.18 [1.23-3.85]	
Round 2 data – Poland								
IL21	2	0	*rs2221903*	C	0.31 / 0.38	**0.04 / 0.02**	0.75 [0.58-0.96]	
CYP24A1	1	0	*rs4809959*	G	0.56 / 0.49	0.11 / **0.03**	1.29 [1.02-1.64]	
GATA3	3	0	*rs444929*	C	0.18 / 0.21	**0.03** / 0.12	0.79 [0.59-1.06]	

Table 15: Summary of significant associations in the German, Swedish, Italian and Polish cohorts in round 2 of genotyping.

Only significant results ($P < 0.05$) are shown. *Low LD = r^2 <0.40, moderate LD = r^2 0.40–0.79, significant LD = r^2 >0.79.

	UK	Norway	Germany	Italy	Poland	Sweden
UK		16.7%	21.4%	73.3%	53.3%	46.7%
Norway			7.7%	64.3%	30.8%	14.3%
Germany				50.0%	5.3%	0.0%
Italy					60.0%	57.1%
Poland						20.0%
Sweden						

Table 16: Genetic heterogeneity between the six different European control cohorts included in the 20 candidate gene study.

Allele frequencies were compared between all control populations using a χ^2 test. The percentage of comparisons where a statistically significant result was seen ($P < 0.05$) is shown in the table. A low percentage indicates that there were few markers at which allele frequencies differed between two populations of interest suggesting that there is little genetic heterogeneity between those two populations. A high percentage indicates that allele frequencies between two populations differed significantly at multiple markers, suggesting significant genetic heterogeneity between those two populations. Between the German and Swedish control cohorts, allele frequencies did not differ significantly at a single marker, suggesting that these populations are relatively genetically similar. By contrast, the UK and Italian control cohorts differed significantly at 73% of markers tested, suggesting significant genetic heterogeneity between these two cohorts.

4.4.4 META-ANALYSIS

The RevMan 5 software package was used to estimate an odds ratio, an I^2 statistic of heterogeneity and a two-sided P value, applying a conservative random effects model, at each locus tested in round 2 of the study (21 SNPs). The random effects model was selected to account for the heterogeneity between the cohorts. Data from all six European cohorts were included (1955 cases and 1936 controls). Any SNP, in any cohort, not meeting the quality control criteria (call rate of less than 95% in either the case or control population, or out of HWE (P<0.01) in the control population) was excluded.

In total, four SNPs in three genes were associated with AAD (Table 17). Maximal association was seen with two SNPs in moderate LD (r^2 0.59) in the *STAT4* gene. Maximal association was observed at marker G (*rs4274624,* P <0.0001, OR 1.27 [95% CI 1.12-1.42]). The analysis was repeated in Stata[237] to gain an accurate P value estimate: the P value was calculated as 0.00016. The slight discrepancy between the results from Revman 5 and those from Stata is because Revman 5 uses allele data only for meta-analysis, whereas Stata uses genotype data. As the differences were very small, the results from the two programs were comparable. Association was also noted at marker F in the *STAT4* gene (*rs10931481,* P 0.0007, OR 1.23 [95% CI 1.09-1.39]) (Figure 30).

The *rs4646536* SNP in *CYP27B1* was also associated in the meta-analysis (random effects model: P 0.03, OR 0.90 [95% CI 0.82-0.99]). This marker is in moderate LD with another genotyped SNP, *rs10876993* (r^2 0.45). However, no association was seen with this SNP and AAD (Figure 31). Finally, *rs3802604* (marker E), an independent SNP in *GATA3,* was also associated with AAD (random effects model: P 0.03, OR 0.90 [95% CI 0.82-0.99]) (Figure 32).

When any individuals who were not 21OH autoantibody positive were excluded and the meta-analysis repeated (1204 21OH+ cases: 53 from UK, 290 from Norway, 73 from Poland, 154 from Germany, 266 from Italy and 368 from

Sweden; 1936 controls were available for comparison), three SNPs were associated (Table 17). Using a random effects model, maximal association was again observed at marker G in *STAT4* (*rs4274624,* P 0.0003, 0.76 [95% CI 0.65 – 0.88]). At marker E in *GATA3*, association was observed (*rs3802604,* P 0.04, OR 1.12 [95% CI 1.00-1.25]). In addition, *rs13017599* in *REL* was associated in this cohort (P 0.03, OR 0.88 [95% CI 0.78 – 0.99]) although the result for this SNP in the cohort as a whole did not reach statistical significance (P 0.05).

		All AAD (maximum 1955 cases, 1936 controls)		21-OH + (maximum 1204 cases, 1936 controls)		Notes
Gene	**SNP**	**P value**	**I^2 (%)**	**P value**	**I^2 (%)**	
CYP27B1	rs10876994	0.19	0	0.25	0	UK & Norway excluded (1264 cases; 861 21OH+, 1221 controls)
CYP27B1	rs4646536	**0.03**	0	0.16	9	
STAT4	rs10931481	**0.0007**	0	0.07	64	UK & Norway excluded (1264 cases; 861 21OH+, 1221 controls)
STAT4	rs4274624	**0.00016**	20	**0.0003**	28	
STAT4	rs4853543	0.8	0	0.97	7	UK & Norway excluded (1264 cases; 861 21OH+, 1221 controls)
REL	rs13017599	0.05	0	**0.03**	0	UK & Poland excluded (1371 cases; 1078 21OH+, 1305 controls)
GATA3	rs569421	0.65	48	0.78	0	
GATA3	rs3802604	**0.03**	0	**0.04**	0	
GATA3	rs444929	0.1	58	0.42	55	

Table 17: Summary of significant meta-analysis results, applying a random effects model, for round 2 of the twenty candidate gene study.

Data are shown for all studied SNPs in genes where at least one result was significant (P <0.05, shown in yellow and bold). The results are shown for the cohort as a whole and for the cohort excluding any individuals who are 21OH autoantibody negative. The “notes” column details any cohorts excluded for quality control reasons.

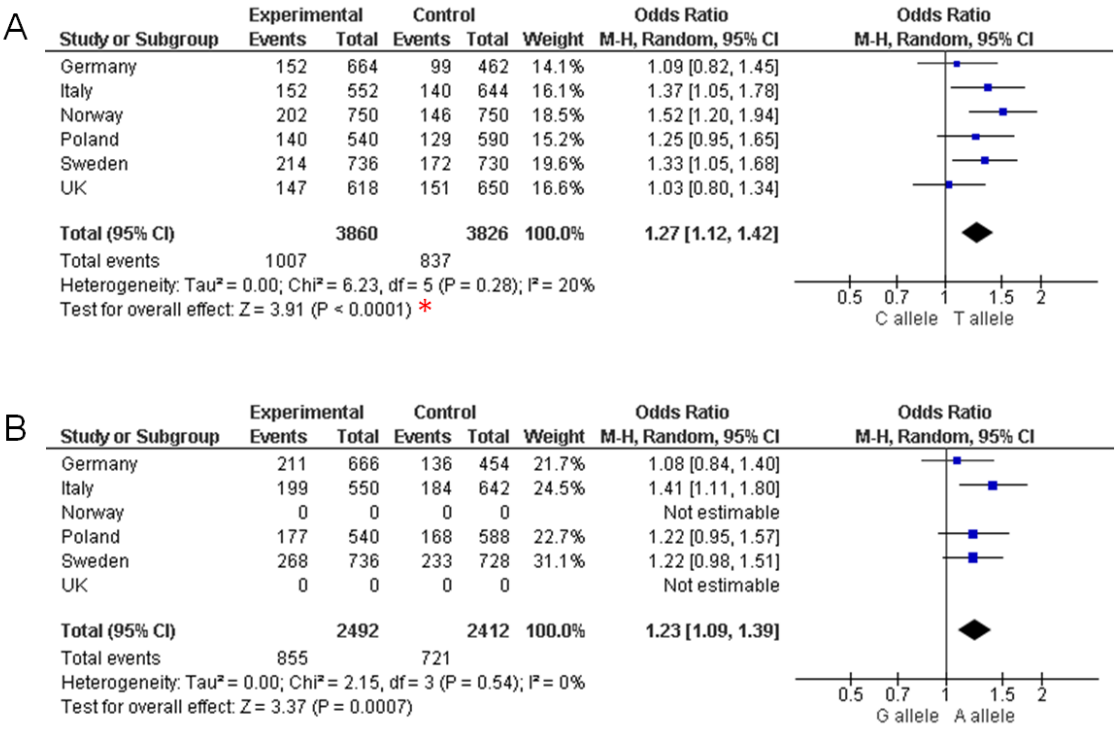


Figure 30: Forest plots of meta-analysis results for two markers, rs4274624 (panel A) and rs10931481 (panel B) in the STAT4 gene.

Data from the UK and Norway at marker *rs10931481* did not meet the quality control criteria and are therefore missing. Pooled analysis showed some heterogeneity among the cohorts at *rs4274624* (P 0.28, I^2 20%) and no heterogeneity at *rs10931481* (P 0.54. I^2 0%). Using a random effects model, the meta-analysis confirms association between the T allele at SNP *rs4274624*, and the A allele at SNP *rs10931481* and AAD in different European populations, with an odds ratio (OR) of 1.27 [95% CI 1.12 – 1.42], P <0.0001 and 1.23 [95% CI 1.09 – 1.39], P 0.0007 respectively. *P value 0.00016 when data analysed under a random effects model in Stata.

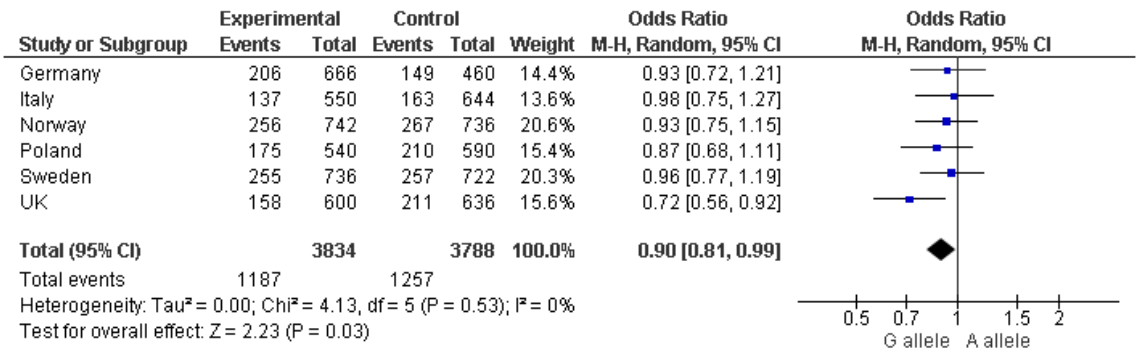


Figure 31: Forest plot of meta-analysis results for a single marker, rs4646536, in the CYP27B1 gene in AAD.

Pooled analysis showed no heterogeneity among the cohorts (P 0.53, I^2 0%). Applying a random effects model, the meta-analysis confirms association between the G allele at this SNP and AAD in different European populations, with an OR of 0.90 [95% CI 0.81 – 0.99), P 0.03.

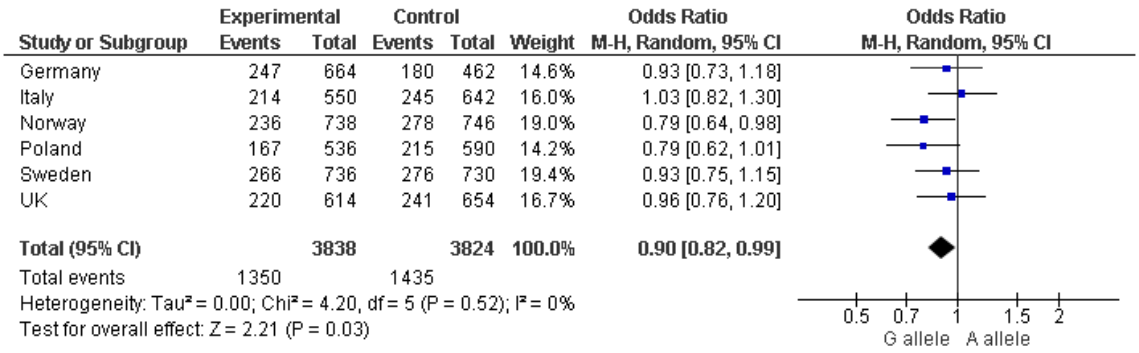


Figure 32: Forest plot of meta-analysis results for a single marker, rs3802604 SNP in the GATA3 gene in AAD.

Pooled analysis showed no heterogeneity among the cohorts (P 0.52, I^2 0%). Using a random effects model, the meta-analysis confirms association between the G allele at this SNP and AAD in different European populations, with an OR of 0.90 [95% CI 0.82 – 0.99), P 0.03.

4.4.5 REPLICATION STUDIES – THE *rs4698861 NF-κB1* POLYMORPHISM IN UK GRAVES' DISEASE

To determine whether the finding of association between alleles at marker E, *rs4698861* in the *NF-κB1* gene, and AAD in the UK cohort could be replicated in another more common autoimmune disease, Graves' disease, in the UK population, this SNP was genotyped by RFLP in 392 UK GD cases for comparison with Sequenom UK control data from round 1.

4.4.5.1 RESULTS – NF-κB1 IN UK GRAVES' DISEASE

The G/G genotype was present in 35/392 (8.9%) UK Graves' cases compared to 45/321 (14.0%) UK controls, while the A/G heterozygote was present in 173/392 (44.1%) Graves' disease cases, compared to 150/321 (46.7%) controls ($P_{genotype}$ 0.035). The minor G allele appeared to be protective for Graves' as it was present in only 243/784 (31.0%) of cases, compared to 240/642 (37.4%) controls (P_{allele} 0.011, OR 0.75 [95% CI 0.60 – 0.94]).

4.4.6 REPLICATION STUDIES – *GATA3* POLYMORPHISMS AND AAD SUSCEPTIBILITY

Polymorphisms in the *GATA3* gene have not previously been associated with autoimmune conditions, thus this represented a novel finding. I therefore decided to investigate other SNPs in this gene, using the Sequenom platform, in the UK and Norwegian AAD cohorts to determine the extent of the association in AAD. In addition, I genotyped these SNPs in a cohort of UK Graves' disease patients, in individuals with type 1 diabetes from Norway and in a cohort of patients with rheumatoid arthritis from New Zealand, to determine whether any association found was specific to AAD or extended to these other, more common, autoimmune conditions. All genotype data can be found in electronic appendix B.

HapMap was used to select SNPs in all of the common LD blocks of the *GATA3* gene and a Sequenom plex was designed, including 15 SNPs in total. The Sequenom platform was then used to genotype these SNPs in UK and Norwegian AAD cohorts (335 UK AAD, 302 UK controls; 352 Norwegian AAD, 1353 controls), UK Graves' disease patients (283 cases), Norwegian type 1 diabetes patients (1195 cases, 1353 controls) and in rheumatoid arthritis patients from New Zealand (650 cases, 452 controls). SNPs with a call rate of less than 95% were excluded from the analysis, as were SNPs out of HWE (P <0.01) in the control cohorts.

4.4.6.1 DATA QUALITY CONTROL

In the UK AAD cohort, one SNP (*rs2229359*) was excluded on the basis of a low call rate of 90% in the cases. In the Norwegian AAD and type 1 diabetes analyses where a single control group was shared, two SNPs, *rs2275806* and *rs1058240,* were excluded on the basis of low call rates in the control group. In the UK Graves' disease cohort, one SNP (*rs9746*) was excluded, again due to a low call rate in the cases. This SNP also genotyped poorly in the rheumatoid arthritis cases and controls from New Zealand and was also excluded from analysis in that cohort.

4.4.6.2 RESULTS - GATA3 IN UK AAD

4.4.6.2.1 ASSOCIATION ANALYSIS

Of the 14 SNPs included in the analysis (Figure 33), maximal association was observed at marker I, an intronic variant (*rs422628*). At this SNP, the CC genotype was seen less commonly in AAD cases (13/327; 4.0%) compared to controls (28/287; 9.8%. $P_{genotype}$ 0.013). The CT heterozygote was seen in similar proportions in cases and controls (38.5% and 38.7% respectively). A similar decrease in the frequency of the C allele was seen in AAD cases compared to controls at this SNP (23.2% versus 29.1% respectively, P_{allele} 0.01, OR 0.74 [95% CI 0.57 – 0.95]). When the cohort was subdivided into those with iAAD (n=157) and those with APS2 (n=178), the association was only with iAAD ($P_{genotype}$ 0.001, P_{allele} 0.0005, OR 0.55 [95% CI 0.39 – 0.77]). This SNP is in

moderate LD with neighbouring intronic marker H (*rs444929*, r^2 0.78) and association was also noted with genotypes and alleles at this SNP ($P_{genotype}$ 0.017, P_{allele} 0.019, OR for C allele 0.73 [95% CI 0.56 – 0.95]). At this marker, the association was also with the iAAD subgroup ($P_{genotype}$ 0.011, P_{allele} 0.0055, OR 0.62 [95% CI 0.44 – 0.87]). Markers H and I are in moderate LD with marker N in the 3' UTR (*rs1058240,* r^2 0.79). Here, although a similar trend was seen, statistical significance was not reached ($P_{genotype}$ 0.052, P_{allele} 0.081, OR for G allele 0.79 [95% CI 0.60 – 1.03]). However, in the subgroup analysis, association was noted with the iAAD subgroup at this SNP ($P_{genotype}$ 0.11, P_{allele} 0.0073, OR for G allele 0.62 [95% CI 0.43 – 0.88]).

An association with genotypes and alleles with AAD was also seen at intronic marker G (*rs569421:* $P_{genotype}$ 0.029, P_{allele} 0.0096, OR for C allele 1.42 [95% CI 1.09 – 1.85]). Again, the association was with iAAD ($P_{genotype}$ 0.048, P_{allele} 0.013, OR 1.50 [95% CI 1.09 – 2.07]) and not with APS2. This marker is in significant LD with marker J (*rs406103,* r^2 0.82) and nominal association was seen with alleles only at this marker ($P_{genotype}$ 0.08, P_{allele} 0.026, OR for T allele 1.36 [95% CI 1.04 – 1.78]). When the cohort was subdivided into iAAD and APS2 no significant association was seen ($P_{genotype}$ 0.17, P_{allele} 0.07 with iAAD; $P_{genotype}$ 0.11, P_{allele} 0.055 with APS2). At marker M in the 3' UTR (*rs9746*), also in moderate LD with both marker G (r^2 0.51) and J (r^2 0.61), a similar nominal association was seen with alleles only ($P_{genotype}$ 0.097, P_{allele} 0.048, OR for G allele 1.39 [95% CI 1.00 – 1.94]). When the cohort was subdivided, there was nominal association with iAAD with both genotypes and alleles ($P_{genotype}$ 0.049, P_{allele} 0.029, OR for G allele 1.55 [95% CI 1.04 – 2.30]).

Finally, at marker B (*rs2275806*), an upstream variant, an association was seen with genotypes only ($P_{genotype}$ 0.022, P_{allele} 0.06). When the cohort was subdivided, nominal association was seen with genotypes only with iAAD ($P_{genotype}$ 0.036, P_{allele} 0.088). This SNP is in moderate LD with neighbouring intronic markers D, E and F (*rs3781094,* r^2 0.62; *rs3802604,* r^2 0.68; *rs570613,* r^2 0.47). Association was not seen at any of these markers.

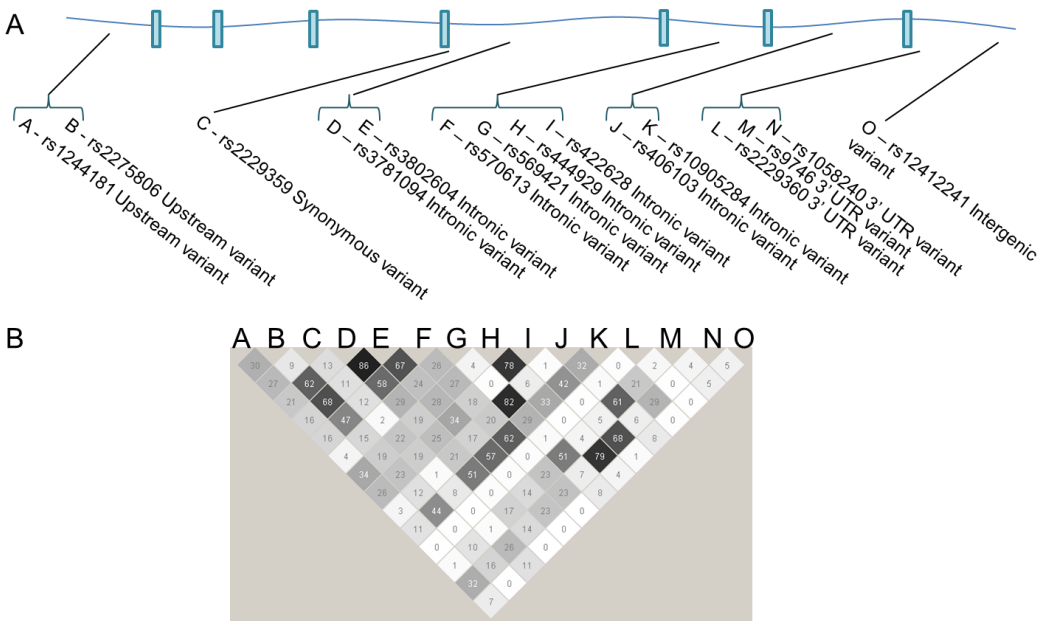


Figure 33: Schematic representation of the GATA3 locus (panel A) and pairwise linkage disequilibrium (r2) measures between 15 SNPs genotyped in and around the GATA3 gene (panel B).

In panel A, exons are depicted by the blue boxes and intronic sequence is illustrated by the line. The 15 genotyped variants are shown with their approximate locations. In panel B, LD measures between the 15 SNPs are shown in a Haploview LD plot generated using genotype data derived from Caucasian individuals from HapMap. White boxes represent the lowest r^2 values and black boxes represent the highest r^2 values. There is significant LD between markers B, D, E, F and K, markers G, J and M and between markers H, I and N.

4.4.6.2.2 HAPLOTYPE ANALYSIS

As multiple markers were found to be associated with AAD in the UK cohort, the UNPHASED program[220] was used to estimate haplotype frequencies and association. In the analysis, the SNP where maximal association with alleles was observed (SNP G, *rs569421*) was initially conditioned on. The next most associated marker was then selected (marker M, *rs9746*) and this was conditioned on, in addition to marker G and so on, until all the association was accounted for (Figure 34). A four marker haplotype, comprising markers D-G-I-M (*rs3781094-rs569421-rs422628-rs9746*), was found to be significantly associated with AAD, with an overall P value of 1.72×10^{-12}. The major haplotype at these markers, C–T–T–A, was found with frequencies of approximately 53% in both cases and controls. The A–C–T–A haplotype appeared to be protective for AAD, being present in just 1.9% of cases compared to 10.9% of controls (OR 0.17 [95% CI 0.082 – 0.36]).

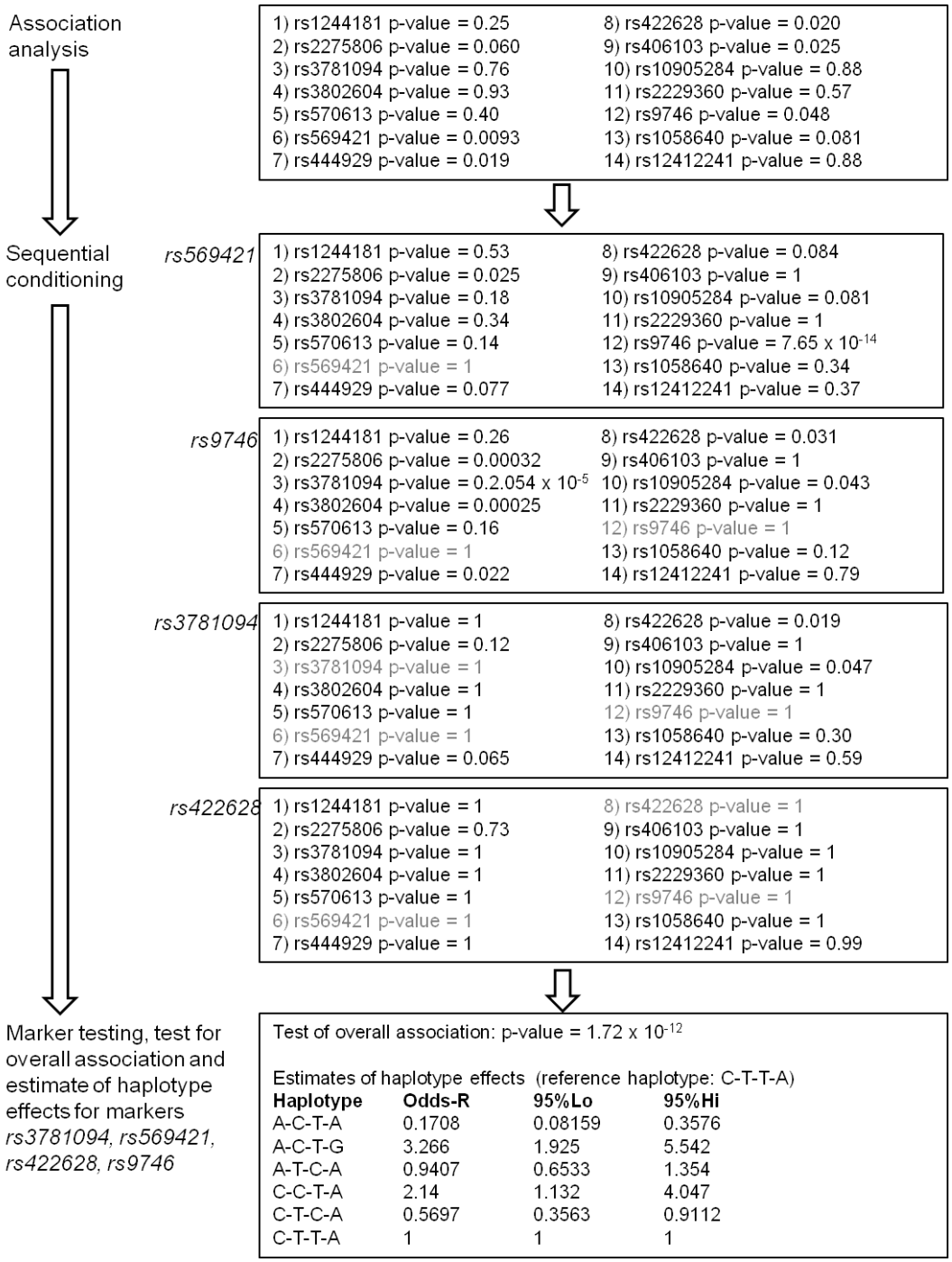


Figure 34: Flowchart of method for UNPHASED haplotype analysis.

Haplotype analysis of *GATA3* SNPs genotyped in the UK AAD cohort is used as an example. Following association analysis, sequential conditioning occurs until all association has been accounted for. Haplotype analysis can then be undertaken for the associated SNPs, in this case *rs3781094, rs569421, rs422628* and *rs9746,* using the most common haplotype, C–T–T–A, as the reference.

4.4.6.3 RESULTS – GATA3 IN NORWEGIAN AAD

In the Norwegian AAD cohort, nominal association was noted with two intronic markers in moderate LD; D (*rs3781094*) and F (*rs570613*) (r^2 0.58) (Figure 33). Alleles only were associated with AAD at marker D ($P_{genotype}$ 0.11, P_{allele} 0.042, OR for A allele 0.82 [95% CI 0.68 – 0.99]) but when the cohort was subdivided into iAAD (n=149) and APS2 (n=203), no significant association was noted. Genotypes only were associated with AAD at marker F ($P_{genotype}$ 0.049, P_{allele} 0.32). No association was observed with alleles at this marker. Again, no significant association was observed with either the iAAD or the APS2 subcohorts.

Alleles at marker E (*rs3802604)*, which were modestly associated with AAD in the Norwegian cohort in round 1 of this study (382 AAD cases, 380 controls; $P_{genotype}$ 0.04, P_{allele} 0.032), were not associated with AAD in this analysis which included a larger control cohort for comparison ($P_{genotype}$ 0.34, P_{allele} 0.15). Indeed, if the genotyping results for AAD cases from round 1 at this marker are compared with the genotyping results for the larger control cohort in this analysis, no association is noted ($P_{genotype}$ 0.46, P_{allele} 0.22) suggesting that the round 1 result may have been a spurious finding due to the small control cohort size.

4.4.6.4 RESULTS – GATA3 META-ANALYSIS IN AAD

The UK and Norwegian data were analysed together in a meta-analysis, using the Revman 5 program and applying a random effects model. There was significant genetic heterogeneity at five markers, indicated by an I^2 of greater than 60%. No association was noted at any markers in the meta-analysis.

In the previous rounds of genotyping, marker E (*rs3802604*) was associated with AAD in a meta-analysis of the data from the six European cohorts (P 0.03). If the data from the larger Norwegian control cohort are used for the six cohort

meta-analysis in place of the Norwegian control data from round 1, the result no longer reaches statistical significance (P 0.06).

4.4.6.5 RESULTS – GATA3 IN UK GRAVES' DISEASE

In the UK Graves' disease cohort, association was observed with both genotypes and alleles at marker G (*rs569421*) (Figure 33). The CC genotype was seen in 17/276 (6.2%) of cases compared to in 13/300 (4.3%) controls ($P_{genotype}$ 0.049). A similar increase in C allele frequency was observed in cases compared to controls (25.7% versus 19.8%, OR for C allele 1.40 [95% CI 1.06 – 1.85], P_{allele} 0.017). There is moderate to significant LD between this SNP and markers J and M (*rs406103,* r^2 0.82; *rs9746,* r^2 0.51), however, no association was seen at these markers.

At marker F (*rs570613*) an association with Graves' was observed with genotypes only ($P_{genotype}$ 0.018, P_{allele} 0.17). This SNP is in moderate LD with four others (markers B, D, F and K r^2 0.47 – 0.67), however, none of these SNPs were associated in this analysis.

4.4.6.6 RESULTS – GATA3 IN NORWEGIAN TYPE 1 DIABETES MELLITUS

In the Norwegian type 1 diabetes cohort, nominal association was noted with alleles only at a single independent marker in the 3' UTR (marker L, *rs2229360*; $P_{genotype}$ 0.063, P_{allele} 0.024, OR 1.56 [95% CI 1.06 – 2.29]).

4.4.6.7 RESULTS – GATA3 IN NEW ZEALAND RHEUMATOID ARTHRITIS

In the New Zealand rheumatoid arthritis cohort, association was seen at intronic marker E (*rs3802604*) with genotypes and alleles (Figure 33). The GG genotype was seen in 109/640 (17.0%) of cases compared to 53/441 (12.0%) of controls ($P_{genotype}$ 0.033). Similarly, the G allele was seen in 40.3% of cases and in 34.8% of controls (P_{allele} 0.0096, OR 1.27 [95% CI 1.06 – 1.51]). At the four markers in moderate LD with this SNP (markers B, D, F and K), no association was noted.

Association with alleles only at marker L in the 3' UTR (*rs2229360*) was also observed ($P_{genotype}$ 0.098, P_{allele} 0.026, OR for T allele 2.90 [95% CI 1.09 – 7.71]).

4.4.6.8 RESULTS – GATA3 META-ANALYSIS IN AUTOIMMUNE DISEASES

The genotype data for the 15 *GATA3* SNPs were analysed together in a meta-analysis applying a random effects model. Initially, all data were combined. As the same control cohorts were used for comparison with the UK and Norwegian case cohorts, a second analysis was conducted, excluding the smallest cohorts (i.e. the UK Graves' disease cohort and the Norwegian AAD cohort) to eliminate overlap of the control groups.

At ten loci, significant heterogeneity was noted (I^2 values 41 to 74%). In the meta-analysis of all five autoimmune disease cohorts, no marker was associated with disease. This was also the case if the UK Graves' and Norwegian AAD cohorts were removed to eliminate control group overlap. This result, in conjunction with the single cohort analyses, suggest that *GATA3* does not confer susceptibility to autoimmune diseases in general, although it does appear to be exerting an effect in the UK AAD cohort.

4.5 DISCUSSION

This candidate gene case-control association study is the largest genetic study performed in AAD to date and includes almost two thousand affected participants from six European countries. This study has been made possible through collaboration within the Euradrenal consortium funded by a European Union Framework Programme 7 grant. It has generated some interesting and novel data, highlighting significant genetic heterogeneity between Caucasian European populations and implicating a number of biomolecular pathways in the pathogenesis of this rare autoimmune condition.

4.5.1 GENETIC HETEROGENEITY BETWEEN EUROPEAN COHORTS

This study has revealed significant differences in allele frequencies between healthy Caucasian individuals from the six European countries involved (Table 16). In general, as would perhaps be predicted, allele frequencies tend to differ less between countries that are geographically close when compared to countries that are distant. For example, allele frequencies between the control cohorts from adjacent countries Germany and Poland differed significantly at only 1 of 19 (5.3%) of the loci tested. Likewise, allele frequencies between the control cohorts from Norway and Sweden differed at three of 21 loci (14.3%). Conversely, allele frequencies differed greatly between geographically distant control populations, such as the UK and Sweden, where significant differences in allele frequencies were noted at almost half of the loci tested. However, allele frequencies between Sweden and Germany did not differ at any of the loci tested, even though these countries are relatively far apart. This reflects a greater degree of common ancestry between these two populations: the Swedish empire during the 17th and early 18th centuries included large portions of northern Europe, including northern Germany. Allele frequencies tended to be more comparable among the northern European countries (UK, Germany, Poland, Norway and Sweden) compared to allele frequencies between these countries and Italy, the only southern European country included in this study. The Italian control cohort was genetically dissimilar to the other controls cohorts at 50% or more of the loci tested. This emphasises the importance of carefully

matching controls to the population under study, and in appropriately accounting for this heterogeneity in pooled statistical analyses, such as meta-analysis.

