The Genetics of Anti-Neutrophil Cytoplasmic Antibody Associated Vasculitis (AAV)

Limy Wong

Pembroke College, University of Cambridge
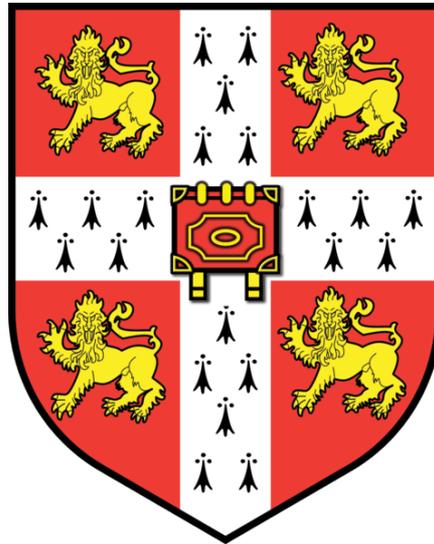
Department of Medicine

September 2019

Supervisors: Professor Kenneth GC Smith & Dr Paul A Lyons

This dissertation is submitted for the degree of Doctor of Philosophy.

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

The dissertation does not exceed the prescribed word limit specified by the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine.

Limy Wong
September 2019

ABSTRACT

The Genetics of Anti-Neutrophil Cytoplasmic Antibody Associated Vasculitis (AAV)

Limy Wong

Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV) is a multi-systemic autoimmune disorder with evidence of circulating pathogenic ANCA. There are two main antigenic targets: proteinase 3 (PR3) and myeloperoxidase (MPO). Previous genome-wide association studies (GWAS) have provided evidence that PR3-AAV and MPO-AAV are genetically distinct autoimmune syndromes, though only three loci specific to PR3-AAV and one to MPO-AAV have been identified to date.

With the European Vasculitis Genetics Consortium, we conducted a larger GWAS, powered to discover additional risk loci in both PR3-AAV and MPO-AAV independently. A meta-analysis of two European cohorts was conducted, comprising 1,610 PR3-AAV cases, 870 MPO-AAV cases and 11,947 controls. For *PTPN22* (rs6679677), a further replication cohort, previously genotyped using the Sequenom MassARRAY platform, and comprising 1,122 PR3-AAV and 347 MPO-AAV cases and 1,531 controls was included in the combined analysis.

This is the largest genome-wide association study of AAV to date and we have identified a total of 12 AAV susceptibility loci. Previously genome-wide significant loci were confirmed, including HLA class II, *SERPINA1*, *PRTN3* and *PTPN22*. Seven new genome-wide significant loci were identified: three associated with PR3-AAV (*BCL2L11-MIR4435-2HG*, *EBF3-MGMT*, *IGHV1-69*), two with MPO-AAV (*BACH2*, *ANKRD11-SPG7*) and two shared by both (*CTLA-4* and *DGUOK-TET3*). Further analyses based on common variants suggested that a substantial component of the genetic architecture was shared between PR3-AAV and MPO-AAV, similar to that observed between ulcerative colitis and Crohn's disease.

In addition, Mendelian randomisation analysis confirmed that a higher eosinophil count increased the risk of PR3-AAV but not MPO-AAV, and this effect might, in part, be modulated by *MIR4435-2HG* through prolongation of eosinophil survival. *MIR4435-2HG* encodes a long non-coding RNA that plays a critical role in the regulation of *BCL2L11* transcription (a Bcl2 family member essential for controlling apoptosis) in myeloid cells and hence their lifespan.

We have also identified a missense variant in *IGHV1-69* (rs11845244) that leads to a loss in neutralising function of antibodies generated against the NEAT2 domain of *Staphylococcus aureus*. This therefore provides a plausible host genetic factor in determining the susceptibility to infectious disease and as a potential driver of PR3-AAV. Overall, this study provides key novel insights into disease biology for AAV and potential therapeutic targets.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LISTS OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A1AT	α1-antitrypsin
AAV	Anti-neutrophil cytoplasmic antibody associated vasculitis
ACR	American College of Rheumatology
ANCA	Anti-neutrophil cytoplasmic antibody
ANOVA	One-way analysis of variance
APC	Antigen presenting cells
AS	Ankylosing spondylitis
ASE	Allele-specific expression
BCL2L11	Bcl-2-like protein 11
BH	Bcl-2 homology
BLyS	B lymphocyte stimulator
bp	Base pair
B_{reg}	Regulatory B cells
BTNL2	Butyrophilin-like protein 2
C5aR	Complement component fragment 5a receptor
c-ANCA	Cytoplasmic anti-neutrophil cytoplasmic antibody
CD	Cluster of differentiation
cDNA	Complementary deoxyribonucleic acid
CDR	Complementarity determining region
CG	Candidate gene approach studies
CHCC	Chapel Hill Consensus Conference on nomenclature of vasculitides
CI	Confidence interval
CKD	Chronic kidney disease
CLL	Chronic lymphocytic leukaemia
CNV	Copy number variation
cPR3	Complementary proteinase 3
Csk	C-terminal Src kinase
C_T	Cycle threshold
CYC	Cyclophosphamide
DGUOK	Deoxyguanosine kinase
DNA	Deoxyribonucleic acid

EBF3	Early B-cell factor 3
E. coli	*Escherichia coli*
EETs	Eosinophil extracellular DNA traps
EGPA	Eosinophilic granulomatous with polyangiitis
EMEA	European Medicines Agency
EPIC	The European Prospective Investigation of Cancer
eQTL	Expression quantitative trait loci
ER	Endoplasmic reticulum
ERAP	Endoplasmic reticulum aminopeptidase
ESKD	End-stage kidney disease
EVGC	European Vasculitis Genetics Consortium
GCA	Giant cell arteritis
GCTA	Genome-wide complex trait analysis
gDNA	Genomic deoxyribonucleic acid
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GPA	Granulomatosis with polyangiitis
GRM	Genetic relationship matrix
GTEx	Genotype-Tissue Expression Consortium
GWAS	Genome-wide association study
$\mathbf{h^2_{SNP}}$	SNP-based heritability
HA	Hemagglutinin
HBV	Hepatitis B virus
HES	Hypereosinophilic syndrome
HLA	Human leukocyte antigen
hLAMP-2	Human lysosome associated membrane protein-2
hMPO	Human myeloperoxidase
IFN	Interferon
Ig	Immunoglobulin
IL-2	Interleukin-2
IMDM	Iscove modified Dulbecco medium
IMGT	International immunogenetic information system®
IsD	Iron-regulated surface determinant system

ITAMs	Immunoreceptor tyrosine-based activation motifs
IVW	Inverse-variance weighted
JIA	Juvenile idiopathic arthritis
LAMP-2	Lysosomal associated membrane protein-2
LD	Linkage disequilibrium
LMM	Linear mixed model
LPS	Lipopolysaccharide
MAF	Minor allele frequency
MBN	Myeloblastin
MHC	Major histocompatibility complex
MPA	Microscopic polyangiitis
MPO	Myeloperoxidase
mPR3	Membrane proteinase 3
MPRA	Massively parallel reporter assay
MR	Mendelian randomisation
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis
mtDNA	Mitochondrial DNA
na	Not applicable
NCGN	Necrotising crescentic glomerulonephritis
NEAT	NEAr-iron Transporter
NETs	Neutrophil extracellular traps
NHGRI-EBI	National Human Genome Research Institute-European Bioinformatics Institute
NOD	Non-obese diabetic
OR	Odds ratio
PAMPs	Pathogen associated molecular patterns
PAN	Polyarteritis nodosa
p-ANCA	Perinuclear anti-neutrophil cytoplasmic antibody
PBC	Primary biliary cirrhosis
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PC	Principal component

PCA	Principal component analysis
PCHi-C	Promoter capture Hi-C
PCR	Polymerase chain reaction
PEP	PEST domain-enriched tyrosine phosphatase
Pi	Protease inhibitor
PI	Propidium iodide
PMN	Polymorphonuclear cell
PPP	Platelet poor plasma
pQTL	Protein quantitative trait loci
PR3	Proteinase 3
PRP	Platelet rich plasma
PRR	Pattern recognition receptor
PSC	Primary sclerosing cholangitis
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
REML	Restricted maximum likelihood
rmPR3	Recombinant murine proteinase 3
RNA	Ribonucleic acid
ROS	Reactive oxygen species
S. aureus	*Staphylococcus aureus*
SCG	Spontaneous crescentic glomerulonephritis
SCID	Severe combined immune deficient
SE	Standard error
SEM	Standard error of mean
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SVV	Small-vessel vasculitis
TAK	Takayasu arteritis
TAP	Transporter associated antigen processing protein
TCR	T-cell receptor
TET3	Tet methylcytosine dioxygenase 3

TGF-β	Transforming growth factor-beta
TLR	Toll-like receptor
TNF	Tumour necrosis factor
T$_{reg}$	Regulatory T cells
T1D	Type 1 diabetes
UC	Ulcerative colitis
VCF	Variant call format
VCRC	US Vasculitis Clinical Research Consortium
WG	Wegener's granulomatosis
WKY	Wistar-Kyoto
WTCCC	Wellcome Trust Case Control Consortium

1 Introduction

1.1 The history of systemic vasculitis

Although polyarteritis nodosa (PAN) was first described as a disease entity by Carl von Rokitansky in 1852 (1), the first description of necrotising vasculitis was attributed to Adolf Kussmaul and Rudolf Maier in 1866, who reported a case of systemic syndrome affecting the nervous system and kidneys of a young man. They described the gross and microscopic pathology of the disease which consisted of focal, inflammatory arterial nodules in the medium and small-sized arteries and named it "periarteritis nodosa" (2). It was not until 1903 that Ferrari in Trieste, Italy suggested the name "polyarteritis nodosa" to emphasise the transmural inflammation of the vessel wall that characterises this vasculitis (3), which remains in the revised nomenclature of systemic vasculitis today as the increasingly rare form of vasculitis affecting the medium-sized vessels with a strong association with hepatitis B virus (HBV). Later, two further reports, one from F J Wohlwill in Hamburg who described a case of what is now known as microscopic polyangiitis (MPA) in 1923 (4) and the second by Heinz Klinger (5), who reported a case of granulomatous respiratory inflammation with necrotising glomerulonephritis in 1931, both preceded the first case series of Wegener's granulomatosis (WG) by Friedrich Wegener in 1939, under the name "singular rhinogenic granulomatosis" (6). In fact, it was Jacob Churg along with Gabriel Godman who attributed this condition to Wegener's in 1954 (7), though the eponym has recently been dropped in the wake of emergence of evidence linking Wegener to the Nazi regime during World War II (8). Churg and Godman were also the first to describe that microscopic polyarteritis, WG and allergic angiitis and granulomatosis (Churg-Strauss syndrome) were different from PAN based primarily on the pathological features (7), which was later supported by the findings of many groups that the former three conditions were associated with anti-neutrophil cytoplasmic antibodies (ANCA) (9-14) but not the latter one (15), further endorsing the categorisation of vasculitis on the basis of biomarkers (Figure 1.1).

Around the same time of the discovery of vasculitis affecting the medium-sized arteries, Schönlein (16), Henoch (17, 18) and Osler (19, 20) identified a form of purpura that was accompanied by arthralgias, abdominal pain and nephritis. The histological examination of these purpuric lesions demonstrated inflammation of small dermal vessels with extensive leukocytoclasia and IgA-dominant immune deposits in the vessel wall (21), which is now termed IgA vasculitis. As most of the clinical and pathological manifestations of systemic vasculitides were identified before the end of twentieth century, our understanding of the aetiology and pathogenesis of these diseases have improved considerably in the recent decades, so as the classification and nomenclature systems which have undergone numerous revisions since.
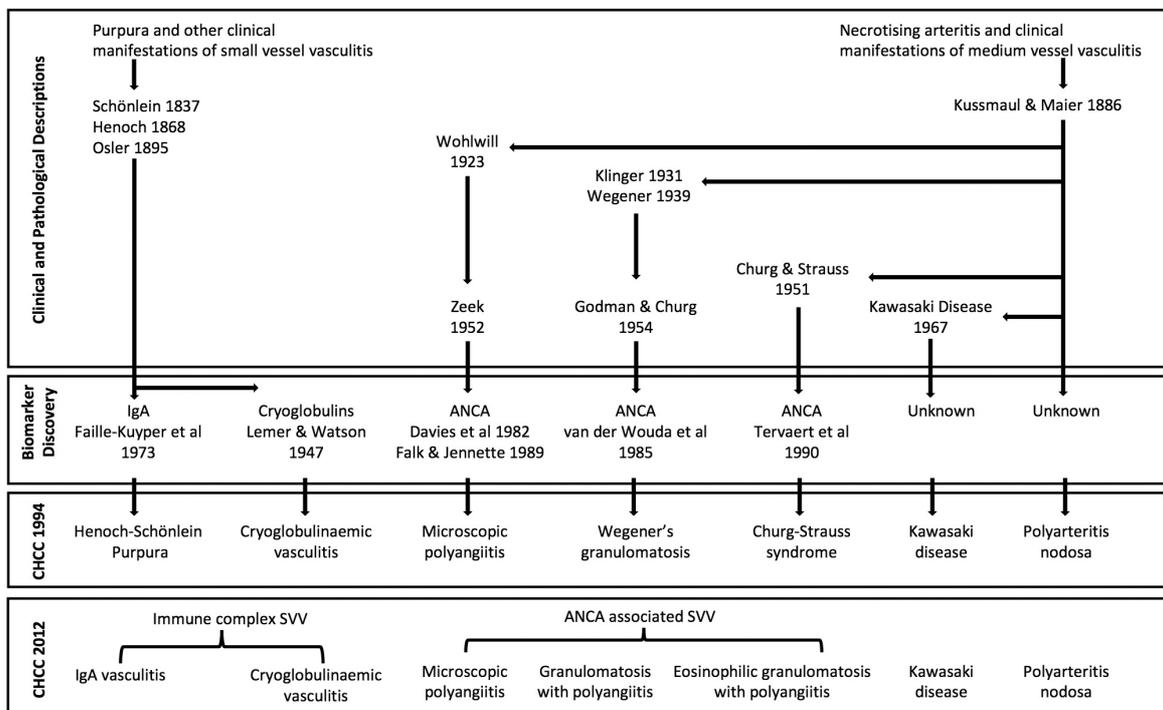
Figure 1.1: The historical evolution of the nomenclature of necrotising systemic vasculitis.