The design of this study, with rounds of genotyping based on the results from previous analyses, makes the assumption that loci contributing to AAD in the UK and Norwegian cohorts in round 1 might also be contributing to disease in the other cohorts in round 2. Conversely, the study design also assumes that loci not associated with AAD in the UK and Norwegian cohorts will also not be contributing to AAD susceptibility in the other cohorts. Thus SNPs associated with AAD in the UK and/or Norway in round 1 were taken forward to round 2 for genotyping while those not associated were discarded. The study was designed in this manner due to time and budget constraints. In addition, the DNA samples for the UK and Norwegian case and control cohorts were already available at the outset of the study, while other samples had to be collected. Therefore, it made logistical sense to use these samples for round 1. However, the significant genetic heterogeneity observed between the control cohorts means that this underlying assumption may not be correct. It is therefore possible that SNPs genotyped in round 1 which were discarded for round 2 may have yielded significant and interesting results in the round 2 cohorts had they been genotyped. This is likely to be the case for the Italian cohort in particular which was genetically less similar to the other cohorts.

4.5.2 CANDIDATE GENE ASSOCIATIONS

The sizes of the individual cohorts included in this study are comparable to many cohorts included in previously published genetic studies of AAD, however power calculations estimated prior to undertaking genotyping demonstrate that each individual cohort would have more than 60% power to detect a locus with an odds ratio of 1.6, assuming a MAF of 0.3. If a MAF of 0.2 and 0.1 are assumed, the power drops significantly (Figure 27). Criticisms levelled at previously published studies include this lack of power and also the failure to correct for multiple testing. Indeed, although the individual cohort results generated from this study are interesting and highlight many loci that are

potentially contributing to AAD aetiology, if the effects of multiple testing are taken into account, due to the relatively small cohort sizes, many of the findings would fail to meet the corrected threshold for statistical significance.

There are a number of approaches which can be used to allow for multiple comparisons within a data set[443]. These can be useful for interpreting results and reducing the possibility of reporting a false positive result. Using a sequential study design, such as employed in this study, is one recognised way of verifying a positive finding. In the sequential method, which was first applied in the field of genetics by Morton in the 1950s[88], data are collected and/or analysed in stages, thus eliminating false positive initial results through failure to replicate them at a later stage in the study. However, even when this method is used to design a study, the testing of multiple markers should still be considered when interpreting the results. There are a number of ways of doing this.

Permutation testing is a computational simulation-based method which can be used to calculate significance levels for SNPs using the data derived from genotyping. In this method, data are re-analysed again and again up to a maximum number of permutations. On each occasion, the case and control labels are randomly assigned within the whole data set. Observed P values can then be compared to P values generated during the repeated permutations[444]. This method can be carried out using a number of computer programs including PLINK[217], but it can be computationally intensive. It also looks at each SNP individually and does not allow for markers being in LD.

The false discovery rate (FDR) method[445] works on the basis that, when a well-defined statistical test is repeated again and again, there will be an expected number of false positive results generated. The proportion of false positives expected can therefore be estimated using computer software to generate a threshold of statistical significance. This is a less computationally intensive method than permutation testing but again does not allow for multiple markers in LD.

An alternative method is the Bonferroni method[223] which offers a conceptually much more simple way of correcting for multiple testing. If the desired significance level is set at P 0.05 and 100 markers are tested, any result with a P value of 0.0005 (0.05/100) would be considered statistically significant. This is a conservative approach and some argue that it is too stringent, and increases the likelihood of a false negative result[443]. In addition, this method assumes that all loci tested are independent and therefore, like permutation testing, does not account for some markers being in LD. Therefore, an adaptation of the Bonferroni method can be used which allows for LD between markers. An arbitrary threshold can be selected above which SNPs are said to be in LD, for example an r^2 ≥0.40. If markers are in LD, they account for a single locus and if they are not in LD, they are considered independent. The significance level is then corrected for the number of loci tested. For example, if 100 tests are performed but there is significant LD between markers and therefore, in real terms, only 50 independent loci are tested, the corrected level of significance would be P 0.001 (0.05/50). As this study is designed to look at multiple SNPs within candidate genes, with some independent SNPs and some in LD being tested, the adapted Bonferroni method for determining significance levels has been applied.

In round 1, 101 SNPs were genotyped. Taking into account LD patterns between SNPs, using data derived from Caucasian individuals from HapMap, taking an r^2 cut off <0.40 to signify independence, 64 of these can be considered independent markers. Correcting for the 64 independent loci results in a corrected significance level of P <0.00078 (0.05/64). Similarly, in round 2, where 21 SNPs were analysed but only 15 of these represent independent loci, a significance threshold of P <0.0033 (0.05/15) can be applied. Combining the results in a meta-analysis provides a more powerful means of data analysis. Including all individuals with AAD, this study has more than 80% power to detect a locus with an odds ratio of 1.4, assuming a MAF of 0.1 or more. The same correction for multiple testing can be made for the meta-analysis results as were made for the data results for round 2.

Allowing for multiple comparisons in round 1 (64 loci tested, P <0.00078), only 1 SNP from the Norwegian cohort (*rs4274624* in *STAT4*; P_{allele} 0.00045) and two from the UK cohort (*rs230532* and *rs4698861* in *NF-κB1*; P_{allele} 0.00041 and 0.00017 respectively) would meet the threshold for association. Even if a more stringent Bonferroni correction is applied, allowing for 101 SNPs tested (P <0.0005), these SNPs would remain associated. In round 2, allowing for testing of 15 independent loci (P < 0.0033), only 1 marker in the Italian cohort (*rs11171806* in *IL23A*; P_{allele} 0.0028) would be significantly associated with AAD. If a more stringent significance threshold was applied, allowing for 21 tests (<P 0.0024), this result would not be considered significant. In the meta-analysis, taking the AAD cohort as a whole and allowing for multiple comparisons, alleles at only two SNPs, both in the *STAT4* gene, would be associated with AAD (*rs4274624* P 0.0001; *rs10931481* P 0.0007) and the more modest associations observed with alleles at *CYP27B1*, *GATA3* and *REL* would no longer be considered significant.

In general, a robust approach to correcting for multiple comparisons is justified in order to reduce false positive results. Nonetheless, some true results may be discarded as a consequence. It is possible that this is the case for the multiple SNPs at the *CD28-CTLA4-ICOS* which were found to be associated with AAD in the Norwegian population in round 1. Polymorphisms at the *CTLA4* locus have previously been investigated in AAD in a number of small cohorts, with association reported in studies involving Italian[446], Norwegian[171] and UK subjects[116, 170]. However, conflicting results reporting no association have also been published from a Spanish cohort study[447] and from a study in the UK population, which reported no association except in those with the *HLA-DQA1* allele[448]. The *CTLA4* SNP that is most consistently associated with AAD is the *rs231775* marker (also known as *CTLA4 +49A/G* and *Ala17*). A meta-analysis of published study results for this SNP has previously been conducted by Brozzetti *et al*[446], to include 537 AAD cases and 1528 controls. Using both random and fixed effects models, they reported an association with AAD (P <0.0001), with an odds ratio of 1.48 [95% CI 1.28 – 1.71]. The two models gave very similar results as there was little heterogeneity between the studies. If the UK and Norwegian round 1 candidate gene data from the current study are

added to this meta-analysis, and the data from Vaidya *et al* removed from the dataset due to overlap between this and the UK cohort, using a random effects model (I^2 51%), the overall association is highly significant (P 0.0004, OR 1.34 [95% CI 1.14 – 1.58]) (Figure 35). This indicates that the finding of multiple SNPs associated with AAD at the *CD28-CTLA4-ICOS* locus in the Norwegian cohort might be a true finding, in line with previous studies, despite not reaching statistical significance when multiple testing is considered. Nevertheless, correction for multiple comparisons has been made in this study because, on balance, it is important that candidate gene association study findings are robust and replicable.

In summary, following correction for multiple testing, this study implicates 3 genes in AAD susceptibility: *NF-κB1* in the UK population, *IL23A* in the Italian population and *STAT4* in the cohort as a whole, with marked association noted in the Norwegian population in particular. These findings provide novel insights into the underlying genetic aetiology of AAD and implicate a number of biomolecular pathways in the pathogenesis of this rare condition. The STAT4 and IL23A proteins play a role in $CD4^+$ cell fate: the STAT4 transcription factor is known to be vital for the T_H1 response and also plays a role in T_H17 differentiation while IL23A is a subunit of the IL23 cytokine which drives differentiation of naïve T_H lymphocytes to T_H17 cells. NF-κB1 is a crucial component of the NF-κB pathway which allows a vigorous and rapid immune and inflammatory response to numerous potentially harmful stimuli.

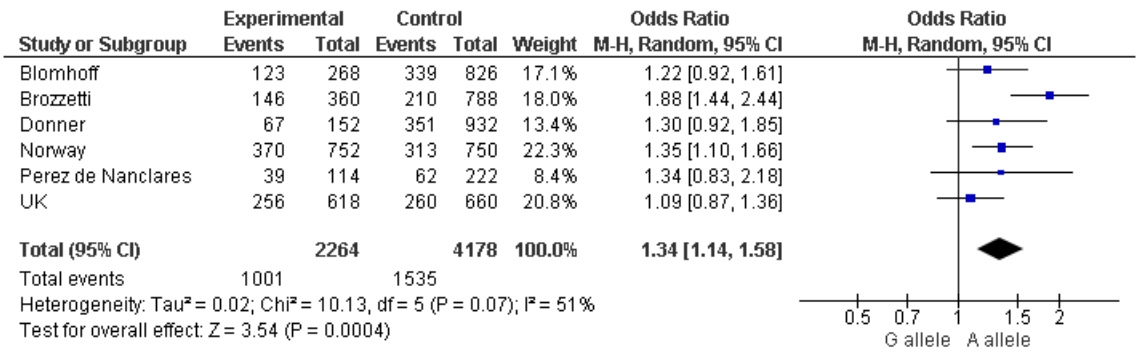


Figure 35: Forest plot of meta-analysis results for the rs231775 SNP in the CTLA4 gene.

The UK and Norwegian data for this SNP, from round 1 of the genotyping process, have been added to data previously collated by Brozzetti *et al,* European Journal of Endocrinology 2010[446]. The Vaidya study has been removed from the dataset due to overlap with the UK AAD cohort. Pooled analysis showed heterogeneity among the studies (I^2 51%); therefore a random effects model was applied. The meta-analysis confirms association between the A allele at *rs231775* and AAD, with an OR of 1.34 [95% CI 1.14 – 1.58], P 0.0004.

4.5.2.1 THE NF-κB PATHWAY IS ASSOCIATED WITH AAD AND GRAVES' DISEASE IN THE UK POPULATION

In this study, six SNPs in and around the *NF-κB1* gene were genotyped (Figure 28). Alleles at three of these SNPs have been associated with AAD for the first time in the UK cohort. This finding adds to the growing literature of autoimmune conditions known to be associated with variants in this gene[449]. An *NF-κB1* promoter polymorphism *(-94ins/del ATTG*, *rs28720239*) has previously been associated with ulcerative colitis[450] and Graves' disease[406], while alleles of a CA repeat downstream of *NF-κB1* in a regulatory region have been linked to susceptibility to type 1 diabetes[405]. In the UK AAD cohort, alleles at three SNPs in a haplotype (markers B, C and E - *rs10026278*, *rs230532* and *rs4698861*) were associated with disease, with the minor allele at each SNP appearing to be protective for AAD (OR 0.63 – 0.70). The association was consistently with APS2 rather than with iAAD. As those individuals with APS2 most commonly have autoimmune thyroid disease, the next logical step was to investigate the SNP which accounts for the association, *rs4698861* (marker E), in a cohort of individuals with Graves' disease from the UK. This marker is found downstream of the *NF-κB1* gene. Using RFLP, we noted association with Graves' disease and alleles at this marker, with the G allele again appearing to confer disease protection (OR 0.75 [95% CI 0.60 – 0.94]). This result provides independent confirmation that variants in this gene contribute to Graves' disease susceptibility in individuals from the UK.

In this study, individuals with Graves' disease were genotyped by RFLP and compared to control data from an earlier part of the study derived from the Sequenom platform. Ideally, genotype data to be compared between cases and controls should be generated using a single platform. This is because genotyping errors can markedly influence results. If a single platform is used, any bias introduced by the method of genotyping that is not detected in the quality control procedures will, in theory, be shared between both cases and controls. There is very little data comparing Sequenom to RFLP, as the two methods tend to be used in different contexts. Sequenom is a high-throughput

application designed for genotyping multiple SNPs together in a large number of samples, whereas RFLP is a more labour-intensive method and is used to look at one SNP at a time. However, one previous study has compared results gathered from analysing 73 human DNA samples on six different genotyping platforms and has demonstrated that, for biallelic SNPs, the inter-assay error rates are actually relatively similar[451]. In this study, a triallelic SNP was genotyped using Sequencing, Sequenom and LightCycler qPCR chemistry to detect the three alleles, G, T and A, while biallelic assays were designed for Taqman qPCR chemistry, RFLP and conventional allelic discrimination PCR to detect the G and T alleles. Excluding the six samples with the rare A allele which could not be detected by the RFLP assay, only two genotyping errors occurred (TT genotypes called incorrectly as GT and GG), allowing a genotype detection error rate of 3.0% to be calculated. Conversely, Sequenom did not incorrectly call any genotypes (genotype detection error rate 0%) and could be used to detect the rare A allele, but four of the 73 samples (5.5%) failed to genotype by this method (Figure 36). The above results suggest that the methods are reasonably comparable. The RFLP genotype error rate will be partly dependent upon the assay used and the quality of the post-digest products, as some PCR digests result in crisp bands which are easy to interpret, while others do not. The RFLP assay selected to genotype SNP E (*rs4698861*) was selected as it gave easily interpretable post-digest products. In addition, two individuals independently assessed the RFLP products and called the genotypes, so that unclear genotypes could be repeated or discarded. These strategies together should reduce the rate of erroneous genotypes, supporting the validity of the above association, despite the limitations in the methodology.

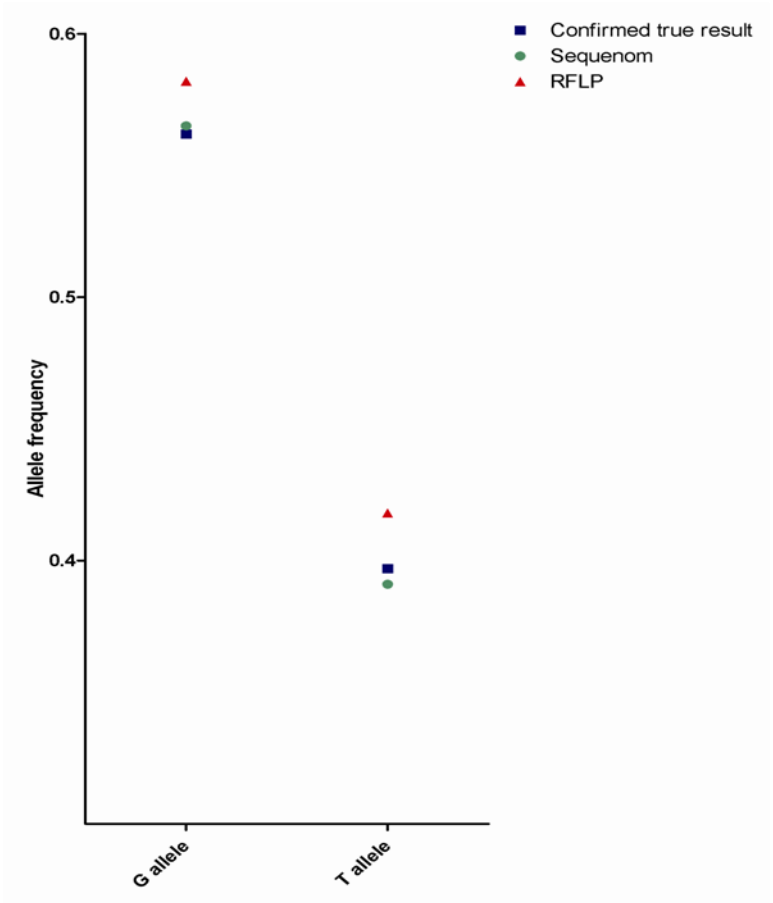


Figure 36: Comparison of allele frequency results generated by RFLP and Sequenom genotyping methods.

Graph generated from data taken from Huebner *et al*, Cancer Epidemiol Biomarkers Prev, 2007[451]), comparing allele frequencies derived from the Sequenom iPlex platform and RFLP compared to the true allele frequency, derived from triangulation of results from at least four methods (RFLP and Sequenom results shown only). The two platforms give comparable results to the true result.

Alleles at the associated markers in *NF-κB1* are relatively common variants, seen in a significant proportion of the healthy population. As such, these synonymous polymorphisms are not likely to be contributing to disease susceptibility themselves. Instead, they are likely to be "tagging" a susceptibility allele that is in LD with them. Interestingly, the three SNPs associated with AAD, which include marker E (*rs4698861*) which was also associated with GD, are all contained within a region of extended LD. This region also contains the previously associated *-94ins/del ATTG* variant and it is possible that the associated SNPs in this study are tagging this variant. Alleles of *-94ins/del ATTG* have previously been shown to influence nuclear protein binding to the *NF-κB1* promoter, with the ATTG deletion allele having markedly reduced promoter activity compared to the wild type[450]. It is therefore possible that this functional polymorphism, in LD with the associated SNPs in this study, could be subtly disrupting the NF-κB pathway, altering the immune response and thus rendering individuals susceptible to autoimmune disease.

4.5.2.2 AN IL23A VARIANT IS ASSOCIATED WITH AAD IN THE ITALIAN POPULATION

Polymorphisms in the *IL23A* gene have previously been implicated in susceptibility to psoriasis[325] and psoriatic arthropathy[452] in Europeans. In these previous studies, significant LD between *IL23A* and the adjacent gene *STAT2* has been noted and it is not clear at present which is conferring susceptibility or whether both are acting as markers for another locus. This candidate gene study suggests a role for the *IL23A/STAT2* locus in AAD in the Italian population. In the Italian cohort, the A allele at *rs11171806,* a synonymous variant in the *IL23A* gene, was associated with disease (P 0.0028, OR 2.37 [95% CI 1.32 – 4.23]). This marker is in LD with both the genotyped SNPs in *STAT2* (*rs2066808*, *rs2066807* r^2>0.90) and, although they did not remain significantly associated once correction for multiple testing was made, these were also associated with AAD in the Italian cohort (*rs2066808* P_{allele} 0.014, OR 1.93 [95% CI 1.13 – 3.28]; *rs2066807* P_{allele} 0.0063, OR 2.18 [95% CI 1.23 – 3.85]). These variants are found in healthy controls and are synonymous. Therefore, like the variants in the *NF-κB1* gene, they are likely to be tagging a

susceptibility locus rather than contributing to disease propensity directly themselves. A number of non-synonymous variants are found in and around the *IL23A* and *STAT2* genes and it is possible that these, or a non-coding sequence variant exerting a regulatory effect, might be contributing to disease susceptibility in this case.

4.5.2.3 STAT4 POLYMORPHISMS ARE ASSOCIATED WITH AAD IN THE EUROPEAN POPULATION

Variants in the *STAT4* gene have previously been associated with systemic autoimmune conditions. An association with *STAT4* SNPs and rheumatoid arthritis has been observed in North American Caucasians[321] and in Korean Asians[453], while associations with both SLE[321] and primary Sjögren's syndrome[454] have been observed in European Caucasians in separate studies. This study extends the association to an organ-specific autoimmune condition, AAD.

The association with markers in *STAT4* and AAD was initially noted in the Norwegian cohort in round 1, where the minor C allele of marker G (*rs4274624)* was present in 27.2% of AAD cases compared with just 19.5% of controls ($P_{genotype}$ 0.00084, P_{allele} 0.0004, OR 1.55 [95% CI 1.21 – 1.97]) (Figure 29). The C allele at this marker was also associated with AAD in the Swedish and Italian cohorts in round 2 (Table 15), although the associations were modest and did not withstand correction for multiple testing. In the Italian cohort, the G allele at marker F (*rs10931481*), in moderate LD (r^2 0.59) with marker G, was also associated with disease, but again, this finding was not significant when corrections were made for multiple comparisons. In a meta-analysis, using a random effects model, both marker G and marker F were associated with AAD (marker G, *rs4274624* P 0.00016, OR 1.27 for C allele [95% CI 1.12-1.42]; marker F, *rs10931481* P 0.0007, OR for G allele 1.23 [95% CI 1.09-1.39]) (Figure 30).

When using case-control cohorts from different countries, the possibility of heterogeneity between cases due to diagnostic differences must be considered. To be included in this study, individuals with AAD had to have biochemical evidence of adrenal failure, with a maximum serum cortisol of less than 550nmol/L 1 hour following intravenous administration of synthetic ACTH analogue (synacthen 250µg) and a raised ACTH level. In addition, every effort was made to exclude those with secondary causes of adrenal insufficiency and those with APS1. Exclusions were made based on history and examination findings and they were applied to create relatively homogeneous AAD cohorts, containing individuals with similar phenotypes regardless of their country of origin. Unfortunately, despite these efforts, it is inevitable that there will be some heterogeneity between cases due to diagnostic uncertainty in a small minority of cases. This is one limitation universal to nearly all genetic studies where there is more than one possible underlying cause for the disease. In order to further define the AAD phenotype and increase AAD cohort homogeneity, a subgroup analysis including only individuals with 21OH autoantibodies can be studied. While 21OH autoantibody subgroup analyses were not performed for the individual AAD cohorts due to their small sizes, a 21OH autoantibody subgroup analysis was included in the meta-analysis, where any 21OH autoantibody negative individual was excluded. In this analysis, marker G in *STAT4* remained significantly associated (*rs4274624,* P 0.0003).

In a study by Remmers *et al*[321], the minor allele at *STAT4* marker *rs7574865* was significantly associated with both rheumatoid arthritis (P 4.64 x 10^{-8}, OR 1.27 [95% CI 1.16 – 1.36]) and SLE (P 1.87 x 10^{-9}, OR 1.55 [95% CI 1.34 – 1.79] in a meta-analysis. The minor allele at this SNP, in addition to three others in intron 3 of *STAT4*, was also associated with rheumatoid arthritis in the Korean population (P 0.0065, OR 1.27 [95% CI 1.11 – 1.45])[453] and with primary Sjögren's syndrome in a small study (P 0.01, OR 1.47 [95% CI 1.09 – 1.97])[454]. The marker most associated with AAD in the meta-analysis performed in this study, marker G (*rs4274624*), is in significant LD (r^2 0.90) with SNP *rs7574865.* Marker F (*rs10931481),* also associated with AAD in the meta-analysis, is in moderate LD with both marker G (r^2 0.59) and *rs7574865* (r^2 0.53) and was further associated with rheumatoid arthritis and SLE directly in the

study by Remmers (P 0.005, 0.025 respectively)[321], but to a lesser degree than *rs7574865.* These SNPs are all within a large intron in the *STAT4* gene which raises the possibility that they are tagging a variant which, rather than disrupting protein structure and/or function directly as deleterious mutations in the coding regions might, may result in splice variation or disrupt non-coding regulatory components to result in disease susceptibility.

4.5.3 THE *GATA3* LOCUS AND ITS ROLE IN SUSCEPTIBILITY TO AUTOIMMUNITY

From the data gathered in round 1, one further locus appeared to be exerting an effect on AAD susceptibility in both the UK and Norwegian AAD cohorts. Consistently, the association was with iAAD rather than APS2. Different markers were associated in the two cohorts: markers G and H (*rs569421* and *rs444929*) in the UK cohort and marker E (*rs3802604*) in the Norwegian cohort (Figure 33). However, given the significant genetic heterogeneity between the control groups from the two countries, this result was not surprising. When multiple testing corrections were applied to the round 1 data set, statistical significance at this locus was not quite achieved. Nonetheless, the association of multiple markers in *GATA3* with iAAD appeared convincing. The function of GATA3 as a key regulator of T lymphocyte development, in particular T_H2 cell differentiation and T_H2[345] and T_{Reg} function[348], make it an excellent candidate for autoimmune disease and was a persuasive argument for investigating this locus further. In addition, as *GATA3* polymorphisms have not previously been associated with autoimmune disease, this would represent a novel finding in terms of both AAD and autoimmunity in general. Therefore, this locus was chosen for further exploration. In the *GATA3* replication study, 15 SNPs were genotyped, in cohorts of UK AAD, Norwegian AAD, UK Graves', Norwegian type 1 diabetes and New Zealand rheumatoid arthritis cohorts. The 15 SNPs genotyped represent five independent loci.

4.5.3.1 GATA3 INFLUENCES SUSCEPTIBILITY TO AAD IN THE UK BUT NOT THE NORWEGIAN POPULATION

In the UK cohort, the pattern of association noted between the four previously genotyped *GATA3* SNPs and AAD in round 1 was replicated. Genotyping additional markers added more information to the round 1 findings. In the UK AAD cohort as a whole, association was noted at six of the 14 markers which passed the quality control checks (Figure 33). At each of the six associated markers, the association was with iAAD and not with APS2. Allowing for five comparisons (P 0.05/5 = 0.01), alleles at two intronic markers, (markers G and I, *rs569421* and *rs422628* respectively), would be considered significantly associated with AAD (P_{allele} 0.0096 and 0.01 respectively). In addition, a four marker haplotype, comprising SNPs D-G-I-M (*rs3781094-rs569421-rs422628-rs9746*), was found to be significantly associated with AAD in the UK subjects (P 1.72×10^{-12}), with the A-C-T-A haplotype conferring significant disease protection (OR 0.17 [95% CI 0.082 – 0.36]).

For the *GATA3* replication study, a large Norwegian control cohort, comprising 1353 individuals, became available for use. This was significantly larger than the control cohort available for genotyping in round 1, which comprised 380 individuals. In round 1, four SNPs in *GATA3* were genotyped in the Norwegian case-control cohorts and association was noted with genotypes and alleles at marker E only (*rs3802604;* $P_{genotype}$ 0.04, P_{allele} 0.032). In the *GATA3* replication study, comparing 352 AAD cases to 1353 controls, no such association was observed at this marker ($P_{genotype}$ 0.34, P_{allele} 0.15). To investigate this discrepancy, the round 1 Norwegian control cohort genotype and allele frequency data at this marker were compared to the data collected for the 1353 Norwegian controls in the follow-on study, using a χ^2 test. Genotype frequencies between the two Norwegian control cohorts differed significantly at this marker (P 0.034). Furthermore, if the genotyping results for AAD cases from round 1 at this marker are compared with the results for the larger control cohort in this analysis, no association is noted ($P_{genotype}$ 0.46, P_{allele} 0.22) suggesting that the round 1 result may have been a spurious finding, likely reflecting random sampling error in the smaller control cohort. This explains why the result failed

to replicate in the *GATA3* follow on study. In the Norwegian cohort, across the 13 *GATA3* SNPs which met the inclusion criteria, modest association with alleles only was noted at marker D (*rs3781094,* P 0.042) and with genotypes only at marker F (*rs570613,* P 0.049). However, these results do not remain significant when multiple comparisons are considered.

A meta-analysis of the UK and Norwegian data revealed significant heterogeneity between the two populations. In this analysis, no single marker was associated with AAD.

In conclusion, this study has demonstrated significant association between *GATA3* alleles and AAD in the UK cohort, but has failed to replicate this finding in a Norwegian AAD cohort. The first possible explanation for this observation is that both findings are true and that there is true genetic heterogeneity between the populations, meaning that polymorphisms at the *GATA3* locus are conferring disease susceptibility in UK individuals but that it is not a susceptibility locus for AAD in Norwegian individuals.

While Caucasian individuals of European origin have historically often been studied as a single entity in genetic analyses, there are large cohort studies which suggest that allele frequencies vary considerably within this group, and even between different regions within countries, lending support to the case for using carefully matched controls for genetic studies. For example, a study by Cross *et al* published in 2010[455] compared allele frequencies at 51 SNPs in 19,027 self-reported white Caucasians. The cohort was divided into those from Scandinavia, the UK, Germany and Eastern Europe. Between these four European regions, minor allele frequencies differed significantly at 19 (37.3%) SNPs (P <0.05). The difference was particularly marked (P <0.0001) at 5 (9.8%) of the 51 SNPs analysed (Figure 37). This supports the hypothesis that a susceptibility allele in the UK cohort might not be replicated in the Norwegian population. Furthermore, this has previously been observed in genetic studies in AAD in UK and Norwegian cohorts, where a significant association has been demonstrated in one cohort but not the other[170, 456]. One explanation for this is

selection pressure. There are many selection pressures, however infection is a particularly strong driver of selection and, combined with differing environmental factors, contributes significantly to genetic diversity between countries[457].

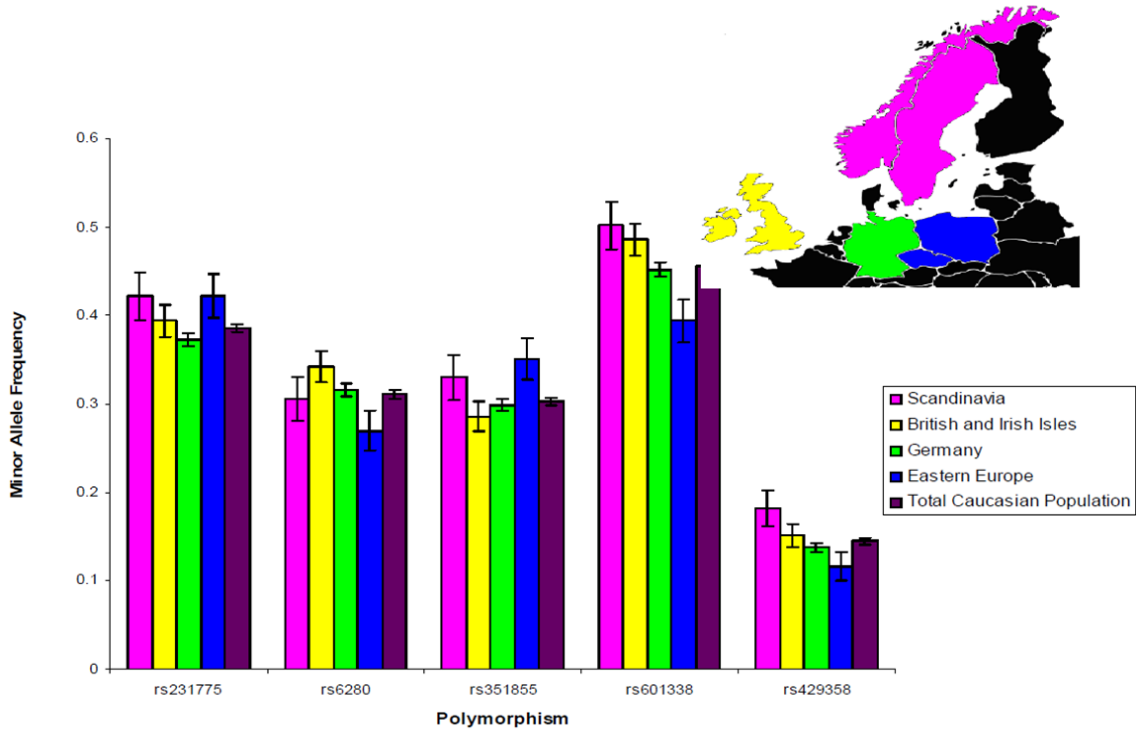


Figure 37: Differences in minor allele frequencies between subgroups of European Caucasians at five loci.

Graph to illustrate differences (P <0.0001) between minor allele frequencies at five SNPs between subgroups of European Caucasians. This figure is reproduced from page 12 of "Population based allele frequencies of disease associated polymorphisms in the Personalized Medicine Research Project" by Cross *et al*, published in BMC Genetics in 2010 (volume 11)[455], with permission from the authors.

Exposure to different pathogens between two populations can result in significant differences in allele frequencies. This could result in two populations, for example the UK and Norway, having different susceptibility loci for a disease of interest[457], particularly in a complex disease such as AAD where multiple factors result in disease.

The second explanation is that the association in the UK AAD cohort is a false positive finding, and that the differences observed between the case and control cohorts are due to inadequate sample size and random sampling error. A comparison of the minor allele frequencies between the UK AAD cohort controls (n=302) and a much larger control cohort of 3000 UK controls available from the WTCCC[458] (WTCCC genotype data for *GATA3* SNPs kindly provided by Mandy Phipps-Green, assistant research fellow, Merriman lab, Department of Biochemistry, University of Otago, New Zealand) was therefore undertaken to determine whether the controls used for this study were comparable to the larger control cohort at the two SNPs which remained associated after correction for multiple testing, markers G and I (*rs569421* and *rs422628)*.

There was no significant difference at marker G between the UK AAD controls and the WTCCC controls (MAF AAD controls 0.20, MAF WTCCC controls 0.17; P 0.072). However, at marker I, a statistically significant difference was seen, with the UK AAD controls having a MAF of 0.29 compared to the WTCCC cohort which had a MAF of 0.24 (P 0.0093). Indeed, if the UK AAD case data at these SNPs are compared directly with the WTCCC control data, a strong association is seen at marker G (*rs569421;* $P_{genotype}$ <0.00001, P_{allele} <0.00001, OR 1.72 [95% CI 1.43 – 2.07]) but not at marker I (*rs422628;* $P_{genotype}$ 0.22, P_{allele} 0.59, OR 0.95 [95% CI 0.78 – 1.15]). This suggests that the association with alleles at marker I with AAD in this study may be spurious, due to sampling error, but that the association with alleles at marker G is a true finding, as it remains strongly associated when a larger control cohort comprising 3000 UK individuals is used for comparison. Marker G is in strong LD (r^2 0.82) with another intronic variant, marker J (*rs406103)*. Alleles at this marker were also associated with AAD in the UK cohort (P_{allele} 0.026) but this result did not

withstand correction for multiple testing. If the UK AAD data at this marker are compared to the imputed control data for this marker from the WTCCC, the association is highly significant ($P_{genotype}$ <0.00001, P_{allele} <0.00001, OR 1.58 [95% CI 1.31 – 1.91]), supporting the hypothesis that the association of UK AAD with *GATA3* polymorphisms is a true association and not a spurious one.