The upper panel depicts the investigators who made significant contributions to the description of clinical and pathological patterns of vasculitis. The middle panel shows the discovery of biomarkers that demonstrate different pathogenic categories of vasculitis. Lastly, the bottom panels provide the diagnostic names proposed by the 1994 and 2012 Chapel Hill Consensus Conference (CHCC) on Nomenclature of Vasculitides. This diagram was adapted and modified from Jennette JC, *Presse Med* 2013 (22). ANCA, anti-neutrophil cytoplasmic antibody; SVV, small-vessel vasculitis.

1.2 The classification and nomenclature of systemic vasculitis

Classification of disease is the process of categorising illnesses in a larger framework of medical conditions to facilitate compilation of mortality and morbidity statistics for epidemiological purposes, highlighting the associations between diseases as well as creating a framework for diagnosis and management. It can be based on topology (bodily region or system), anatomy, aetiology, histology, physiology, clinical characteristics or a combination of these features. It differs from nomenclature which is essentially a system of names and descriptions of diseases. Both the classification and nomenclature of systemic vasculitis have undergone numerous revisions over the years which reflect the complex and heterogeneous nature of the illnesses.

1.2.1 American College of Rheumatology (ACR) criteria

In 1990, the ACR published criteria for the classification of seven types of systemic vasculitis which include giant cell arteritis (GCA), Takayasu arteritis (TAK), eosinophil granulomatosis with polyangiitis (EGPA, a.k.a Churg-Strauss Syndrome), granulomatosis with polyangiitis (GPA, a.k.a Wegener's granulomatosis), PAN, IgA vasculitis and hypersensitivity vasculitis (23). This represents the first most comprehensive set of classification criteria though it has several limitations. Firstly, MPA was not included as one of the named conditions as it was not widely recognised in the 1980s. Secondly, the criteria were developed before the use of ANCA testing which is now readily available and has become a fundamental aspect in the diagnosis and classification of ANCA-associated vasculitis (AAV). Furthermore, the introduction of new radiological diagnostic techniques (e.g. computed tomography and magnetic resonance imaging) has contributed to a better distinction between different types of vasculitis. Interestingly, a recent evaluation of the classification criteria by Seeliger et al showed that while the sensitivity of the 1990 ACR Criteria for vasculitis at diagnosis was 67.1%, but the specificity has remained high (24).

1.2.2 Chapel Hill Consensus Conference (CHCC) on the Nomenclature of Vasculitides

The CHCC in 1994 was aimed to construct the names and definitions of the most common forms of vasculitides including MPA, but not for the intention of classification or diagnosis. A

revision was made in 2012 and it retained the primary reliance on the calibre of the vessels involved, but added several important categories comprising of variable vessel vasculitis. These include Behcet's disease and Cogan's syndrome, single organ vasculitis, vasculitis associated with systemic disease and vasculitis associated with a probable aetiology as well as subcategories which include division of SVV into immune complex, ANCA-associated, anti-glomerular basement membrane and hypocomplementemic urticarial vasculitis. It also recommended that the ANCA reactivity should be specified, i.e. myeloperoxidase (MPO)-ANCA, proteinase 3 (PR3)-ANCA or ANCA-negative AAV (25).

1.2.3 European Medicines Agency (EMEA) Algorithm

Given the overlapping and conflicting classifications from the existing schemes such as ACR, CHCC definitions as well as Lanham Criteria for EGPA, the EMEA algorithm was developed for the purpose of classification in epidemiological studies (26). All patients must fulfil the criteria for a clinical diagnosis of AAV or PAN prior entering the algorithm, which is optimally used for classification rather than diagnosis of acute cases and patients should ideally be followed up for at least 3 months where possible. It includes a list of surrogate markers for the diagnosis of GPA consisting of laboratory, radiological, histological and clinical manifestations of the upper and lower respiratory tracts. The major advantage of this algorithm is its ability to unequivocally classify the patients within the subcategory of AAV down to a single diagnosis and it has been independently validated on a large cohort of Chinese patients (27).

The existing classification criteria are useful but there is no doubt that there are some shortcomings given the evolving diagnostic tools, discovery of novel biomarkers, better understanding of the disease pathogenesis and recent genetic studies which have highlighted the difference between GPA and MPA. Interestingly, Mahr and colleagues employed an unsupervised machine learning algorithm - cluster analysis to identify subgroups of AAV patients. They have successfully defined five clusters with distinct mortality and relapse rates using four variables consisting of ANCA specificity, renal, cardiac and gastrointestinal involvement (28), suggesting this subgrouping might be aetiologically and therapeutically relevant and requires further validation. A global study - The Diagnostic and Classification Criteria for Vasculitis Study (DCVAS) which began in 2010, is currently underway to develop

and improve the diagnostic criteria for systemic vasculitides. Over 100 sites across 30 countries in Asia, Australasia, Europe, North America, South America and the UK are involved and the results of this study will be much anticipated (29).

1.3   Epidemiology of ANCA-associated vasculitis

The incidence of AAV across different populations is approximately 20 per million with a peak age of diagnosis between 65 to 74 years. The differences in the geographical and ethnic distributions in the clinical features of patients with primary systemic vasculitis have been well reported. In Europe, GPA appears to be more common at high latitudes (Norway, United Kingdom and Germany), whereas MPA shows the reverse pattern, being encountered more frequently in southern Europe (Lugo, Spain), Kuwait and other Asian countries including Japan, Korea and China (30-35). The annual incidence of AAV was previously reported to be similar between Japan and the UK, but the clinical phenotypes were different. Among patients with AAV, MPA was the predominant subtype accounting for about 83% of the patients in Japan while in the UK, more than half of the patients were of the GPA subtype. As for the pattern of ANCA positivity, more than 80% of the Japanese patients were perinuclear ANCA (p-ANCA) or MPO-ANCA positive, whereas two-thirds of UK patients were cytoplasmic ANCA (c-ANCA) or PR3-ANCA positive (36).

In New Zealand, GPA was twice as common in Europeans than those of Maori or Asian background. A latitude-dependent incidence gradient was observed (37). Similarly in a French multi-ethnic population study, the prevalence of AAV was twice as common in individuals of European as compared to those of non-European ancestry (38). GPA is relatively uncommon in African Americans (39). The occurrence of different phenotypes is likely to be influenced by both the environmental factors and different genetic susceptibility in the diverse ethnicity groups (40). Some of the environmental risk factors that have been previously reported include low vitamin D levels as a results of decreased sun exposure (41), infection - particularly chronic nasal carriage of *Staphylococcus aureus* (*S. aureus*) which is related to an increased risk of relapse in GPA (42) and fimbriated pathogens which have been identified as potential triggers of autoimmunity to lysosomal associated membrane protein-2 (LAMP-2) (43), silica

and high occupational solvent exposure (44), asbestos exposure (45), medications such as propylthiouracil (46, 47), hydralazine (48), anti-tumour necrosis factor agents (49) and more recently, levamisole-adulterated cocaine (50).

1.4 The genetics of ANCA-associated vasculitis

The most obvious challenge to conducting a rigorous genetic study in AAV is the small number of eligible patients due to the rarity of the condition and therefore, limited study power. There are several pointers towards a genetic component, mainly from studies of familial occurrence (51-55), human leukocyte antigen (HLA) type of patients and differences in the incidence of AAV in various ethnic groups. A number of candidate gene approach studies have been performed over the years but the success of this approach depends largely on the chosen genes or pathways studied. The understanding of genetics in AAV has only substantially improved following the genome-wide association studies (GWAS) conducted by the European Vasculitis Genetics Consortium (EVGC) and the US Vasculitis Clinical Research Consortium (VCRC) which have thus far been limited to individuals of European ancestry.

1.4.1 Human leukocyte antigen region associations

The HLA region or major histocompatibility complex is located in the short arm of chromosome 6p21.3 and spans over 4 Mb containing the highly polymorphic classical HLA class I and II loci, as well as other genes that are involved in the immunological function. This region has been rigorously studied due to its pivotal role in immune response in the setting of transplantation, strong association to various autoimmune disorders (56) as well as susceptibility and resistance to a range of bacterial and viral infections (57). One of the most striking features observed is that autoimmune diseases with characteristic antibodies are generally associated with HLA class II alleles, whereas seronegative diseases most often associate with HLA class I alleles, such as *HLA-B27* in ankylosing spondylitis (AS) (58). The genetic associations between HLA and autoimmune diseases have been reported since the early 1970s (59). Up until now, it has been confirmed that HLA has the strongest effect size in the GWAS of autoimmune diseases, as compared to any other non-HLA loci. However, the molecular mechanism that might underpin the HLA-autoimmunity association has remained

elusive in majority of the diseases and the presence of extensive linkage disequilibrium (LD) in this region poses another challenge to identifying the causative genes. One good example is the discovery of epistatic interaction between *HLA-B27* and *HLA-B40* with endoplasmic reticulum aminopeptidase 1 (*ERAP1*) in AS (58, 60). *ERAP1* has dual functions: it works together with endoplasmic reticulum aminopeptidase 2 (*ERAP2*) in endoplasmic reticulum to trim and present peptides on HLA class I molecules and it is involved in cytokine receptor shedding such as tumour necrosis factor receptor I, IL-1 receptor 2 and IL-6Rα. This interaction supports the disease model where aberrant trimming of peptides or antigen presentation by *ERAP1* are involved in the pathogenesis AS together with abnormalities in cell surface receptor shedding which could affect the cellular responses to the relevant cytokines (58, 61).

The human HLA is divided into three regions. The class I region consists of the classical class I genes (*HLA-A*, *HLA-B and HLA-C*) and two clusters of non-classical class I genes. The class II region contains genes that have functions related to antigen processing (*TAP*, *LMP*) or presentation (*HLA-DR*, *HLA-DQ* and *HLD-DP*) and the class III region is the most gene-dense region in the human genome comprising of many innate immunity related genes including cytokines tumour necrosis factor (TNF) and complement as well as non-immune related genes. The major differences between MHC class I and class II genes and functions are summarised in Table 1.1 (62, 63). The International Immunogenetic Information system® (IMGT) HLA database is a very useful resource that contains a compilation of sequences of the human HLA that are updated on a frequent basis as well as tools for comparing and naming HLA alleles and proteins (https://www.ebi.ac.uk/ipd/imgt/hla/). Both the candidate gene studies and GWAS for HLA associations in AAV are summarised in Table 1.2 (64-78). The strongest evidence of HLA association in AAV is *HLA-DPB1*. This association was first reported in a cohort of 150 German GPA patients (*HLA-DPB1*04:01*, p = 1.51 x 10^{-10}, odds ratio (OR) of 3.91) (67) and subsequently replicated in an independent cohort of 108 GPA patients with a corrected p-value, p_c = 6.40 x 10^{-8} (66). The association remained significant in the combined analysis with a p_c = 1.26 x 10^{-22} and was later confirmed by the three GWAS; the most significant single nucleotide polymorphism (SNP) in the HLA region from the EVGC GWAS was rs3117242, encoding *HLA-DPB1* (p = 1.50 x 10^{-71}, OR = 3.67) (70) and subsequent stepwise

logistic regression analysis confirmed that there was one independent signal in this region. Similar *HLA-DP* associations were also replicated in both GWAS performed by VCRC (71, 78).

Table 1.1: The major differences between HLA class I and class II genes

	HLA Class I	HLA Class II
Structure	One membrane-spanning α chain (heavy chain) produced by the HLA genes and one β chain (light chain) produced by the β2-microglobulin gene	Two membrane-spanning chains (α and β) of similar size and produced by the HLA genes
Gene regions	HLA-A, -B, -C	HLA-DR, -DP, -DQ
Type of antigen presenting cells (APCs)	Present in all nucleated cells	Only on specialised APCs including macrophages, dendritic cells and B cells
Nature of antigen presentation	Endogenous antigens originate from cytoplasm	Exogenous antigens originate extracellularly
Size of peptide	8-10 amino acid peptides	13-25 amino acid peptides
Responsive T cells	Present antigen to cytotoxic T cells (CD8+)	Present antigen to helper T cells (CD4+)
Co-receptor responsible	Bind with CD8 co-receptors on cytotoxic T cells	Bind with CD4 co-receptors on helper T cells
Source of protein antigens	Cytosolic proteins, mainly synthesised in the cell	Endosomal/lysosomal proteins, mostly internalised from extracellular environment
Enzyme responsible for peptide generation	Cytosolic proteasome	Endosomal and lysosomal proteases
Site of peptide loading on HLA	ER	Specialised vesicular compartment
Molecules involved in transport of peptides and loading on HLA	Chaperones, TAP in ER	Chaperones in ER, invariant chain in ER (Ii), Golgi and HLA class II compartment
Final outcome	Target cells for destruction	Regulation of effector cells e.g. induces antibody production and attracts other immune cells to area of infection

APC, antigen presenting cells; CD, cluster of differentiation; ER, endoplasmic reticulum; HLA, human leukocyte antigen; TAP, transporter associated with antigen processing protein (62, 63).

Table 1.2: Previous candidate gene studies and GWAS on HLA associations in AAV

Study, Year	Population	HLA	Cases/Controls	OR	P-value
Spencer, 1992	British	*HLA-DQw7*	34 GPA & 25 MPA/1,103	2.90	$<2.50 \times 10^{-3}$
		HLA-DR3		0.31	$<1.00 \times 10^{-2}$
Papiha, 1992	British	*HLA-DR1*	27 GPA/105	2.30	4.00×10^{-2}
Hagen, 1995	Dutch	*HLA-DR6*	224 AAV/2,443	0.43	$<4.00 \times 10^{-3}$
		HLA-DR13		0.32	$<2.00 \times 10^{-3}$
Gencik, 1999	German	*HLA-DRB1*13*	76 PR3/na	0.40	$<5.00 \times 10^{-2}$
		*HLA-DQB1*06:03*		0.20	$<5.00 \times 10^{-3}$
Tsuchiya, 2003	Japanese	*HLA-DRB1*09:01*	64 MPO/na	2.44	1.40×10^{-3}
Jagiello, 2004	German	*HLA-DPB1*04:01*	150 PR3/100	3.91	1.51×10^{-10}
		*HLA-DPB1*04:01/ RXRB03*		6.41	7.13×10^{-17}
Tsuchiya, 2006	Japanese	*HLA-DQB1*03:03*	50 MPA/77	2.11	1.70×10^{-2}
		*HLA-DRB1*09:01- HLA-DQB1*03:03*		2.36	2.30×10^{-2}
Heckmann, 2007	German	*HLA-DPB1*04:01*	282 GPA/380	3.01	2.31×10^{-21}
Stassen, 2009	Dutch	*HLA-DR6*	304 AAV/9,872	0.30	$<1.00 \times 10^{-4}$
		HLA-DR4		1.70	$<1.00 \times 10^{-4}$
Luo, 2011	Chinese	*HLA-DRB1*14:54*	152 AAV/200	0.02	1.70×10^{-2}
		*HLA-DRB1*11:01*	107 MPA/200	2.53	2.30×10^{-2}
†Lyons, 2012	European	*HLA-DP (rs3117242)*	1,521 PR3/6,858	7.03	6.20×10^{-89}
		HLA-DQ (rs5000634)	556 MPO/6,858	0.65	2.10×10^{-8}
†Xie, 2013	North American	*HLA-DPB1 (rs9277554)*	750 GPA/1,820	0.24	1.92×10^{-50}
Wu, 2015	Chinese	*HLA-DPB1 (rs3117242)*	100 GPA/485	2.09	6.24×10^{-5}
Kawasaki, 2016	Japanese	*HLA-DRB1*09:01*	377 MPO/596	1.57	2.10×10^{-4}
		*HLA-DRB1*13:02*		0.42	2.30×10^{-5}
†Merkel, 2016	North American	*HLA-DPB1 (rs141530233)*	1,361 PR3/4,723	6.19	1.33×10^{-106}
		HLA-DQA2 (rs7454108)	378 MPO/4,723	2.73	5.03×10^{-25}

AAV, ANCA-associated vasculitis; GPA, granulomatosis with polyangiitis; HLA, human leukocyte antigen; MPA, microscopic polyangiitis; MPO, myeloperoxidase; na, not applicable; OR, odds ratio; PR3, proteinase 3.

†SNPs showing the most significant association in the GWAS.

Apart from the GWAS, the patient cohort might not be completely independent from each other for those from the same ancestry background (64-78).

Strikingly, the sub-analysis of the EVGC GWAS demonstrated that there was a clear distinction in the genetic associations between GPA and MPA where patients were classified according to the EMEA algorithm. More importantly, the *HLA-DP* association was stronger in the PR3-ANCA positive subgroup (p = 6.20 x 10^{-89}), independent of the clinical diagnosis, while this was not observed in the MPO-ANCA positive subgroup. In the latter group, a further analysis discovered an association in the *HLA-DQ* region (p = 2.10 x 10^{-8}) which had probably been masked in the primary analysis given the small number of MPO-ANCA positive patients. This finding was replicated in an independent cohort of Italian patients as well as in VCRC GWAS.

Comparably in the Han Chinese cohort, rs3117242 (*HLA-DPB1*) conferred an increased risk for GPA with an OR of 2.09, 95% CI 1.51-2.88 (77). While in the Japanese MPO-AAV cohort, a significant association was detected with *HLA-DRB1*09:01* where the carrier frequency was increased in MPO-AAV patients as compared to healthy Japanese controls (68, 75, 76). *HLA-DRB1*09:01* is one of the most common *HLA-DRB1* alleles in Asians but is rare in Caucasian populations. Interestingly, it has also been shown to be associated with multiple autoimmune diseases, including type 1 diabetes (T1D), rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) (79). Therefore, *HLA-DRB1*09:01* haplotype appears to be an Asian-specific genetic factor for multiple autoimmune diseases.

1.4.2   Non-HLA associations

SERPINA1

α1-antitrypsin (A1AT) is encoded by the *SERPINA1* (serpin family A member 1) gene. It is an acute phase protein that belongs to the serine protease inhibitor family with capability of inactivating many different proteases including PR3. Deficiency of A1AT as a result of carriage of pathogenic alleles of Pi-gene (protease inhibitor) causes chronic uninhibited breakdown of lung tissues and eventually leads to characteristic manifestation of pulmonary emphysema. The most common and normally functioning allelic form is PiM and so the healthy human phenotype is designated as PiMM. There are more than 100 genetic A1AT variants, among which PiZ and PiS are the most common and clinically significant alleles. A1AT deficiency clinically manifests in individuals carrying mutations of both gene Pi alleles, especially in the PiZZ variant, whereas in the heterozygous state, the defect is partly compensated by the

normal allele that is found in individuals with PiMZ and PiMS phenotypes. A list of the previous candidate gene approach studies and GWAS of *SERPINA1* association in AAV is summarised in Table 1.3 (70, 71, 78, 80-83).

The balance of A1AT and PR3 has been suggested to play a critical role in GPA pathogenesis. In the normal physiological state, A1AT binds to PR3 and antagonises its pro-inflammatory activity. However, when there is an excessive amount of PR3 in the blood serum due to a defect of the *SERPINA1* gene or inherited predisposition for an enhanced expression of the *PRTN3* gene, this may trigger the synthesis of anti-PR3 ANCA and promote PR3-mediated proteolytic vessel damage. Interestingly, it has been reported that an elderly individual with GPA was successfully treated with A1AT replacement therapy and steroids with regression of skin lesions and normalization of renal function (84).

Table 1.3: Previous candidate gene studies and GWAS of *SERPINA1* association in AAV

Study, Year	Population	SERPINA1	No. of Cases	OR	P-value
Elzouki, 1994	Swedish	PiZ allele	88 AAV (66 GPA)	6.00	1.00×10^{-4}
Borgmann, 2001	German	PiZ allele	79 GPA	3.50	$<1.00 \times 10^{-4}$
Mahr, 2010	German	PiZ and PiS alleles	433 GPA	14.58	2.00×10^{-3}
†Lyons, 2012	European	§rs7151526	1,521 PR3	0.53	5.60×10^{-12}
†Xie, 2013	North American	rs1956707	987 GPA	na	9.00×10^{-3}
Pervakova, 2016	Russian	PiZ and PiF alleles	38 GPA	9.68	3.42×10^{-2}
†Merkel, 2016	North American	§rs28929474	1,361 PR3	2.43	1.29×10^{-13}

AAV, ANCA-associated vasculitis; GPA, granulomatosis with polyangiitis; GWAS, genome-wide association studies; na, not applicable; OR, odds ratio; PR3, proteinase 3.

†SNPs showing the most significant association in the GWAS.

§These associations were stronger when PR3-AAV was compared with healthy controls, rather than by clinical syndrome (i.e. GPA). In contrast, no association was observed when MPO-AAV was compared with healthy controls.

CTLA4

CTLA4 encodes for cytotoxic T-lymphocyte associated protein 4, which is also known as cluster of differentiation 152 (CD152). It is a transmembrane protein of the immunoglobulin gene superfamily with a single variable domain and it shares 76% sequence homology with CD28, which is an important co-stimulator of T cells and it binds to the same ligands, CD80 (B7-1) and CD86 (B7-2) (85). *CTLA4* is expressed by T lymphocytes following activation and it is known to bind to CD80 and CD86 with markedly higher affinity than CD28. However, it functions as a CD28 antagonist and has an essential inhibitory role in the regulation of immune responses. Hence, the B7-CD28/CTLA-4 pathway is crucial in regulating T-cell activation and tolerance. This is most clearly demonstrated by the phenotype of *CTLA4* deficient mice, which die at a young age from the consequences of profound uninhibited T-cell proliferation, leading to lymphocytic infiltration and destruction of major organs (86). In addition, Ipilimumab (CTLA-4 blockade) which is now used in mainstream Oncology practice, has demonstrated promising results in improving survival of melanoma through tumour-specific immune attack, but has also been associated with autoimmune related adverse events affecting multiple organs of the body including the gastrointestinal tract, skin, lungs, kidneys, joints, liver, endocrine system and central nervous system (87).

The *CTLA4* gene locus is polymorphic in the human population and has been identified to be associated with several autoimmune diseases, e.g. RA (88-90), T1D (91-93), Graves' disease and Hashimoto's thyroiditis (93-96), coeliac disease (97), alopecia (98), Crohn's disease, ulcerative colitis (UC), primary sclerosing cholangitis (PSC), psoriasis and AS (99). A list of the previous candidate gene approach studies and GWAS of *CTLA4* association in AAV is summarised in Table 1.4 (70, 76, 78, 100-108). The published results are conflicting, which is perhaps not surprising given the small number of patients included in each cohort. This locus also did not achieve genome-wide significance ($p < 5.0 \times 10^{-8}$) in both the EVGC and VCRC GWAS. Interestingly, Steiner et al previously demonstrated that *CTLA4* was more highly expressed by unstimulated peripheral blood CD4+ T cells from GPA patients than the healthy control group. However, upon mitogen stimulation, T cells from GPA patients were defective in their expression of *CTLA4*, similar to findings observed in human immunodeficiency virus infection, suggestive of underlying T-cell function impairment (109).

Table 1.4: Candidate gene studies and GWAS of *CTLA4* association in AAV

Study, Year	Population	CTLA4	Cases/Controls	OR	P-value
Huang, 2000	Swedish	*CTLA4* (AT) 86 repeat	32 GPA/109	na	$<1.60 \times 10^{-3}$
Giscombe, 2002	Swedish	rs5742909	32 GPA/122	2.04	2.59×10^{-2}
Tsuchiya, 2003	Japanese	na	69 AAV/na	na	ns
Zhou, 2004	North American	*CTLA4* (AT) 86 repeat	117 GPA/123	0.38	1.00×10^{-3}
Spriewald, 2005	German	rs231775	32 GPA/91	na	ns
		rs5742909			
Slot, 2008	Dutch	rs231775	102 AAV/188	1.53	1.00×10^{-2}
		rs733618		na	ns
		rs4553808			
		rs5742909			
Kamesh, 2009	British	rs231775	222 AAV/629	1.32	1.00×10^{-2}
		rs3087243		0.67	1.00×10^{-4}
Carr, 2009	British	rs3087243	575 AAV/9,040	0.84	6.40×10^{-3}
†Lyons, 2012	European	rs16840252	2,267 AAV/6,858	0.87	1.28×10^{-5}
Chung, 2012	North American	rs231775	880 GPA/1,969	0.82	§1.00×10^{-3}
		rs3087243		0.79	§9.83×10^{-5}
†Xie, 2013	North American	rs231726	987 GPA/2,731	na	3.00×10^{-3}
		rs3096851		na	3.00×10^{-3}
Wu, 2017	Chinese	rs231775	176 AAV/485	1.04	ns
		rs5742909		1.02	ns

AAV, ANCA-associated vasculitis; GPA, granulomatosis with polyangiitis; GWAS, genome-wide association studies; na, not applicable; ns, not significant; OR, odds ratio.