4.5.3.2 GATA3 IN AUTOIMMUNE DISEASE SUSCEPTIBILITY

As association with *GATA3* polymorphisms and AAD in the UK population had been established and appeared convincing, the 15 *GATA3* SNPs in this follow-on study were also genotyped in three non-AAD autoimmune cohorts (UK Graves' disease, Norwegian type 1 diabetes and New Zealand rheumatoid arthritis). The purpose of this work was to establish whether *GATA3* could be a pleiotropic susceptibility locus for autoimmunity (that is one gene influencing multiple phenotypes). Following correction for testing five loci (0.05/5 = P <0.01), only one intronic SNP, marker E (*rs3802604)*, remained associated in the rheumatoid arthritis cohort (P_{allele} 0.0096) (Figure 33). When LD patterns in the New Zealand controls were studied using Haploview[216], significant LD between marker E and neighbouring intronic markers D and G was observed *(r^2* 0.88 and r^2 0.66 respectively). However, no significant association was seen at these other two SNPs, suggesting that the association at marker E in this cohort might be spurious. A meta-analysis was performed across the disease cohorts studied at the 15 *GATA3* SNPs to look for a marker which might be influencing autoimmunity in general, but no association was found.

There are many examples of pleiotropic genes in autoimmune diseases. The *MHC* region is the most widely replicated example and appears to be a universal autoimmunity locus. Other examples include the *CTLA4* and *PTPN22* loci, which appear to influence AAD[178, 179], Graves' disease[177, 459], type 1 diabetes[162, 176, 211] and rheumatoid arthritis[168, 460] susceptibility, among other conditions. Furthermore, this study has provided evidence that the *NF-κB1* gene is a susceptibility locus for both AAD and Graves' disease in the UK population. However, there are also numerous examples in the literature where a susceptibility locus for one autoimmune condition is investigated in another and

no association is detected. In some cases, the failure to replicate a finding is ascribed to a lack of study power. However, in other instances this does not appear to be the case: both the positive and the negative findings appear to be true. A recent study by Ramos *et al*[461] looked in detail at this phenomenon in the context of SLE, a rare multisystem autoimmune condition, in Europeans. They studied 446 non-*MHC* variants that had been associated in genome-wide studies with one or more of 17 autoimmune conditions in a cohort of 1500 SLE cases and 5706 controls. Some findings were then replicated in a separate cohort comprising 2085 SLE cases and 2854 controls. The investigators found a number of pleiotropic loci which appeared to be contributing to susceptibility to SLE and other autoimmune conditions, but also found a number of loci which appeared to be unique to SLE. Moreover, they generated some interesting results from a hierarchical clustering analysis which aimed to determine which autoimmune conditions are most similar. They found the most genetic similarity between type 1 diabetes and rheumatoid arthritis, and between Crohn's disease and ulcerative colitis. In contrast, they discovered that SLE is the most genetically distinct autoimmune disease of those studied, which did not include AAD as no genome-wide studies have been conducted in this condition to date. Therefore, *GATA3* may be an example of a gene which contributes significant susceptibility to a single autoimmune condition in one population (i.e. AAD in the UK), but does not appear to be influencing susceptibility to the other conditions studied. Conversely, AAD may be genetically dissimilar to the autoimmune conditions selected in the *GATA3* replication study for further investigation. If alternative autoimmune cohorts were selected, for example vitiligo or premature autoimmune ovarian failure, one or more of these diseases might show association with *GATA3* polymorphisms.

4.6 CONCLUSIONS AND FUTURE DIRECTIONS

This is the largest genetic study performed in AAD to date, including almost 2000 affected participants from six European countries. The results demonstrate significant genetic heterogeneity between the participating European countries. In addition, this study provides novel insight into the genetic aetiology of AAD, implicating *IL23A* in susceptibility to AAD in Italians, *NF-κB1* and *GATA3* in susceptibility to AAD in the UK population and *STAT4* in susceptibility to AAD in a meta-analysis of all cohorts.

The functional significance of these associations must now be studied. For example, it would be interesting to take *STAT4*, which harboured the most associated variants and look for functional differences between individuals with, and without AAD, with different genotypes at the most associated SNPs. Quantitative PCR could be used to determine whether genotype influences levels of *STAT4* expression and Western blotting could be used to determine whether individuals with differing genotypes produce differing amounts of the STAT4 protein. As STAT4 is essential for the T_H1 response, further studies could also aim to quantify the relative proportions of T_H1 and T_H2 cells in individuals with differing *STAT4* polymorphism genotypes using flow cytometry. The production of the primary T_H1 cytokine, IFN-γ, stimulated and unstimulated, could be studied in individuals with different genotypes using an intracellular cytokine flow cytometry method secretion assay or an Elispot assay.

Finally, a well-powered genome-wide association study might provide unique and interesting insights into the genetic architecture of AAD. This would also allow AAD to be compared to other autoimmune conditions, using a hierarchical clustering method similar to that used by Ramos[461] to determine which condition, if any, is most similar to AAD. This information would be valuable in directing future research efforts: positive genetic and functional findings in a condition closely related to AAD genetically could be prioritised for further investigation ahead of those found in autoimmune conditions less related to AAD. Unfortunately, while a significant collaborative European cohort now

exists, a replication cohort is currently lacking and this would be needed in order to validate any initial genome-wide screen findings. Therefore the onus is on clinicians and researchers alike to engage individuals with AAD in genetic research, in order to make further progress in the field.

CHAPTER 5 – A DISCOVERY-DRIVEN APPROACH TO THE INVESTIGATION OF AAD – A GENOME-WIDE STUDY OF MULTIPLEX AAD FAMILIES

5.1 BACKGROUND

AAD has a high genetic load compared to other autoimmune conditions, and yet we know relatively little about its genetic aetiology. This is because it is a rare disease which means that large AAD cohorts suitable for a powerful genetic study are scarce. If a large cohort of unrelated AAD patients and healthy controls could be collected together, a genome-wide association study could be conducted to shed light on the underlying genetic aetiology of AAD. This approach has been used for other complex autoimmune conditions, for example type 1 diabetes[201]. However, to date, even collaborative efforts to collect cohorts have not resulted in large enough sample sizes to generate sufficient study power for an initial genome-wide study and a replication study. Linkage studies are an alternative, powerful means of identifying genetic susceptibility loci. Multiplex AAD families (families comprising two or more individuals with AAD) could be used as an alternative study group in a linkage analysis, an approach which has never been applied to AAD before.

Carefully phenotyped multiplex AAD families are likely to be highly informative for genetic investigation by linkage. In addition, they could also be used for an intrafamilial association study, taking affected family members as cases and comparing them to the unaffected family members who act as controls in this study design. To avoid false positive results, linkage analysis requires a relatively sparse marker map made up of carefully selected, informative SNPs. Conversely, a dense marker map is needed for association analysis. Genotyping by SNP microarrays provides data that, with appropriate formatting and management, can be used for both linkage and association analysis, thus allowing the maximum amount of information to be gained from a single study.

5.2 AIM

I aimed to perform a genome-wide linkage and association analysis on multiplex AAD families, using the Affymetrix Genome-wide human SNP array 6.0 genotyping platform, searching for novel genetic loci for further investigation.

5.3 SUMMARY OF STUDY DESIGN

24 multiplex AAD families from the UK and Norway (Figure 5, Figure 6), comprising 121 individuals in total, were identified to be included in a linkage analysis. DNA from all individuals was sent for genotyping on the Affymetrix SNP 6.0 array and raw data was returned and formatted for analysis. Following strict quality control measures, the selected markers were thinned to produce a marker map suitable for linkage analysis in Merlin[236]. Parametric and non-parametric linkage analyses were performed on the autosomes and then these analyses were repeated for the X chromosome using MINX, a version of Merlin designed for the analysis of X chromosome markers. A further linkage analysis was then performed, where the study individuals were coded as being either 21OH autoantibody positive cases (36 individuals) or 21OH autoantibody negative controls (69 individuals). Autoantibody status was unknown for 12 individuals and these were therefore excluded from this analysis.

Following linkage analysis, the quality controlled but unthinned data set, constituting a much denser marker map, was used for an association analysis in EMMAX. Two analyses were conducted. The first took affected family members as cases and compared them to unaffected family members as controls. The second again used affected family members as cases but compared them to control genotype data from more than 2000 healthy individuals genotyped as part of the WTCCC 1958 UK birth cohort. Linkage and association data were then compared and two linkage regions prioritised for a validation experiment. For this, 64 SNPs were selected from within two regions of interest. Primer sequences for these assays can be found in electronic appendix C. These SNPs were genotyped on the Sequenom platform for use in an association study in UK, Norwegian and Swedish unrelated AAD case-control cohorts.

5.4 RESULTS – LINKAGE ANALYSIS IN MULTIPLEX AAD FAMILIES

5.4.1 LINKAGE STUDY POWER

SLINK[[235] was used to estimate study power. Under a rare dominant model (Table 8), assuming a disease allele frequency in the population of 1 in 10,000 and assuming a disease penetrance of 0.001 (0.1%) if 0 risk alleles are present and 0.999 (99.9%) if 1 or 2 risk alleles are present and assuming 75% of families are linked (i.e. 25% heterogeneity), the study has 77% power to detect a locus with an HLOD of 3.0 or greater, and 98% power to detect a locus with an HLOD score of 2.0 or greater. Allowing for greater levels of heterogeneity significantly reduced the study power (Table 18).

5.4.2 LINKAGE ANALYSIS – QUALITY CONTROL RESULTS

5.4.2.1 LOCUS AND INDIVIDUAL MISSINGNESS

The mean genotyping call rate per SNP was 99.3% (minimum 0%, maximum 100%, median 100%). Of the 909,622 SNPs genotyped, 80,150 (8.8%) had a call rate of less than 99% and were excluded. The mean call rate per individual genotyped was 99.3% (minimum 88.5%, maximum 99.8%, median 99.5%). Only one individual had a call rate of less than 97.5%. This individual was excluded, leaving 120 people in 24 families to be analysed (Figure 38).

5.4.2.2 SEX CHECK

Initially, seven individuals were identified whose genetic sex did not match their allocated sex in the pedigree file. DNA from these individuals was used in an amelogenin PCR assay designed to determine genetic sex. In all seven cases, the genetic sex by amelogenin PCR matched the genetic sex determined from the array data (Figure 39). On further investigation, four had been misassigned when our collaborators had drawn their pedigree diagrams and the pedigree file was updated to correct this error. In the case of three individuals (all from one Norwegian family), the problem could not be resolved. This family was excluded

from further analysis, leaving 117 individuals in 23 families in the cohort to be analysed.

5.4.2.3 MENDELIAN ERROR RATES

At each SNP, the mean number of Mendelian errors was low at 0.037 (minimum 0, maximum 5, median 0). The number of Mendelian errors within each individual were also low (44 individuals had one or more errors). Of these 44 individuals, the mean number of errors was 166.5 (minimum 9, maximum 1006, median 66). Within families, the number of Mendelian errors was also low: nine families had one or more Mendelian errors. Of these nine families, the mean number of errors was 267.9 (minimum 21, maximum 1050, median 176.5).

5.4.2.4 HETEROZYGOSITY RATES

The mean calculated heterozygosity rate was 0.32 (minimum 0.31, maximum 0.35, median 0.32). This was plotted against the call rate of each sample. There were no anomalous results and therefore no individuals were excluded on the basis of heterozygosity rate (Figure 38).

5.4.2.5 ALLELE SHARING

There was no excess IBD allele sharing in unrelated individuals (mean alleles shared IBD less than 0.2 with a standard error of 0 to 0.47), while parent-offspring pairs shared an appropriate proportion of alleles IBD (mean of 1 allele shared IBD with a standard error of 0 to 0.3) as did full sibling pairs (mean of 1 allele shared IBD with a standard error of 0.6 to 0.8) (Figure 40).

Following quality control measures, 23 families comprising 117 individuals remained and were included in the final analysis. Of these, 50 were cases (29 females, 21 males) and 67 were controls (34 females, 33 males).

	All families linked (no heterogeneity)	75% of families linked (25% heterogeneity)	50% of families linked (50% heterogeneity)	25% of families linked (75% heterogeneity)
Power to detect HLOD >3	99%	77%	30%	1%
Power to detect HLOD >2	100%	98%	52%	13%
Power to detect HLOD >1	100%	100%	81%	35%

Table 18: Linkage study power estimates generated from the SLINK program, assuming differing levels of heterogeneity between the families.

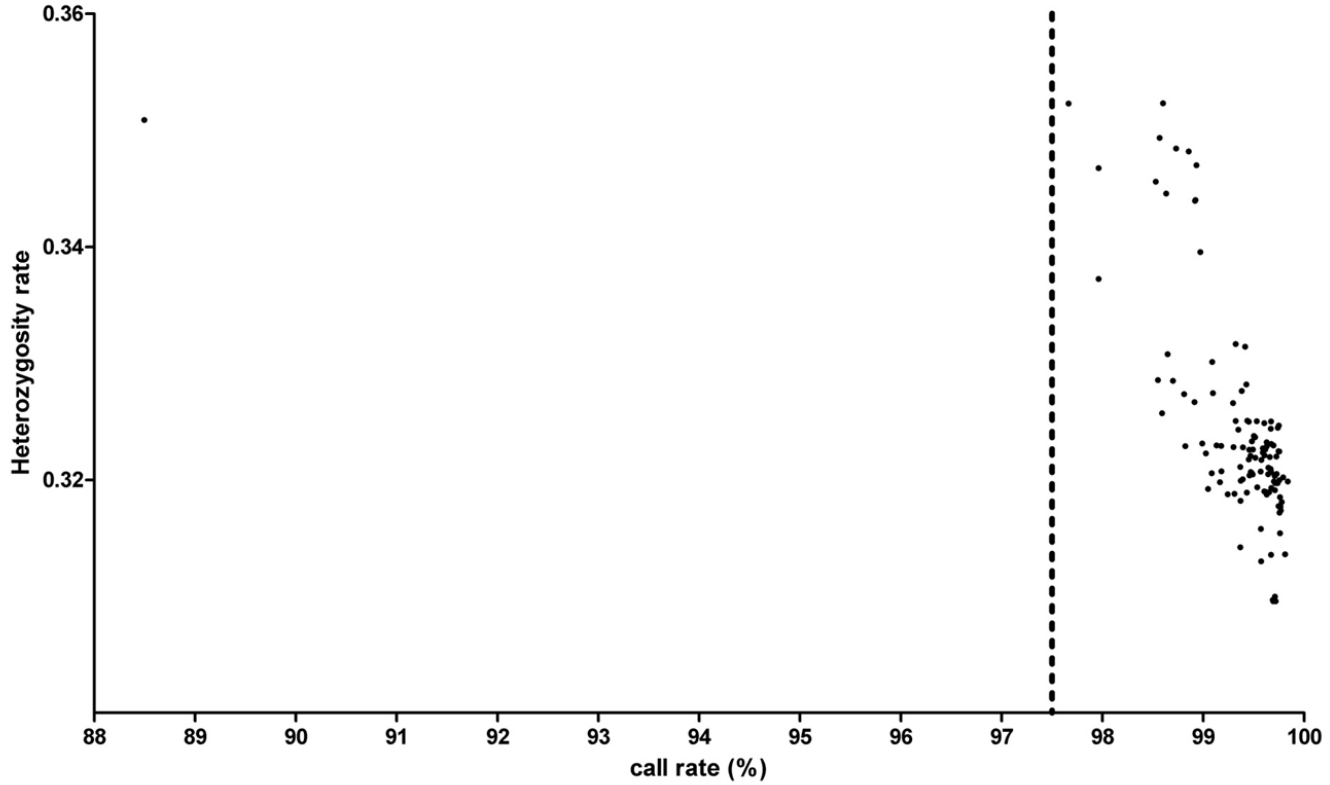


Figure 38: Heterozygosity and genotyping call rates for each individual genotyped for the linkage study in AAD.

Heterozygosity rate is shown on the y axis and the genotyping call rate on the x axis for each family member genotyped. The individual with a call rate of less than 89% was excluded from the analysis.

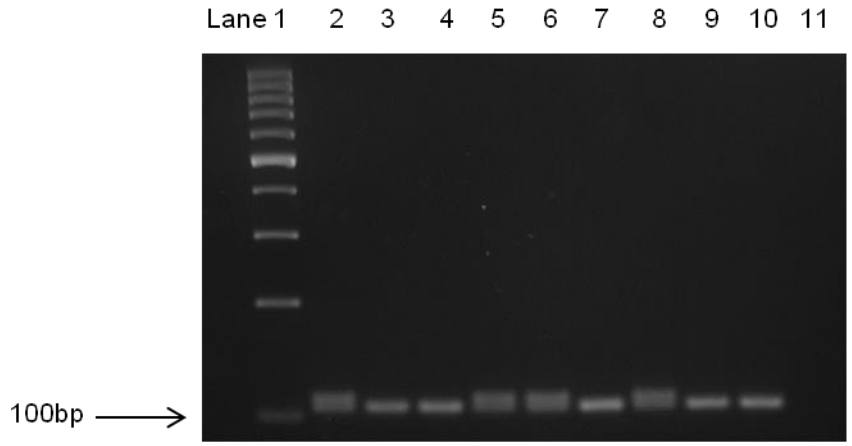


Figure 39: Amelogenin sex differentiation PCR assay gel image.

4% agarose gel electrophoresis of PCR products from the amelogenin assay. A 100bp ladder is shown in lane 1 and a no template control, run as a negative control, is shown in lane 11. Positive controls of known gender are shown in lane 2 (a male sample with two bands close together of 106 and 112bp) and lane 3 (a female sample with a single 106bp band). Lanes 4 to 10 are samples who had miss-assigned genders (gender by PCR: female, male, male, female, male, female, female respectively).

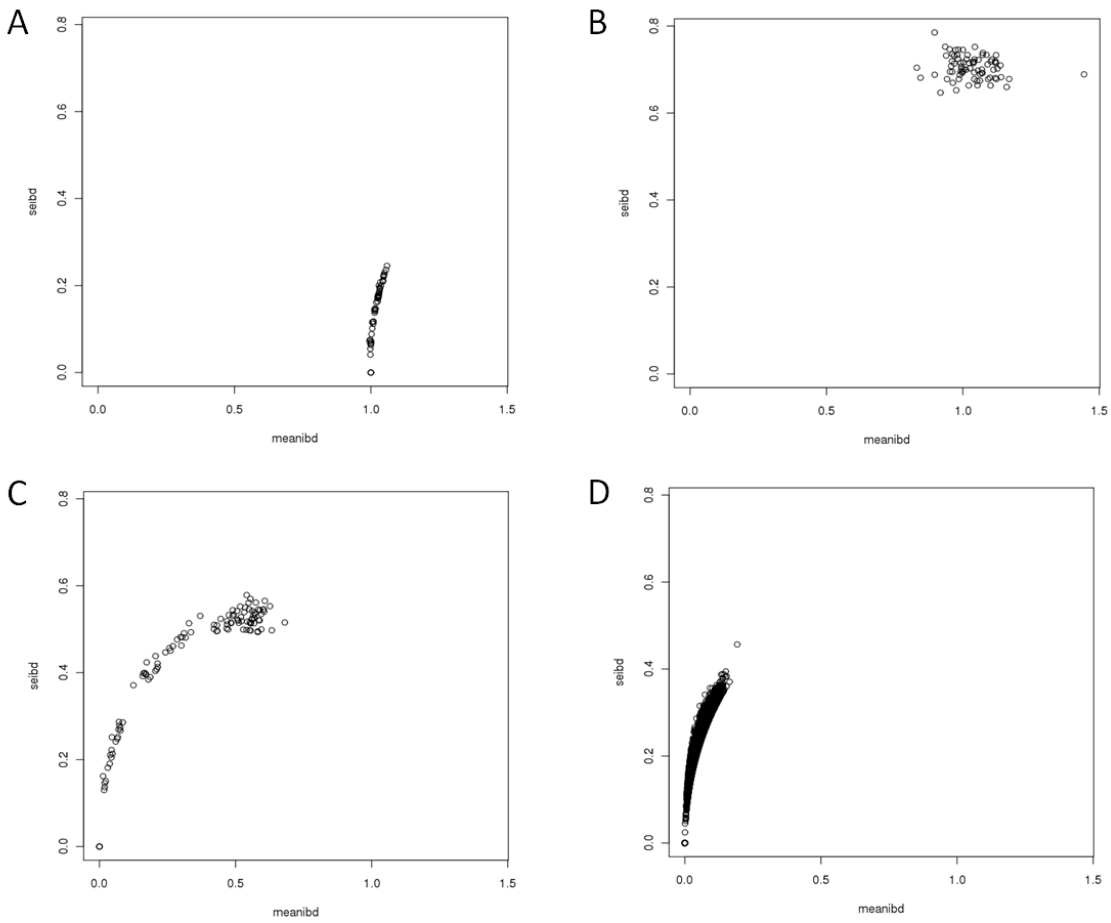


Figure 40: Allele sharing identical by descent (IBD) among individuals within the multiplex AAD family pedigrees.

Graphs demonstrating IBD allele sharing among individuals within the AAD pedigrees. The mean number of alleles shared IBD is shown on the x axis and the standard error is shown on the y axis. Panel A demonstrates that parent-offspring pairs share 1 allele IBD with a low standard error. Panel B shows that full sibling pairs share 1 allele IBD but with a higher standard error as 0, 1 or 2 alleles can be shared IBD. Panel C shows that between pairs in other relationships e.g. cousins *etc*, the degree of allele sharing IBD is lower. Panel D shows that unrelated individuals share on average 0 alleles IBD.

5.4.3 LINKAGE ANALYSIS RESULTS – MARKER MAP INFORMATION CONTENT

Linkage analysis was performed using three marker maps of varying density. A dense map of 36,775 markers provided excellent information content but produced large files that were difficult to manipulate. Using a marker map of two SNPs per cM (7429 SNPs) meant that some information content was lost. Using a map of four SNPs per cM (14,771 SNPs) gave a good level of information content, between 80 and 100% per chromosome, while generating manageable files (Figure 41). All SNP positions are derived from the Ensembl database (GRCh37)[215].

5.4.4 LINKAGE ANALYSIS RESULTS – AAD TAKEN AS THE TRAIT OF INTEREST

5.4.4.1 PARAMETRIC LINKAGE ANALYSIS OF THE AUTOSOMES

Applying a rare dominant model, assuming a disease allele frequency in the population of 1 in 10,000 (0.0001) and assuming a disease penetrance of 0.001 (0.1%) if 0 risk alleles are present and 0.999 (99.9%) if 1 or 2 risk alleles are present, three loci on chromosomes 18, 9 and 7, had LOD scores of greater than 2.0 (Figure 42).

The maximum LOD score was observed within a linkage peak on chromosome 18, between 116.5 and 121.9cM (75241668 – 77950543bp). Within this peak, a maximum LOD and HLOD score of 3.00 was seen at marker *SNP_A-8291421* (*rs1113678*, 76554812bp).

On chromosome 9, a linkage peak was observed between 36.0 and 40.4cM (17486802 – 19751149bp), with a maximum LOD and HLOD score within this peak of 2.90 at marker *SNP_A-1996138* (*rs10123624*, 19025385bp).

On chromosome 7, a maximum LOD and HLOD of 2.88 was seen at marker *SNP_A-4232044* (*rs10263367*) at position 70082089bp within a linkage peak

spanning 82.4 – 86.2cM (70020160 – 73809454bp). A second, smaller linkage peak between 69.8 and 71.7cM (47565504 – 49876993bp) was also seen on this chromosome. Within this peak, a maximum HLOD of 2.09 at marker *SNP_A-2279338* (*rs13228770,* 49457067bp) was observed (estimated proportion of linked families (α) 0.66). At this locus, the maximum LOD score was 1.28.

When a rare co-dominant model was applied, assuming a disease allele frequency in the population of 1 in 10,000 (0.0001) and assuming a disease penetrance of 0.001 (0.1%) if 0 risk alleles are present, 0.75 (75%) if 1 risk allele is present and 0.999 (99.9%) if 2 alleles are present, two loci had an HLOD score of greater than 2.0 (Figure 43). On chromosome 2, a linkage peak between 32.4 and 35.5cM (12675092 – 15457535bp) was seen. The maximum HLOD within this peak was 2.57 (α 0.80, LOD 1.43) at marker *SNP_A-8683599* (*rs2380452*, 12897058bp). On chromosome 9, a small linkage peak was seen between 145.5 and 146.6cM (135482340 – 136043697bp) with a maximum HLOD of 2.08 (α 0.75, LOD -0.97) at *SNP_A-2265585*, *rs10901207* at 135618325bp.

When a rare recessive model was applied, assuming a disease allele frequency in the population of 1 in 10,000 (0.0001) and assuming a disease penetrance of 0.001 (0.1%) if 0 or 1 risk alleles are present and 0.999 (99.9%) if 2 risk alleles are present, no loci had a LOD score of greater than 2.0.

5.4.4.2 NON-PARAMETRIC LINKAGE ANALYSIS OF THE AUTOSOMES

In a non-parametric analysis, which effectively excludes the parent-offspring pairs (n=5), one locus on chromosome 6 had a LOD score of greater than 3.0 (Figure 44). Here, a large linkage peak was seen from 46.0 to 55.4cM (22375648 – 35968100bp). The maximum LOD score, applying the Kong and Cox exponential model searching for a large increase in allele sharing in a small number of families, was 3.01 at 51.5cM (*SNP_A-1923640*, *rs2072633* at 31919578bp). At this locus, the linear LOD score (designed to identify small increases in allele sharing spread across a large number of families) was 3.13.

5.4.4.3 PARAMETRIC AND NON-PARAMETRIC LINKAGE ANALYSIS OF THE X CHROMOSOME

The X chromosome was analysed, using the same parametric models as above and by non-parametric analysis. No loci had a LOD score of greater than 1.

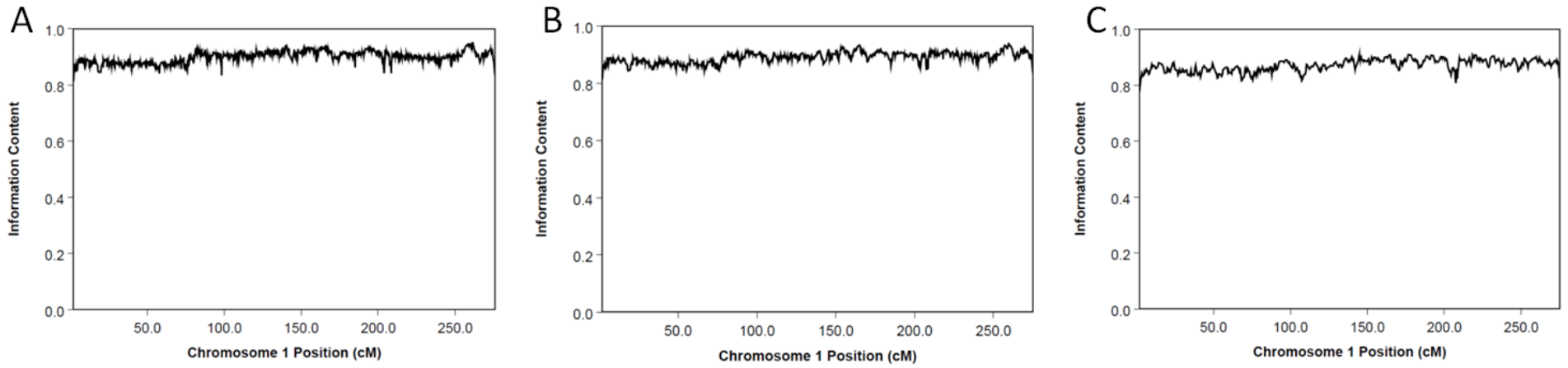


Figure 41: Information content in the linkage analysis using three marker maps of differing densities.

Information content graphs for chromosome 1 using an unthinned marker map of 36,775 SNPs (panel A) compared to a marker map thinned to four SNPs per cM, comprising 14,771 SNPs (panel B) and to two SNPs per cM, comprising 7429 SNPs (panel C). An information content above 0.8 (80%) indicates sufficient coverage. There is little difference in information content gained when using a dense marker map compared to a thinned marker map.

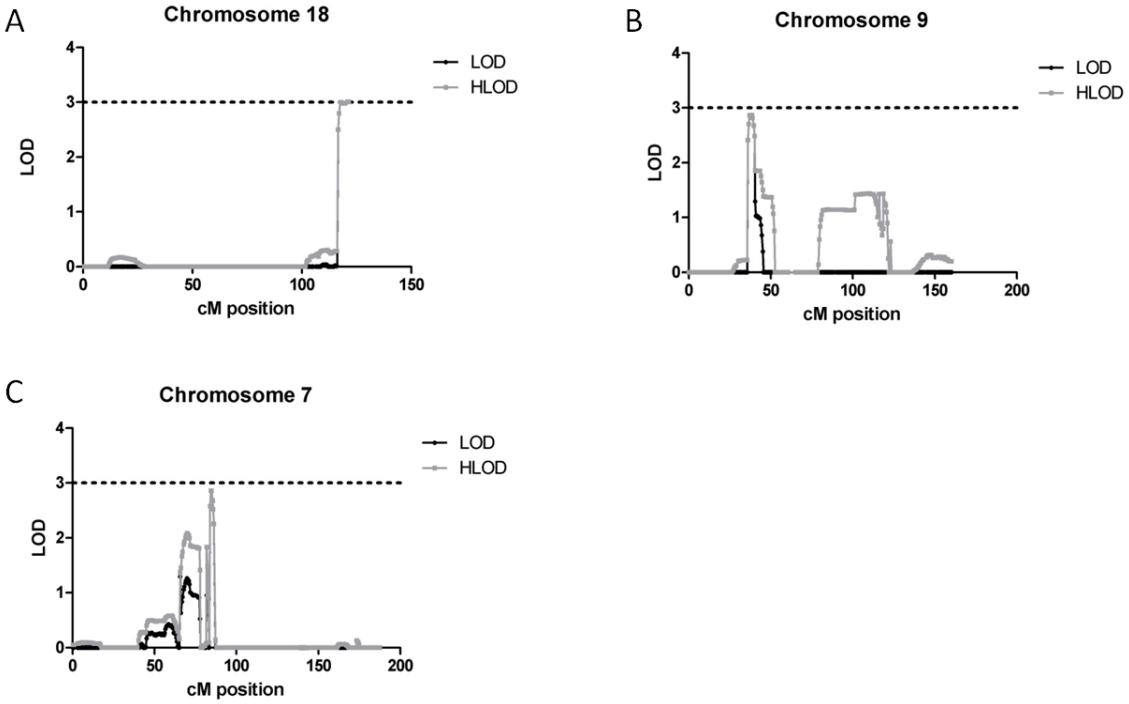


Figure 42: Graphical representation of parametric linkage results in the multiplex AAD families assuming a rare dominant model.

Graphs showing LOD (black lines) and HLOD (grey lines) scores generated in a parametric linkage analysis, assuming a rare dominant model. The LOD/HLOD score is on the y axis and the cM position is on the x axis. The dotted black line shows the LOD/HLOD threshold of 3.0, taken as convincing evidence of linkage. Linkage peaks of LOD/HLOD greater than 2.0 were observed on chromosomes 18 (panel A), 9 (panel B) and 7 (panel C). An HLOD score of 3.0 was observed on chromosome 18.

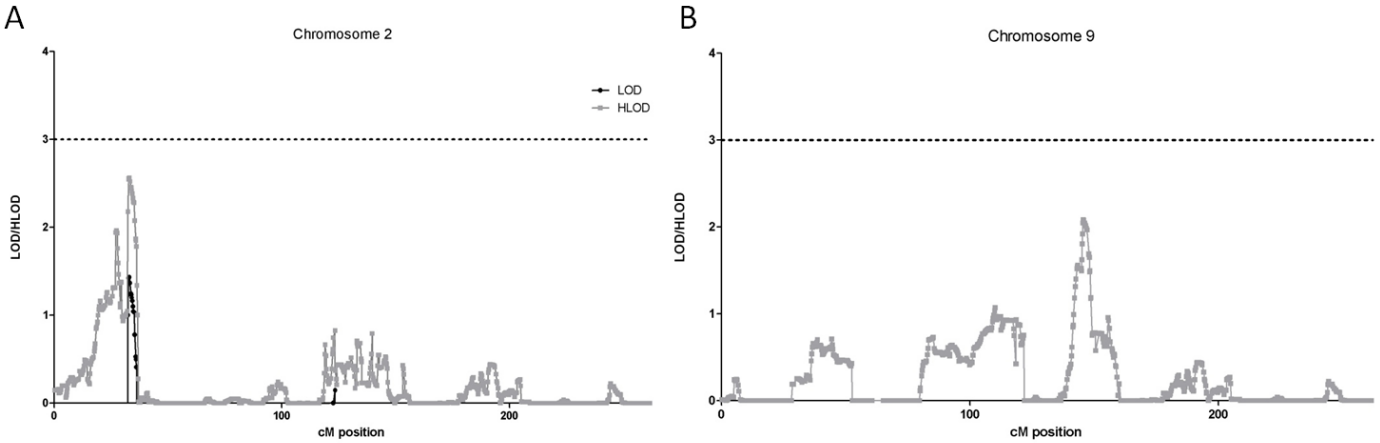


Figure 43: Graphical representation of parametric linkage results in the multiplex AAD families assuming a rare co-dominant model.

Graphs showing LOD (black lines) and HLOD (grey lines) scores generated in a parametric linkage analysis, assuming a rare co-dominant model. The LOD/HLOD score is shown on the y axis and the cM position is on the x axis. The dotted black line shows the LOD/HLOD threshold of 3.0, taken as convincing evidence of linkage. Linkage peaks of LOD/HLOD greater than 2.0 were observed on chromosomes 2 (panel A) and 9 (panel B). No linkage peaks with a LOD/HLOD of 3.0 or greater were seen using this model.