†SNPs showing the most significant level of association in the GWAS.

§A set of 384 SNPs was analysed in this study and the p-values had not been adjusted for multiple testing.

PRTN3

PR3 is encoded by *PRTN3* gene. It is a serine protease located in the azurophilic granules of the neutrophils. In the normal physiological state, PR3 is mainly inhibited by A1AT. There are two alternative transcripts of *PRTN3* – canonical (*PRTN3-201*) and alternative (*PRTN3-202*) which initiates transcription at the intron 1 promoter. The canonical isoform is involved in anti-bacterial activity (110) while the alternative isoform encodes a 24 kD protein with a

sequence similar to that previously described for myeloblastin (111), which is involved in the proliferation and differentiation of neutrophils (112). The enzyme PR3 is translocated to the cell membrane upon stimulation with a low dose of pro-inflammatory cytokines such as TNF (113, 114) or by isolation procedure (115). This key process is called priming and it is the membrane-bound form of the PR3 that is able to interact with the ANCA, precipitating neutrophil activation and endothelial adhesion. Halbwachs-Mecarelli and colleagues previously demonstrated that there is a bimodal distribution of membrane PR3 expression (mPR3+/-) and the proportion of mPR3+ and mPR3- cells remains highly stable despite priming and over prolonged periods of time in a given individual (116). This finding was subsequently replicated (117) and more importantly, the authors also reported that a high concordance of mPR3+ percentage exists in monozygotic twins but not in dizygotic twins, which strongly suggests that the membrane expression of PR3 is genetically determined.

Over the last two decades, several observations linking PR3 to AAV have been made and these include (i) circulating leukocytes from patients with AAV displayed an increased transcription of the *PRTN3* gene (118, 119), (ii) AAV patients had increased levels of circulating PR3 in the plasma (120), (iii) the proportion of neutrophils with high membrane PR3 expression was increased in AAV patients (115), (iv) deficiency of A1AT seems to predispose to development of PR3-AAV and (v) a SNP in the promotor region of the *PRTN3* gene (-564A/G or rs62132295) was found to be associated with GPA (121) and this was further replicated in two GWAS (70, 71). In addition, rs62132293 which is in high LD with rs62132295 (r^2 = 0.92, D' = 1.0) has recently been reported to be an expression quantitative trait loci (eQTL) in neutrophils where the PR3-AAV risk allele (G) homozygotes had a higher expression of *PRTN3* than individuals with CC or CG genotypes (71).

In contrast to the aforementioned findings, Pieters K et al interrogated whether -564 A/G polymorphism caused quantitative differences in PR3 expression using a luciferase reporter assay. They transfected four different cell lines (promyelocytic HL60, NB-4 cells, monocytic U937 cells and HeLa) with three variants of the *PRTN3* promoter (-564A/G, -621 A/G and wild-type) which were cloned into a luciferase reporter vector. They demonstrated a good correlation between endogenous PR3 messenger ribonucleic acid (mRNA) expression and the promoter activity, but no significant difference in activity observed between the wild-type

and polymorphic variants which argued against the -564 A/G polymorphism being responsible for the increased expression levels of *PRTN3* seen in myeloid cells from GPA patients (122). Abdgawad M and colleagues also reported similar findings involving a cohort of 63 AAV patients and 107 healthy controls. They found that AAV patients had increased plasma concentrations of PR3 as well as mPR3-positive neutrophils compared to healthy controls, but found no skewed distribution of the -564 A/G polymorphism in the *PRTN3* gene. They concluded that increased plasma PR3 and high expression of mPR3 are associated with AAV, but not as a consequence of the -564 A/G polymorphism of the *PRTN3* gene promoter (123).

PTPN22

Protein tyrosine phosphatase, non-receptor type 22, also known as lymphocyte tyrosine phosphatase, is a protein that is encoded by the PTPN22 gene and is primarily expressed in lymphoid tissues. The important role of *PTPN22* as a risk factor for autoimmunity was first recognized with the association of a missense SNP (1858C>T; rs2476601; R620W) with an increased risk of T1D (91-93, 124, 125) and it encodes an arginine to tryptophan substitution at amino acid 620 (Arg620Trp) in the first proline-rich motif of the PTPN22 protein. This susceptibility locus has subsequently been found to be associated with numerous autoimmune diseases including RA (90, 126, 127), SLE (128, 129), juvenile idiopathic arthritis (JIA) (130), autoimmune thyroid diseases (93, 96) and Crohn's disease (131). The frequency of the PTPN22 1858T allele varies among different populations where it is most commonly found in Northern and Eastern Europe (>10%) and is encountered less frequently in Southern Europe (2-3% of the populations). The frequency in the US, Australia and New Zealand white populations mirrors those of the Northern Europeans (6-10%). However, it is rare in Native American (<1%), African (<1%), Middle Eastern (0-3%) and Asian populations (<1%) (132). The published genetic association results for *PTPN22* in AAV are summarised in Table 1.5 (48, 70, 71, 78, 100, 133-135).

PTPN22 has an N-terminal catalytic domain, an interdomain region and a C-terminal domain with four proline-rich regions facilitating interaction with other proteins. In the adaptive immune system, it functions as a regulator of tyrosine phosphorylation-based cell signal transduction by removing phosphate groups from tyrosine residues on intracellular protein, which is the natural counterpart of protein tyrosine kinases and inhibit T-cell activation (136).

While in the innate immune system, PTPN22 selectively promote type I interferon (IFN) production by enhancing the downstream signalling of pattern recognition receptors (137). T-cell activation via the engagement of T-cell receptor (TCR) initiates a cascade of signalling events which induces proliferation, mobilisation and differentiation of T cells. This involves a sequence of tyrosine-based phosphorylation events mediated by kinases of the Src family (Lck and Fyn) and the Syk family (ZAP-70), which cause phosphorylation of the immunoreceptor tyrosine-based activation motifs (ITAMs) of the TCR-associated CD3ζ-chain and trigger multiple intracellular signalling pathways (138). The binding of C-terminal Src kinase (Csk) to the proline rich region of PTPN22 is thought to further augment the inhibitory tuning of Src family kinases and reduce TCR signal transduction (139). However, the negative effect of this interaction has been challenged as other studies have found that the PTPN22/Csk complex could limit the activity of PTPN22 via promoting phosphorylation of an inhibitory residue on PTPN22 (140) or by altering its localisation within cell membrane lipid rafts (141).

A series of approaches including transfected cell lines, primary human lymphocytes and animal models have been employed to understand how the genetic risk variant (PTPN22 Arg620Trp) contributes to development of autoimmunity. In Jurkat T cells lines, co-transfection of the autoimmune-predisposing allele (PTPN22 Trp620) with Csk resulted in reduced interaction with Csk as compared to wild-type (124) and an increase in TCR responses (142). On the contrary, when PTPN22 Trp620 was transfected without the addition of Csk, a blunting of the TCR signal was observed (143). These differences with respect to the impact on TCR signalling have raised questions as to whether PTPN22 Trp620 is a gain or a loss of function allele. Rawlings and Buckner speculated that the differences in signalling analyses could be due to the relative expression levels of PTPN22 and Csk, where in the setting of limiting Csk, the localisation of PTPN22 to the cell membrane may be enhanced and therefore resulting in increased inhibitory activity while in cell lines where PTPN22 and Csk are expressed to a comparable level, the loss of interaction between the PTPN22 Trp620 and Csk may emphasise its loss of function in inhibition of TCR signalling (144).

Table 1.5: Candidate gene studies and GWAS of *PTPN22* association in AAV

Study, Year	Population	PTPN22	Cases/Controls	OR	P-value
Jagiello, 2005	German	rs2476601 [A]	199 GPA/399	1.75	2.00×10^{-3}
Carr, 2009	British	rs2476601 [A]	641 AAV/9,115	1.40	1.40×10^{-4}
Martorana, 2012	Italian	rs2476601 [A]	143 GPA/945	1.92	5.00×10^{-3}
			102 MPA/945	1.59	1.07×10^{-1}
			99 EGPA/945	0.52	1.51×10^{-1}
Cao, 2012	North American	rs2476601 [A]	230 AAV/1,178	1.45	3.68×10^{-2}
Chung, 2012	North American	rs2476601 [A]	895 GPA/1,976	1.35	2.00×10^{-3}
†Lyons, 2012	European	rs6679677 [A]	1,353 AAV/1,599	0.80	2.60×10^{-1}
†Xie, 2013	North American	rs2476601 [T]	750 GPA/1,820	na	6.00×10^{-3}
†Merkel, 2016	North American	rs2476601 [A]	1,986 AAV/4,723	1.36	1.86×10^{-7}
			1,361 PR3/4,723	1.33	3.19×10^{-5}
			378 MPO/4,723	1.64	5.85×10^{-6}
		rs6679677 [A]	1,986 AAV/4,723	1.40	1.88×10^{-8}
			1,361 PR3/4,723	1.36	7.89×10^{-6}
			378 MPO/4,723	1.71	8.83×10^{-7}

AAV, ANCA-associated vasculitis; EGPA, eosinophilic granulomatosis with polyangiitis; GPA, granulomatosis with polyangiitis; GWAS, genome-wide association studies; MPA, microscopic polyangiitis; MPO, myeloperoxidase; na, not applicable; OR, odds ratio; PR3, proteinase 3.

†SNPs showing the most significant level of association in the GWAS.

In human lymphocytes, several important observations have been made regarding the role of PTPN22 Trp620 in immune responses. Vang et al demonstrated that T cells from T1D patients carrying PTPN22 Trp620 produce less interleukin-2 (IL-2) upon TCR stimulation (143) and this was subsequently confirmed in a larger cohort (145). A consistent finding was demonstrated in healthy controls where T cells from individuals carrying the risk allele (PTPN22 Trp620) were found to have blunted Ca^{2+} flux, IL-2 and IL-10 production following anti-CD3 stimulation and the effects were most pronounced in the memory T cell compartment (146). Other downstream effects of PTPN22 Trp620 on T cell maturation include the expansion of CD4 memory T cell compartment (146) and favouring Th1 responses

(147). More recently, Ferreira et al reported the association of PTPN22 Trp620 with increased frequencies of thymically-derived Tregs in the peripheral blood of SLE patients, as well as increased PD-1 expression on both Tregs and effector T cells (148). In addition, studies of normal human B cells indicate that PTPN22 Trp620 might directly contribute to the development of autoantibodies and autoimmunity through its influence on BCR signalling and B cell development (149, 150).

The murine ortholog of *PTPN22* encodes the protein PEST domain-enriched tyrosine phosphatase (PEP). An early study from Zikherman J et al demonstrated that the PEP$^{-/-}$ mice have increased memory and effector T cells and spontaneous germinal center production but do not develop systemic autoimmunity. However, crossbreeding of the PEP$^{-/-}$ mice with non-autoimmune background mice (C56BL/6) that express an activating mutation within CD45 results in a SLE-like syndrome (142). Recently, Zhang et al generated knock-in mice expressing PEP R619W (a mutant analogous to PTPN22 R620W) and they developed enlargement of the thymus and spleen, accompanied by expanded memory/effector T cell populations and hyperresponsiveness in both lymphocyte and dendritic cells, consistent with a loss-of-function effect (151). Similarly, Dai and colleagues also generated knock-in mice expressing PEP R619W which displayed T-cell expansion, transitional, germinal center and age-related B-cell expansion, accompanied by autoantibody production and development of systemic autoimmunity (152).

Overall, both individuals with PTPN22 Trp620 and knock-in animals exhibit an expansion of effector memory T cells. However, the discordant response to antigen-receptor triggered signalling between human and murine studies where human memory T and B cells from individuals carrying the risk variant exhibit significantly blunted signals upon intracellular calcium flux and the induction of tyrosine-phosphorylated signalling effectors while exaggerated signals are observed in thymocytes, naïve T cells and effector T cells derived from knock-in mice, still require further investigation. Furthermore, the role of PTPN22 Trp620 in shaping the B cell repertoire and B cell selection has yet to be elucidated.

IRF5

Interferon regulatory factor 5 (IRF5) is a regulator of type 1 IFN gene expression, involved in the production of type 1 IFN and pro-inflammatory cytokines such as IL-6, IL-12, IL-17, IL-23, TNF, IFN-gamma-inducible protein 10, MCP-1 and RANTES in response to toll-like receptor (TLR) ligands such as lipopolysaccharide (LPS) or unmethylated DNA (153, 154). Increased expression of type 1 IFN and IFN-related genes have been reported in SLE (155-157), and are considered to have a significant role in the pathogenesis of the disease. In addition, genetic variants in *IRF5* have been strongly linked to SLE in multiple independent case-control cohorts from different ancestry backgrounds (129, 158-161). It is also implicated in other autoimmune diseases such as RA (127), primary biliary cirrhosis (PBC) (162, 163), systemic sclerosis (164, 165), Sjogren's syndrome (166) and UC (167).

Graham et al reported a haplotype of the *IRF5* gene to be a risk factor for SLE. More interestingly, this haplotype was shown to be functionally relevant. Firstly, rs2004640 T allele alters a consensus splice donor site leading to expression of *IRF5* isoforms bearing an alternative exon 1 - exon 1B (168). Another independent cis-variant, rs10954213 which is located in the 3' UTR region, alters the polyadenylation of *IRF5* mRNA and hence its stability. Lastly, rs60344245 (30-bp- in-frame indel) which is located in a proline-, glutamic acid-, serine- and threonine-rich region known to affect the stability and function of protein, might have an impact on the transcription initiating effect of *IRF5* on its target genes, while rs2070197 is a further surrogate marker of common haplotypes formed by the other three polymorphisms (169). This haplotype (combinations of the three functional polymorphisms) provides a good mechanistic explanation of higher expression of *IRF5*, feeding forward to the type 1 IFN pathway and contributing to the pathogenesis of SLE, although it has yet to be determined whether this is mediated in an additive or epistatic manner.

Two previous studies investigated polymorphisms within *IRF5* in AVV patients with contradictory results. In a German cohort consisting of 601 GPA cases and 827 controls, Wieczorek et al identified a four SNP haplotype of *IRF5* as described earlier, which conferred a protective effect in GPA patients (rs2004640/rs60344245/rs2070197/rs10954213; G/ins/T/G; OR = 0.73, p_c = 0.0012, CI 0.62-0.85) a similar trend to that observed in SLE (170). On the other hand, Kawasaki et al found that the frequency of rs10954213 G allele was

increased in MPO-ANCA AAV (n = 232) as compared to healthy controls (n = 710, additive model, p = 0.023, OR = 1.27) (171). This locus did not achieve genome-wide significance in all three GWAS of AAV.

FCRG3B

Copy number variations (CNVs) are defined as the presence of variable copies of genomic regions as a result of deletion or duplication, with size ranging from kilobases to megabases in different individuals. CNVs have been estimated to account for about 13% of the human genome and it can be inherited or sporadic (172). CNVs can alter the gene dosage of the affected region and therefore influence the risk of developing a range of diseases. Association between *FCGR3B* CNV and a number of autoimmune diseases (e.g. SLE (173) and RA (174)) have been reported in the European population.

FcγRIIIb is a member of a family of low-affinity Fcγ receptors for IgG immune complexes where its expression is largely restricted to neutrophils and is anchored to the outer layer of plasma membrane via a phosphatidylinositol linkage (175). FcγRIIIb can also be shed from the cell surface upon activation. FcγR plays crucial role in initiation and regulation of antibody-mediated immune response as well as recognition and clearance of immune complexes. There are two isoforms of FcγRIIIb: FcγRIIIb-HNA1a (NA1) which has been associated with SLE (176) and FcγRIIIb-HNA1b (NA2) which has been associated with AAV (177). Wilcocks et al previously demonstrated that a higher *FCGR3B* copy number was observed in AAV patients but a lower copy number in SLE patients as compared to healthy controls (178). This finding is aligned with the pathophysiology of the disease given that ANCA mediates its pathogenic effect by activating neutrophils via the Fcγ receptor and should therefore be worsened by a higher *FCGR3B* copy number.

TLR9

Toll-like receptor 9 (TLR9) is a member of the TLR family, also known as pattern recognition receptors (PRRs) which recognise molecular structures that are broadly conserved among the pathogens, known as pathogen associated molecular patterns (PAMPs). The interaction between TLRs and PAMPs triggers intracellular signals and promotes inflammatory response

such as via type 1 IFN and pro-inflammatory cytokine production (179). Husmann et al previously investigated the associations of four SNPs (rs352162, rs352140, rs352139 and rs5743836) within *TLR9* in AAV patients comprising of 646 GPA, 164 EGPA and 53 MPA. They found that all four SNPs were strongly associated with GPA and interestingly, significant associations were also identified in MPA but each with the contrary allele. A much stronger association was observed when they compared cases to controls according to the ANCA status and again a reciprocal effect was seen in MPO-AAV (180). These findings were replicated in a cohort of 426 Dutch and British cases. The authors provided yet another piece of evidence that PR3-AAV differs from MPO-AAV genetically and that TLR9 signalling might be involved in disease pathogenesis, in favour of an infection model triggering autoimmunity.

MPO

A common *MPO* promoter polymorphism (-463G/A, also known as rs2333227) was previously identified to be linked to Alzheimer's disease (181, 182), lupus nephritis (183) and a range of cancers (184, 185). The -463 G/A polymorphism is within an Alu-encoded hormone response element, which includes overlapping binding sites for SP1 transcription factor, retinoic acid receptor and thyroid hormone receptor. -463G is within the SP1 binding site, while -463A largely inhibits SP1 binding, creating a binding site for oestrogen receptor alpha. In myeloid leukemia cells, the GG genotype was associated with two to three-fold higher levels of MPO mRNA and protein expression than in the GA or AA genotypes (186). Moreover, -463G was associated with stronger promoter activity than -463A in cellular transfection assays.

Reynolds and his colleagues previously looked at patients with different types of AAV for clinical associations with -463G/A SNP. They demonstrated that GG genotype was associated with an increased incidence of vasculitis in female MPO-AAV only while there was an association of the GA/AA genotypes with an increased relapse rate in patients with MPO-AAV (187). The authors suggested that the G and A alleles might be differentially regulated in males and females given that the association was only observed in females and this was supported by a Japanese study which demonstrated that the MPO activity was 2-fold higher in females than males (188), may be due to oestrogen receptor binding to the *MPO* promoter.

CD18

The β2 integrin (CD18) is a protein that is encoded by the *ITGB2* gene which is located at chromosome 21q22.3. It is the common β-subunit of several different heterodimers: lymphocyte function-associated antigen 1 (LFA-1), macrophage-1 antigen (Mac-1), integrin alphaXbeta2 and integrin alphaDbeta2. Integrins are known to participate in cell adhesion as well as cell-surface mediated signalling. Mutations in *ITGB2* results in leukocyte adhesion deficiency type 1 (LAD-1) where individuals present in early infancy with failure of umbilical cord separation, cutaneous, respiratory and gingival infections due to impairment of neutrophil and monocyte cell migration and phagocytosis (189). Gencik et al first reported a variant in the exon 11 of *ITGB2*, thymidine to cytosine transversion affecting the restriction enzyme AvaII recognition sequence which confers a higher risk of MPO-AAV (190). This was followed by a second study from the same group where they screened the entire coding and regulatory regions of the *ITGB2* gene to identify the potential causal variants that are involved in the pathogenesis of AAV. A total of 10 SNPs were identified, of which four showed significant associations with MPO-AAV including the former one (191). One of these SNPs is located in the alternative transcription start site which could influence the expression of the gene.

Others

A list of other polymorphisms that have been reported in small series to play a role in AAV is summarised in Table 1.6 (72, 100, 106, 192-202). Most of the variants are involved in the innate and adaptive immune systems. However, none of them had been replicated in the GWAS.

Table 1.6: A summary of the genetic associations that were reported in AAV

Study, Year	Population	Gene (variant)	Cases / Controls	OR	P-value
Papiha, 1992	British	Complement C4B	27 GPA/163	2.56	4.00×10^{-2}
Finn, 1994	British	Complement C3F	63 AAV/na	2.60	$<2.5 \times 10^{-2}$
Persson U, 1999	Swedish	Complement C3F	37 PR3/101	1.88	$<5.00 \times 10^{-2}$
Murakozy, 2001	German	IL-10-1082	39 GPA/72	0.72	$<5.00 \times 10^{-2}$
Zhou, 2002	Swedish	IL-10 promoter microsatellite	32 GPA/109	na	$<5.00 \times 10^{-2}$
Spriewald, 2005	German	IL-10-1082, -819, -592 haplotype	32 GPA/91	na	ns
		IFNG +874 (A/T)		3.14	[§]2.7×10^{-2}
		TNF promoter -238 (G/A)		5.01	[§]2.9×10^{-2}
Miyashita, 2006	Japanese	KIR2DS3	43 MPA/239	0.24	3.8×10^{-2}
Mamegano, 2008	Japanese	LILRA2 (rs2241524)	50 MPA/284	2.52	4.9×10^{-2}
Wieczorek, 2009	German	CD226	642 GPA/1,226	1.23	3.1×10^{-3}
Carr, 2009	British	CD226 (rs763361)	641 AAV/9,115	0.90	2.1×10^{-1}
		IL-2RA (rs41295061)	675 AAV/8,936	0.77	1.22×10^{-2}
Wieczorek, 2010	German	LEPR (rs8179183)	460 GPA/878	0.72	3.40×10^{-3}
		GHSR haplotype (rs519384, rs572169, rs509035)		1.30	6.70×10^{-3}
Zhou, 2012	Chinese	DEFB4 CNV	112 AAV/523	na	9.00×10^{-3}
Wu, 2014	Chinese	TLR2 (rs3804100)	76 MPA/501	na	1.80×10^{-2}

AAV, ANCA-associated vasculitis; CD226, cluster of differentiation 226; CNV, copy number variation; DEFB4, defensin beta 4A; GHSR, growth hormone secretagogue receptor; GPA, granulomatosis with polyangiitis; IFNG, interferon gamma; IL2RA, interleukin-2 receptor alpha chain; IL-10, interleukin-10; KIR2DS3, killer cell immunoglobulin-like receptor, two Ig domains and short cytoplasmic tail 3; LEPR, leptin receptor; LILRA2, leukocyte immunoglobulin-like receptor A2; MPA, microscopic polyangiitis; na, not applicable; ns, not significant; OR, odds ratio; PR3, proteinase 3; TLR2, toll-like receptor 2; TNF, tumour necrosis factor.

[§]These p-values had not been subjected to multiple testing correction.

1.5   The role of ANCA in the pathogenesis of AAV

Since the first description of ANCA by Davies et al in 1982 (9), followed by the identification of MPO and PR3 as the antigenic targets which paved the foundation of our current understanding of AAV, ANCA has become an important diagnostic marker and in most laboratories, screening for ANCA is done by an indirect immunofluorescence technique using human buffy coat cells as substrate. Sera positive for ANCA are further tested by direct enzyme immunoassay (EIA) to reveal their specificity for either purified PR3 or MPO as a target antigen, both of which can give rise to a c-ANCA or p-ANCA staining pattern respectively (203). There is a significant overlap of the ANCA patterns between GPA and MPA. PR3-ANCA are detected in 70-80% of the GPA patients, whereas about 10% of GPA present MPO-ANCA, which are more characteristic of MPA. In addition, approximately 10% of patients have no detectable ANCA (204). The cause of injury in ANCA-negative patients remains uncertain and whether this could be due to antibodies which cannot be detected using currently available immunoassays or epitope specificity remains unknown (205, 206). Several other possibilities have been raised including inhibition of assays for anti-MPO antibodies by caeruloplasmin fragments in serum (207), podocyte specific non-immune triggers (e.g. epidermal growth factor receptor and VHL/HIF oxygen-sensing pathway) leading to rapidly progressive glomerulonephritis which have been demonstrated in animal models (208, 209) and autoantibodies against human LAMP-2 (hLAMP-2), which was discovered by Kain et al in the late 90s. The autoantibodies against hLAMP-2 were found in untreated AAV patients presenting with pauci-immune necrotising crescentic glomerulonephritis (NCGN), both ANCA-positive and ANCA-negative (43, 210). However, the role of this autoantibody in AAV is not fully understood and further discussion is elaborated in section 1.6.2.

AAV predominantly affects the small vessels in the body (arteries, arterioles, venules and veins). In the acute phase, AAV is characterised a necrotizing vasculitis with infiltrating neutrophils and monocytes. Within days, the neutrophils undergo leukocytoclasis and disappear, to be replaced by mononuclear leukocytes including macrophages, monocytes and T lymphocytes. Histologically, AAV is characterised by the absence of immunoglobulin deposition in vessel walls, which indicates that the pathogenesis of AAV is different from the

pathogenesis of other vasculitides that are associated with extensive vessel wall deposition of immunoglobulin, indicative of immune complex vasculitis or anti-glomerular basement disease. Therefore, ANCA was originally thought to be an epiphenomenon, accompanying but not driving the disease, although this paradigm has been gradually overturned in recent years following the emergence of evidence supporting ANCA as the key player in AAV pathogenesis.