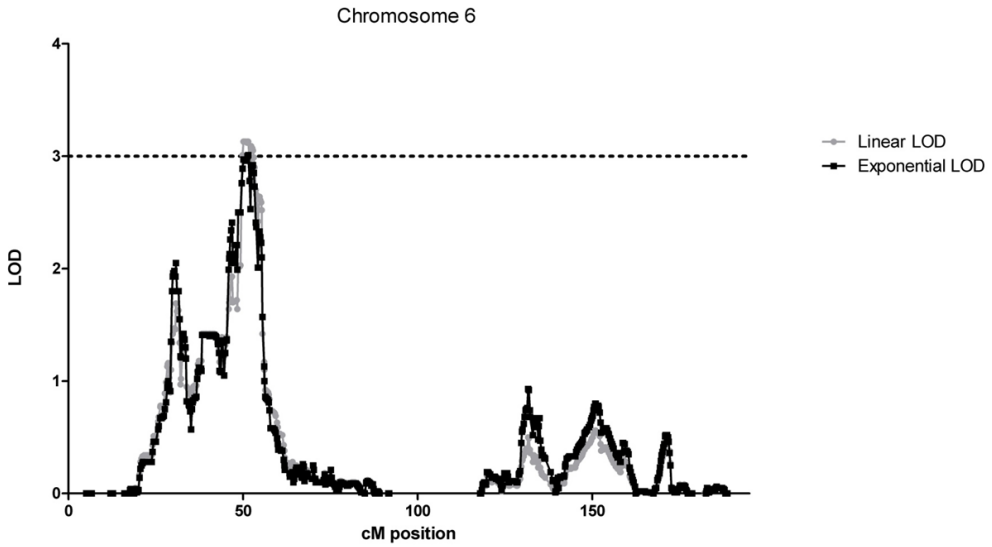


Figure 44: Graphical representation of non-parametric linkage results in the multiplex AAD families.

Graph showing exponential LOD (black lines) and linear LOD (grey lines) scores generated in a non-parametric linkage analysis. The LOD score is on the y axis and the cM position is on the x axis. The dotted black line shows the LOD threshold of 3.0, taken as convincing evidence of linkage. A linkage peak of LOD greater than 3.0 was observed on chromosome 6 only.

5.4.5 LINKAGE ANALYSIS RESULTS – 21OH AUTOANTIBODY POSITIVITY AS THE TRAIT OF INTEREST

5.4.5.1 PARAMETRIC LINKAGE ANALYSIS OF THE AUTOSOMES

Applying a rare dominant model, and using a four SNP per cM marker map, three loci on chromosomes 3, 18 and 1 had LOD or HLOD scores of 2 or more (Figure 45). On chromosome 3, a maximum HLOD of 3.25 (α 0.68, LOD 1.02) was seen at 19.1cM (*SNP_A-1839084*, *rs1948153*, 6481324bp), in a large linkage peak spanning 0.4 to 19.39cM (98655 – 6725175bp). A linkage peak was also observed on chromosome 18, spanning 117.4 to 120.8cM (75814302 – 76966335bp) with a max HLOD of 2.14 at 119.6cM (α 0.50, LOD -3.14, *SNP_A-8291421*, *rs1113678*, 76554812bp). Additionally, a smaller linkage peak was also seen on chromosome 1, spanning 32.80 to 35.80cM (18346165 – 19049946bp) with a maximum HLOD of 2.06 (α 0.69, LOD 1.10) observed at 35.3cM (*SNP_A-2304622*, *rs1934057*, 18962095bp).

Applying a rare co-dominant model, linkage peaks were observed in the same regions of chromosomes 3 and 18 as seen with the dominant model. On chromosome 3, a linkage peak was observed at 14.3 to 19.4cM (4665565 – 6725175bp), with a maximum HLOD of 2.22 at 19.10cM (*SNP_A-1839084*, *rs1948153*, 6481324bp, α 0.69, LOD 1.44). In addition, a narrower peak was observed at 32.4 to 33.7cM (13720097 – 14628404bp) with a maximum HLOD and LOD of 3.12 observed at 33.4cM (*SNP_A-8468450*, *rs4402920*, 14478231bp). On chromosome 18, a linkage peak was seen between 117.4 and 119.6cM, with a maximum HLOD of 2.13 observed at 117.5cM (*SNP_A-8660306*, *rs9946731*, 75851340bp, α 0.57, LOD -2.46).

Applying a rare recessive model, no loci had a LOD or HLOD score of greater than 2.

5.4.5.2 NON-PARAMETRIC LINKAGE ANALYSIS OF THE AUTOSOMES

In a non-parametric linkage analysis, two narrow linkage peaks were observed on chromosome 3 (Figure 46). The first was seen between 32.8 and 33.7cM (14180552 – 14628404bp), with a maximum exponential LOD of 2.52 at 33.39cM (*SNP_A-8468450*, *rs4402920*, 14478231bp). The linear LOD here was 2.06. The second peak was observed between 51.8 and 52.1cM (28700022 – 28887146bp) where a maximum exponential LOD of 2.38 was seen at 51.92cM (*SNP_A-8447998*, *rs2888033*, 28760544bp). Here, the linear LOD was 2.28.

5.4.5.3 PARAMETRIC AND NON-PARAMETRIC LINKAGE ANALYSIS OF THE X CHROMOSOME

The X chromosome was analysed, using the same parametric models as above and by non-parametric linkage analysis. No loci had a LOD score of greater than 1.

Following linkage analysis, the Ensembl[215] and HapMap[102] databases were used to look for plausible candidate genes located within the regions of linkage.

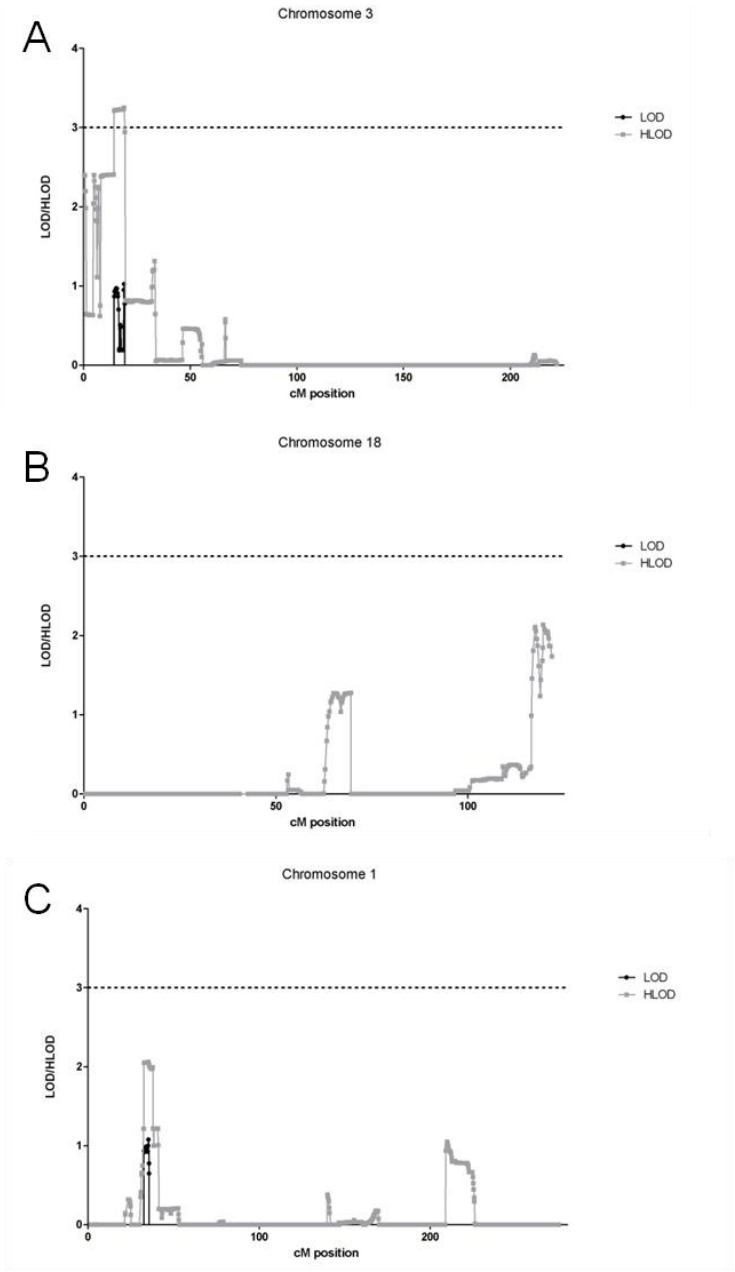


Figure 45: Graphical representation of parametric linkage results, taking 21OH status as the trait, assuming a rare dominant model.

Graphs showing LOD (black lines) and HLOD (grey lines) scores generated in a parametric linkage analysis taking 21OH autoantibody positive status as the trait, assuming a rare dominant model. The LOD/HLOD score is on the y axis and the cM position is on the x axis. The dotted black line shows the LOD/HLOD threshold of 3.0, taken as convincing evidence of linkage. Linkage peaks of LOD/HLOD greater than 2.0 were observed on chromosomes 3 (panel A), 18 (panel B) and 1 (panel C). The HLOD score for the linkage peak on chromosome 3 exceeded 3.0.

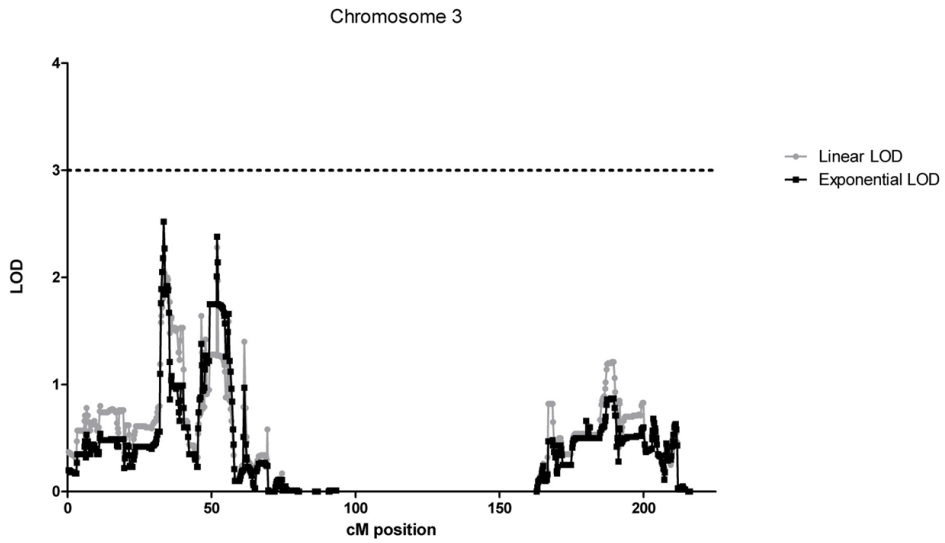


Figure 46: Graphical representation of non-parametric linkage results, taking 21OH status as the trait.

Graph showing Exponential LOD (black lines) and linear LOD (grey lines) scores generated in a non-parametric linkage analysis taking 21OH autoantibody positivity as the trait. The LOD score is on the y axis and the cM position is on the x axis. The dotted black line shows the LOD threshold of 3.0, taken as convincing evidence of linkage. A linkage peak of LOD greater than 2.0 was observed on chromosome 3 only.

5.5 RESULTS – ASSOCIATION ANALYSIS IN MULTIPLEX AAD FAMILIES

5.5.1 ASSOCIATION STUDY POWER

A power calculation was performed for both association analyses planned using the QUANTO program[442]. In the case of 50 AAD cases versus 67 unaffected relatives, assuming a MAF of 0.3 and an α of 0.05, the study has 71% power to detect a locus with an odds ratio of 2.0 and 22% power to detect a locus with a more modest odds ratio of 1.4. Under the same assumptions, if a genome-wide α of $5x10^{-7}$ is assumed to allow for multiple testing, the study has no power to detect a locus with an odds ratio of 2.0 or less, but has 9% and 36% power to detect loci with an odds ratio of 3.0 and 4.0 respectively.

For the second analysis, comparing the 50 AAD cases with 2706 1958 birth cohort controls under the same assumptions (MAF 0.3, α 0.05), the study has 92% power to detect a locus with an odds ratio of 2.0 and 36% power to detect a locus with a more modest odds ratio of 1.4. Again, under the same assumptions but applying a genome wide α of $5x10^{-7}$, the study has only 2% power to detect a locus with an odds ratio of 2.0 but has 52% and 92% power to detect loci with odds ratios of 3.0 and 4.0 respectively.

5.5.2 ASSOCIATION ANALYSIS QUALITY CONTROL

Association analyses were performed using the EMMAX program[246]. For these analyses, SNPs were excluded if they had a call rate of <99% or a minor allele frequency of <5%. SNPs were also excluded if they were out of HWE in the control cohort. For this, two thresholds were applied: a stringent threshold ($P < 0.01$) gave a final data set of 592,843 SNPs in total while the less stringent threshold ($P < 1.0 \times 10^{-8}$) gave a final data set of 595,118 SNPs.

5.5.3 ASSOCIATION ANALYSIS RESULTS – 50 AAD FAMILY CASES VERSUS 67 CONTROLS

An initial comparison of the results revealed no difference in the most associated SNPs with both HWE thresholds, therefore the results with the threshold of $P < 1.0 \times 10^{-8}$ are presented. R was used to generate a quantile-quantile (QQ plot) (Figure 47), where observed statistics ($-\log_{10}$ P value generated in EMMAX) were plotted against the expected values under the null hypothesis. This plot revealed that the observed results were significantly deflated compared to the expected results, but still followed a linear pattern. I hypothesise that this is because of the structure of the population tested. EMMAX accounts for relatedness within populations and expects that those individuals sharing a phenotype i.e. cases, will be genetically more similar than those who are discordant for a phenotype. In this study design, those discordant for the phenotype i.e. cases and controls, are more genetically similar than would be expected as they come from many small pedigrees. Therefore, the population tested in this case does not fit the model assumed by EMMAX. This has resulted in a deflation of the statistical results which should be considered in the interpretation of the results.

The threshold used to denote genome-wide significance in large genome-wide association studies is $P < 0.0000005$ (5×10^{-7}). The association study in the AAD families did not reveal any SNPs that met the threshold for genome-wide significance (Figure 48), even accounting for the deflation of the results. Due to the lack of study power and the weakness of the associations detected in this analysis, the results were disregarded.

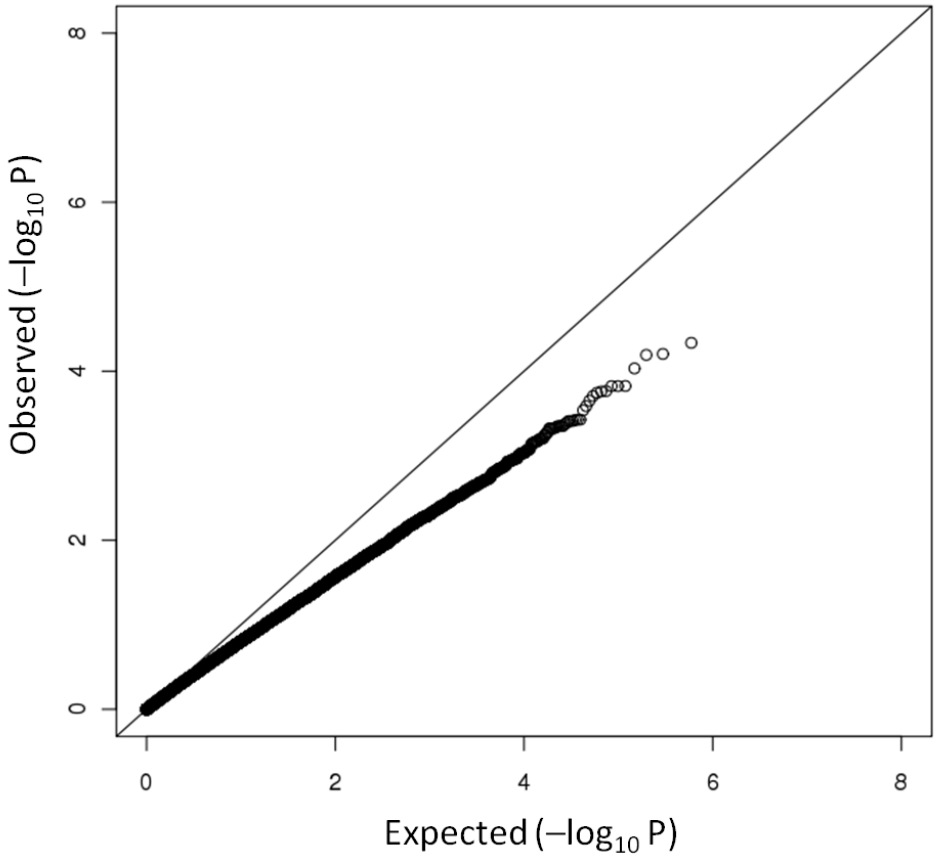


Figure 47: Quantile-quantile (QQ) plot for an association analysis comparing 50 multiplex AAD family cases versus 67 family controls.

QQ plot of ordered observed statistics (-$\log_{10}$ P value) on the y axis versus expected values under the null hypothesis on the x axis. The line which intercepts the x and y axes at zero represents the relationship between expected and observed values under the null hypothesis. The observed results are deflated compared to the expected results, as the data do not fit the model assumed by the software program used, EMMAX.

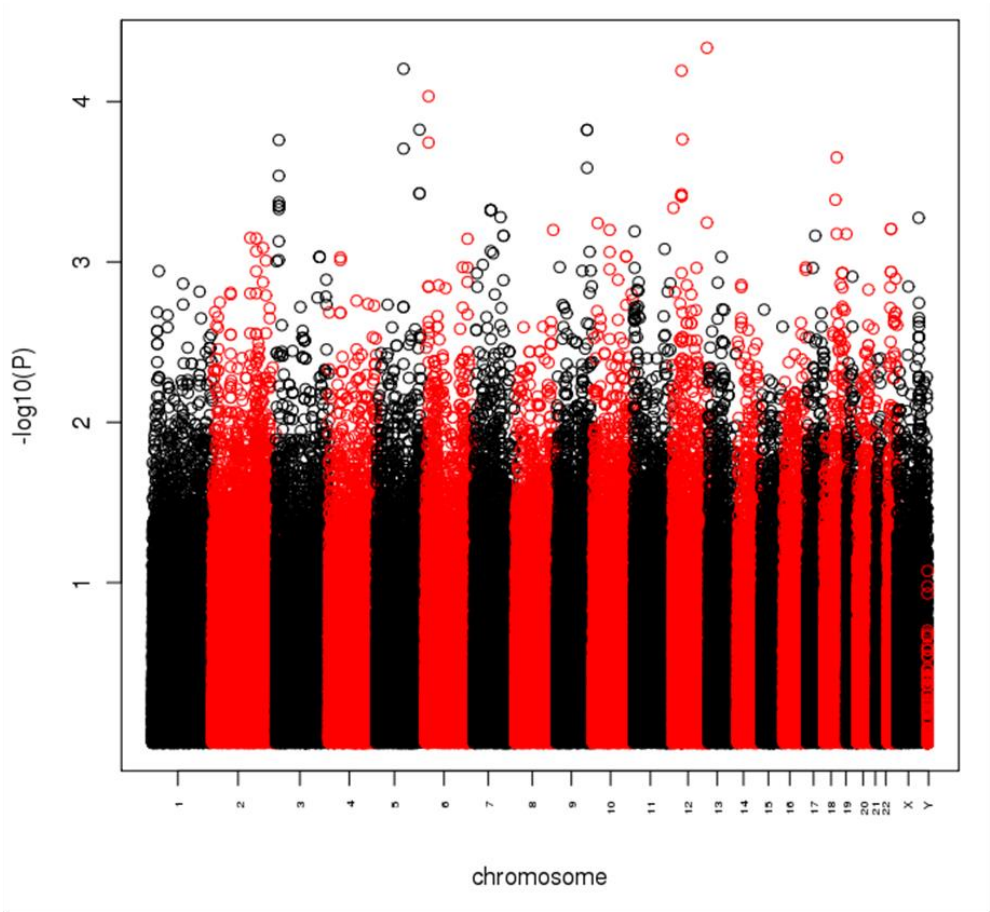


Figure 48: Manhattan plot showing results from an association analysis of 50 AAD family cases versus 67 family controls.

In this Manhattan plot, the chromosome position is shown on the x axis. The y axis shows the $-\log_{10}$ of the P value; –log10(P). Each SNP is represented by a coloured circle. In a genome-wide study such as this, a –log10(P) of seven, equivalent to a P value of 0.0000001, would be considered statistically significant. In this analysis, no SNPs met this threshold of significance.

5.5.4 ASSOCIATION ANALYSIS RESULTS – 50 AAD FAMILY CASES VERSUS 1958 BIRTH COHORT CONTROLS

EMMAX was used to perform an association analysis between the 50 affected family members acting as cases and 2706 controls (1396 males, 1310 females) from the 1958 UK birth cohort available from the WTCCC. 551,634 SNPs met the quality control criteria and were used for the analysis. The QQ plot, drawn in R (Figure 49), did not demonstrate significant deflation. This is due to the fact that, in this analysis, the population studied does fit the model assumed by EMMAX. The cases are from many small families and so are genetically more similar compared to unrelated people. Furthermore, the cases are unrelated to the WTCCC controls. Therefore, the groups that are discordant for the phenotype are more genetically dissimilar, as fits the EMMAX assumption.

17 SNPs were associated with a P value of less than 5×10^{-7} and these are shown in Table 19. This is graphically represented by the Manhattan plot (Figure 50) and regional data was visualised using LocusZoom[247]. A cluster of five associated intergenic SNPs in tight LD (r^2>0.8), spanning 228kb on chromosome 2 was seen (Figure 51), with maximal association at *rs10495950* (P 9.7×10^{-8}). In addition, a pair of SNPs on chromosome 6 were also associated (Figure 52). These were in the HLA region, a region of extended LD, and a number of associated SNPs were noted here, all in moderate to significant LD with the most associated SNP in this analysis, *rs2187668*. As clusters of SNPs reaching, or almost reaching, genome-wide statistical significance were seen in these two regions, these results were thought to represent true association.

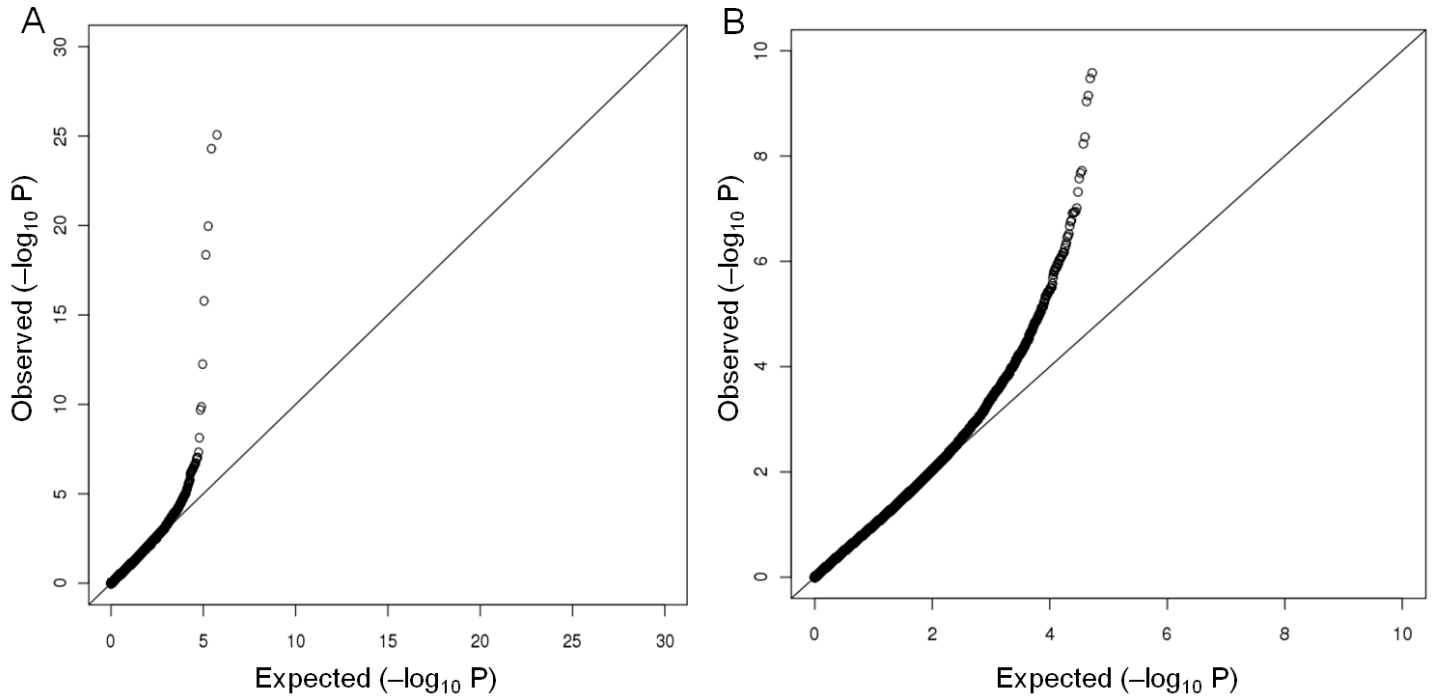


Figure 49: Quantile-quantile (QQ) plots for an association analysis comparing 50 AAD family cases versus 2706 WTCCC 1958 birth cohort controls.

QQ plots of ordered observed statistics (-$\log_{10}$ P value) on the y axis versus expected values under the null hypothesis on the x axis. The line which intercepts the x and y axes at zero represents the relationship between expected and observed values under the null hypothesis. Panel A represents the QQ plot for all data. Panel B represents the QQ plot with the six most outlying results removed.

Chr	rs ID	Position (HapMap)	EMMAX P value	Position
2	rs11681243	48104650	1.7×10^{-7}	Intergenic
2	rs7561696	48157345	1.8×10^{-7}	Intergenic
2	rs13431982	48286415	3.3×10^{-7}	Intergenic
2	rs10495950	48288258	9.7×10^{-8}	Intergenic
2	rs17325209	48332704	2.1×10^{-7}	Intergenic
2	rs1406242	151151406	5.9×10^{-20}	Intergenic
3	rs7632505	124220987	4.3×10^{-9}	SEMA5B (semaphorin 5B)
3	rs16840699	134920281	2.6×10^{-10}	Intergenic
6	rs1150753	32167835	4.4×10^{-7}	TNXB (Tenascin B isoform 1)
6	rs2187668	32713852	3.0×10^{-7}	HLA-DQA1
6	rs15680	58380385	1.5×10^{-12}	Glucuronidase beta-like 2)
8	rs1426192	21392622	3.4×10^{-10}	Intergenic
13	rs873294	109965408	5.4×10^{-24}	Intergenic
14	rs8007744	27399226	1.1×10^{-18}	Intergenic
18	rs7229302	11498894	9.1×10^{-16}	Intergenic
21	rs3787764	37154098	6.2×10^{-25}	HLCS (Holocarboxylase synthetase)

Table 19: Association analysis results between 50 AAD family cases and 2706 1958 birth cohort controls.

All SNPs with a P value less than 5×10^{-7} are shown in chromosome order. The final column shows the SNPs in relation to genes and non-coding RNAs. Intergenic SNPs are those not in genes or non-coding RNAs.

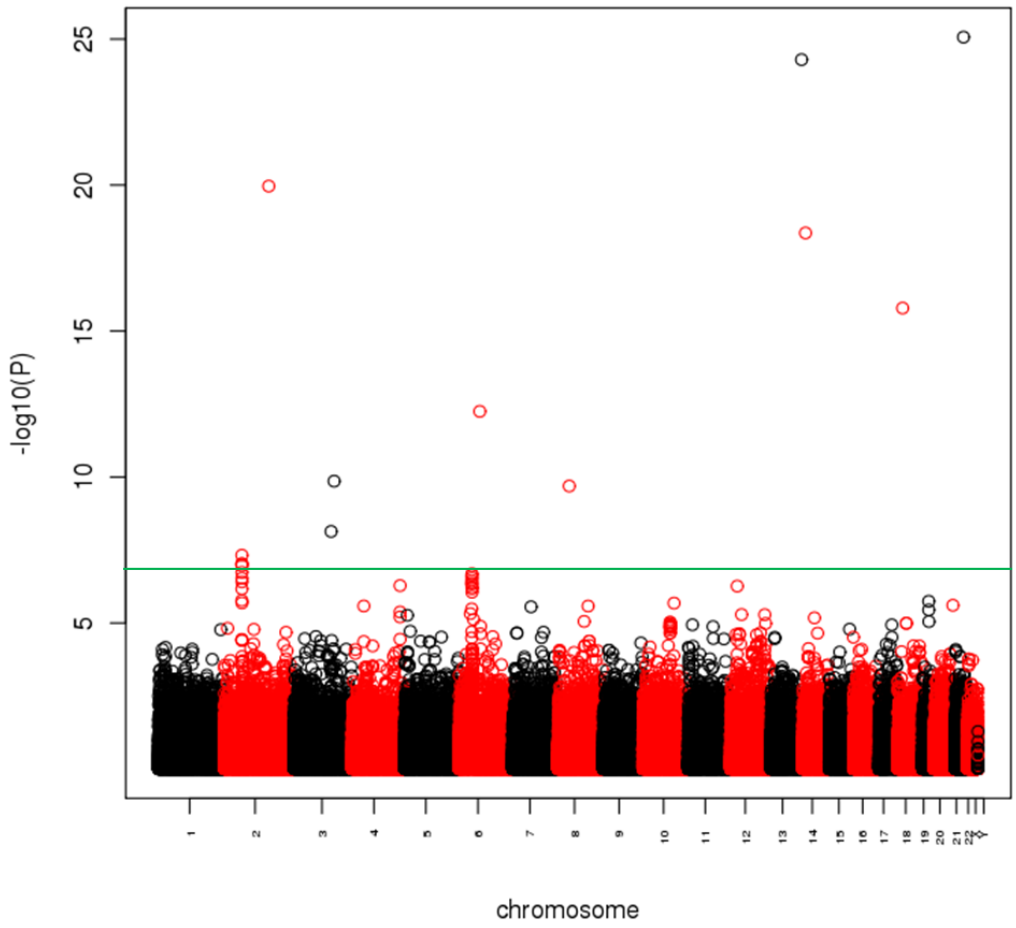


Figure 50: Manhattan plot showing results from an association analysis of 50 AAD family cases versus 2706 WTCCC 1958 birth cohort controls.

In this Manhattan plot, the chromosome position is shown on the x axis. The y axis shows the $-\log_{10}$ P value; –log10(P). Each SNP is represented by a coloured circle. In a genome-wide study such as this, a –log10(P) of seven, equivalent to a P value of 0.0000001, would be considered statistically significant (shown by the green line). In this analysis, a number of SNPs exceed this threshold, including a cluster of SNPs on chromosome 2 and a pair on chromosome 6.

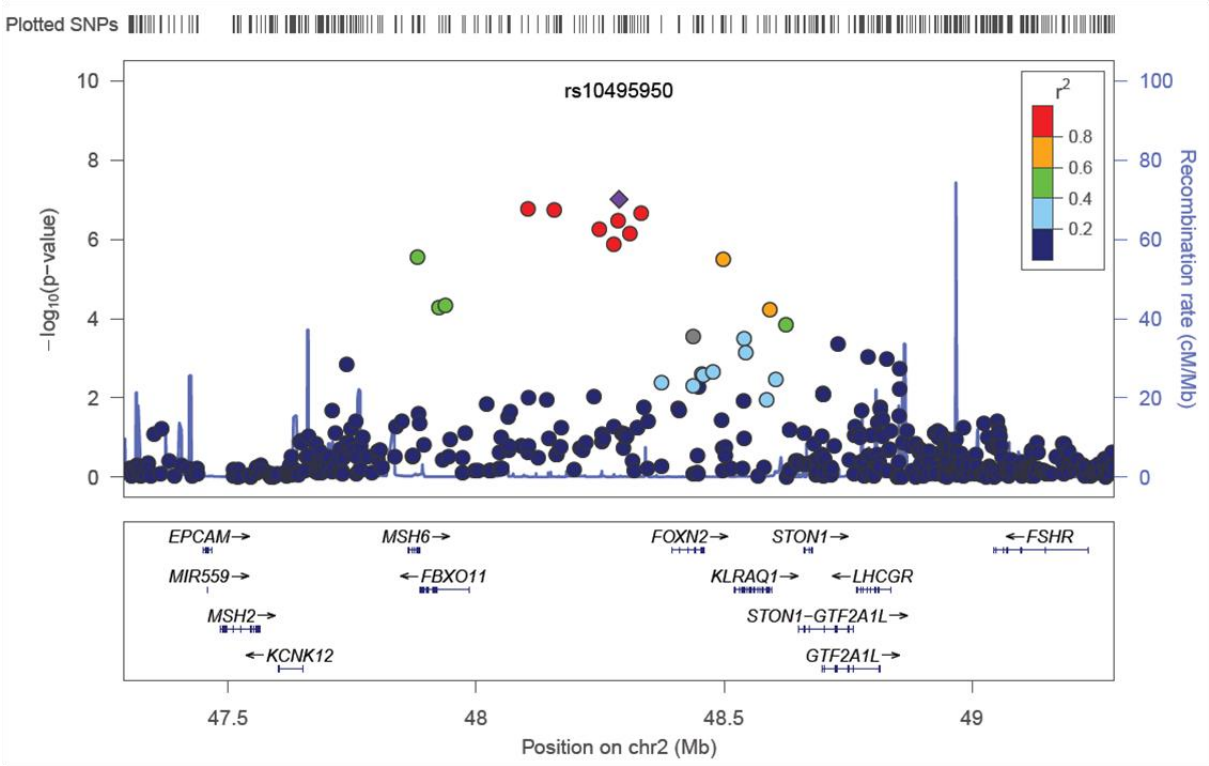


Figure 51: Associated SNPs on chromosome 2 in 50 multiplex AAD cases compared to 2706 WTCCC 1958 birth cohort controls.

LocusZoom plot showing a region on chromosome 2 containing an associated SNP, *rs10495950*, denoted by the purple diamond. This plot shows that this SNP is significantly associated with AAD in an analysis of 50 multiplex AAD cases and 2706 healthy controls, as are a cluster of other SNPs in the same intergenic region which are in significant LD with *rs10495950* (r^2 >0.80). These SNPs are in a region of low recombination and LD extends to SNPs in neighbouring genes *MSH6*, *FBXO11, FOXN2, KLRAQ1* and *STON1.*

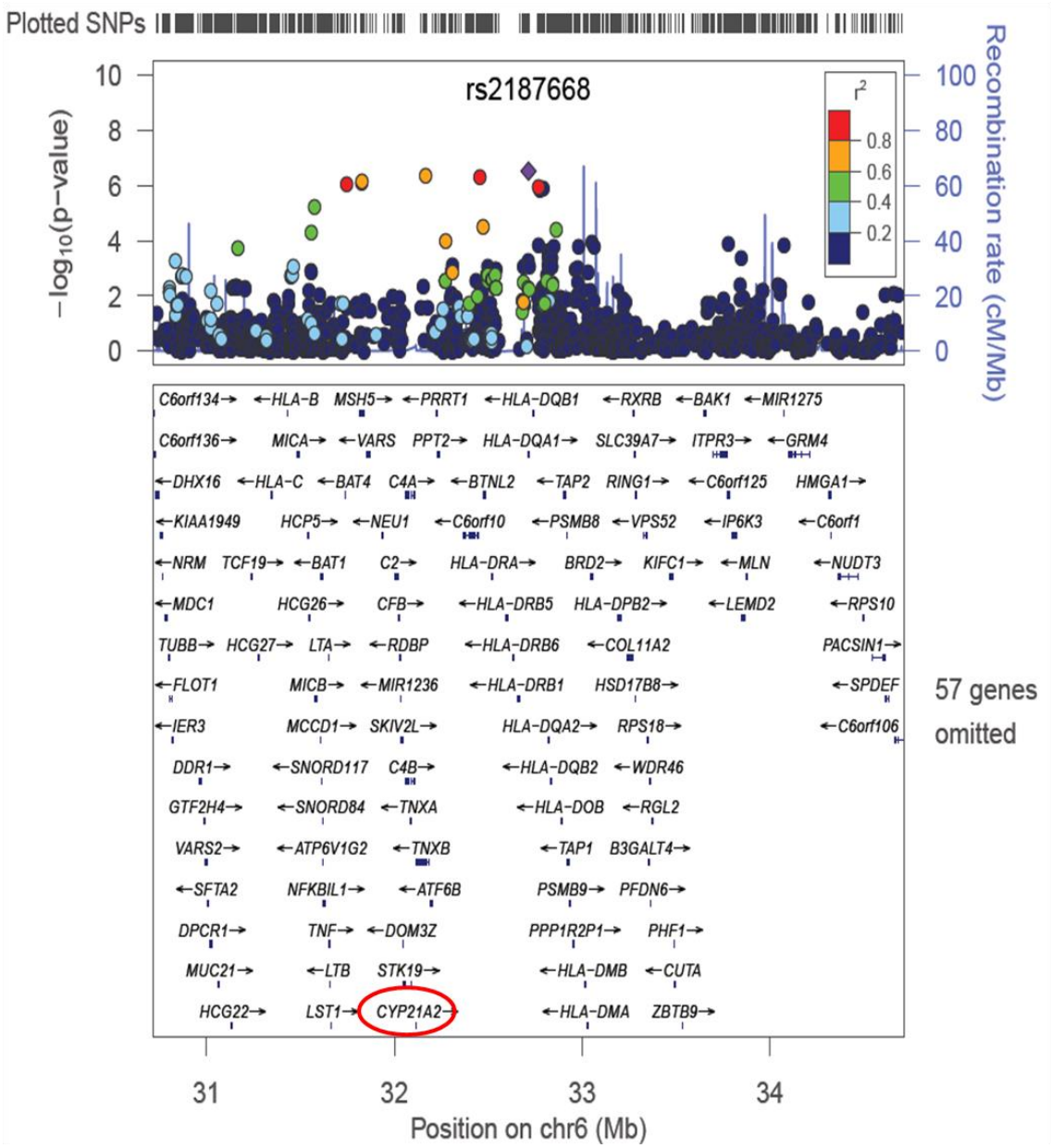


Figure 52: Associated SNPs on chromosome 6 in 50 multiplex AAD cases compared to 2706 WTCCC 1958 birth cohort controls.

LocusZoom plot showing the HLA region on chromosome 6 containing an associated SNP, *rs2187668*, denoted by the purple diamond. The plot shows that this SNP is significantly associated with AAD in an analysis of 50 multiplex AAD cases and 2706 healthy controls, as are a cluster of other SNPs in the same region which are in moderate to significant LD with *rs2187668* (r^2 >0.40). These SNPs are in a region of low recombination which is very gene-dense (57 genes are omitted from the bottom panel) and includes the *CYP21A2* gene (indicated by the red circle) which encodes the steroid 21OH enzyme, a key component of the steroidogenesis pathway. LD extends to SNPs in neighbouring genes.

The linkage and association data were then compared to determine whether any of the most associated SNPs were located within the linkage peaks found (Table 20). There was overlap between the linkage peak on chromosome 6 generated from the non-parametric analysis, taking AAD as the trait, with a pair of associated SNPs on chromosome 6 found in the genome-wide association analysis between the 50 AAD family cases and 2706 WTCCC 1958 birth cohort controls (Table 20). These regions both correspond to HLA, already known to be associated with AAD. This finding validates the approaches taken. The other region of interest found in the 50 AAD family cases and 2706 WTCCC 1958 birth cohort controls association study, on chromosome 2, did not correspond with the linkage peak on chromosome 2 generated from the parametric linkage analysis using a co-dominant model. The most associated marker on chromosome 2 in the association study was some 35Mb upstream of the most associated marker in the linkage analysis. Therefore, the association analysis did not help to focus the areas of interest generated by the linkage analyses.

Study	Trait of interest	Analysis	Chromosomal segment linked/associated	Most linked/associated marker (position)	Maximum LOD/HLOD or P value
Linkage study	AAD	Non-parametric	Chr 6: 22.4-36.0 Mb	*rs2072633* (32.0 Mb)	Linear LOD 3.13/exponential LOD 3.01
		Parametric, rare dominant	Chr 18: 75.2-78.0 Mb	*rs1113678* (76.6 Mb)	LOD/HLOD 3.00
			Chr 9: 17.5-19.8 Mb	*rs10123624* (19.0 Mb)	LOD/HLOD 2.90
			Chr 7: 70.0-73.8 Mb	*rs10263367* (70.1 Mb)	LOD/HLOD 2.88
		Parametric, rare co-dominant	Chr 2: 12.7-15.5 Mb	*rs2380452* (12.9 Mb)	LOD 1.43/HLOD 2.57
	21OH+	Parametric, rare dominant	Chr 3: 0.99-6.7 Mb	*rs1948153* (6.5 Mb)	LOD 1.02/HLOD 3.25
Association study	AAD	50 AAD, 2706 WTCCC controls	Chr 2: 48.1-48.3 Mb	*rs10495950* (48.3 Mb)	P 9.7×10^{-8}
			Chr 6: 32.2-32.7 Mb	*rs2187668* (32.7 Mb)	P 3.0×10^{-7}
Validation study meta-analysis	AAD	Maximum 1097 AAD, 1117 controls	Chr 18: 74.5-76.1 Mb	*rs7236339* (75.7 Mb)	P 0.004
	21OH+		Chr 18: 74.5-76.1 Mb	*rs7231100* (74.5 Mb)	P 0.004
			Chr 7: 69.4-70.9 Mb	*rs12698902* (69.5 Mb)	P 0.01

Table 20: Summary of significant results from the AAD linkage, association and validation study analyses.

For the linkage study, only loci with a maximum HLOD of greater than or equal to 2.50 are shown. For the association study, only loci where at least two SNPs reached genome-wide significance ($P < 5 \times 10^{-7}$) are shown. For the validation study, all results are presented.

5.6 VALIDATION STUDY IN EUROPEAN UNRELATED AAD CASE-CONTROL COHORTS

In summary, taking the results generated from the analysis of AAD as the trait of interest, linkage peaks with a LOD or HLOD of greater than 2.0 were observed on chromosomes 7, 9 and 18 applying a dominant model, and on 2 and 9 applying a co-dominant model in a parametric analysis. In the non-parametric analysis, one locus with a LOD of greater than 3.0 on chromosome 6 was observed. Pairs or small clusters of SNPs reaching genome-wide significance were only seen on chromosomes 2 and 6. There was overlap between the linkage peak and association results on chromosome 6 only. A validation experiment was planned, aiming to investigate SNPs underlying some of the linkage peaks in a study of unrelated AAD cases and controls. As the linkage peaks were large and numerous, to make the validation experiment financially feasible, two peaks were prioritised for further investigation in a case-control association study design.

Initially, the linkage peak on chromosome 6 was eliminated from further analysis because of its close proximity to the MHC. MHC is already known to be strongly associated with AAD and autoimmunity. LD extends a large distance from MHC itself, which would make it difficult to determine whether any significant results were exerting an independent effect or arising due to LD with MHC alleles. The linkage peak on chromosome 18 was selected as the first area for more detailed analysis, as this had the greatest LOD score and the region of linkage was close to a plausible candidate gene, *nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1* (*NFATC1*). *NFATC1* encodes a transcription factor which plays a central role in gene transcription during the immune response and T cell activation and differentiation[422, 462]. The peak on chromosome 7 was then also chosen for further investigation, as this had a strong LOD score and the region of interest is very gene-rich, containing some plausible candidates for AAD.

A Sequenom assay was then designed, in three plexes, to include SNPs representing as many LD blocks as possible from in and around the regions of linkage on chromosomes 18 (75241668 – 77639585bp) and 7 (71185100 – 73809454bp). Sequenom primer sequences for the 64 selected SNPs can be found in electronic appendix C. LD plots of all SNPs genotyped on chromosomes 18 and 7 are included in electronic appendix D. These 64 SNPs were genotyped by CIGMR, Manchester, in AAD case and control cohorts from the UK (346 AAD, 367 controls) and Norway (384 cases, 384 controls). A cohort from Sweden was also included. In this cohort, due to a shortage of DNA, one plex (21 SNPs in total), was genotyped in 345 AAD cases and 344 controls while the other two plexes (43 SNPs) were genotyped in 367 cases and 366 controls. The data analysis was performed using contingency tables to calculate χ^2 statistics and P values. The full genotyping data for this analysis can be found in electronic appendix E.

5.6.1.1 VALIDATION STUDY POWER

Using QUANTO[442], power calculations for this validation study were performed. Assuming a MAF of 0.3 and α of 0.00083 to account for multiple testing, each individual cohort has more than 80% power to detect a locus with an odds ratio of 1.6 or more (Table 21). Combining the cohorts in a meta-analysis, under the same assumptions, this study has 97% power to detect a locus with an odds ratio 1.4.

5.6.1.2 DATA QUALITY CONTROL

Any SNP with a call rate of less than 95%, and any SNP out of HWE in the control cohort (P <0.01), was excluded from analysis. In the UK cohort, two SNPs were excluded based on a low call rate in the control cohort (*rs1051978* on chromosome 18 and *rs7808818* on chromosome 7). In the Norwegian cohort, one SNP (*rs4717599* on chromosome 7) was out of HWE in the control cohort (P 0.0097) and was therefore excluded. In the Swedish cohort, one SNP (*rs2067534* on chromosome 18) was excluded on the basis of a call rate of 94% in the control cohort (see electronic appendix D).

Cohort (number of AAD cases)	Case:control ratio	OR 1.2	OR 1.4	OR 1.6	OR 1.8
Combined (1097)	1:1.02	30%	97%	100%	100%
Norway (384)	1:1	5%	41%	85%	98%
Sweden (367)	1:1	4%	38%	82%	98%
UK (346)	1:1.06	4%	36%	81%	97%

Table 21: Table showing outcomes of power calculations for the validation study.

Power is shown as a percentage for each individual cohort and the cohort as a whole, assuming a minor allele frequency of 0.3, a disease prevalence of 1/10000 (0.0001) and an α of 0.00083 to account for multiple testing.

5.6.1.3 INDIVIDUAL COHORT RESULTS

The significant results from the validation experiment are summarised in Table 22.

5.6.1.4 META-ANALYSIS

Meta-analysis was performed using Revman 5[221]. A random effects model was used to allow for heterogeneity between the cohorts. Any SNP not meeting the quality control criteria was excluded. In the cohort as a whole, three independent SNPs, all on chromosome 18, were associated. Maximal association was seen at *rs7236339* (P 0.004), with association also noted at *rs754093* (P 0.02) and *rs8091998* (P 0.04). *rs754093* was noted to be in the *NFATC1* gene while the other two markers are intergenic (Table 23).

In the 21OH autoantibody positive cohort, four SNPs were associated; three independent SNPs on chromosome 18 and one SNP on chromosome 7. On chromosome 18, maximal association was seen at *rs7231100* (P 0.004), with less striking association also seen at *rs1960120* (P 0.02) and *rs11081569* (P 0.03). These markers were all noted to be intergenic. A single SNP on chromosome 7, *rs12698902*, located in the *autism susceptibility candidate 2* (*AUTS2)* gene, was also associated with AAD (P 0.01).

The results of the linkage and association analyses, in addition to the validation study results, are summarised in Table 20.

Chr	SNP ID	UK cohort (346 AAD, 367 controls)		Norwegian cohort (384 AAD, 384 controls)		Swedish cohort (367 AAD, 366 controls; except SNPs marked * 345 AAD, 344 controls)	
		P genotype	P allele	P genotype	P allele	P genotype	P allele
18	*rs7231100**	N/S	N/S	**0.042**	**0.026**	N/S	N/S
18	*rs2941794*	N/S	N/S	**0.046**	0.38	N/S	N/S
18	*rs6506869*	**0.0097**	0.39	N/S	N/S	N/S	N/S
18	*rs2085985**	**0.014**	0.41	N/S	N/S	N/S	N/S
18	*RS754093*	N/S	N/S	N/S	N/S	**0.033**	0.07
18	*rs3826573*	N/S	N/S	N/S	N/S	**0.0053**	**0.0016**
18	*rs1960120**	N/S	N/S	N/S	N/S	0.058	**0.017**
18	*rs7236339**	N/S	N/S	0.073	**0.032**	**0.031**	0.28
7	*rs12672930*	N/S	N/S	**0.029**	**0.034**	N/S	N/S
7	*rs12698902*	N/S	N/S	N/S	N/S	**0.02**	**0.0063**
7	*rs10486872*	0.093	**0.037**	N/S	N/S	N/S	N/S
7	*rs10237317*	0.1	**0.044**	N/S	N/S	N/S	N/S
7	*rs38319*	N/S	N/S	0.12	**0.038**	N/S	N/S
7	*rs38307*	0.1	**0.035**	**0.0037**	**0.001**	N/S	N/S

Table 22: Validation study significant association results for the UK, Norwegian and Swedish AAD cohorts.

Of the 64 SNPs genotyped in this study, 14 (eight on chromosome 18 and six on chromosome 7) were associated with AAD in one or more of the cohorts. Non-significant results (P >0.05) are shown as N/S. SNPs marked with an asterisk (*) were genotyped in 345 Swedish AAD and 344 controls due to a shortage of DNA.

Chromosome	rs ID	AAD	21OH+ only	Information
		P value (I^2)	P value (I^2)	
18	*rs7231100*	0.19 (59%)	0.004 (0%)	Intergenic
18	*rs754093*	0.02 (0%)	0.12 (0%)	NFATC1
18	*rs1960120*	0.28 (46%)	0.02 (0%)	Intergenic
18	*rs8091998*	0.04 (0%)	0.69 (8%)	Intergenic
18	*rs7236339*	0.004 (0%)	0.05 (0%)	Intergenic
18	*rs11081569*	0.07 (0%)	0.03 (0%)	Intergenic
7	*rs12698902*	0.13 (45%)	0.01 (0%)	AUTS2

Table 23: Meta-analysis, applying a random effects model, of chromosome 18 and chromosome 7 validation study genotyping results from UK, Norwegian and Swedish AAD cohorts.

Association is seen with alleles at three SNPs on chromosome 18 (shown in yellow) in the AAD cohort as a whole. When all 21OH autoantibody negative people were excluded, alleles at three SNPs on this chromosome, and a further SNP on chromosome 7, were associated (shown in pink).

5.7 DISCUSSION

5.7.1 LINKAGE ANALYSIS

This study is the first linkage analysis in AAD and takes advantage of a unique sample resource of multiplex AAD families from both the UK and Norway. It has generated some interesting results which will need further investigation and replication.

This study implicates a number of chromosomal regions in susceptibility to AAD. In a non-parametric analysis, a linkage peak with a linear LOD of 3.13 was seen on chromosome 6, corresponding to the HLA region which is a known susceptibility locus for AAD. The detection of a linkage peak at this region of chromosome 6 demonstrates that the study is sufficiently powerful to detect a locus with an odds ratio previously estimated to be between 3 and 15[55, 463], and validates the approach used.

When a parametric analysis was performed, applying a dominant model, a single linkage peak on chromosome 18 had a LOD/HLOD score of 3.00 (at 76.6Mb) confirming linkage, while two other loci, on chromosomes 9 and 7, had LOD/HLOD scores of just less than 3.0 (maximum LOD/HLOD 2.90 at 19.0Mb on chromosome 9; maximum LOD/HLOD 2.88 at 70.1Mb on chromosome 7), which is highly suggestive of linkage. When this analysis was repeated using a co-dominant model, one linkage peak on chromosome 2 was observed with an HLOD greater than 2.5, suggestive of linkage (HLOD 2.57 at 12.9Mb, α 0.80). Assessing the pedigrees included in this analysis, no unifying mode of disease inheritance was apparent across all families. Some families displayed an autosomal dominant mode of inheritance, with affected individuals seen in multiple generations, while others obeyed a recessive pattern, with generations being skipped. In some families, where information was only available for a single generation, for example a trio of siblings, the mode of inheritance was unclear.

In any linkage analysis, this apparent heterogeneity can be allowed for in one of two ways: either by using a non-parametric approach, or by using a parametric approach and assessing the proportion of linked families, represented by α. The non-parametric approach, which assesses sharing of alleles IBD and therefore by necessity must exclude affected parent-offspring pairs, loses some power as the result of the reduced number of kindreds suitable for the analysis. The exclusion of affected parent-offspring pairs, suggestive of a dominant pattern of inheritance, means that the non-parametric method often gives similar results to using a recessive model in a parametric analysis. Indeed, although not significant, a region of linkage with a maximum LOD/HLOD of 1.66 was seen on chromosome 6 in the parametric linkage analysis applying a recessive model, supporting this observation. In contrast, the parametric approach, using multiple models, allows all pedigrees to be included, thus increasing study power, and the calculation of the HLOD allows for genetic heterogeneity between the kindreds. At the most significant linkage peaks on chromosomes 18, 9 and 7, all families were contributing to the results and therefore the observed LOD and HLOD scores were the same. Applying different models in the parametric linkage analysis gave very different results, demonstrating clearly that parametric linkage analysis is sensitive to the model specified. On balance, assessing the pedigree information available, the dominant model would seem to be the most appropriate. Results should be interpreted in light of the multiple analyses used in this study.

The kindreds selected for this study were relatively small, comprising just a few individuals. In general, larger, multi-generational families tend to be more informative for linkage analysis and if smaller kindreds are to be used, more will be needed in order to generate sufficient study power. The small nature of the multiplex AAD kindreds collected for this study reflects the fact that AAD is a condition which largely affects people of Caucasian extraction, where families tend to be small, and has its major peak of onset in middle-age. Therefore, kindreds available for analysis tend to be small and, as the probands are often in their fourth or fifth decade of life, their parents may be deceased and therefore unavailable for analysis. Only five three-generation families (i.e. grandparents, parents and children) were available for inclusion, while five

kindreds included just a single generation (i.e. a group of siblings). Kindreds included in this linkage study were carefully assessed in order to select a relatively homogeneous AAD population. To limit the possibility of including kindreds with a non-autoimmune aetiology to their Addison's, such as tuberculosis or adrenoleukodystrophy, and to exclude those with monogenic AAD due to APS1, families included were carefully questioned. A history of mucocutaneous candidiasis, hypoparathyroidism and dental problems was sought to exclude those with APS1 and a history of tuberculosis or other unusual phenotypic features which would suggest a diagnosis other than AAD was sought from included family members. Features in families that were considered suggestive of an autoimmune aetiology included the presence of other autoimmune conditions in those with AAD and in their first degree relatives. In particular, a personal or family history of autoimmune thyroid disease or type 1 diabetes was sought as these conditions are commonly seen in conjunction with AAD, and in relatives of people with AAD.

21OH autoantibody status was given low diagnostic priority in terms of confirming the diagnosis of AAD in affected cases, and in terms of excluding AAD in "healthy" relative controls. As the hallmark for autoimmune disease, assigning 21OH autoantibody status a low diagnostic priority may seem counterintuitive. However, the 21OH status of individuals may be misleading. The 21OH autoantibody assay is not currently in routine diagnostic use and, although careful quality control measures are employed during the use of this assay, including batching the samples with both positive and negative controls, there is the possibility of a spurious (most commonly a false negative) result, as with any other assay. A recent study has demonstrated that there is some inevitable intra and inter-laboratory assay variation[464]. In terms of its use in confirming the diagnosis of an autoimmune aetiology in those with Addison's, 21OH autoantibodies are useful when positive, but relatively unhelpful when negative. The reasons for this are twofold. Firstly, a minority of individuals with known AAD are found to be 21OH autoantibody negative[465]. These individuals may have autoantibodies to other adrenal components which are not routinely tested for. Secondly, in individuals with AAD, 21OH autoantibody titres have been found to reduce over time as the adrenal cortex is gradually destroyed and

the autoimmune process burns out. In one study, 21OH autoantibodies were detected in 92% of individuals tested within 2 years of diagnosis and in 78% of those tested after 2 years of diagnosis[465]. Therefore, in individuals who have been diagnosed with Addison's disease for a number of years, a negative autoantibody result is not unexpected and certainly does not exclude the diagnosis of AAD.

In terms of its use in healthy controls with no clinical signs or symptoms of adrenal insufficiency, the significance of a positive 21OH autoantibody assay result is currently unknown. Simplistically, you could regard a negative 21OH result in this population as normal and a positive result as abnormal, meaning that a 21OH autoantibody positive control should be excluded. However, in a review[466] which collated results from multiple studies, in 6488 healthy controls, without any evidence of autoimmune disease, a total of 37 (0.57%) had positive 21OH or adrenal cortex autoantibodies, with studies reporting incidences between 0 and 1.6%[54, 467]. In those people who do have positive autoantibodies, the risk of progression to overt adrenal failure has been calculated. In one longitudinal follow-up study, of 100 individuals who were adrenal autoantibody positive, 31 developed AAD over a follow-up period of up to 21 years[46]. This allowed a cumulative risk of 48.5% to be calculated for 21OH autoantibody positive individuals[46]. The risk of progression was increased in children compared with adults and in those with high titres compared to those with moderate or low autoantibody titres. Crucially however, not all individuals with autoantibodies go on to develop progressive disease, with some remaining antibody positive but with normal adrenal function long term, and with some reverting to being autoantibody negative[45, 46]. Together, these findings demonstrate that autoantibody status cannot be reliably used to predict progression to AAD in people who are well and autoantibody positive, or to confirm the diagnosis of AAD in individuals with disease who are autoantibody negative.

In our kindreds, eight individuals with AAD were found to be 21OH autoantibody negative. These individuals came from five families (three from the UK, two

from Norway) and had a mean duration of disease 10.3 years (median 6.5 years, minimum 2 years, maximum 37 years). Given that all had been diagnosed with AAD for at least 2 years, and some for much longer, these negative 21OH assay results are not entirely surprising and do not exclude a diagnosis of AAD. In addition, those who were 21OH autoantibody negative, and those who could not be tested due to lack of serum, all had either a personal history of autoimmune thyroid disease or type 1 diabetes, a history of these conditions in a first degree relative or a close relative with positive 21OH autoantibodies, with the exception of two sisters from UK family 6. These factors combine to make a diagnosis of AAD likely, even in those without confirmed 21OH autoantibodies. The two siblings who did not meet the above criteria, did however both have ankylosing spondylitis in addition to Addison's, a disease with a probable underlying autoimmune aetiology associated with HLA alleles. Therefore, this family was not excluded from the linkage analysis.

Five "healthy" relatives, with no clinical evidence of adrenal insufficiency, from five separate kindreds, were found to be 21OH autoantibody positive. This finding provoked a second linkage analysis, looking at 21OH autoantibody status as the trait of interest. Unfortunately, serum samples, and therefore 21OH autoantibody results, were not available for 12 individuals, 11 of whom had AAD, therefore considerable power was lost in this analysis. However, there was a single linkage peak with a LOD of greater than 3.0 in the parametric analysis applying a dominant model. This was on chromosome 3 where a maximum HLOD of 3.23 at 6.5Mb provided strong evidence for linkage. The detection of a novel linkage peak on chromosome 3 in this analysis, which was not observed when AAD was assessed as the trait, suggests that 21OH autoantibody status may have different genetic determinants to AAD. It can be hypothesised that these differences perhaps determine which individuals with 21OH autoantibodies go on to develop AAD, which remain autoantibody positive but do not progress to disease and which clear their autoantibodies and remain autoantibody and disease-free in the long term. As 21OH autoantibodies are more common early in disease, genes associated with 21OH positivity are perhaps those responsible for disease initiation, while the genetic determinants

of disease progression may be entirely distinct. This hypothesis requires further investigation.

In order to ensure that linkage results generated were valid and replicable, considerable effort was made to select a homogenous cohort of carefully phenotyped kindreds for this study. Despite this, one major limitation to any linkage study in later-onset diseases such as AAD is that while an individual's genetic makeup is relatively stable, rendering them susceptible to a particular disease, their phenotype can change. In the kindreds studied here, it is possible that some unaffected relatives, whether currently 21OH autoantibody positive or negative, may later develop AAD. Furthermore, the 21OH autoantibody status of some individuals may change over time, with people with AAD becoming autoantibody negative as the time interval from diagnosis increases and some people who are antibody negative becoming positive. It is impossible to control for this effectively. Excluding children under the age of 18 as healthy controls helps to minimise this as an unaffected child could simply not have had the opportunity to develop disease if they are genetically predisposed to do so. Careful long-term follow up will be useful to determine the outcome of all of the people involved in this study, in particular those who are 21OH autoantibody positive but currently clinically unaffected by AAD.

5.7.2 GENOME-WIDE ASSOCIATION ANALYSES

The linkage peaks observed in this study were large and contained many genes, some coding for proteins of unknown function and some which appeared to be plausible candidates for AAD. A full list of all genes underlying the linkage peaks, taken from HapMap[102], can be found in electronic appendix F. Two genome-wide association analyses were therefore performed using the genotyping data from the multiplex AAD families in an attempt to narrow down the areas of interest. Initially, affected AAD individuals from within the families were used as cases and compared to the unaffected relatives as controls. This analysis did not reveal any significant associations and this was not surprising as this analysis was very underpowered. The second analysis took the affected AAD individuals within the families as cases and compared them to more than

2000 controls from the WTCCC UK 1958 birth cohort to increase the study power. A number of SNPs were associated in this analysis. Any associated SNPs seen in isolation were disregarded, as these are likely to represent chance findings or spurious genotyping results. However a pair of SNPs on chromosome 6, 546kb apart, corresponding to the HLA, were associated (*rs2187668,* P 3.0 x 10^{-7} ; *rs1150753,* P 4.4 x 10^{-7}) (Figure 52) as were a cluster of five SNPs spanning 230kb on chromosome 2, where maximal association was seen with SNP *rs10495950* (P 9.7 x 10^{-8}) (Figure 51). The finding of a pair of SNPs corresponding to the HLA region again validates this method. The cluster of associated SNPs on chromosome 2 was approximately 30Mb upstream of the linkage peak seen on the same chromosome in the parametric analysis using a co-dominant model: the two regions therefore did not correspond. These associated SNPs were noted to be between the *F-box only protein 11* (*FBXO11)* and *Forkhead box N2 (FOXN2*) genes. These genes are 118 and 63kb distant from the nearest associated SNP respectively.

The FBXO11 protein product contains domains that allow it to methylate arginine residues[468]. Methylation is known to be important in multiple cellular processes and arginine methylation in particular is a common post-translational modification thought to be important in regulating protein function[469, 470]. The FBXO11 protein has been shown to suppress the function of p53, a ubiquitous tumour suppressor protein, through neddylation, a process very similar to ubiquitination[471]. In addition, it is also known to regulate the TGFβ signalling pathway[472, 473] which is important for cell growth, differentiation, apoptosis and T_{Reg} function. These diverse functions make it a plausible potential candidate gene for AAD. The *FOXN2* gene encodes a forkhead domain DNA binding protein, also known as the Human T-cell leukaemia virus enhancer factor. As its alternative name suggests, it is thought to regulate transcription of the human T-cell leukaemia virus long terminal repeat[474], thus perhaps playing a role in T cell leukaemia and lymphoma. As a DNA-binding protein, it is likely to have other functions, although these are poorly understood at present. Studies in mice have shown that the Human T-cell leukaemia virus enhancer factor is expressed in a number of tissues during embryogenesis, suggesting that it

might have a role in development[475]. Unfortunately, the associated SNPs on chromosome 2 observed between these two genes are in a region of low recombination and therefore extended LD, which includes not only the *FBXO11* and *FOXN2* genes, but also *mutS homolog 6 (MSH6)*, a highly conserved DNA repair gene[476], *protein phosphatase 1, regulatory subunit 21 (PPP1R21),* a gene which encodes a protein thought to regulate protein phosphatase 1 which is involved in numerous cellular processes including cell signalling and DNA damage repair[477] and *stonin 1 (STON1),* a member of the stonin protein family, members of which are involved in molecule trafficking and endocytosis. Based on known function alone, the *FBXO11* gene is probably the most likely candidate in AAD. However, as our understanding of protein function is incomplete, the true association could lie with any of these genes, or indeed with a non-coding RNA exerting a regulatory effect. This region therefore requires further investigation.

A significant limitation in the design of the genome-wide association study, comparing the 50 familial AAD cases with the UK WTCCC controls, is that more than half of the AAD cases (54%) originate from Norway. Comparing Norwegian AAD cases to controls from the UK could potentially result in some spurious findings due to inherent genetic differences between the populations (population stratification). Indeed, the candidate gene study included in this thesis did demonstrate some genetic heterogeneity between the UK and Norwegian control cohorts. The UK WTCCC controls were selected for this study as the data are freely available, the demographics of the control population are well described and the data have undergone strict quality control measures and are therefore known to be robust. Conversely, a similar Norwegian control cohort that could be used for comparison is not freely available. Furthermore, if the 50 familial AAD cases were to be divided into individual UK and Norwegian cohorts for separate analyses, the study would lose significant power. The use of the UK WTCCC controls therefore represents a compromise which should be considered when interpreting the results.

With this limitation in mind, given that the linkage analysis was more powerful from the outset than the genome-wide association analyses and as there was no overlap between the regions of linkage and the associated SNPs, with the exception of the HLA region on chromosome 6, the decision was taken to use the preliminary linkage analysis results to direct further investigations to detect novel susceptibility loci for AAD, with a view to investigating the association with chromosome 2 at a later date.

5.7.3 VALIDATION STUDY ASSOCATION ANALYSIS

Following assimilation of the linkage and association data, the linkage peaks were scrutinised further, with particular attention being paid to any likely candidate genes underlying the peaks. The linkage peaks on chromosomes 18 and 7 were chosen for a case-control validation study in UK, Norwegian and Swedish AAD cohorts. Chromosome 18 was chosen as this was the only linkage region, with the exception of chromosome 6, meeting the LOD threshold of 3.0 when AAD was examined as a trait. There was little difference between the amplitude of the linkage peaks on chromosomes 9 and 7. However, chromosome 7 was chosen for the validation study as the region of interest was more gene-rich. The meta-analysis was performed in two ways. Initially, the cohort was analysed as a whole, including 1097 AAD cases and 1117 controls, using a random effects model to allow for heterogeneity. Then, a second analysis was performed, excluding any individuals who were not 21OH autoantibody positive (including a maximum 702 AAD cases and 1117 controls). In this study, 60 independent loci were tested and the effects of multiple testing must be considered when interpreting the results. If a Bonferroni correction is applied based on 60 independent tests, a P value of 0.00083 (0.05/60) would be considered statistically significant and any result not meeting this threshold could be considered a chance finding as a result of multiple testing.

In the meta-analysis of the cohort as a whole, modest association was observed with three independent SNPs on chromosome 18 and AAD; however, none of these would meet the corrected P value threshold.

Two of these SNPs on chromosome 18 are 18kb apart (*rs8091998* (P 0.04), *rs7236339* (P 0.004). *rs8091998* is found approximately 47kb upstream from the *carboxy-terminal domain RNA polymerase II, polypeptide A phosphatase, subunit 1 (CTDP1*) gene which encodes a protein that can phosphorylate RNA polymerase II and therefore influence transcription[478], and *rs7236339* is approximately 44kb downsteam from the *potassium voltage-gated channel subfamily G, member 2* (*KCNG2*) gene which encodes a subunit of the potassium voltage-gated ion channel. Importantly, there is no LD between these associated markers, and no LD between these and any SNPs within the two neighbouring genes. As there is no LD between the associated markers and SNPs in *CTDP1* or *KCNG2,* it is very unlikely that this observed association is due to a causative mutation in either of these genes. Three other SNPs in the region between *CTDP1* and *KCNG2* were also genotyped in this validation study: these were not in LD with the associated SNPs, and no association was observed with these markers and disease.

When the meta-analysis was repeated, excluding 21OH autoantibody negative individuals, two different independent SNPs (*rs1960120,* P 0.02; *rs11081569,* P 0.03), in the same intergenic region, were associated with AAD. These SNPs are 87kb apart: *rs1960120* is 21kb upstream of *CTDP1* while *rs11081569* is just 1kb downstream from *KCNG2.* Again, there is no LD between these two SNPs, and no LD between these and any SNPs within the neighbouring genes. It is possible that these are chance findings as the P values do not remain significant when multiple testing is considered. However, the finding of multiple, associated, independent SNPs in this intergenic region could also indicate association between AAD and a causative non-coding variant located between the two genes which is being tagged by these associated markers.