1.5.1 Clinical evidence supporting the pathogenicity of ANCA

AAV is characterised by vascular inflammation that is highly correlated with the presence of circulating ANCA. In general, high levels of ANCA are present in patients with active disease and the titres fall during therapy. Nevertheless, this is not an absolute correlation using the current clinical assays for ANCA and there is insufficient data to support its use as a biomarker for disease activity as rising titres have been observed in up to 40% of stable AAV patients (211). Bansal and Tobin reported a case of transplacental transfer of MPO-ANCA from a mother to a 33-week gestational age neonate resulting in neonatal pulmonary haemorrhage and renal involvement that was successfully treated with high-dose steroid therapy and plasma exchange (212). This case report is frequently cited as the most direct clinical evidence for the pathogenicity of MPO-ANCA in humans; though no further similar cases have been reported and the delivery of a healthy newborn despite transplacental transfer of high level of MPO-ANCA from a mother with MPA was previously described (213). The latter case suggested that anti-MPO alone might not be sufficient to exert the pathogenic effect and an inflammatory stimulus or tissue injury is required for the development of clinical vasculitis.

Another observation in humans that supports a pathogenic role of ANCA is the association of AAV development with specific drugs, such as propylthiouracil, minocycline, pimagedine, hydralazine, allopurinol, D-penicillamine, phenytoin, isotretinoin and lemivasole-adulterated cocaine (214). The clinical manifestations can range from cutaneous and mild systemic symptoms to invasive involvement of several organs and it is often associated with anti-MPO and p-ANCA rather than anti-PR3 and c-ANCA. Although there is no standardised approach to the treatment of drug-induced vasculitis, complete clinical resolution typically occurs after drug withdrawal, immunosuppressive treatment and is usually followed by a decline in ANCA titres.

Lastly, the effectiveness of B-cell targeting therapy (e.g. Rituximab) for remission induction strongly supports a pathogenic role for ANCA. In the RAVE trial, Rituximab was shown to be noninferior to daily cyclophosphamide (CYC) regimen for induction of remission in severe AAV and superior to CYC in relapsing patients (215). Similarly, in the RITUXVAS trial, Rituximab in combination with two pulses of CYC was as effective as a standard CYC regimen in inducing remission in newly diagnosed AAV patients with severe renal involvement (216).

1.5.2 Experimental data supporting the pathogenicity of ANCA

The spontaneous crescentic glomerulonephritis-Kinjoh (SCG-Kinjoh) mice develop a high degree of glomerular crescents and high titres of anti-MPO antibodies (217). However, this autoimmune-prone strain also produces other antibodies that might contribute to pathology combined with significant immune deposits found in the kidneys, rendering it not ideal as a model of tissue injury in AAV that is pauci-immune by definition. A pivotal breakthrough in this field was the demonstration of passively transferred anti-MPO immunoglobulin G (IgG) causing focal pauci-immune NCGN in wild-type mice over the duration of a week by Xiao et al (218), suggesting that antibodies alone are sufficient to induce crescentic glomerulonephritis in the absence of contribution from the cellular immune response, supporting the pathogenicity of MPO-ANCA. Although the degree of disease observed was mild, this model was further refined by Huugen et al by the administration of LPS (219), simulating the synergistic effect of infection in exacerbating autoimmune phenomena where the effect was primarily driven by TNF. Another autoimmune model has been developed using Wistar-Kyoto (WKY) rats by Little et al (220). By immunising the WKY rats with human MPO (hMPO), without further intervention, they developed antibodies specific for hMPO within two weeks and cross-reacted with rat MPO, resulting in vasculitis and pauci-immune NCGN by 6-8 weeks.

Interestingly, no spontaneous models of glomerulonephritis or vasculitis associated with anti-PR3 antibodies have been reported. Furthermore, similar approaches as described were employed to generate an animal model for PR3-AAV but proved unsuccessful in initial attempts, which have primarily been attributed to the differences in the structure and expression pattern of PR3 in humans and rodents as well as the absence of PR3 on the surface of mouse neutrophils (221). It was later demonstrated by van der Geld et al that immunisation

of mice and rats with chimeric human/mouse PR3 induced high titres of anti-PR3 antibodies, but not disease (222). It was not until recently that Little and colleagues generated humanised immunodeficient, non-obese diabetic (NOD)/severe combined immune deficient (SCID), IL-2 receptor knockout mice, which received human haemopoietic stem cells and developed a human-mouse chimeric immune system. Upon passive transfer of PR3-ANCA containing IgG derived from patients with severe systemic vasculitis, these mice developed mild glomerulonephritis and lung haemorrhage which supports the pathogenicity of PR3-ANCA (223).

Another interesting observation was made by Primo et al where the transfer of splenocytes from the autoimmune-prone NOD mouse following immunisation with recombinant murine PR3 (rmPR3) led to NCGN in immunodeficient NOD-SCID recipients, but no disease was observed in NOD-SCID mice which received splenocytes from rmPR3-immunised C57BL/6 control mice (224), suggesting a susceptible genetic background might be required for the development of disease in addition to the presence of ANCA. Recent developments in the generation of animal models for AAV have been successful, although none can explain why the immune tolerance to self-antigens is broken in the first instance, these models are a useful tool for investigating the molecular mechanisms of downstream events once the immune tolerance is broken.

Both MPO and PR3 are typically found in the cytoplasmic azurophilic neutrophil granules. Several in vitro studies showed that ANCA can activate neutrophils and mononuclear phagocytes by binding to their cognate antigen on the cell surface after priming with TNF (113, 225), causing an oxidative burst with release of reactive oxygen species (ROS), degranulation with release of MPO, elastase and PR3 (113, 223, 225), secretion of pro-inflammatory cytokines (226) and upregulation of cell adhesion molecules, which promotes attachment of neutrophils to vascular endothelial cells (227, 228), contributing to acute vascular injury. Therefore, a pathogenic model has been proposed where neutrophils might be activated by circulating ANCA to cause endothelial damage and organ dysfunction in the presence of pro-inflammatory signals.

1.5.3 Genetic studies supporting the pathogenicity of ANCA

Recent GWAS identified strong associations of AAV with SNPs located at the MHC class II region, *SERPINA1* and *PRTN3* (70, 71, 78). These SNPs were more strongly associated with GPA than MPA patients, and interestingly, the associations were even more significant with anti-PR3 positive vasculitis than anti-MPO. In the subgroup analysis, PR3-AAV was associated with *HLA-DP* while MPO-AAV was associated with *HLA-DQ* (70). These findings strongly support a pathogenic model for PR3-AAV in which an altered availability of PR3 autoantigen (as recently demonstrated by Merkel et al where the PR3-AAV risk allele (rs62132293:G) is associated with a higher expression of *PRTN3* in neutrophils) (71), together with impaired functional suppression by A1AT which is a natural physiological inhibitor of PR3, prompt the generation of anti-PR3 antibodies through permissive HLA presentation and eventually leading to AAV development. However, the genetic evidence for MPO-AAV is still lacking.

1.6 The possible mechanisms of ANCA emergence

As with most autoimmune diseases, the exact aetiologies of AAV are not known, but are undoubtedly multifactorial with a combination of genetic predisposition, environmental exposure and characteristics of the immune system likely responsible. There is increasing evidence to support the role of bacterial infection as the trigger that breaks tolerance, resulting in autoimmunity. The three main hypotheses that dominate the field are discussed below.

1.6.1 Complementary PR3 (cPR3) and anti-idiotype antibodies

Complementary protein is a protein that results from transcription of the anti-sense strand of the gene encoding for the original protein while anti-idiotype response is defined by the presence of an antibody that recognises the complementarity-determining region of another antibody, generating circulating immune complexes which lead to abolishment of self-nonself distinction and eventually, autoimmune processes (229). In 2004, Falk and colleagues inadvertently identified antibodies that could react with cPR3 peptide in the serum of GPA patients (230) and immunisation of mice with cPR3 peptides could induce production of both

anti-cPR3 and anti-PR3 antibodies. They demonstrated that antibodies to PR3 and cPR3 were bound to each other in both human and mice, supporting an idiotypic relationship. Another observation made by the former group was that the immunodominant part of cPR3 showed partial homology with various microbial peptides, including *S. aureus*. This could explain the relationship observed between chronic *S. aureus* colonisation and increased relapse rates in AAV (42) and the reduction in relapse following treatment with co-trimoxazole (231). However, these findings have not been confirmed by another group (232) and immunisation of Wistar rats with *S. aureus* did not induce vasculitis or NCGN (233). Though intriguingly, the rats in the control arm that were immunised with *Escherichia coli* (*E. coli*) did develop vasculitis and c-ANCA.

1.6.2 Anti-LAMP-2 and molecular mimicry

LAMP-2 is a heavily glycosylated membrane protein expressed in lysosomes and on the cell surface. It plays important roles in transporting intracellular proteins, maintaining cellular homeostasis and response to stress (234). Anti-hLAMP-2 has been described as a sensitive marker for AAV with pauci-immune NCGN. Kain et al reported that circulating anti-hLAMP-2 antibodies were present in 80-90% of patients with AAV, including ANCA-negative patients and were absent in the sera of healthy controls (43, 235, 236). The autoantibodies rapidly become undetectable following immunosuppressive treatment and during remission phase, suggesting a potential role as a biomarker for active disease. More importantly, they showed that passive transfer of rabbit anti-hLAMP-2 IgG could induce glomerulonephritis in WKY rats by cross-reacting with rat LAMP-2 (43). LAMP-2 IgG could activate primed neutrophils and induce apoptosis in endothelial cells in a similar fashion to ANCA MPO and PR3 IgG. Furthermore, they characterised two epitopes that were recognised by the autoantibodies to hLAMP-2 and one of which (LAMP-2 epitope P_{41-49}) has complete homology with FimH, a bacterial adhesion molecule. Rats immunized with FimH produced antibodies to rat and human LAMP-2 and developed pauci-immune NCGN. Considering all available findings, infection with fimbriated bacteria bearing FimH, such as *E. coli* and *Klebsiella pneumoniae*, might trigger cross-reactive autoimmune responses to LAMP-2 and cause AAV. The clinical significance of anti-LAMP-2 has yet to be confirmed as the majority of the data originate from

a single group and Roth et al recently reported a much lower prevalence of autoantibodies to LAMP-2 in AAV patients at 21% (237).

1.6.3   NETosis

Neutrophil extracellular traps (NETs) are released by neutrophils upon activation, consisting of granule proteins and chromatin combined together to form an extracellular structure to enhance their antimicrobial properties, serving as a physical barrier to trap invading microbes and keeping potentially noxious proteins like proteases from diffusing away and inducing adjacent tissues damage (238). Kessenbrock and colleagues demonstrated that even in the absence of microbial infection, NETs formation was observed in neutrophils incubated with purified IgG from AAV patients but not in those incubated with sera for healthy controls. Both antigenic targets, PR3 and MPO, were identified in the fibrous DNA deposits (239). Moreover, components of NETs were found in the vicinity of neutrophil infiltrates in the affected glomeruli and interstitium of AAV patients. Therefore, it has been suggested that NETs might play a role in the development of autoimmune diseases due to the persistent exposure of antigen-chromatin complexes to the immune system and that this could be enhanced by bacterial infection such as *S. aureus*, which has been shown to strongly induce NETs formation (240).

1.7   Immune cells and pathways involved in the pathogenesis of AAV

Both the innate and adaptive immune systems play important roles in the pathogenesis of AAV. As discussed in section 1.5.2, neutrophils represent both the target cells of the autoimmune response and key effector cells triggering endothelial cell damage. Mice lacking circulating neutrophils have been shown to be protected from anti-MPO IgG-induced NCGN (241), while all control mice developed NCGN. In addition, patients with active GPA were found to have elevated serum B lymphocyte stimulator (BLyS, a.k.a. B-cell activating factor) levels (242), which is a member of the TNF ligand superfamily cytokines that is known to be essential for B-cell differentiation, proliferation and immunoglobulin production. Although the authors did not further investigate the mechanisms and the cell types responsible for this elevation, it has been speculated that ANCA-activated neutrophils could be an important

source of BLyS (243). This is further supported by an in vitro study by Holden et al which demonstrated ANCA induced production and release of BLyS from activated neutrophils, promoting B-cell survival (244).

There has also been growing interest in recent years of the role of T cells in AAV, where evidence has emerged supporting its role in the pathogenesis of AAV. Firstly, ANCA production is T-cell dependent where B cells require stimulation from antigen-specific CD4+ T cells to generate high affinity class-switched antibodies (245). Secondly, in addition to granulomas, extensive T-cell infiltration in different renal compartments are commonly found in renal biopsies of acute AAV patients, suggesting a pathogenic contribution of T-cell mediated immune response (246-248). Furthermore, T-cell activation markers such as soluble IL-2 receptor and soluble CD30 have been shown to be increased in the plasma of PR3-AAV patients as compared to healthy controls and the levels of these markers correlated strongly with disease activity (249). McKinney et al recently identified and validated a novel transcriptional signature in CD8+ T cells with prognostic and clinical relevance in AAV patients and that was enriched for genes within the IL-7 and TCR signalling pathways (250, 251). Furthermore, several studies have consistently demonstrated defective regulatory T cell (T_{reg}) suppressive function in AAV patients despite conflicting results regarding the number of T_{reg} in the peripheral blood (252-254). Lastly, in the MPO-induced glomerulonephritis murine model, the severity of NCGN was attenuated in the absence of CD8+ T cells (255) and CD4+ T cells (256).