The third chromosome 18 SNP also associated with AAD in the full AAD cohort meta-analysis was *rs754093* (P 0.02). This SNP is located in the *NFATC1* gene. *NFATC1* is an excellent candidate gene for autoimmune conditions in general, including AAD, as it encodes a transcription factor which is a member of the NFAT family of proteins. The NFAT proteins are expressed on cells of the

immune system and rapidly induce gene expression, which allows a fast and vigorous immune response[419-421]. NFAT proteins have a REL-homology region and an NFAT-homology region and can bind specific DNA sequences[399, 422]. They are present in the cytoplasm but rapidly translocate to the cell nucleus when surface T cell receptors, or other receptors coupled to calcium mobilisation, are stimulated. The translocation process is controlled by calcineurin which interacts with the NFAT domain. In the nucleus, they form transcription complexes with NFAT proteins and other molecules, and regulate gene expression in response to T cell activation during immune and inflammatory responses[422]. *NFATC1* is a gene that comprises ten exons, spanning approximately 130kb. In Caucasians, the gene can be loosely divided into six LD blocks. In this study, eight independent SNPs in *NFATC1* were genotyped, but only one was associated with AAD and the association was modest. The individual cohort results did not reveal anything of added interest in this region. Again, this could be a chance finding, but could also be an indication that the *NFATC1* gene is exerting a modest influence on AAD susceptibility. Further investigation into this locus would be needed to confirm or refute this hypothesis.

In the meta-analysis of 21OH autoantibody positive subjects, in addition to the two intergenic chromosome 18 SNPs discussed above, a further SNP on chromosome 18, *rs7231100,* was also associated (P 0.004).This SNP lies between the *galanin receptor 1* (*GALR1*) gene, which encodes a G-protein coupled receptor for the neuropeptide galanin[479, 480], and the *Sal-like 3* (*SALL3*) gene, which encodes a conserved zinc finger protein[481] thought to be important in developmental processes and in regulating DNA methylation[482, 483]. These two genes are almost 2Mb apart and the associated SNP it is not in LD with any SNPs within these genes. This SNP lies 54kb from the *rs2002842* SNP which has been associated with rheumatoid arthritis in a previous genome wide association study[484]. However, these SNPs are not in LD and the *rs2002842* SNP was investigated in this study and was not found to be associated with AAD, either in the cohort as a whole, or in the 21OH autoantibody positive only cohort (P 0.70 and 0.47 respectively).

Finally, a single SNP in the *AUTS2* gene on chromosome 7, *rs12698902*, was associated with AAD (P 0.01) in the 21OH autoantibody positive analysis. The *AUTS2* gene was initially identified by a group studying a pair of monozygotic twins with autism and developmental delay[485] and as such, research into the gene and its protein product have, to date, largely focussed on neurodevelopmental disorders and been driven by researchers in this field. It is a large gene that spans approximately 1.2Mb and is composed of 19 exons. The physiological role of the AUTS2 protein is unknown. From studies in mice, AUTS2 has been found to be a nuclear protein which is highly expressed in the cerebral cortex and cerebellum[486]. Mutations in this gene have been associated with developmental delay[487] and autism[488], while genome-wide association studies have linked this region to both epilepsy[489] and alcohol intake[486]. Finally, a recent study has linked *AUTS2* polymorphisms to heroin dependence and demonstrated reduced levels of expression in heroin addicts[490]. These findings make the *AUTS2* gene an unlikely functional candidate for AAD at present, however it is likely that the AUTS2 protein has broader functions and is more widely expressed than is currently appreciated, and therefore it is possible that it might have a role to play in susceptibility to an autoimmune disorder. Of the seven independent SNPs in *AUTS2* genotyped in this study, only one was associated with AAD in the meta-analysis of the 21OH autoantibody positive cohort. Again, this could be a chance finding or it could indicate that *AUTS2* is associated with susceptibility to AAD. The likelihood of this being a real association is increased by the observation that there is association with different *AUTS2* SNPs within the individual cohorts genotyped (Table 22). Although the meta-analysis P value result at this SNP does not meet the threshold set when multiple testing is considered, the pattern of association between the cohorts with SNPs in the *AUTS2* gene does indicate that there might be a role for *AUTS2* in susceptibility to AAD.

The main limitation of this validation study is that the linkage peaks are very large and the resulting haplotype coverage provided by the 64 SNPs genotyped was therefore inadequate. Although the Sequenom platform allows the user to

multiplex assays, a maximum of 40 assays can be genotyped in one plex, and assay compatibility issues often mean that only 20 to 25 SNPs can be genotyped together. Ideally, all the variation underneath the two linkage peaks of interest would have been captured by genotyping one SNP per LD block. However, the chromosome 18 linkage peak spans 2.7 million base pairs and the chromosome 7 peak spans 3.8 million base pairs. To attain adequate coverage, even with careful SNP selection, many hundreds more SNPs would need to be genotyped.

Another significant flaw of the validation study pertains to a fundamental assumption made when designing the study: that individuals with AAD occurring sporadically will have the same underlying genetic basis to their disease as individuals with the familial form. If this is not the case and variants conferring disease in multiplex AAD families are entirely different to variants conferring susceptibility to AAD in sporadic cases, then searching under regions linked to familial AAD in sporadic AAD case-control cohorts would be a futile exercise. The finding of linkage with the HLA region on chromosome 6 suggests that the multiplex AAD families do at least share some susceptibility loci with sporadic cases. However, there are some loci that are known to confer susceptibility to sporadic AAD which do not correspond to any of the linkage peaks seen in the families. Replicated loci known to confer susceptibility to sporadic AAD include the *CTLA4* gene on chromosome 2, which is some 189 million bases from the linkage peak observed in the parametric co-dominant analysis and 156 million bases from the cluster of associated SNPs on chromosome 2 from the association analysis. The *PTPN22* gene on chromosome 1 has also been associated with AAD in a number of cohorts but no linkage was seen in any analyses to the chromosomal region containing this susceptibility locus. However, the *CTLA4* and *PTPN22* loci confer only modest risk, with odds ratios calculated between 1 and 2, therefore these might not be detected by linkage analysis.

5.8 CONCLUSIONS AND FUTURE DIRECTIONS

The linkage study has implicated a number of novel chromosomal regions in the pathogenesis of AAD in multiplex AAD families. However, the linkage peaks generated are large; neither the genome-wide association analyses performed nor the regional case-control association results, gathered through genotyping SNPs underlying the linkage peaks, have helped to narrow down the areas of interest in order to locate a causative locus. The genome-wide analysis looking at the 50 familial AAD cases compared to the UK 1958 birth cohort available from the WTCCC revealed a cluster of intergenic SNPs on chromosome 2 which were associated with AAD. The case-control validation study demonstrated some modest association between an intergenic region between the *CTDP1* and the *KCNG2* genes on chromosome 18, the *NFATC1* gene on chromosome 18 and the *AUTS2* gene on chromosome 7 and AAD. However, none of the P values obtained in the cohorts individually, or in the meta-analysis, met the threshold applied when multiple testing is taken into consideration, making these results less convincing. These findings warrant further investigation and a number of approaches could be adopted.

To replicate and validate the linkage study findings, a carefully selected replication cohort, made up of additional multiplex AAD kindreds, is ideally required. As these kindreds are relatively rare, this replication cohort may take some years to assemble. An alternative approach would be to sequence selected members of the multiplex AAD families, looking for deleterious variants. Whole genome sequencing or whole exome sequencing technology, which is now falling in price and increasing in availability, could be used for this purpose. Indeed, a number of members of the multiplex AAD families used in this study are now being investigated by whole exome sequencing, by our Norwegian collaborators, for this reason. The results of this work are not yet available.

To investigate further the cluster of associated SNPs on chromosome 2 in a region spanning 1Mb, detected by the genome-wide association study in the 50

familial AAD cases and 2706 WTCCC controls, one of two approaches could be taken. A custom microarray could be designed to include SNPs in the region. These arrays usually include multiples of 96 assays selected by the researcher to provide the maximum amount of information for the region of interest. Another approach which could be used is targeted resequencing, where a subset of the genome is sequenced: for example, a region of interest, a single chromosome or a group of exons. This technique commonly uses either array-based hybridisation capture methods, where numerous oligonucleotides complementary to the region of interest are synthesised *in situ* on a DNA microarray[491, 492], or solution-based hybridisation capture, where oligonucleotide capture probes in solution are used to target regions of interest[493]. Both custom microarrays and targeted resequencing offer good regional coverage, however the advantage of targeted resequencing over a custom array is that it can be easily used to detect rare variants and structural changes which might be contributing to AAD disease susceptibility.

In sporadic AAD cases, the results of interest from the validation study could be taken forward to a further candidate region study, looking in more detail at the *NFATC1* and *AUTS2* genes in addition to the intergenic region noted to be associated on chromosome 18. Again, either a custom microarray or targeted resequencing approach could be used in this instance to provide good coverage of the genes and surrounding regions and to determine whether these loci are contributing to AAD susceptibility.

Finally, a genome-wide association study in sporadic AAD could be considered. This would require multi-centre cooperation to gather enough DNA samples for an initial study and a replication study. This approach would be expensive, but potentially very interesting.

This work has implicated some novel chromosomal regions in the aetiology of both familial and sporadic AAD, but should be viewed as a hypothesis-generating study. These findings must now be replicated and refined in order to increase our understanding of the pathophysiology of AAD.

CONCLUDING REMARKS

Despite being highly heritable, the underlying genetic aetiology to AAD is poorly understood. The rarity of AAD, combined with the lack of a good animal model of disease and the complex nature of its inheritance, makes it an interesting but challenging condition to investigate genetically.

Prior to this series of studies, a number of candidate genes had been investigated in AAD by association analysis and implicated in disease susceptibility, including the HLA region and the *CTLA4* and *PTPN22* genes, which are the most widely replicated susceptibility loci in AAD. However, the studies previously conducted had been on relatively small cohorts and large, collaborative studies were lacking. In addition, a linkage study in AAD had never been conducted due to the rarity of multiplex families containing two or more individuals with AAD.

This study has made use of unique sample resources and close collaborations with researchers across Europe to investigate the underlying genetic aetiology of AAD using a number of complementary approaches. This study was conducted with the single aim of discovering new genetic disease determinants. The hypothesis-driven approach, which has previously been applied to the study of AAD, has been employed in this work to investigate a number of promising candidate genes by association analysis in the largest cohort of AAD patients ever assembled. In addition, the candidate gene approach has been combined with a hypothesis-free, discovery-driven approach, using genome-wide SNP genotyping technology in multiplex AAD families to perform the first linkage and genome-wide association study in AAD.

In this study, PCR-based approaches have been undertaken to investigate perhaps the most obvious candidate locus in AAD, the steroid 21-hydroxylase locus on chromosome 6. This locus contains the *CYP21A2* gene which encodes the 21-hydroxylase enzyme, the primary autoantigen in AAD, and also contains a highly homologous pseudogene, *CYP21A1P*, which contains a deleterious 8bp deletion rendering the predicted protein product truncated and non-functional. Attempts have been previously made to study this locus in AAD.

However, *CYP21A2* is in a region of extended linkage disequilibrium, in close proximity to HLA, a known susceptibility locus in AAD, and this has made interpretation of the results of these studies difficult. In work which led up to this study, our group discovered that individuals with AAD are more likely to have no genomic copies of the *CYP21A1P* pseudogene when compared to controls. This observation could be attributed to LD with HLA. However, we hypothesise that this is not the case and that *CYP21A1P* transcripts and/or CYP21A1P protein products may be expressed in human thymus to promote and induce tolerance to components of the steroidogenic enzymes. Absence of these transcripts in individuals who do not have any copies of the *CYP21A1P* pseudogene in their gDNA could therefore result in increased susceptibility to AAD. As a first step in investigating this hypothesis, a cohort of individuals with Graves' disease, for which HLA is also a susceptibility locus, were also genotyped for the presence or absence of *CYP21A1P*. While *CYP21A1P* was absent more often in Graves' patients compared to controls, the association was far stronger with AAD. This suggests that although absence of *CYP21A1P* in individuals with AAD may be partly explained by LD with HLA, this deletion might also be conferring an independent effect on disease susceptibility in support of the initial hypothesis. Expression analysis, combining qPCR and tissue *in situ* hybridisation, were then successfully used to identify *CYP21A1P* transcripts in both fetal adrenal tissue and in thymic tissue obtained from infants and children. The presence of these transcripts are again in keeping with our hypothesis and may indicate a role for the *CYP21A1P* pseudogene in induction of immune tolerance. This intriguing finding will now need to be investigated further. We plan to use Western blotting to determine whether a truncated CYP21A1P protein product is expressed in these tissues. The presence of a pseudogene protein product would add further weight to the hypothesis that *CYP21A1P* has a role in establishing and maintaining immune tolerance to steroidogenic machinery and that absence of *CYP21A1P* predisposes to AAD.

Taking a broader candidate gene approach, in this study, association analysis of twenty candidate genes in six European AAD cohorts was performed. Prior to this candidate gene study, loci implicated in the aetiology of AAD, and replicated in two cohorts, included *MHC*, *MICA*, *MHC2TA*, *CYP27B1*, *NLRP-1,*

PD-L1, *CTLA4* and *PTPN22.* Following correction for multiple testing, the results from the 20 candidate gene study suggest a role for polymorphisms at three loci already associated with other autoimmune diseases in AAD and a single novel locus, not previously associated with autoimmunity. Of the three loci previously associated with autoimmune disease, two of these, *NF-κB1* and *IL23A,* appear to influence AAD disease susceptibility only in certain populations: *NF-κB1* in individuals from the UK and *IL23A* in Italian individuals. The third locus, *STAT4,* appears to influence disease susceptibility more universally in Europeans, as indicated by meta-analysis. A single locus not previously associated with autoimmunity, *GATA3,* was also associated with AAD, but only in the UK population. Further investigation of case-control cohorts with other autoimmune conditions, including type 1 diabetes, Graves' disease and rheumatoid arthritis, failed to demonstrate any association with polymorphisms in *GATA3*. This suggests that this locus may not be influencing autoimmunity broadly, but may be specific to AAD susceptibility. This large candidate gene study has therefore added significantly to the growing list of susceptibility loci for AAD. In addition, this study has demonstrated significant heterogeneity between Caucasian European control cohorts. This finding is significant as it implies that, across individuals of European origin, often assumed to be genetically similar, the underlying genetic aetiology to AAD, and indeed other complex genetic disorders, may be subtly different. This could lead to difficulty replicating findings across cohorts from different countries and highlights the importance of carefully matching cases and controls in studies.

Applying a discovery-driven, genome-wide approach to the investigation of a complex disease can result in interesting and unpredictable findings as no prior hypothesis is required. In this study, genome-wide SNP genotyping technology has been used to conduct the first linkage analysis in AAD, including 23 families from the UK and Norway with two or more individuals with AAD. A genome-wide association study, taking the 50 individuals with AAD from these families and comparing their genotype data to that from the publicly available Wellcome Trust 1958 UK Birth Cohort control group was also conducted for the first time in this disease. A non-parametric linkage analysis identified a single region of linkage on chromosome 6, corresponding to the HLA region. SNPs in this same

region on chromosome 6 were also associated in the genome-wide association analysis. The linkage peak and cluster of associated SNPs, which corresponded to HLA, a known AAD susceptibility locus, validated the two approaches taken. In addition, a parametric linkage analysis, applying a rare dominant model, identified three further chromosomal regions linked to disease, on chromosomes 7, 9 and 18, which are of interest. A further linkage peak, on chromosome 3, was identified when 21OH autoantibody positivity was analysed as the trait of interest in a parametric analysis applying a dominant model. This intriguing finding suggests that the genetic determinants of overt disease may be different to those which determine autoantibody status and progression from autoantibody positivity to subclinical to overt disease. Finally, a cluster of SNPs on chromosome 2 was also identified in the genome-wide association study, suggesting possible aetiological variants in these regions.

To follow up these findings, we investigated 64 SNPs underlying the linked regions on chromosomes 7 and 18 in unrelated AAD case and control cohorts from the UK, Norway and Sweden. Unfortunately, this experiment did not reveal any susceptibility variants when multiple testing had been accounted for. This result was disappointing but perhaps predictable, as the linkage peaks were large, owing to the resolution of linkage analysis, and the majority of haplotypes in the two regions of interest were not represented in the data set. It is also possible that genetic determinants of familial AAD are different to those which confer susceptibility in the sporadic form of disease. Nevertheless, this work has generated some very interesting preliminary data and the findings now need to be replicated and investigated further. To this end, members of our multiplex AAD families are currently being investigated by whole exome sequencing by our Norwegian collaborators and the results of this are eagerly anticipated. In addition, collaborators in Sweden are currently collecting multiplex AAD families for a replication linkage study. Together, this further work should contribute significantly to our understanding of the genetic aetiology of AAD.

Despite the challenges associated with investigating a rare, complex genetic disorder such as AAD, it is very important that progress is made in the field.

Ultimately, it is hoped that, through uncovering the genetic determinants of AAD and other related autoimmune conditions, and understanding how these result in disease, we will gain a greater understanding of the pathogenesis of AAD and autoimmunity in general. This, in turn, will lead to new therapeutic interventions being developed. Advances in genetic technologies and analysis methods are bringing this possibility ever closer, providing endless opportunities for learning more about heritable but non-Mendelian conditions such as AAD. The whole human genome can now be sequenced to search for rare variants that were, until recently, difficult and expensive to investigate and this could prove very useful in searching for variants contributing to AAD. The combination of genomic and functional approaches, which aim to investigate how sequence variants confer disease susceptibility, offer exciting and unique insights into disease pathogenesis, with the possibility of developing new drug targets.

New therapies are much needed for AAD. Current clinical management focuses on replacing deficient steroid hormones, treating Addisonian crises when they arise and managing the consequences of long-term minor steroid over-replacement. While this can be considered adequate, it is far from optimal: individuals with AAD have a poorer quality of life and increased mortality when compared to unaffected, age-matched peers. Rather than managing problems as they arise, a more satisfactory approach would be to attempt to alter the natural history of the disease. This could be achieved, for example, by modulating the immune response in an attempt to halt the autoimmune process which causes AAD, perhaps resulting in restoration of tolerance to native adrenal antigens. Indeed, attempts at altering the natural history of AAD by using B cell depletion therapy in affected individuals early in the disease, while some residual steroidogenic function remains, have already been attempted in our research group with some promising initial results[494]. However, in order to develop more new therapeutic approaches, novel therapeutic targets are needed. It is only by investigating the underlying aetiology of this disease that we will come to understand its pathogenesis more fully. It is through these insights that the gap between bench and bedside will be bridged and we will better understand how to treat individuals with AAD optimally and how to counsel their relatives of their own risk of AAD in the future.

REFERENCES

1. Mitchell, A.L. and S.H. Pearce, *Autoimmune Addison disease: pathophysiology and genetic complexity.* Nat Rev Endocrinol, 2012. **8**(5): p. 306-16.
2. Hollander, G.A. and P. Peterson, *Learning to be tolerant: how T cells keep out of trouble.* J Intern Med, 2009. **265**(5): p. 541-61.
3. Itoh, M., et al., *Thymus and autoimmunity: production of CD25+CD4+ naturally anergic and suppressive T cells as a key function of the thymus in maintaining immunologic self-tolerance.* J Immunol, 1999. **162**(9): p. 5317-26.
4. Kuroda, N., et al., *Development of autoimmunity against transcriptionally unrepressed target antigen in the thymus of Aire-deficient mice.* J Immunol, 2005. **174**(4): p. 1862-70.
5. Ramsey, C., et al., *Aire deficient mice develop multiple features of APECED phenotype and show altered immune response.* Hum Mol Genet, 2002. **11**(4): p. 397-409.
6. Finnish-German APECED Consortium, *An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains.* Nat Genet, 1997. **17**(4): p. 399-403.
7. Sakaguchi, S., et al., *Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases.* J Immunol, 1995. **155**: p. 1151-1164.
8. Fontenot, J.D., M.A. Gavin, and A.Y. Rudensky, *Foxp3 programs the development and function of CD4+CD25+ regulatory T cells.* Nat Immunol, 2003. **4**(4): p. 330-6.
9. Hori, S., T. Nomura, and S. Sakaguchi, *Control of regulatory T cell development by the transcription factor Foxp3.* Science, 2003. **299**(5609): p. 1057-61.
10. Takahashi, T., et al., *Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state.* Int Immunol, 1998. **10**(12): p. 1969-80.
11. Read, S., V. Malmstrom, and F. Powrie, *Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation.* J Exp Med, 2000. **192**(2): p. 295-302.
12. Takahashi, T., et al., *Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4.* J Exp Med, 2000. **192**(2): p. 303-10.
13. Dieckmann, D., et al., *Ex vivo isolation and characterization of CD4(+)CD25(+) T cells with regulatory properties from human blood.* J Exp Med, 2001. **193**(11): p. 1303-10.

14. Annacker, O., et al., *CD25+ CD4+ T cells regulate the expansion of peripheral CD4 T cells through the production of IL-10.* J Immunol, 2001. **166**(5): p. 3008-18.
15. Maloy, K.J., et al., *CD4+CD25+ T(R) cells suppress innate immune pathology through cytokine-dependent mechanisms.* J Exp Med, 2003. **197**(1): p. 111-9.
16. Sakaguchi, S., et al., *Organ-specific autoimmune diseases induced in mice by elimination of T cell subset. I. Evidence for the active participation of T cells in natural self-tolerance; deficit of a T cell subset as a possible cause of autoimmune disease.* J Exp Med, 1985. **161**(1): p. 72-87.
17. Bennett, C.L., et al., *The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3.* Nat Genet, 2001. **27**(1): p. 20-1.
18. Hoyne, G.F., *Mechanisms that regulate peripheral immune responses to control organ-specific autoimmunity.* Clin Dev Immunol, 2011. **2011**: p. 294968.
19. Shlomchik, M.J., *Activating systemic autoimmunity: B's, T's, and tolls.* Curr Opin Immunol, 2009. **21**(6): p. 626-33.
20. von Boehmer, H. and F. Melchers, *Checkpoints in lymphocyte development and autoimmune disease.* Nat Immunol, 2010. **11**(1): p. 14-20.
21. Frauwirth, K.A., M.L. Alegre, and C.B. Thompson, *Induction of T cell anergy in the absence of CTLA-4/B7 interaction.* J Immunol, 2000. **164**(6): p. 2987-93.
22. Schwartz, R.H., *T cell anergy.* Annu Rev Immunol, 2003. **21**: p. 305-34.
23. Goodnow, C.C., et al., *Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice.* Nature, 1988. **334**(6184): p. 676-82.
24. Yarkoni, Y., A. Getahun, and J.C. Cambier, *Molecular underpinning of B-cell anergy.* Immunol Rev, 2010. **237**(1): p. 249-63.
25. Green, D.R., N. Droin, and M. Pinkoski, *Activation-induced cell death in T cells.* Immunol Rev, 2003. **193**: p. 70-81.
26. Lenardo, M., et al., *Mature T lymphocyte apoptosis--immune regulation in a dynamic and unpredictable antigenic environment.* Annu Rev Immunol, 1999. **17**: p. 221-53.
27. Ten, S., M. New, and N. Maclaren, *Clinical review 130: Addison's disease 2001.* J Clin Endocrinol Metab, 2001. **86**(7): p. 2909-22.
28. Addison, T., *On The Constitutional And Local Effects Of Disease Of The Supra-Renal Capsules.* 1855.
29. Nerup, J., *Addison's disease - a review of some clinical, pathological and immunological features.* Dan Med Bull, 1974. **21**(6): p. 201-17.
30. Soderbergh, A., et al., *Adrenal autoantibodies and organ-specific autoimmunity in patients with Addison's disease.* Clin Endocrinol (Oxf), 1996. **45**(4): p. 453-60.
31. Kong, M.F. and W. Jeffcoate, *Eighty-six cases of Addison's disease.* Clin Endocrinol (Oxf), 1994. **41**(6): p. 757-61.

32. Zelissen, P.M., E.J. Bast, and R.J. Croughs, *Associated autoimmunity in Addison's disease.* J Autoimmun, 1995. **8**(1): p. 121-30.
33. Laureti, S., et al., *Is the prevalence of Addison's disease underestimated?* J Clin Endocrinol Metab, 1999. **84**(5): p. 1762.
34. Lovas, K. and E.S. Husebye, *High prevalence and increasing incidence of Addison's disease in western Norway.* Clin Endocrinol (Oxf), 2002. **56**(6): p. 787-91.
35. Winqvist, O., F.A. Karlsson, and O. Kampe, *21-Hydroxylase, a major autoantigen in idiopathic Addison's disease.* Lancet, 1992. **339**(8809): p. 1559-62.
36. Nikoshkov, A., et al., *A conformation-dependent epitope in Addison's disease and other endocrinological autoimmune diseases maps to a carboxyl-terminal functional domain of human steroid 21-hydroxylase.* J Immunol, 1999. **162**(4): p. 2422-6.
37. Furmaniak, J., et al., *Autoimmune Addison's disease--evidence for a role of steroid 21-hydroxylase autoantibodies in adrenal insufficiency.* J Clin Endocrinol Metab, 1994. **79**(5): p. 1517-21.
38. Nikfarjam, L., et al., *Mechanism of inhibition of cytochrome P450 C21 enzyme activity by autoantibodies from patients with Addison's disease.* Eur J Endocrinol, 2005. **152**(1): p. 95-101.
39. Boscaro, M., et al., *Hormonal responses during various phases of autoimmune adrenal failure: no evidence for 21-hydroxylase enzyme activity inhibition in vivo.* J Clin Endocrinol Metab, 1996. **81**(8): p. 2801-4.
40. UniProt Consortium, *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Res, 2012. **40**(Database issue): p. D71-5.
41. Krohn, K., et al., *Identification by molecular cloning of an autoantigen associated with Addison's disease as steroid 17 alpha-hydroxylase.* Lancet, 1992. **339**(8796): p. 770-3.
42. Winqvist, O., et al., *Two different cytochrome P450 enzymes are the adrenal antigens in autoimmune polyendocrine syndrome type I and Addison's disease.* J Clin Invest, 1993. **92**(5): p. 2377-85.
43. Lovas, K., J.H. Loge, and E.S. Husebye, *Subjective health status in Norwegian patients with Addison's disease.* Clin Endocrinol (Oxf), 2002. **56**(5): p. 581-8.
44. Bergthorsdottir, R., et al., *Premature mortality in patients with Addison's disease: a population-based study.* J Clin Endocrinol Metab, 2006. **91**(12): p. 4849-53.
45. Betterle, C., et al., *The natural history of adrenal function in autoimmune patients with adrenal autoantibodies.* J Endocrinol, 1988. **117**(3): p. 467-75.
46. Coco, G., et al., *Estimated risk for developing autoimmune Addison's disease in patients with adrenal cortex autoantibodies.* J Clin Endocrinol Metab, 2006. **91**(5): p. 1637-45.
47. Torrejon, S., et al., *Long-lasting subclinical Addison's disease.* Exp Clin Endocrinol Diabetes, 2007. **115**(8): p. 530-2.

48. Smans, L.C. and P.M. Zelissen, *Partial recovery of adrenal function in a patient with autoimmune Addison's disease.* J Endocrinol Invest, 2008. **31**(7): p. 672-4.
49. Chakera, A.J. and B. Vaidya, *Spontaneously resolving Addison's disease.* QJM, 2012. **105**(11): p. 1113-5.
50. De Bellis, A., et al., *Remission of subclinical adrenocortical failure in subjects with adrenal autoantibodies.* J Clin Endocrinol Metab, 1993. **76**(4): p. 1002-7.
51. De Bellis, A.A., et al., *Time course of 21-hydroxylase antibodies and long-term remission of subclinical autoimmune adrenalitis after corticosteroid therapy: case report.* J Clin Endocrinol Metab, 2001. **86**(2): p. 675-8.
52. al Sabri, A.M., N. Smith, and A. Busuttil, *Sudden death due to auto-immune Addison's disease in a 12-year-old girl.* Int J Legal Med, 1997. **110**(5): p. 278-80.
53. Hayashi, Y., et al., *Focal lymphocytic infiltration in the adrenal cortex of the elderly: immunohistological analysis of infiltrating lymphocytes.* Clin Exp Immunol, 1989. **77**(1): p. 101-5.
54. Falorni, A., et al., *21-hydroxylase autoantibodies in adult patients with endocrine autoimmune diseases are highly specific for Addison's disease. Belgian Diabetes Registry.* Clin Exp Immunol, 1997. **107**(2): p. 341-6.
55. Myhre, A.G., et al., *Autoimmune adrenocortical failure in Norway autoantibodies and human leukocyte antigen class II associations related to clinical features.* J Clin Endocrinol Metab, 2002. **87**(2): p. 618-23.
56. Nagamine, K., et al., *Positional cloning of the APECED gene.* Nat Genet, 1997. **17**(4): p. 393-8.
57. Pearce, S.H., et al., *A common and recurrent 13-bp deletion in the autoimmune regulator gene in British kindreds with autoimmune polyendocrinopathy type 1.* Am J Hum Genet, 1998. **63**(6): p. 1675-84.
58. Pearce, S.H. and T.D. Cheetham, *Autoimmune polyendocrinopathy syndrome type 1: treat with kid gloves.* Clin Endocrinol (Oxf), 2001. **54**(4): p. 433-5.
59. Zlotogora, J. and M.S. Shapiro, *Polyglandular autoimmune syndrome type I among Iranian Jews.* J Med Genet, 1992. **29**(11): p. 824-6.
60. Rosatelli, M.C., et al., *A common mutation in Sardinian autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy patients.* Hum Genet, 1998. **103**(4): p. 428-34.
61. Perheentupa, J., *Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy.* J Clin Endocrinol Metab, 2006. **91**(8): p. 2843-50.
62. Wolff, A.S., et al., *Autoimmune polyendocrine syndrome type 1 in Norway: phenotypic variation, autoantibodies, and novel mutations in the autoimmune regulator gene.* J Clin Endocrinol Metab, 2007. **92**(2): p. 595-603.
63. Neufeld, M., N.K. Maclaren, and R.M. Blizzard, *Two types of autoimmune Addison's disease associated with different polyglandular*

autoimmune (PGA) syndromes. Medicine (Baltimore), 1981. **60**(5): p. 355-62.

64. Ahonen, P., et al., *Clinical variation of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) in a series of 68 patients.* N Engl J Med, 1990. **322**(26): p. 1829-36.
65. Betterle, C., N.A. Greggio, and M. Volpato, *Clinical review 93: Autoimmune polyglandular syndrome type 1.* J Clin Endocrinol Metab, 1998. **83**(4): p. 1049-55.
66. Betterle, C., et al., *Type 2 polyglandular autoimmune disease (Schmidt's syndrome).* J Pediatr Endocrinol Metab, 1996. **9 Suppl 1**: p. 113-23.
67. Dittmar, M. and G.J. Kahaly, *Polyglandular autoimmune syndromes: immunogenetics and long-term follow-up.* J Clin Endocrinol Metab, 2003. **88**(7): p. 2983-92.
68. Heggarty, H., *Addison's disease in identical twins.* Br Med J, 1968. **1**(5591): p. 559.
69. Simmonds, J.P. and J. Lister, *Auto-immune Addison's disease in identical twins.* Postgrad Med J, 1978. **54**(634): p. 552-4.
70. Smith, M.E., J. Gough, and O.P. Galpin, *ADDISON'S DISEASE IN IDENTICAL TWINS.* Br Med J, 1963. **2**(5368): p. 1316.
71. Russell, G.A., et al., *Autoimmune Addison's disease and thyrotoxic thyroiditis presenting as encephalopathy in twins.* Arch Dis Child, 1991. **66**(3): p. 350-2.
72. Fairchild, R.S., R.N. Schimke, and N.I. Abdou, *Immunoregulation abnormalities in familial Addison's disease.* J Clin Endocrinol Metab, 1980. **51**(5): p. 1074-7.
73. Hewitt, P.H., *Addison's disease occurring in sisters.* Br Med J, 1957. **2**(5060): p. 1530-1.
74. Penrose, L.S., *The genetical background of common diseases.* Acta Genet Stat Med, 1953. **4**(2-3): p. 257-65.
75. Risch, N., *Assessing the role of HLA-linked and unlinked determinants of disease.* Am J Hum Genet, 1987. **40**(1): p. 1-14.
76. Vaidya, B., P. Kendall-Taylor, and S.H. Pearce, *The genetics of autoimmune thyroid disease.* J Clin Endocrinol Metab, 2002. **87**(12): p. 5385-97.
77. Kuster, W., et al., *The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives.* Am J Med Genet, 1989. **32**(1): p. 105-8.
78. Tysk, C., et al., *Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking.* Gut, 1988. **29**(7): p. 990-6.
79. Hemminki, K., et al., *Familial association between type 1 diabetes and other autoimmune and related diseases.* Diabetologia, 2009. **52**(9): p. 1820-8.
80. Vyse, T.J. and J.A. Todd, *Genetic analysis of autoimmune disease.* Cell, 1996. **85**(3): p. 311-8.
81. Lander, E.S. and N.J. Schork, *Genetic dissection of complex traits.* Science, 1994. **265**(5181): p. 2037-48.

82. Kruglyak, L., et al., *Parametric and nonparametric linkage analysis: a unified multipoint approach.* Am J Hum Genet, 1996. **58**(6): p. 1347-63.
83. Ott, J., *Analysis of Human Genetic Linkage*. 1996: John Hopkins University Press, Baltimore.
84. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
85. Healy, D.G., *Case-control studies in the genomic era: a clinician's guide.* Lancet Neurol, 2006. **5**(8): p. 701-7.
86. National Institutes of Health office of rare diseases research. *Glossary*. Available from: http://rarediseases.info.nih.gov.
87. Chotai, J., *On the lod score method in linkage analysis.* Ann Hum Genet, 1984. **48**(Pt 4): p. 359-78.
88. Morton, N.E., *Sequential tests for the detection of linkage.* Am J Hum Genet, 1955. **7**(3): p. 277-318.
89. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.* Nat Genet, 1995. **11**(3): p. 241-7.
90. Abreu, P.C., S.E. Hodge, and D.A. Greenberg, *Quantification of type I error probabilities for heterogeneity LOD scores.* Genet Epidemiol, 2002. **22**(2): p. 156-69.
91. Altmuller, J., et al., *Genomewide scans of complex human diseases: true linkage is hard to find.* Am J Hum Genet, 2001. **69**(5): p. 936-50.
92. Ott, J. and J. Hoh, *Statistical approaches to gene mapping.* Am J Hum Genet, 2000. **67**(2): p. 289-94.
93. Morton, N.E., *The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type.* Am J Hum Genet, 1956. **8**(2): p. 80-96.
94. Clerget-Darpoux, F., C. Bonaiti-Pellie, and J. Hochez, *Effects of misspecifying genetic parameters in lod score analysis.* Biometrics, 1986. **42**(2): p. 393-9.
95. Blackwelder, W.C. and R.C. Elston, *A comparison of sib-pair linkage tests for disease susceptibility loci.* Genet Epidemiol, 1985. **2**(1): p. 85-97.
96. Risch, N., *Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs.* Am J Hum Genet, 1990. **46**(2): p. 242-53.
97. Whittemore, A.S. and J. Halpern, *A class of tests for linkage using affected pedigree members.* Biometrics, 1994. **50**(1): p. 118-27.
98. Curtis, D. and P.C. Sham, *Using risk calculation to implement an extended relative pair analysis.* Ann Hum Genet, 1994. **58**(Pt 2): p. 151-62.
99. Curtis, D. and P.C. Sham, *Model-free linkage analysis using likelihoods.* Am J Hum Genet, 1995. **57**(3): p. 703-16.
100. Risch, N., *A note on multiple testing procedures in linkage analysis.* Am J Hum Genet, 1991. **48**(6): p. 1058-64.

101. Davies, J.L., et al., *A genome-wide search for human type 1 diabetes susceptibility genes.* Nature, 1994. **371**(6493): p. 130-6.
102. The International HapMap Consortium, *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
103. 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.
104. Curtis, D., *Use of siblings as controls in case-control association studies.* Ann Hum Genet, 1997. **61**(Pt 4): p. 319-33.
105. Cardon, L.R. and J.I. Bell, *Association study designs for complex diseases.* Nat Rev Genet, 2001. **2**(2): p. 91-9.
106. Gambaro, G., F. Anglani, and A. D'Angelo, *Association studies of genetic polymorphisms and complex disease.* Lancet, 2000. **355**(9200): p. 308-11.
107. Pearce, S.H. and T.R. Merriman, *Genetic progress towards the molecular basis of autoimmunity.* Trends Mol Med, 2006. **12**(2): p. 90-8.
108. Terwilliger, J.D. and K.M. Weiss, *Linkage disequilibrium mapping of complex disease: fantasy or reality?* Curr Opin Biotechnol, 1998. **9**(6): p. 578-94.
109. Weiss, K.M. and J.D. Terwilliger, *How many diseases does it take to map a gene with SNPs?* Nat Genet, 2000. **26**(2): p. 151-7.
110. Kennedy, D., *The old file-drawer problem.* Science, 2004. **305**(5683): p. 451.
111. Pompanon, F., et al., *Genotyping errors: causes, consequences and solutions.* Nat Rev Genet, 2005. **6**(11): p. 847-59.
112. Salanti, G., et al., *Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power.* Eur J Hum Genet, 2005. **13**(7): p. 840-8.
113. Allahabadia, A., et al., *Lack of association between polymorphism of the thyrotropin receptor gene and Graves' disease in United Kingdom and Hong Kong Chinese patients: case control and family-based studies.* Thyroid, 1998. **8**(9): p. 777-80.
114. Patterson, M. and L. Cardon, *Replication publication.* PLoS Biol, 2005. **3**(9): p. e327.
115. Hattersley, A.T. and M.I. McCarthy, *What makes a good genetic association study?* Lancet, 2005. **366**(9493): p. 1315-23.
116. Vaidya, B., et al., *Association analysis of the cytotoxic T lymphocyte antigen-4 (CTLA-4) and autoimmune regulator-1 (AIRE-1) genes in sporadic autoimmune Addison's disease.* J Clin Endocrinol Metab, 2000. **85**(2): p. 688-91.
117. Todd, J.A., et al., *Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.* Nat Genet, 2007. **39**(7): p. 857-64.
118. Begovich, A.B., et al., *A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis.* Am J Hum Genet, 2004. **75**(2): p. 330-7.

119. Levy, S., et al., *The diploid genome sequence of an individual human.* PLoS Biol, 2007. **5**(10): p. e254.
120. Wheeler, D.A., et al., *The complete genome of an individual by massively parallel DNA sequencing.* Nature, 2008. **452**(7189): p. 872-6.
121. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
122. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
123. de Vries, H.G., et al., *Prevalence of delta F508 cystic fibrosis carriers in The Netherlands: logistic regression on sex, age, region of residence and number of offspring.* Hum Genet, 1997. **99**(1): p. 74-9.
124. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
125. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
126. Gonzaga-Jauregui, C., J.R. Lupski, and R.A. Gibbs, *Human genome sequencing in health and disease.* Annu Rev Med, 2012. **63**: p. 35-61.
127. Mardis, E.R., *The $1,000 genome, the $100,000 analysis?* Genome Med, 2010. **2**(11): p. 84.
128. 1000 Genomes Project Consortium, *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
129. Gibson, G., *Rare and common variants: twenty arguments.* Nat Rev Genet, 2012. **13**(2): p. 135-45.
130. Botstein, D. and N. Risch, *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* Nat Genet, 2003. **33 Suppl**: p. 228-37.
131. Pritchard, J.K. and N.J. Cox, *The allelic architecture of human disease genes: common disease-common variant...or not?* Hum Mol Genet, 2002. **11**(20): p. 2417-23.
132. Reich, D.E. and E.S. Lander, *On the allelic spectrum of human disease.* Trends Genet, 2001. **17**(9): p. 502-10.
133. Lango-Allen, H., et al., *Hundreds of variants clustered in genomic loci and biological pathways affect human height.* Nature, 2010. **467**(7317): p. 832-8.
134. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.
135. Bodmer, W. and C. Bonilla, *Common and rare variants in multifactorial susceptibility to common diseases.* Nat Genet, 2008. **40**(6): p. 695-701.
136. Cirulli, E.T. and D.B. Goldstein, *Uncovering the roles of rare variants in common disease through whole-genome sequencing.* Nat Rev Genet, 2010. **11**(6): p. 415-25.
137. Eichler, E.E., et al., *Missing heritability and strategies for finding the underlying causes of complex disease.* Nat Rev Genet, 2010. **11**(6): p. 446-50.

138. Feinberg, A.P., *Phenotypic plasticity and the epigenetics of human disease.* Nature, 2007. **447**(7143): p. 433-40.
139. Drexhage, H., *Autoimmune Adrenocortical Failure*, in *Autoimmune Endocrinopathies*, R. Volpe, Editor. 1999, Humana Press.
140. Greco, D.S., *Hypoadrenocorticism in small animals.* Clin Tech Small Anim Pract, 2007. **22**(1): p. 32-5.
141. Peterson, M.E., P.P. Kintzer, and P.H. Kass, *Pretreatment clinical and laboratory findings in dogs with hypoadrenocorticism: 225 cases (1979-1993).* J Am Vet Med Assoc, 1996. **208**(1): p. 85-91.
142. Famula, T.R., J.M. Belanger, and A.M. Oberbauer, *Heritability and complex segregation analysis of hypoadrenocorticism in the standard poodle.* J Small Anim Pract, 2003. **44**(1): p. 8-12.
143. Oberbauer, A.M., et al., *Inheritance of hypoadrenocorticism in bearded collies.* Am J Vet Res, 2002. **63**(5): p. 643-7.
144. Pedersen, N.C., *A review of immunologic diseases of the dog.* Vet Immunol Immunopathol, 1999. **69**(2-4): p. 251-342.
145. Smallwood, L.J. and J.A. Barsanti, *Hypoadrenocorticism in a family of leonbergers.* J Am Anim Hosp Assoc, 1995. **31**(4): p. 301-5.
146. Bowen, D., M. Schaer, and W. Riley, *Autoimmune polyglandular syndrome in a dog: a case report.* J Am Anim Hosp Assoc, 1985. **22**: p. 649-654.
147. Maclaren, N.K. and W.J. Riley, *Inherited susceptibility to autoimmune Addison's disease is linked to human leukocyte antigens-DR3 and/or DR4, except when associated with type I autoimmune polyglandular syndrome.* J Clin Endocrinol Metab, 1986. **62**(3): p. 455-9.
148. Boehm, B.O., et al., *The HLA-DQ beta non-Asp-57 allele: a predictor of future insulin-dependent diabetes mellitus in patients with autoimmune Addison's disease.* Tissue Antigens, 1991. **37**(3): p. 130-2.
149. Huang, W., et al., *Although DR3-DQB1*0201 may be associated with multiple component diseases of the autoimmune polyglandular syndromes, the human leukocyte antigen DR4-DQB1*0302 haplotype is implicated only in beta-cell autoimmunity.* J Clin Endocrinol Metab, 1996. **81**(7): p. 2559-63.
150. Partanen, J., et al., *Major histocompatibility complex class II and III in Addison's disease MHC alleles do not predict autoantibody specificity and 21-hydroxylase gene polymorphism has no independent role in disease susceptibility.* Human Immunology, 1994. **41**(2): p. 135-140.
151. Park, Y.S., et al., *Additional association of intra-MHC genes, MICA and D6S273, with Addison's disease.* Tissue Antigens, 2002. **60**(2): p. 155-63.
152. Yu, L., et al., *DRB1*04 and DQ alleles: expression of 21-hydroxylase autoantibodies and risk of progression to Addison's disease.* J Clin Endocrinol Metab, 1999. **84**(1): p. 328-35.
153. Peterson, P., et al., *Steroid 21-hydroxylase gene polymorphism in Addison's disease patients.* Tissue Antigens, 1995. **46**(1): p. 63-7.

154. Gambelunghe, G., et al., *Microsatellite polymorphism of the MHC class I chain-related (MIC-A and MIC-B) genes marks the risk for autoimmune Addison's disease.* J Clin Endocrinol Metab, 1999. **84**(10): p. 3701-7.
155. Triolo, T.M., et al., *Homozygosity of the polymorphism MICA5.1 identifies extreme risk of progression to overt adrenal insufficiency among 21-hydroxylase antibody-positive patients with type 1 diabetes.* J Clin Endocrinol Metab, 2009. **94**(11): p. 4517-23.
156. Dziembowska, M., et al., *Three novel mutations of the CIITA gene in MHC class II-deficient patients with a severe immunodeficiency.* Immunogenetics, 2002. **53**(10-11): p. 821-9.
157. Eyre, S., et al., *Investigation of the MHC2TA gene, associated with rheumatoid arthritis in a Swedish population, in a UK rheumatoid arthritis cohort.* Arthritis Rheum, 2006. **54**(11): p. 3417-22.
158. Koizumi, K., et al., *Single nucleotide polymorphisms in the gene encoding the major histocompatibility complex class II transactivator (CIITA) in systemic lupus erythematosus.* Ann Rheum Dis, 2005. **64**(6): p. 947-50.
159. Ghaderi, M., et al., *MHC2TA single nucleotide polymorphism and genetic risk for autoimmune adrenal insufficiency.* J Clin Endocrinol Metab, 2006. **91**(10): p. 4107-11.
160. Skinningsrud, B., et al., *Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency.* J Clin Endocrinol Metab, 2008. **93**(9): p. 3310-7.
161. Brunet, J.F., et al., *A new member of the immunoglobulin superfamily--CTLA-4.* Nature, 1987. **328**(6127): p. 267-70.
162. Marron, M.P., et al., *Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups.* Hum Mol Genet, 1997. **6**(8): p. 1275-82.
163. Nistico, L., et al., *The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry.* Hum Mol Genet, 1996. **5**(7): p. 1075-80.
164. Akamizu, T., et al., *Association of autoimmune thyroid disease with microsatellite markers for the thyrotropin receptor gene and CTLA-4 in Japanese patients.* Thyroid, 2000. **10**(10): p. 851-8.
165. Kotsa, K., P.F. Watson, and A.P. Weetman, *A CTLA-4 gene polymorphism is associated with both Graves disease and autoimmune hypothyroidism.* Clin Endocrinol (Oxf), 1997. **46**(5): p. 551-4.
166. Sale, M.M., et al., *Association of autoimmune thyroid disease with a microsatellite marker for the thyrotropin receptor gene and CTLA-4 in a Japanese population.* Proc Assoc Am Physicians, 1997. **109**(5): p. 453-61.
167. Yanagawa, T., et al., *CTLA-4 gene polymorphism associated with Graves' disease in a Caucasian population.* J Clin Endocrinol Metab, 1995. **80**(1): p. 41-5.
168. Seidl, C., et al., *CTLA4 codon 17 dimorphism in patients with rheumatoid arthritis.* Tissue Antigens, 1998. **51**(1): p. 62-6.

169. Vaidya, B., et al., *An association between the CTLA4 exon 1 polymorphism and early rheumatoid arthritis with autoimmune endocrinopathies.* Rheumatology (Oxford), 2002. **41**(2): p. 180-3.
170. Kemp, E.H., et al., *A cytotoxic T lymphocyte antigen-4 (CTLA-4) gene polymorphism is associated with autoimmune Addison's disease in English patients.* Clin Endocrinol (Oxf), 1998. **49**(5): p. 609-13.
171. Blomhoff, A., et al., *Polymorphisms in the cytotoxic T lymphocyte antigen-4 gene region confer susceptibility to Addison's disease.* J Clin Endocrinol Metab, 2004. **89**(7): p. 3474-6.
172. Gerold, K.D., et al., *The soluble CTLA-4 splice variant protects from type 1 diabetes and potentiates regulatory T-cell function.* Diabetes, 2011. **60**(7): p. 1955-63.
173. Ueda, H., et al., *Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease.* Nature, 2003. **423**(6939): p. 506-11.
174. Daroszewski, J., et al., *Soluble CTLA-4 receptor an immunological marker of Graves' disease and severity of ophthalmopathy is associated with CTLA-4 Jo31 and CT60 gene polymorphisms.* Eur J Endocrinol, 2009. **161**(5): p. 787-93.
175. Oaks, M.K. and K.M. Hallett, *Cutting edge: a soluble form of CTLA-4 in patients with autoimmune thyroid disease.* J Immunol, 2000. **164**(10): p. 5015-8.
176. Kahles, H., et al., *Sex-specific association of PTPN22 1858T with type 1 diabetes but not with Hashimoto's thyroiditis or Addison's disease in the German population.* Eur J Endocrinol, 2005. **153**(6): p. 895-9.
177. Velaga, M.R., et al., *The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease.* J Clin Endocrinol Metab, 2004. **89**(11): p. 5862-5.
178. Skinningsrud, B., et al., *Mutation screening of PTPN22: association of the 1858T-allele with Addison's disease.* Eur J Hum Genet, 2008. **16**(8): p. 977-82.
179. Roycroft, M., et al., *The tryptophan 620 allele of the lymphoid tyrosine phosphatase (PTPN22) gene predisposes to autoimmune Addison's disease.* Clin Endocrinol (Oxf), 2009. **70**(3): p. 358-62.
180. Arechiga, A.F., et al., *Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling.* J Immunol, 2009. **182**(6): p. 3343-7.
181. Rieck, M., et al., *Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes.* J Immunol, 2007. **179**(7): p. 4704-10.
182. Vang, T., et al., *Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant.* Nat Genet, 2005. **37**(12): p. 1317-9.
183. Zhang, J., et al., *The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness.* Nat Genet, 2011. **43**(9): p. 902-7.

184. Brown, J.A., et al., *Blockade of programmed death-1 ligands on dendritic cells enhances T cell activation and cytokine production.* J Immunol, 2003. **170**(3): p. 1257-66.
185. Hayashi, M., et al., *Association of an A/C single nucleotide polymorphism in programmed cell death-ligand 1 gene with Graves' disease in Japanese patients.* Eur J Endocrinol, 2008. **158**(6): p. 817-22.
186. Mitchell, A.L., et al., *Programmed death ligand 1 (PD-L1) gene variants contribute to autoimmune Addison's disease and Graves' disease susceptibility.* J Clin Endocrinol Metab, 2009. **94**(12): p. 5139-45.
187. Shibuya, A., et al., *DNAM-1, a novel adhesion molecule involved in the cytolytic function of T lymphocytes.* Immunity, 1996. **4**(6): p. 573-81.
188. Hafler, J.P., et al., *CD226 Gly307Ser association with multiple autoimmune diseases.* Genes Immun, 2009. **10**(1): p. 5-10.
189. Lofgren, S.E., et al., *A 3'-untranslated region variant is associated with impaired expression of CD226 in T and natural killer T cells and is associated with susceptibility to systemic lupus erythematosus.* Arthritis Rheum, 2010. **62**(11): p. 3404-14.
190. Gan, E.H., et al., *The role of a nonsynonymous CD226 (DNAX-accessory molecule-1) variant (Gly 307Ser) in isolated Addison's disease and autoimmune polyendocrinopathy type 2 pathogenesis.* Clin Endocrinol (Oxf), 2011. **75**(2): p. 165-8.
191. Capon, F., et al., *Fine mapping of the PSORS4 psoriasis susceptibility region on chromosome 1q21.* J Invest Dermatol, 2001. **116**(5): p. 728-30.
192. Kochi, Y., et al., *A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities.* Nat Genet, 2005. **37**(5): p. 478-85.
193. Kyogoku, C., et al., *Fcgamma receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: contribution of FCGR2B to genetic susceptibility.* Arthritis Rheum, 2002. **46**(5): p. 1242-54.
194. Ehrhardt, G.R., et al., *Fc receptor-like proteins (FCRL): immunomodulators of B cell function.* Adv Exp Med Biol, 2007. **596**: p. 155-62.
195. Owen, C.J., et al., *Analysis of the Fc receptor-like-3 (FCRL3) locus in Caucasians with autoimmune disorders suggests a complex pattern of disease association.* J Clin Endocrinol Metab, 2007. **92**(3): p. 1106-11.
196. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 599-603.
197. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease.* Nature, 2001. **411**(6837): p. 603-6.
198. Jin, Y., et al., *Genetic variations in NALP1 are associated with generalized vitiligo in a Romanian population.* J Invest Dermatol, 2007. **127**(11): p. 2558-62.
199. Magitta, N.F., et al., *A coding polymorphism in NALP1 confers risk for autoimmune Addison's disease and type 1 diabetes.* Genes Immun, 2009. **10**(2): p. 120-4.
200. Zurawek, M., et al., *A coding variant in NLRP1 is associated with autoimmune Addison's disease.* Hum Immunol, 2010. **71**(5): p. 530-4.

201. Hakonarson, H., et al., *A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene.* Nature, 2007. **448**(7153): p. 591-4.
202. Dubois, P.C., et al., *Multiple common variants for celiac disease influencing immune gene expression.* Nat Genet, 2010. **42**(4): p. 295-302.
203. Marquez, A., et al., *Specific association of a CLEC16A/KIAA0350 polymorphism with NOD2/CARD15(-) Crohn's disease patients.* Eur J Hum Genet, 2009. **17**(10): p. 1304-8.
204. Skinningsrud, B., et al., *A CLEC16A variant confers risk for juvenile idiopathic arthritis and anti-CCP negative rheumatoid arthritis.* Ann Rheum Dis, 2009.
205. Zoledziewska, M., et al., *Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia.* Genes Immun, 2009. **10**(1): p. 15-7.
206. Mathieu, C., et al., *Prevention of autoimmune diabetes in NOD mice by 1,25 dihydroxyvitamin D3.* Diabetologia, 1994. **37**(6): p. 552-8.
207. Zella, J.B. and H.F. DeLuca, *Vitamin D and autoimmune diabetes.* J Cell Biochem, 2003. **88**(2): p. 216-22.
208. Pani, M.A., et al., *A polymorphism within the vitamin D-binding protein gene is associated with Graves' disease but not with Hashimoto's thyroiditis.* J Clin Endocrinol Metab, 2002. **87**(6): p. 2564-7.
209. Lopez, E.R., et al., *A promoter polymorphism of the CYP27B1 gene is associated with Addison's disease, Hashimoto's thyroiditis, Graves' disease and type 1 diabetes mellitus in Germans.* Eur J Endocrinol, 2004. **151**(2): p. 193-7.
210. Jennings, C.E., et al., *A haplotype of the CYP27B1 promoter is associated with autoimmune Addison's disease but not with Graves' disease in a UK population.* J Mol Endocrinol, 2005. **34**(3): p. 859-63.
211. Bjornvold, M., et al., *Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes.* Diabetologia, 2008. **51**(4): p. 589-96.
212. EURODIAB ACE Study Group, *Variation and trends in incidence of childhood diabetes in Europe. .* Lancet, 2000. **355**(9207): p. 873-6.
213. Husebye, E.S., et al., *Autoantibodies against aromatic L-amino acid decarboxylase in autoimmune polyendocrine syndrome type I.* J Clin Endocrinol Metab, 1997. **82**(1): p. 147-50.
214. Gabriel, S., L. Ziaugra, and D. Tabbaa, *SNP genotyping using the Sequenom MassARRAY iPLEX platform.* Curr Protoc Hum Genet, 2009. **Chapter 2**: p. Unit 2 12.
215. Hubbard, T., et al., *The Ensembl genome database project.* Nucleic Acids Res, 2002. **30**(1): p. 38-41.
216. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
217. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

218. Woolf, B., *On estimating the relation between blood group and disease.* Ann Hum Genet, 1955. **19**(4): p. 251-3.
219. Bland, J.M. and D.G. Altman, *Statistics notes. The odds ratio.* BMJ, 2000. **320**(7247): p. 1468.
220. Dudbridge, F., *Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data.* Hum Hered, 2008. **66**(2): p. 87-98.
221. The Cochrane Collaboration, *Review Manager (RevMan) [Computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre.* 2011.
222. Wittes, J., *Clinical trials must cope better with multiplicity.* Nat Med, 2012. **18**(11): p. 1607.
223. Dunn, O.J., *Multiple Comparisons Among Means.* Journal of the American Statistical Association, 1961. **56**(293): p. 52-64.
224. Kumar, R., et al., *The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA.* Bioorg Med Chem Lett, 1998. **8**(16): p. 2219-22.
225. Petersen, M., et al., *The conformations of locked nucleic acids (LNA).* J Mol Recognit, 2000. **13**(1): p. 44-53.
226. McTigue, P.M., R.J. Peterson, and J.D. Kahn, *Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation.* Biochemistry, 2004. **43**(18): p. 5388-405.
227. Jacobsen, N., et al., *LNA-enhanced detection of single nucleotide polymorphisms in the apolipoprotein E.* Nucleic Acids Res, 2002. **30**(19): p. e100.
228. Johnson, M.P., L.M. Haupt, and L.R. Griffiths, *Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR.* Nucleic Acids Res, 2004. **32**(6): p. e55.
229. Darnell, D.K., et al., *MicroRNA expression during chick embryo development.* Dev Dyn, 2006. **235**(11): p. 3156-65.
230. Kloosterman, W.P., et al., *In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes.* Nat Methods, 2006. **3**(1): p. 27-9.
231. Wienholds, E., et al., *MicroRNA expression in zebrafish embryonic development.* Science, 2005. **309**(5732): p. 310-1.
232. Rozen, S. and H. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.* Methods Mol Biol, 2000. **132**: p. 365-86.
233. Terwilliger, J.D. and J. Ott, *Handbook of Human Genetic Linkage*. 1994: Johns Hopkins University Press, Baltimore.
234. Ott, J., *Computer-simulation methods in human linkage analysis.* Proc Natl Acad Sci U S A, 1989. **86**(11): p. 4175-8.
235. Weeks, D.E., J. Ott, and G.M. Lathrop, *SLINK: A general simulation program for linkage analysis.* American Journal of Human Genetics, 1990. **47**: p. A204.
236. Abecasis, G.R., et al., *Merlin--rapid analysis of dense genetic maps using sparse gene flow trees.* Nat Genet, 2002. **30**(1): p. 97-101.

237. StataCorp, *Stata Statistical Software: Release 12*. 2011, College Station, StataCorp LP: Texas.
238. Matise, T.C., et al., *A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set.* Am J Hum Genet, 2003. **73**(2): p. 271-84.
239. Hacia, J.G., et al., *Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays.* Genome Res, 1998. **8**(12): p. 1245-58.
240. Affymetrix. *Genome-Wide Human SNP Array 6.0*. [cited 2010 20/10/2010]; Available from: www.affymetrix.com.
241. Weale, M.E., *Quality control for genome-wide association studies*, in *Genetic Variation: Methods and Protocols*. Springer. p. 341-72.
242. R Core Team, *R: A Language and Environment for Statistical Computing*. 2012, Vienna, Austria: R Foundation for Statistical Computing
243. Howey, R. and H.J. Cordell. *MapThin*. 2011 01/07/2012]; Available from: http://www.staff.ncl.ac.uk/richard.howey/mapthin/.
244. Hodge, S.E., V.J. Vieland, and D.A. Greenberg, *HLODs remain powerful tools for detection of linkage in the presence of genetic heterogeneity.* Am J Hum Genet, 2002. **70**(2): p. 556-9.
245. Vieland, V.J. and M. Logue, *HLODs, trait models, and ascertainment: implications of admixture for parameter estimation and linkage detection.* Hum Hered, 2002. **53**(1): p. 23-35.
246. Kang, H.M., et al., *Variance component model to account for sample structure in genome-wide association studies.* Nat Genet, 2010. **42**(4): p. 348-54.
247. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics, 2010. **26**(18): p. 2336-7.
248. Jacob, F. and J. Monod, *Genetic regulatory mechanisms in the synthesis of proteins.* J Mol Biol, 1961. **3**: p. 318-56.
249. Jacq, C., J.R. Miller, and G.G. Brownlee, *A pseudogene structure in 5S DNA of Xenopus laevis.* Cell, 1977. **12**(1): p. 109-20.
250. Vanin, E.F., *Processed pseudogenes: characteristics and evolution.* Annu Rev Genet, 1985. **19**: p. 253-72.
251. Wilde, C.D., *Pseudogenes.* CRC Crit Rev Biochem, 1986. **19**(4): p. 323-52.
252. McCarrey, J.R., *Molecular evolution of the human Pgk-2 retroposon.* Nucleic Acids Res, 1990. **18**(4): p. 949-55.
253. Little, P.F., *Globin pseudogenes.* Cell, 1982. **28**(4): p. 683-4.
254. Blanchong, C.A., et al., *Deficiencies of human complement component C4A and C4B and heterozygosity in length variants of RP-C4-CYP21-TNX (RCCX) modules in caucasians. The load of RCCX genetic diversity on major histocompatibility complex-associated disease.* J Exp Med, 2000. **191**(12): p. 2183-96.
255. Takemori, S. and S. Kominami, *The role of cytochromes P-450 in adrenal steroidogenesis.* Trends Biochem. Sci., 1984(9): p. 393-396.

256. Rodrigues, N.R., et al., *Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia.* EMBO J, 1987. **6**(6): p. 1653-61.
257. Goto, M., et al., *In humans, early cortisol biosynthesis provides a mechanism to safeguard female sexual development.* J Clin Invest, 2006. **116**(4): p. 953-60.
258. Pezzi, V., et al., *Profiling transcript levels for steroidogenic enzymes in fetal tissues.* J Steroid Biochem Mol Biol, 2003. **87**(2-3): p. 181-9.
259. Higashi, Y., et al., *Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene.* Proc Natl Acad Sci U S A, 1986. **83**(9): p. 2841-5.
260. Pruitt, K.D., et al., *The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes.* Genome Res, 2009. **19**(7): p. 1316-23.
261. Donohoue, P.A., et al., *Gene conversion in salt-losing congenital adrenal hyperplasia with absent complement C4B protein.* J Clin Endocrinol Metab, 1986. **62**(5): p. 995-1002.
262. Werkmeister, J.W., et al., *Frequent deletion and duplication of the steroid 21-hydroxylase genes.* Am J Hum Genet, 1986. **39**(4): p. 461-9.
263. White, P.C., et al., *Characterization of frequent deletions causing steroid 21-hydroxylase deficiency.* Proc Natl Acad Sci U S A, 1988. **85**(12): p. 4436-40.
264. Balakirev, E.S. and F.J. Ayala, *Pseudogenes: are they "junk" or functional DNA?* Annu Rev Genet, 2003. **37**: p. 123-51.
265. Bristow, J., et al., *Abundant adrenal-specific transcription of the human P450c21A "pseudogene".* J Biol Chem, 1993. **268**(17): p. 12919-24.
266. Darnell, D.K., et al., *Whole mount in situ hybridization detection of mRNAs using short LNA containing DNA oligonucleotide probes.* RNA, 2010. **16**(3): p. 632-7.
267. Partanen, J., et al., *Major histocompatibility complex class II and III in Addison's disease. MHC alleles do not predict autoantibody specificity and 21-hydroxylase gene polymorphism has no independent role in disease susceptibility.* Hum Immunol, 1994. **41**(2): p. 135-40.
268. Yang, Y., et al., *Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans.* Am J Hum Genet, 2007. **80**(6): p. 1037-54.
269. Bertrams, J., et al., *Gene and haplotype frequencies of the fourth component of complement (C4) in type 1 diabetics and normal controls.* Immunobiology, 1984. **166**(4-5): p. 335-44.
270. Stewart, C.A., et al., *Complete MHC haplotype sequencing for common disease gene mapping.* Genome Res, 2004. **14**(6): p. 1176-87.
271. Saxena, K., et al., *Great genotypic and phenotypic diversities associated with copy-number variations of complement C4 and RP-C4-CYP21-TNX*

(RCCX) modules: a comparison of Asian-Indian and European American populations. Mol Immunol, 2009. **46**(7): p. 1289-303.

272. Hanifi Moghaddam, P., et al., *Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry.* Diabetes, 1998. **47**(2): p. 263-9.
273. Christiansen, F.T., et al., *Major histocompatibility complex (MHC) complement deficiency, ancestral haplotypes and systemic lupus erythematosus (SLE): C4 deficiency explains some but not all of the influence of the MHC.* J Rheumatol, 1991. **18**(9): p. 1350-8.
274. Ahmed, A.R., et al., *Major histocompatibility complex susceptibility genes for dermatitis herpetiformis compared with those for gluten-sensitive enteropathy.* J Exp Med, 1993. **178**(6): p. 2067-75.
275. Hall, M.A., J.S. Lanchbury, and P.J. Ciclitira, *HLA class II region genes and susceptibility to dermatitis herpetiformis: DPB1 and TAP2 associations are secondary to those of the DQ subregion.* Eur J Immunogenet, 1996. **23**(4): p. 285-96.
276. Skinningsrud, B., et al., *Multiple loci in the HLA complex are associated with Addison's disease.* J Clin Endocrinol Metab, 2011. **96**(10): p. E1703-8.
277. Garlepp, M.J., et al., *Rearrangement of 21-hydroxylase genes in disease-associated MHC supratypes.* Immunogenetics, 1986. **23**(2): p. 100-5.
278. Ulgiati, D., et al., *Complete sequence of the complement C4 gene from the HLA-A1, B8, C4AQ0, C4B1, DR3 haplotype.* Immunogenetics, 1996. **43**(4): p. 250-2.
279. Banlaki, Z., et al., *Fine-tuned characterization of RCCX copy number variants and their relationship with extended MHC haplotypes.* Genes Immun, 2012. **13**(7): p. 530-5.
280. Boteva, L., et al., *Genetically determined partial complement C4 deficiency states are not independent risk factors for SLE in UK and Spanish populations.* Am J Hum Genet. **90**(3): p. 445-56.
281. Fernando, M.M., et al., *Identification of two independent risk factors for lupus within the MHC in United Kingdom families.* PLoS Genet, 2007. **3**(11): p. e192.
282. Christiansen, F.T., et al., *Complement allotyping in SLE: association with C4A null.* Aust N Z J Med, 1983. **13**(5): p. 483-8.
283. Doherty, D.G., et al., *Major histocompatibility complex genes and susceptibility to systemic lupus erythematosus in southern Chinese.* Arthritis Rheum, 1992. **35**(6): p. 641-6.
284. Hong, G.H., et al., *Association of complement C4 and HLA-DR alleles with systemic lupus erythematosus in Koreans.* J Rheumatol, 1994. **21**(3): p. 442-7.
285. Price, P., et al., *The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases.* Immunol Rev, 1999. **167**: p. 257-74.

286. Ochs, H.D., et al., *Complement, membrane glycoproteins, and complement receptors: their role in regulation of the immune response.* Clin Immunol Immunopathol, 1986. **40**(1): p. 94-104.
287. Hauptmann, G., G. Tappeiner, and J.A. Schifferli, *Inherited deficiency of the fourth component of human complement.* Immunodefic Rev, 1988. **1**(1): p. 3-22.
288. Schifferli, J.A., Y.C. Ng, and D.K. Peters, *The role of complement and its receptor in the elimination of immune complexes.* N Engl J Med, 1986. **315**(8): p. 488-95.
289. Nicholson, P., et al., *Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors.* Cell Mol Life Sci, 2010. **67**(5): p. 677-700.
290. Ahmed, J.N., et al., *A murine Zic3 transcript with a premature termination codon evades nonsense-mediated decay during axis formation.* Dis Model Mech, 2013. **6**(3): p. 755-67.
291. Danckwardt, S., et al., *Abnormally spliced beta-globin mRNAs: a single point mutation generates transcripts sensitive and insensitive to nonsense-mediated mRNA decay.* Blood, 2002. **99**(5): p. 1811-6.
292. Denecke, J., et al., *An activated 5' cryptic splice site in the human ALG3 gene generates a premature termination codon insensitive to nonsense-mediated mRNA decay in a new case of congenital disorder of glycosylation type Id (CDG-Id).* Hum Mutat, 2004. **23**(5): p. 477-86.
293. Veldhoen, M., et al., *Transforming growth factor-beta 'reprograms' the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset.* Nat Immunol, 2008. **9**(12): p. 1341-6.
294. Eyerich, S., et al., *Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling.* J Clin Invest, 2009. **119**(12): p. 3573-85.
295. Breitfeld, D., et al., *Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production.* J Exp Med, 2000. **192**(11): p. 1545-52.
296. Schaerli, P., et al., *CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function.* J Exp Med, 2000. **192**(11): p. 1553-62.
297. Baecher-Allan, C., et al., *CD4+CD25high regulatory cells in human peripheral blood.* J Immunol, 2001. **167**(3): p. 1245-53.
298. Jonuleit, H., et al., *Identification and functional characterization of human CD4(+)CD25(+) T cells with regulatory properties isolated from peripheral blood.* J Exp Med, 2001. **193**(11): p. 1285-94.
299. Stephens, L.A., et al., *Human CD4(+)CD25(+) thymocytes and peripheral T cells have immune suppressive activity in vitro.* Eur J Immunol, 2001. **31**(4): p. 1247-54.
300. Dong C and Martinez G J, *T cells: the usual subsets*. 2010, Nature Publishing Group.
301. Broere F, et al., *T cell subsets and T cell-mediated immunity*, in *Principles of Immunopharmacology*, Nijkamp FP and Parnham MJ, Editors. 2011, Springer: Basel. p. 15-27.