The importance of B cells in the development of AAV is supported by the clinical benefit of B-cell targeting therapy (section 1.5.1). The anti-CD20 monoclonal antibody rituximab has been widely used in AAV patients. It rapidly depletes CD20$^+$ B lymphocytes, presumably resulting in the interruption of the pathological cell lineage responsible for autoantibodies production and possibly abrogation of cytokine production and antigen presentation by B cells (257). In contrast, a distinct B cell subset with immunoregulatory properties has been termed regulatory B cells (B_{reg}). They are characterised by high CD5 surface expression and IL-10 production (258), which promote the development of T_{reg} and inhibit the differentiation of naïve T cells into Th1 or Th17 cells (259). Several studies have demonstrated a lower

frequency of B$_{reg}$ in the peripheral blood of active AAV patients, without functional impairment, when compared with patients who were in remission and healthy controls (260-262), suggesting that B$_{reg}$ might fail to prevent autoreactive responses and eventually result in autoimmunity.

Given the histopathological feature of pauci-immune glomerulonephritis and the absence of hypocomplementaemia, the complement system was originally thought not to be involved in AAV pathogenesis. However, there is emerging compelling evidence to support the role of alternative complement pathway activation in the development of AAV, where C5 and factor B deficient mice were protected from NCGN in the MPO-AAV murine model, whereas both wild-type and C4 deficient mice developed AAV (263). Additionally, treatment with an anti-C5 monoclonal antibody prior to the administration of anti-MPO IgG in mice ameliorated the development of NCGN (264). The importance of the complement anaphylatoxin C5a, which is a potent neutrophil chemoattractant, has been demonstrated by Schreiber et al where ANCA-stimulated neutrophils released C5a, acting on C5a receptor (C5aR), which in turn recruited and primed more neutrophils, resulting in an amplification loop (265). Importantly, the oral administration of a small molecule C5aR antagonist (CCX168, a.k.a Avacopan) prevented the development of NCGN induced by anti-MPO antibodies in the murine model of AAV (266). In humans, C5aR inhibition with Avacopan was shown in a randomised, placebo-controlled trial (CLEAR trial; https://clinicaltrials.gov/ct2/show/record/NCT01363388) to be an effective replacement for high dose glucocorticoids in AAV patients (267).

1.8   A comparison of PR3-AAV and MPO-AAV

Although patients with GPA and MPA have historically been considered as a single disease entity when it comes to treatment strategies and enrolment into clinical trials, GPA and MPA diverge in several important aspects as highlighted by Franssen et al almost two decades ago (268), such as in the extent of their association with PR3 reactive ANCAs compared to MPO reactive ANCAs, the association of GPA with granulomatous inflammation and the risk of relapse. There is a growing debate as to whether they are in fact part of a single disease spectrum or represent distinct clinical entities. Recent GWAS have clearly demonstrated a

genetic component in the pathogenesis of AAV and suggested a distinction between PR3-AAV and MPO-AAV, which questions whether AAV patients should be stratified based on ANCA serotype rather than by clinical syndrome. That might not only help to reduce selection bias in recruiting patients into clinical trials, but may also be prognostically and therapeutically relevant. The major similarities and differences between PR3-AAV and MPO-AAV are summarised in Table 1.7 (269, 270).

Table 1.7: Similarities and differences between PR3-AAV and MPO-AAV

Feature	PR3-AAV	MPO-AAV
Epidemiology	Northwestern Europe North America	Southern Europe Asia
Age at diagnosis	45-55 years	60-65 years
Pathology	Granuloma	Fibrosis
Organ involvement	Upper airways Lung nodules Higher number of organs involved	Renal Pulmonary fibrosis
Prognosis	Higher risk of relapse	Higher rate of initial treatment failure Increased long-term risk of ESKD
Response to therapy	Rituximab > CYC	Rituximab = CYC
Aetiology	*Staphylococcus aureus* Low vitamin D level	Silica exposure
Genetics	*HLA-DPB1, PRTN3, SERPINA1*	*HLA-DQ*
Pathophysiology In vitro interaction of ANCA with neutrophils	Not different	
Endothelial cells	PR3 induces apoptosis of endothelial cells	MPO induces production of intracellular oxidants
Monocytes	PR3-ANCA induces release of sFlt1	MPO-ANCA does not induce release of sFlt1

AAV, ANCA-associated vasculitis; ANCA, anti-neutrophil cytoplasmic antibody; CYC, cyclophosphamide; ESKD, end-stage kidney disease; MPO, myeloperoxidase; PR3, proteinase 3; sFlt1, soluble fms-like tyrosine kinase-1.

1.9 Complex genetic diseases and GWAS

Both genetic and environmental factors are thought to play important roles in the development and progression of autoimmunity. A better understanding of the genetic basis of autoimmune diseases has the potential to improve therapeutic strategies and prevention. In the 1970s, twin and family studies successfully identified the most important genetic risk of autoimmune diseases – the HLA region (59, 271-274). In the later decades, candidate gene approach studies led to the discovery of several other significant risk loci in autoimmune diseases such as *NOD2* (275, 276), *IL2RA* (277, 278)*, PTPN22* (144) *and CTLA4* (279). The genetic landscape subsequently changed tremendously following the emergence of GWAS in 2006 where hundreds and thousands of risk loci were discovered over the last decade, revealing the polygenetic basis of common diseases and implicating that both adaptive and innate immune systems play a major role in the pathogenesis of many autoimmune diseases, which translates into new therapeutic possibilities. For example, GWAS discoveries have sparked drug repurposing of biological medications targeting components of the IL-23 pathways which have now become the mainstay of treatments for psoriasis, psoriatic arthritis, AS and inflammatory bowel disease (280).

1.9.1 The GWAS Era

GWAS is a powerful tool which utilises hypothesis-free and rigorous statistical association mapping approaches to identify genetic variants that are strongly associated with human traits, predominantly SNPs. The primary goal of GWAS is to better understand the biology of disease. Advances in technology have facilitated researchers to conduct high-throughput SNP analyses in very large case-control populations at lower cost. Briefly, a GWAS is performed by genotyping a set of tagging SNPs using commercially available microarrays, generally over 500K SNPs which are carefully selected to capture a sufficient proportion of the common variations in the genome, enabling the identification of loci associated with disease. Moreover, GWAS of different cohorts can be combined into a meta-analysis despite being genotyped on different platforms, where a homogenous dataset can be derived using imputation methods such as IMPUTE2 (281), MACH (282) or Beagle (283), allowing estimation of unobserved genotypes using haplotype structures provided by the reference panels such

as HapMap or the 1000 Genomes Project (284). Imputation can thus provide more uniform coverage of genetic variation across the genome without the need for resequencing or genotyping, but is restricted to the variants that are available in the reference panel. The HapMap reference set contains samples that have been genotyped on two different platforms, the Perlegen-Affymetrix Human SNP Array 6.0 and the Illumina Human 1M-single BeadChip. The data set consists of 1,184 samples from 11 populations typed up to 2.5 million SNPs. In the 1000 Genomes Project (phase 3), the genomes of 2,504 individuals from 26 populations have been reconstructed using a combination of low-coverage whole-genome sequencing, deep exome sequencing and dense microarray genotyping (285). Overall, the existing HapMap and 1000 Genomes Project reference panels are characterised by a relatively small sample size and for low-frequency variants, a given genotype may not be represented adequately in the reference data set. More recently, a new reference panel from the Haplotype Reference Consortium (HRC) has become available, consisting of 64,976 human haplotypes at 39,235,157 SNPs constructed using whole-genome sequence data from 20 studies with a total of 32,488 individuals of predominantly European ancestry. This permits more accurate genotype imputation at minor allele frequencies as low as 0.1% (see Figure 1.2) and enriched GWAS coverage, allowing the discovery and refinement of causal loci (286).
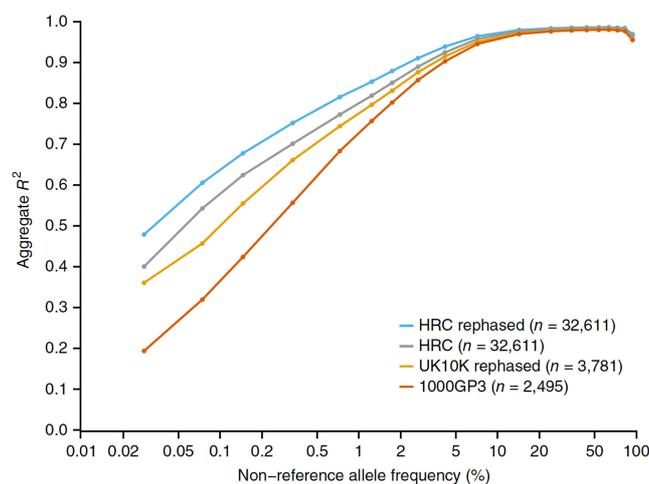
Figure 1.2: The performance of imputation using difference reference panels.

The x-axis shows the non-reference allele frequency of the SNP being imputed on a log scale while the y-axis shows the imputation accuracy measured by aggregate R^2 values when imputing SNP genotypes into ten CEU (Northern Europeans from Utah) samples. This figure was adapted from McCarthy et al (286).

1.9.2 The problem of missing heritability in GWAS

The heterogeneity of human traits and diseases in a population is the result of interactions between environmental influences and the associated genetic components. Heritability in a broad sense refers to the proportion of the phenotypic variance due to genetic variance, which can be further divided into three components: additive (average effect of substituting one allele for the other), dominance (interactions between alleles at the same locus) and epistatic (interactions between alleles at different loci) genetic effects. Heritability in a narrow sense (h^2) is defined as phenotypic variance due to additive genetic effects (287). Total h^2 is traditionally estimated using offspring and parental phenotypes, siblings or monozygotic and dizygotic twins. Despite the success in identifying SNPs that affect the variation in human complex traits and diseases, GWAS has its own fundamental limitations and it has been criticised for being able to explain only a small fraction of the heritability for most complex traits and diseases.

Several hypotheses have been put forward to explain the reasons for the missing heritability and to challenge the "common disease, common variant" hypothesis (288) that forms the initial basis of GWAS, predicting that the genetic component of complex diseases is the sum of the effects of common or frequent genetic variants. Much of the speculation has focused on the "common disease, rare variant" hypothesis, which argues that rare variants with relatively high penetrance (289) are the major contributors of the genetic component explaining the observed phenotypic variation in the population. The strongest argument in favour of the rare allele model comes from evolutionary theory where variants that are deleterious to fitness should be selected against and therefore should not be common. Pritchard argued that the selection processes such as mutation, random genetic drift and purifying selection against susceptibility mutations have acted on the human population during its expansion over the last few centuries and have led to multiple, very recent rare variations contributing to disease that are mildly deleterious but not subjected to overt negative selection for a long time, hence reaching an appreciable frequency and exhibiting extensive allelic heterogeneity (290). Furthermore, several rare structural variants have been identified to confer a substantial risk to autism, mental retardation, epilepsy and schizophrenia (172). On the contrary, Hunt et al performed deep exon sequencing for 25

GWAS risk genes in over 40,000 individuals comprising of 24,892 patients with six autoimmune diseases and found that rare variants within the coding regions have negligible effect in autoimmune disease susceptibility (291).

Another proposed model is the "infinitesimal model" where a large number of small effect common variants across the entire allele frequency spectrum contribute to the genetic component (287, 292). A software tool known as GCTA (Genome-wide Complex Trait Analysis), has recently been developed by Yang et al to estimate additive genetic variance (h^2_{SNP}) explained by all the SNPs simultaneously over the whole genome between unrelated individuals, fitting them as random effects in a mixed linear model, without needing to test for association for every individual SNP with the trait (293). It consists of two steps: (i) generating the genetic relationship matrix (GRM) between individuals, and (ii) estimating the variance explained by all SNPs by a Restricted Maximum Likelihood (REML) analysis of the phenotypes with the GRM (294). However, h^2_{SNP} is not the same as h^2 because not all the SNPs are causal variants and common SNPs in the genotyping arrays are unlikely to tag all causal variants perfectly for a trait. This method was applied to a number of traits such as height (293), body mass index (295), cognitive ability (296, 297) and schizophrenia (298) and indeed a large proportion of heritability can be explained by all the common SNPs, suggesting that most heritability is hidden rather than missing (299) where the effects of these hidden SNPs are too small to reach the stringent GWAS p-value threshold and hence escape identification (293).

Other proposed models include the "broad sense heritability model" which supports genetic variation being the result of the combination of genotype-by-genotype interactions (epistasis), genotype-by-environment interactions and epigenetic effects (300, 301) and the "threshold dependent model" which posits that complex disease is more likely to arise in individuals who have extreme values of multiple intermediate phenotypes that are involved in the disease pathogenesis and often, each of these intermediate phenotypes has a component of genetic inheritance that might not be detectable at the level of the main complex trait (302). The missing heritability could be at least partly attributed to the sum of the heritability of each of these intermediate phenotypes. The concept of "mediated

pleiotropy" was then coined by Solovieff et al to describe a genetic variant that is linked to a complex phenotype because the genetic variant is associated with an intermediate phenotype that in turn is causally related with the outcome complex phenotype (303). This is different from the concept of "biological pleiotropy" where one or more genetic variants are associated with two unrelated phenotypes. In recent years, a new statistical strategy known as Mendelian randomisation (MR) has been developed to identify causal association of an intermediate phenotype with an outcome, by using the SNPs associated with the intermediate phenotype. An example of successful application of MR analysis in autoimmune diseases is the link between vitamin D and multiple sclerosis (MS), with two studies demonstrating that decreased levels of vitamin D are causally associated with an increased risk of MS, supporting the epidemiological observation (304, 305).