302. Oestreich, K.J. and A.S. Weinmann, *Master regulators or lineage-specifying? Changing views on CD4(+) T cell transcription factors.* Nat Rev Immunol, 2012. **12**(11): p. 799-804.
303. Romagnani, S., *The Th1/Th2 paradigm.* Immunol Today, 1997. **18**(6): p. 263-6.
304. Prochazkova, J., K. Pokorna, and V. Holan, *IL-12 inhibits the TGF-beta-dependent T cell developmental programs and skews the TGF-beta-induced differentiation into a Th1-like direction.* Immunobiology, 2012. **217**(1): p. 74-82.
305. Jacobson, N.G., et al., *Interleukin 12 signaling in T helper type 1 (Th1) cells involves tyrosine phosphorylation of signal transducer and activator of transcription (Stat)3 and Stat4.* J Exp Med, 1995. **181**(5): p. 1755-62.
306. Mosmann, T.R., et al., *Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins.* J Immunol, 1986. **136**(7): p. 2348-57.
307. Bonecchi, R., et al., *Differential expression of chemokine receptors and chemotactic responsiveness of type 1 T helper cells (Th1s) and Th2s.* J Exp Med, 1998. **187**(1): p. 129-34.
308. Mosmann, T.R. and R.L. Coffman, *TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties.* Annu Rev Immunol, 1989. **7**: p. 145-73.
309. Szabo, S.J., et al., *A novel transcription factor, T-bet, directs Th1 lineage commitment.* Cell, 2000. **100**(6): p. 655-69.
310. Gollob, J.A., et al., *Altered interleukin-12 responsiveness in Th1 and Th2 cells is associated with the differential activation of STAT5 and STAT1.* Blood, 1998. **91**(4): p. 1341-54.
311. Doodes, P.D., et al., *IFN-gamma regulates the requirement for IL-17 in proteoglycan-induced arthritis.* J Immunol, 2010. **184**(3): p. 1552-9.
312. Sasaki, Y., et al., *Identification of a novel type 1 diabetes susceptibility gene, T-bet.* Hum Genet, 2004. **115**(3): p. 177-84.
313. You, Y., et al., *Association of TBX21 gene haplotypes in a Chinese population with systemic lupus erythematosus.* Scand J Rheumatol, 2010. **39**(3): p. 254-8.
314. Morita, M., et al., *Functional polymorphisms in TBX21 and HLX are associated with development and prognosis of Graves' disease.* Autoimmunity, 2012. **45**(2): p. 129-36.
315. Chae, S.C., S.C. Shim, and H.T. Chung, *Association of TBX21 polymorphisms in a Korean population with rheumatoid arthritis.* Exp Mol Med, 2009. **41**(1): p. 33-41.
316. Horvath, C.M., Z. Wen, and J.E. Darnell, Jr., *A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain.* Genes Dev, 1995. **9**(8): p. 984-94.
317. Xu, X., Y.L. Sun, and T. Hoey, *Cooperative DNA binding and sequence-selective recognition conferred by the STAT amino-terminal domain.* Science, 1996. **273**(5276): p. 794-7.

318. Darnell, J.E., Jr., I.M. Kerr, and G.R. Stark, *Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins.* Science, 1994. **264**(5164): p. 1415-21.
319. Watford, W.T., et al., *Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4.* Immunol Rev, 2004. **202**: p. 139-56.
320. Akira, S., *Functional roles of STAT family proteins: lessons from knockout mice.* Stem Cells, 1999. **17**(3): p. 138-46.
321. Remmers, E.F., et al., *STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus.* N Engl J Med, 2007. **357**(10): p. 977-86.
322. Schindler, C., et al., *Interferon-dependent tyrosine phosphorylation of a latent cytoplasmic transcription factor.* Science, 1992. **257**(5071): p. 809-13.
323. Schindler, C. and J.E. Darnell, Jr., *Transcriptional responses to polypeptide ligands: the JAK-STAT pathway.* Annu Rev Biochem, 1995. **64**: p. 621-51.
324. Veals, S.A., et al., *Subunit of an alpha-interferon-responsive transcription factor is related to interferon regulatory factor and Myb families of DNA-binding proteins.* Mol Cell Biol, 1992. **12**(8): p. 3315-24.
325. Nair, R.P., et al., *Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways.* Nat Genet, 2009. **41**(2): p. 199-204.
326. Kool, M., H. Hammad, and B.N. Lambrecht, *Cellular networks controlling Th2 polarization in allergy and immunity.* F1000 Biol Rep, 2012. **4**: p. 6.
327. Swain, S.L., et al., *The role of IL4 and IL5: characterization of a distinct helper T cell subset that makes IL4 and IL5 (Th2) and requires priming before induction of lymphokine secretion.* Immunol Rev, 1988. **102**: p. 77-105.
328. Fort, M.M., et al., *IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies in vivo.* Immunity, 2001. **15**(6): p. 985-95.
329. Hurst, S.D., et al., *New IL-17 family members promote Th1 or Th2 responses in the lung: in vivo function of the novel cytokine IL-25.* J Immunol, 2002. **169**(1): p. 443-53.
330. Schmitz, J., et al., *IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines.* Immunity, 2005. **23**(5): p. 479-90.
331. Shi, Y., et al., *A novel cytokine receptor-ligand pair. Identification, molecular characterization, and in vivo immunomodulatory activity.* J Biol Chem, 2000. **275**(25): p. 19167-76.
332. Mosmann, T.R., *Regulation of immune responses by T cells with different cytokine secretion phenotypes: role of a new cytokine, cytokine synthesis inhibitory factor (IL10).* Int Arch Allergy Appl Immunol, 1991. **94**(1-4): p. 110-5.
333. Zheng, W. and R.A. Flavell, *The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells.* Cell, 1997. **89**(4): p. 587-96.
334. Kaplan, M.H., et al., *Stat6 is required for mediating responses to IL-4 and for development of Th2 cells.* Immunity, 1996. **4**(3): p. 313-9.

335. Lohoff, M., et al., *Dysregulated T helper cell differentiation in the absence of interferon regulatory factor 4.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11808-12.
336. Rengarajan, J., et al., *Interferon regulatory factor 4 (IRF4) interacts with NFATc2 to modulate interleukin 4 gene expression.* J Exp Med, 2002. **195**(8): p. 1003-12.
337. Liu, Z., et al., *Dec2 promotes Th2 cell differentiation by enhancing IL-2R signaling.* J Immunol, 2009. **183**(10): p. 6320-9.
338. Yang, X.O., et al., *Requirement for the basic helix-loop-helix transcription factor Dec2 in initial TH2 lineage commitment.* Nat Immunol, 2009. **10**(12): p. 1260-6.
339. Ho, I.C., D. Lo, and L.H. Glimcher, *c-maf promotes T helper cell type 2 (Th2) and attenuates Th1 differentiation by both interleukin 4-dependent and -independent mechanisms.* J Exp Med, 1998. **188**(10): p. 1859-66.
340. Orkin, S.H., *GATA-binding transcription factors in hematopoietic cells.* Blood, 1992. **80**(3): p. 575-81.
341. Evans, T., M. Reitman, and G. Felsenfeld, *An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes.* Proc Natl Acad Sci U S A, 1988. **85**(16): p. 5976-80.
342. Merika, M. and S.H. Orkin, *DNA-binding specificity of GATA family transcription factors.* Mol Cell Biol, 1993. **13**(7): p. 3999-4010.
343. Mitchell, P.J. and R. Tjian, *Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.* Science, 1989. **245**(4916): p. 371-8.
344. Ho, I.C., T.S. Tai, and S.Y. Pai, *GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation.* Nat Rev Immunol, 2009. **9**(2): p. 125-35.
345. Hosoya, T., I. Maillard, and J.D. Engel, *From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation.* Immunol Rev, 2011. **238**(1): p. 110-25.
346. Nawijn, M.C., et al., *Enforced expression of GATA-3 in transgenic mice inhibits Th1 differentiation and induces the formation of a T1/ST2-expressing Th2-committed T cell compartment in vivo.* J Immunol, 2001. **167**(2): p. 724-32.
347. Pai, S.Y., M.L. Truitt, and I.C. Ho, *GATA-3 deficiency abrogates the development and maintenance of T helper type 2 cells.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1993-8.
348. Wang, Y., M.A. Su, and Y.Y. Wan, *An essential role of the transcription factor GATA-3 for the function of regulatory T cells.* Immunity, 2011. **35**(3): p. 337-48.
349. Suri-Payer, E., et al., *CD4+CD25+ T cells inhibit both the induction and effector function of autoreactive T cells and represent a unique lineage of immunoregulatory cells.* J Immunol, 1998. **160**(3): p. 1212-8.
350. Wohlfert, E.A., et al., *GATA3 controls Foxp3(+) regulatory T cell fate during inflammation in mice.* J Clin Invest, 2011. **121**(11): p. 4503-15.
351. Valmori, D., et al., *A peripheral circulating compartment of natural naive CD4 Tregs.* J Clin Invest, 2005. **115**(7): p. 1953-62.

352. Seddon, B. and D. Mason, *Peripheral autoantigen induces regulatory T cells that prevent autoimmunity.* J Exp Med, 1999. **189**(5): p. 877-82.
353. Shimizu, J., et al., *Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance.* Nat Immunol, 2002. **3**(2): p. 135-42.
354. Murawski, M.R., et al., *Upregulation of Foxp3 expression in mouse and human Treg is IL-2/STAT5 dependent: implications for the NOD STAT5B mutation in diabetes pathogenesis.* Ann N Y Acad Sci, 2006. **1079**: p. 198-204.
355. Kerdiles, Y.M., et al., *Foxo transcription factors control regulatory T cell development and function.* Immunity, 2010. **33**(6): p. 890-904.
356. Stockis, J., et al., *Comparison of stable human Treg and Th clones by transcriptional profiling.* Eur J Immunol, 2009. **39**(3): p. 869-82.
357. Levings, M.K. and M.G. Roncarolo, *T-regulatory 1 cells: a novel subset of CD4 T cells with immunoregulatory properties.* J Allergy Clin Immunol, 2000. **106**(1 Pt 2): p. S109-12.
358. Van Esch, H., et al., *GATA3 haplo-insufficiency causes human HDR syndrome.* Nature, 2000. **406**(6794): p. 419-22.
359. Zhang, L., et al., *Association of single nucleotide polymorphisms in GATA-3 with allergic rhinitis.* Acta Otolaryngol, 2009. **129**(2): p. 190-4.
360. Cancer Genome Atlas Network, *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
361. Enciso-Mora, V., et al., *A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3).* Nat Genet, 2010. **42**(12): p. 1126-30.
362. Medsger, T.A., Jr., et al., *GATA-3 up-regulation in CD8+ T cells as a biomarker of immune dysfunction in systemic sclerosis, resulting in excessive interleukin-13 production.* Arthritis Rheum, 2011. **63**(6): p. 1738-47.
363. Szeto, C.C., et al., *Monitoring of urinary messenger RNA levels for the prediction of flare in systemic lupus erythematosus.* Clin Chim Acta, 2012. **413**(3-4): p. 448-55.
364. Veldhoen, M., et al., *TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells.* Immunity, 2006. **24**(2): p. 179-89.
365. Parrish-Novak, J., et al., *Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function.* Nature, 2000. **408**(6808): p. 57-63.
366. Dong, C., *TH17 cells in development: an updated view of their molecular identity and genetic programming.* Nat Rev Immunol, 2008. **8**(5): p. 337-48.
367. Langrish, C.L., et al., *IL-23 drives a pathogenic T cell population that induces autoimmune inflammation.* J Exp Med, 2005. **201**(2): p. 233-40.
368. Annunziato, F., et al., *Phenotypic and functional features of human Th17 cells.* J Exp Med, 2007. **204**(8): p. 1849-61.
369. Cosmi, L., et al., *Human interleukin 17-producing cells originate from a CD161+CD4+ T cell precursor.* J Exp Med, 2008. **205**(8): p. 1903-16.

370. Mangan, P.R., et al., *Transforming growth factor-beta induces development of the T(H)17 lineage.* Nature, 2006. **441**(7090): p. 231-4.
371. Zhou, L., et al., *IL-6 programs T(H)-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways.* Nat Immunol, 2007. **8**(9): p. 967-74.
372. Nurieva, R., et al., *Essential autocrine regulation by IL-21 in the generation of inflammatory T cells.* Nature, 2007. **448**(7152): p. 480-3.
373. Aggarwal, S., et al., *Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17.* J Biol Chem, 2003. **278**(3): p. 1910-4.
374. Ivanov, II, et al., *The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells.* Cell, 2006. **126**(6): p. 1121-33.
375. Yang, X.O., et al., *T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR alpha and ROR gamma.* Immunity, 2008. **28**(1): p. 29-39.
376. Yang, X.O., et al., *STAT3 regulates cytokine-mediated generation of inflammatory helper T cells.* J Biol Chem, 2007. **282**(13): p. 9358-63.
377. Hayashi, R., et al., *Influence of IL17A polymorphisms (rs2275913 and rs3748067) on the susceptibility to ulcerative colitis.* Clin Exp Med, 2012.
378. McGovern, D.P., et al., *Genetic epistasis of IL23/IL17 pathway genes in Crohn's disease.* Inflamm Bowel Dis, 2009. **15**(6): p. 883-9.
379. Lew, B.L., et al., *Association between IL17A/IL17RA Gene Polymorphisms and Susceptibility to Alopecia Areata in the Korean Population.* Ann Dermatol, 2012. **24**(1): p. 61-5.
380. He, Z., et al., *Elevated serum levels of interleukin 21 are associated with disease severity in patients with psoriasis.* Br J Dermatol, 2012. **167**(1): p. 191-3.
381. Terrier, B., et al., *Interleukin 21 correlates with T cell and B cell subset alterations in systemic lupus erythematosus.* J Rheumatol, 2012. **39**(9): p. 1819-28.
382. Costelloe, L., J. Jones, and A. Coles, *Secondary autoimmune diseases following alemtuzumab therapy for multiple sclerosis.* Expert Rev Neurother, 2012. **12**(3): p. 335-41.
383. Jia, H.Y., et al., *Association between interleukin 21 and Graves' disease.* Genet Mol Res, 2011. **10**(4): p. 3338-46.
384. Ding, L., et al., *A Single Nucleotide Polymorphism of IL-21 Gene is Associated with Systemic Lupus Erythematosus in a Chinese Population.* Inflammation, 2012.
385. Hughes, T., et al., *Fine-mapping and transethnic genotyping establish IL2/IL21 genetic association with lupus and localize this genetic effect to IL21.* Arthritis Rheum, 2011. **63**(6): p. 1689-97.
386. Mus, A.M., et al., *Interleukin-23 promotes Th17 differentiation by inhibiting T-bet and FoxP3 and is required for elevation of interleukin-22, but not interleukin-21, in autoimmune experimental arthritis.* Arthritis Rheum, 2010. **62**(4): p. 1043-50.

387. Lafferty, K.J., et al., *Immunobiology of tissue transplantation: a return to the passenger leukocyte concept.* Annu Rev Immunol, 1983. **1**: p. 143-73.
388. Lafferty, K.J. and J. Woolnough, *The origin and mechanism of the allograft reaction.* Immunol Rev, 1977. **35**: p. 231-62.
389. Mueller, D.L., M.K. Jenkins, and R.H. Schwartz, *Clonal expansion versus functional clonal inactivation: a costimulatory signalling pathway determines the outcome of T cell antigen receptor occupancy.* Annu Rev Immunol, 1989. **7**: p. 445-80.
390. Rudd, C.E., A. Taylor, and H. Schneider, *CD28 and CTLA-4 coreceptor expression and signal transduction.* Immunol Rev, 2009. **229**(1): p. 12-26.
391. Raychaudhuri, S., et al., *Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk.* Nat Genet, 2009. **41**(12): p. 1313-8.
392. Cao, J., et al., *Increased production of circulating soluble co-stimulatory molecules CTLA-4, CD28 and CD80 in patients with rheumatoid arthritis.* Int Immunopharmacol, 2012. **14**(4): p. 585-592.
393. Hebbar, M., et al., *Detection of circulating soluble CD28 in patients with systemic lupus erythematosus, primary Sjogren's syndrome and systemic sclerosis.* Clin Exp Immunol, 2004. **136**(2): p. 388-92.
394. Wong, C.K., et al., *Aberrant production of soluble costimulatory molecules CTLA-4, CD28, CD80 and CD86 in patients with systemic lupus erythematosus.* Rheumatology (Oxford), 2005. **44**(8): p. 989-94.
395. Kim, Y.O., et al., *Association of the CD28/CTLA4/ICOS polymorphisms with susceptibility to rheumatoid arthritis.* Clin Chem Lab Med, 2010. **48**(3): p. 345-53.
396. Cunninghame Graham, D.S., et al., *Evidence for unique association signals in SLE at the CD28-CTLA4-ICOS locus in a family-based study.* Hum Mol Genet, 2006. **15**(21): p. 3195-205.
397. Baldwin, A.S., Jr., *Series introduction: the transcription factor NF-kappaB and human disease.* J Clin Invest, 2001. **107**(1): p. 3-6.
398. Ghosh, S., M.J. May, and E.B. Kopp, *NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses.* Annu Rev Immunol, 1998. **16**: p. 225-60.
399. Gilmore, T.D., *Introduction to NF-kappaB: players, pathways, perspectives.* Oncogene, 2006. **25**(51): p. 6680-4.
400. Gilmore, T.D., *NF-kappa B, KBF1, dorsal, and related matters.* Cell, 1990. **62**(5): p. 841-3.
401. Baeuerle, P.A. and D. Baltimore, *I kappa B: a specific inhibitor of the NF-kappa B transcription factor.* Science, 1988. **242**(4878): p. 540-6.
402. Hoffmann, A., G. Natoli, and G. Ghosh, *Transcriptional regulation via the NF-kappaB signaling module.* Oncogene, 2006. **25**(51): p. 6706-16.
403. Heron, E., P. Deloukas, and A.P. van Loon, *The complete exon-intron structure of the 156-kb human gene NFKB1, which encodes the p105 and p50 proteins of transcription factors NF-kappa B and I kappa B-*

gamma: implications for NF-kappa B-mediated signal transduction. Genomics, 1995. **30**(3): p. 493-505.

404. Borm, M.E., et al., *A NFKB1 promoter polymorphism is involved in susceptibility to ulcerative colitis.* Int J Immunogenet, 2005. **32**(6): p. 401-5.
405. Hegazy, D.M., et al., *NFkappaB polymorphisms and susceptibility to type 1 diabetes.* Genes Immun, 2001. **2**(6): p. 304-8.
406. Kurylowicz, A., et al., *Association of NFKB1 -94ins/del ATTG promoter polymorphism with susceptibility to and phenotype of Graves' disease.* Genes Immun, 2007. **8**(7): p. 532-8.
407. Banerjee, D., H.C. Liou, and R. Sen, *c-Rel-dependent priming of naive T cells by inflammatory cytokines.* Immunity, 2005. **23**(4): p. 445-58.
408. Grossmann, M., et al., *The anti-apoptotic activities of Rel and RelA required during B-cell maturation involve the regulation of Bcl-2 expression.* EMBO J, 2000. **19**(23): p. 6351-60.
409. Grumont, R.J. and S. Gerondakis, *The subunit composition of NF-kappa B complexes changes during B-cell development.* Cell Growth Differ, 1994. **5**(12): p. 1321-31.
410. Kontgen, F., et al., *Mice lacking the c-rel proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression.* Genes Dev, 1995. **9**(16): p. 1965-77.
411. Mason, N.J., H.C. Liou, and C.A. Hunter, *T cell-intrinsic expression of c-Rel regulates Th1 cell responses essential for resistance to Toxoplasma gondii.* J Immunol, 2004. **172**(6): p. 3704-11.
412. Isomura, I., et al., *c-Rel is required for the development of thymic Foxp3+ CD4 regulatory T cells.* J Exp Med, 2009. **206**(13): p. 3001-14.
413. Chen, G., et al., *The NF-kappaB transcription factor c-Rel is required for Th17 effector cell development in experimental autoimmune encephalomyelitis.* J Immunol, 2011. **187**(9): p. 4483-91.
414. Ruan, Q., et al., *The Th17 immune response is controlled by the Rel-RORgamma-RORgamma T transcriptional axis.* J Exp Med, 2011. **208**(11): p. 2321-33.
415. Gregersen, P.K., et al., *REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis.* Nat Genet, 2009. **41**(7): p. 820-3.
416. Trynka, G., et al., *Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling.* Gut, 2009. **58**(8): p. 1078-83.
417. Ali, F.R., et al., *An investigation of rheumatoid arthritis loci in patients with early-onset psoriasis validates association of the REL gene.* Br J Dermatol, 2012.
418. Strange, A., et al., *A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.* Nat Genet, 2010. **42**(11): p. 985-90.
419. Crabtree, G.R. and N.A. Clipstone, *Signal transmission between the plasma membrane and nucleus of T lymphocytes.* Annu Rev Biochem, 1994. **63**: p. 1045-83.

420. Rooney, J.W., et al., *A common factor regulates both Th1- and Th2-specific cytokine gene expression.* EMBO J, 1994. **13**(3): p. 625-33.
421. Rao, A., *NF-ATp: a transcription factor required for the co-ordinate induction of several cytokine genes.* Immunol Today, 1994. **15**(6): p. 274-81.
422. Rao, A., C. Luo, and P.G. Hogan, *Transcription factors of the NFAT family: regulation and function.* Annu Rev Immunol, 1997. **15**: p. 707-47.
423. Shimada, M., et al., *An approach based on a genome-wide association study reveals candidate loci for narcolepsy.* Hum Genet, 2010. **128**(4): p. 433-41.
424. Zhou, T., et al., *Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis.* PLoS One, 2012. **7**(9): p. e44818.
425. Mellanby, E., *An experimental investigation on rickets. 1919.* Nutrition, 1989. **5**(2): p. 81-6; discussion 87.
426. Paulson, S.K. and H.F. DeLuca, *Subcellular location and properties of rat renal 25-hydroxyvitamin D3-1 alpha-hydroxylase.* J Biol Chem, 1985. **260**(21): p. 11488-92.
427. Bailey, R., et al., *Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes.* Diabetes, 2007. **56**(10): p. 2616-21.
428. Jones, G., S.A. Strugnell, and H.F. DeLuca, *Current understanding of the molecular actions of vitamin D.* Physiol Rev, 1998. **78**(4): p. 1193-231.
429. Knutson, J.C. and H.F. DeLuca, *25-Hydroxyvitamin D3-24-hydroxylase. Subcellular location and properties.* Biochemistry, 1974. **13**(7): p. 1543-8.
430. Cooper, J.D., et al., *Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes.* Diabetes, 2011. **60**(5): p. 1624-31.
431. Jung, K.H., et al., *Associations of vitamin d binding protein gene polymorphisms with the development of peripheral arthritis and uveitis in ankylosing spondylitis.* J Rheumatol, 2011. **38**(10): p. 2224-9.
432. Christakos, S., et al., *Genomic mechanisms involved in the pleiotropic actions of 1,25-dihydroxyvitamin D3.* Biochem J, 1996. **316 (Pt 2)**: p. 361-71.
433. Bhalla, A.K., et al., *Specific high-affinity receptors for 1,25-dihydroxyvitamin D3 in human peripheral blood mononuclear cells: presence in monocytes and induction in T lymphocytes following activation.* J Clin Endocrinol Metab, 1983. **57**(6): p. 1308-10.
434. Bouillon, R., et al., *Paracrine role for calcitriol in the immune system and skin creates new therapeutic possibilities for vitamin D analogs.* Eur J Endocrinol, 1995. **133**(1): p. 7-16.
435. Deluca, H.F. and M.T. Cantorna, *Vitamin D: its role and uses in immunology.* FASEB J, 2001. **15**(14): p. 2579-85.
436. Pani, M.A., et al., *Vitamin D receptor allele combinations influence genetic susceptibility to type 1 diabetes in Germans.* Diabetes, 2000. **49**(3): p. 504-7.
437. Lee, Y.H., et al., *Associations between vitamin D receptor polymorphisms and susceptibility to rheumatoid arthritis and systemic*

lupus erythematosus: a meta-analysis. Mol Biol Rep, 2011. **38**(6): p. 3643-51.

438. Cordin, O., et al., *The DEAD-box protein family of RNA helicases.* Gene, 2006. **367**: p. 17-37.
439. Yoneyama, M., et al., *Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity.* J Immunol, 2005. **175**(5): p. 2851-8.
440. Sutherland, A., et al., *Genomic polymorphism at the interferon-induced helicase (IFIH1) locus contributes to Graves' disease susceptibility.* J Clin Endocrinol Metab, 2007. **92**(8): p. 3338-41.
441. Smyth, D.J., et al., *A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region.* Nat Genet, 2006. **38**(6): p. 617-9.
442. Gauderman, W.J., *Sample size requirements for association studies of gene-gene interaction.* Am J Epidemiol, 2002. **155**(5): p. 478-84.
443. Rice, T.K., N.J. Schork, and D.C. Rao, *Methods for handling multiple testing.* Adv Genet, 2008. **60**: p. 293-308.
444. Simon, J.L. *Resampling: The New Statistics (2nd Edition)*. 1997; Available from: http://www.resample.com/content/text/index.shtml.
445. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
446. Brozzetti, A., et al., *Cytotoxic T lymphocyte antigen-4 Ala17 polymorphism is a genetic marker of autoimmune adrenal insufficiency: Italian association study and meta-analysis of European studies.* Eur J Endocrinol, 2010. **162**(2): p. 361-9.
447. Perez de Nanclares, G., et al., *No evidence of association of CTLA4 polymorphisms with Addison's disease.* Autoimmunity, 2004. **37**(6-7): p. 453-6.
448. Donner, H., et al., *Codon 17 polymorphism of the cytotoxic T lymphocyte antigen 4 gene in Hashimoto's thyroiditis and Addison's disease.* J Clin Endocrinol Metab, 1997. **82**(12): p. 4130-2.
449. Courtois, G. and T.D. Gilmore, *Mutations in the NF-kappaB signaling pathway: implications for human disease.* Oncogene, 2006. **25**(51): p. 6831-43.
450. Karban, A.S., et al., *Functional annotation of a novel NFKB1 promoter polymorphism that increases risk for ulcerative colitis.* Hum Mol Genet, 2004. **13**(1): p. 35-45.
451. Huebner, C., et al., *Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example.* Cancer Epidemiol Biomarkers Prev, 2007. **16**(6): p. 1185-92.
452. Bowes, J., et al., *Confirmation of TNIP1 and IL23A as susceptibility loci for psoriatic arthritis.* Ann Rheum Dis, 2011. **70**(9): p. 1641-4.
453. Lee, H.S., et al., *Association of STAT4 with rheumatoid arthritis in the Korean population.* Mol Med, 2007. **13**(9-10): p. 455-60.
454. Korman, B.D., et al., *Variant form of STAT4 is associated with primary Sjogren's syndrome.* Genes Immun, 2008. **9**(3): p. 267-70.

455. Cross, D.S., et al., *Population based allele frequencies of disease associated polymorphisms in the Personalized Medicine Research Project.* BMC Genet, 2010. **11**: p. 51.
456. Skinningsrud, B., et al., *A CLEC16A variant confers risk for juvenile idiopathic arthritis and anti-cyclic citrullinated peptide antibody negative rheumatoid arthritis.* Ann Rheum Dis, 2009. **69**(8): p. 1471-4.
457. Novembre, J. and E. Han, *Human population structure and the adaptive response to pathogen-induced selection pressures.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1590): p. 878-86.
458. Wellcome Trust Case Control Consortium, *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
459. Vaidya, B., et al., *CTLA4 gene and Graves' disease: association of Graves' disease with the CTLA4 exon 1 and intron 1 polymorphisms, but not with the promoter polymorphism.* Clin Endocrinol (Oxf), 2003. **58**(6): p. 732-5.
460. Rodriguez-Rodriguez, L., et al., *The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in Caucasian case-control samples.* Arthritis Rheum, 2011. **63**(2): p. 365-72.
461. Ramos, P.S., et al., *A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap.* PLoS Genet, 2011. **7**(12): p. e1002406.
462. Horsley, V. and G.K. Pavlath, *NFAT: ubiquitous regulator of cell differentiation and adaptation.* J Cell Biol, 2002. **156**(5): p. 771-4.
463. Gombos, Z., et al., *Analysis of extended human leukocyte antigen haplotype association with Addison's disease in three populations.* Eur J Endocrinol, 2007. **157**(6): p. 757-61.
464. Falorni, A., et al., *Measuring adrenal autoantibody response: interlaboratory concordance in the first international serum exchange for the determination of 21-hydroxylase autoantibodies.* Clin Immunol, 2011. **140**(3): p. 291-9.
465. Betterle, C., et al., *Adrenal-cortex autoantibodies and steroid-producing cells autoantibodies in patients with Addison's disease: comparison of immunofluorescence and immunoprecipitation assays.* J Clin Endocrinol Metab, 1999. **84**(2): p. 618-22.
466. Betterle, C., G. Coco, and R. Zanchetta, *Adrenal cortex autoantibodies in subjects with normal adrenal function.* Best Pract Res Clin Endocrinol Metab, 2005. **19**(1): p. 85-99.
467. Spinner, M.W., et al., *Familial distributions of organ specific antibodies in the blood of patients with Addison's disease and hypoparathyroidism and their relatives.* Clin Exp Immunol, 1969. **5**(5): p. 461-8.
468. Cook, J.R., et al., *FBXO11/PRMT9, a new protein arginine methyltransferase, symmetrically dimethylates arginine residues.* Biochem Biophys Res Commun, 2006. **342**(2): p. 472-81.
469. Blackwell, E. and S. Ceman, *Arginine methylation of RNA-binding proteins regulates cell function and differentiation.* Mol Reprod Dev, 2012. **79**(3): p. 163-75.

470. Di Lorenzo, A. and M.T. Bedford, *Histone arginine methylation.* FEBS Lett, 2011. **585**(13): p. 2024-31.
471. Abida, W.M., et al., *FBXO11 promotes the Neddylation of p53 and inhibits its transcriptional activity.* J Biol Chem, 2007. **282**(3): p. 1797-804.
472. Tateossian, H., et al., *Regulation of TGF-beta signalling by Fbxo11, the gene mutated in the Jeff otitis media mouse mutant.* Pathogenetics, 2009. **2**(1): p. 5.
473. Rye, M.S., et al., *FBXO11, a regulator of the TGFbeta pathway, is associated with severe otitis media in Western Australian children.* Genes Immun, 2011. **12**(5): p. 352-9.
474. Li, C., et al., *Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein HTLF.* Genomics, 1992. **13**(3): p. 658-64.
475. Tribioli, C., R.F. Robledo, and T. Lufkin, *The murine fork head gene Foxn2 is expressed in craniofacial, limb, CNS and somitic tissues during embryogenesis.* Mech Dev, 2002. **118**(1-2): p. 161-3.
476. Edelbrock, M.A., S. Kaliyaperumal, and K.J. Williams, *Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities.* Mutat Res, 2013.
477. Cohen, P.T., *Protein phosphatase 1--targeted in many directions.* J Cell Sci, 2002. **115**(Pt 2): p. 241-56.
478. Maritzen, T., J. Podufall, and V. Haucke, *Stonins--specialized adaptors for synaptic vesicle recycling and beyond?* Traffic, 2010. **11**(1): p. 8-15.
479. Habert-Ortoli, E., et al., *Molecular cloning of a functional human galanin receptor.* Proc Natl Acad Sci U S A, 1994. **91**(21): p. 9780-3.
480. Webling, K.E., et al., *Galanin receptors and ligands.* Front Endocrinol (Lausanne), 2012. **3**: p. 146.
481. Kohlhase, J., et al., *SALL3, a new member of the human spalt-like gene family, maps to 18q23.* Genomics, 1999. **62**(2): p. 216-22.
482. Parrish, M., et al., *Loss of the Sall3 gene leads to palate deficiency, abnormalities in cranial nerves, and perinatal lethality.* Mol Cell Biol, 2004. **24**(16): p. 7102-12.
483. Shikauchi, Y., et al., *SALL3 interacts with DNMT3A and shows the ability to inhibit CpG island methylation in hepatocellular carcinoma.* Mol Cell Biol, 2009. **29**(7): p. 1944-58.
484. Julia, A., et al., *Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility.* Arthritis Rheum, 2008. **58**(8): p. 2275-86.
485. Sultana, R., et al., *Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins.* Genomics, 2002. **80**(2): p. 129-34.
486. Schumann, G., et al., *Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption.* Proc Natl Acad Sci U S A, 2011. **108**(17): p. 7119-24.

487. Kalscheuer, V.M., et al., *Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation.* Hum Genet, 2007. **121**(3-4): p. 501-9.
488. Huang, X.L., et al., *A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient with autism.* Am J Med Genet A, 2010. **152A**(8): p. 2112-4.
489. Mefford, H.C., et al., *Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies.* PLoS Genet, 2010. **6**(5): p. e1000962.
490. Chen, Y.H., et al., *Genetic analysis of AUTS2 as a susceptibility gene of heroin dependence.* Drug Alcohol Depend, 2012.
491. Albert, T.J., et al., *Direct selection of human genomic loci by microarray hybridization.* Nat Methods, 2007. **4**(11): p. 903-5.
492. Kiialainen, A., et al., *Performance of microarray and liquid based capture methods for target enrichment for massively parallel sequencing and SNP discovery.* PLoS One, 2011. **6**(2): p. e16486.
493. Tewhey, R., et al., *Enrichment of sequencing targets from the human genome by solution hybridization.* Genome Biol, 2009. **10**(10): p. R116.
494. Pearce, S.H., et al., *Adrenal steroidogenesis after B lymphocyte depletion therapy in new-onset Addison's disease.* J Clin Endocrinol Metab, 2012. **97**(10): p. E1927-32.