Ultimately, the components of missing heritability remain unknown and might probably be attributed to a combination effect of all the proposed explanations as discussed above (306).

1.9.3   The challenge from GWAS to biology

Since a majority of the identified SNPs do not affect protein coding sequences of genes, one major obstacle in the post-GWAS era is to identify the causal gene(s) or non-coding ribonucleic acids (RNAs) in an unbiased and structured manner as well as the pathways via which the causal variants exerting their biological effects on altering the disease risk. There is temptation to label the gene nearest to the variant with the smallest p-value (the lead "SNP") as the most likely causal gene but the physical distance of a variant to a gene is not a compelling evidence of causality. With the availability of big data derived from large consortia and cohorts, advanced molecular technologies including single-cell genomics and new computational and statistical methods, a transformative era has begun to provide opportunities to bridge this existing knowledge gap. This process can be broadly divided into two main stages: (i) fine mapping of candidate causal variants and (ii) linking the potential variants to likely genes whose perturbation leads to altered disease susceptibility by functional annotation (Figure 1.3).
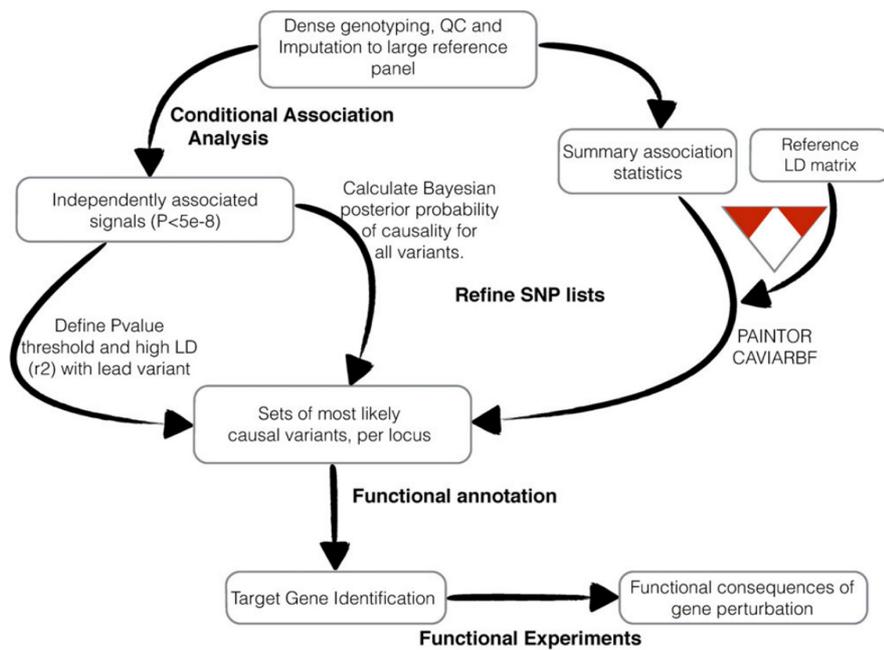
Figure 1.3: A generalised fine-mapping pipeline.

This figure depicts the common steps to progress from the discovery of associated variants to the identification of the potential causal genes and pathways. This figure was adapted from Spain et al (307).

Several statistical methods have been developed for the prioritization of causal variants to explain association signals. One crucial assumption when evaluating the relative evidence of each associated SNPs being the causal is that the true causal SNP is being considered too. A simple approach is to consider all SNPs with a p-value less than a certain threshold (e.g. standard genome-wide significant threshold of 5×10^{-8}), as candidates for causality. This is hardly sensible as p-values are influenced by study-specific factors such as power, which is in turn determined by sample size and locus-specific factors such as minor allele frequency (MAF) and effect size. Therefore, such a threshold is not directly comparable between studies. A slightly more sophisticated method involves considering all SNPs above a certain LD threshold with the lead SNP as potential causal. Although this is less arbitrary than p-value thresholds, it still ignores the properties of the dataset. An alternative approach is the Bayesian method that takes account of the power of the dataset at the cost of additional modelling assumptions and assigns posterior probabilities of causality to each SNP that are comparable across different studies (308). The resolution of fine mapping at a locus can then be assessed by constructing a "credible set" of variants that are most likely to be causal and

be prioritised for further investigation and functional validation. A further refinement of the resolution of fine mapping could be achieved by performing trans-ethnic meta-analysis which harnesses the different LD structures from ethnically diverse populations. For example, a genome-wide trans-ancestry meta-analysis was performed using MANTRA in type 2 diabetes comprising of 26,488 cases and 83,964 controls from European, east Asian, south Asian, Mexican and Mexican American ancestry, demonstrating considerable improvements in refining the association signals in several susceptibility loci – the greatest change was observed at the *JAZF1* locus where the credible set was reduced to only four SNPs with a genomic interval of 16kb (309).

Although protein coding variants are easier to build a case for prioritisation and to design experiments to investigate the functional consequences, they only account for a very small fraction of GWAS hits. Several large efforts such as the ENCODE (Encyclopedia of DNA Elements) (310), BLUEPRINT (311), ROADMAP (312) and FANTOM (313) consortia have greatly transformed our understanding of the roles of the noncoding genome and the mechanisms underpinning variation in gene expression in various cell types, by facilitating the functional annotation of the genome including mapping of post-translational histone modifications, transcription factor binding sites, CpG methylation and areas of open chromatin. A landmark study by Maurano et al demonstrated that there is a significant enrichment of disease- and traits-associated noncoding genome-wide significant variants in the regulatory DNA region marked by deoxyribonuclease I hypersensitive sites, that can alter chromatin accessibility and transcription factor recognition sequences which may explain the phenotype associations (314). Subsequent studies identified the enrichment of autoimmune disease associated SNP in the super-enhancers, which are defined by large clusters of transcriptional enhancers that regulate spatiotemporal gene expression and define cell identity (315, 316). Other public data resources such as the Genotype-Tissue Expression (GTEx) project (317) and whole blood analyses (http://www.eqtlgen.org/cis-eqtls.html) have been extremely useful in generating and compiling the expression quantitative trait loci (eQTL) data for a large of cohort of individuals in multiple tissues under many different conditions (318). However, the challenge remains in autoimmune diseases whereby there is paucity of eQTL studies in purified immune cell types. Recent advances in proteomic

technologies has offered another wealth of genetic associations with plasma protein abundance data (pQTLs) (319). In addition, the development of chromosome-conformation-capture techniques to determine the 3-dimensional structure of chromatin has facilitated the identification of long-range interactions between the regulatory elements containing disease associated SNPs and the promoters of target genes, thus activating or repressing transcription, depending on the nature of the regulatory elements such as enhancers, insulators or silencers (Figure 1.4) (320). The important role of distal enhancer in regulating lineage-specific gene expression has been confirmed by Javierre and colleagues through the identification of networks of interaction in 17 human primary haematopoietic cell types (321). Data obtained with these technologies have markedly improved our knowledge of the 3-dimensional architecture of the genome and their implications on human disease.
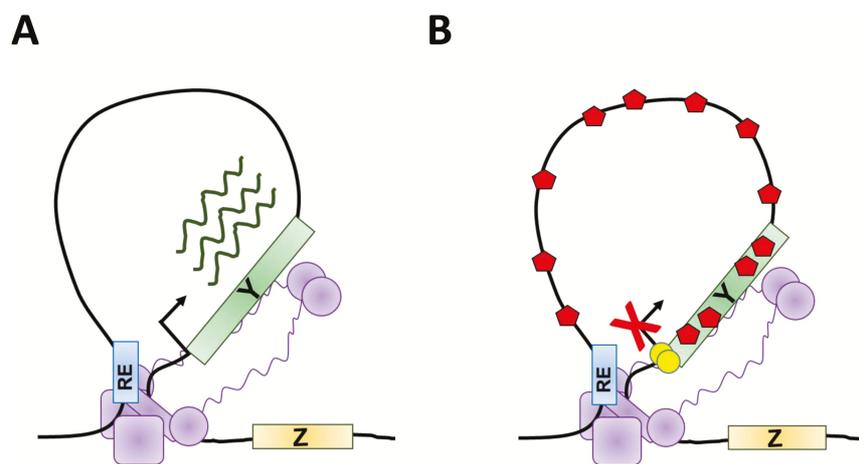
Figure 1.4: The 3-dimensional structure of chromatin.

The formation of DNA loops mediated by genome organising complexes (as indicated in purple) can modulate gene expression. (A) A regulatory element (RE), such as an enhancer, interacts with the promoter of gene Y, resulting in an increased expression of gene Y following the binding of transcription factors while in (B) the RE confers a repressive activity, either by DNA methylation of the promoter of gene Y as indicated by yellow circles or chromatin remodelling proteins that promote repressive histone modifications as indicated by red hexagons. This figure was adapted from Ray G et al (322).

Taken all together, these approaches complement the statistical methods. Each disease-related region in the genome can be fine mapped to obtain a credible set of causal variants. The most likely cell type driving the association can be identified by determining which of

these SNPs lie in the functional region of a particular cell type and the most likely genes that are being regulated can then be determined through eQTL and 3-dimensional chromatin contact. Ultimately, a better understanding of the genetic basis of the biological pathways and processes underlying the aetiology of autoimmune diseases has the potential to improve the success rate and safety of medical therapies development.

1.10 Aim and objectives

Previous GWAS have provided evidence that PR3-AAV and MPO-AAV are genetically distinct autoimmune syndromes, although not conclusive as only three loci specific to PR3-AAV (*HLA-DP*, *SERPINA1* and *PRTN3*) and one to MPO-AAV (*HLA-DQ*) were identified. The aim of this study was to perform a larger GWAS to uncover more disease specific genetic susceptibility loci associated with AAV. Specifically, the objectives were as follows:

I. To perform a larger GWAS powered to examine PR3-AAV independently;

II. To perform a larger GWAS powered to examine MPO-AAV independently;

III. To validate the top hit(s) by performing a meta-analysis of combined samples;

IV. To investigate the genetic relationship between PR3-AAV and MPO-AAV;

V. To elucidate the possible functional underpinnings of the validated top hit(s).

2 Materials and Methods

2.1 Genome-wide association study

2.1.1 Contributions

The study design was conceived by the European Vasculitis Genetics Consortium (EVGC). Cases and controls were ascertained through the EVGC, Wellcome Trust Case Control Consortium (WTCCC) and the European Prospective Investigation of Cancer Consortium, Norfolk (EPIC). Genotyping was performed at various centres described in section 2.1.4. The genotype calling and quality control (QC) were performed by Drs Tim F Rayner and Richard Coulson. The HLA imputation was performed by Dr Stephen Leslie (University of Melbourne, Australia). All other analyses were performed by myself.

2.1.2 Inclusion criteria and study subjects

Case patients had a clinical diagnosis of either GPA or MPA according to the EMEA algorithm, supported by either a positive ANCA assay or a diagnostic biopsy with c-ANCA or p-ANCA. The important issue of diagnostic criteria was previously discussed in the supplementary note of Lyons et al (70).

I included subjects from two European ancestry cohorts as described in Table 2.1. Briefly, GWAS1 was the discovery cohort from Lyons et al with control data from the WTCCC. The summary of patient demographics, clinical data and QC outcomes has been previously described (Table S4, Lyons et al) (70). GWAS2 was a newly recruited cohort consisting of 1,763 cases from 66 centres in 12 different European countries. Following QC, 1,738 remained and of these, 606 (34.9%) were MPO-ANCA positive and 1,132 (65.1%) were PR3-ANCA positive. Genotype data for 6,000 UK controls was obtained from the EPIC consortium (323). A total of 5,465 individuals remained following QC. We also recruited and genotyped 1,223 controls from 6 European countries. The breakdown of the GWAS2 cohort is shown in Table 2.2. Lastly, for rs6679677 (*PTPN22*), another cohort was included where the genotyping was previously

performed using the Sequenom MassARRAY platform. The study was approved by the institutional review board from each participating centre respectively and all individuals provided written informed consent.

Table 2.1: Description of the cohorts included in the meta-analysis

Meta-analysis	Cohort	Genotyping platform	[‡]Cases	[‡]Controls
Stage 1	[†]GWAS1	Affymetrix SNP6	914	5,259
	[§]GWAS2	UK Biobank Axiom array	1,738	6,688
Stage 2	*rs6679677	Sequenom MassARRAY	1,550	1,531

[†]GWAS1 was the discovery cohort from Lyons et al (70), of which 478 were PR3+, 278 were MPO+ and the ANCA status for the remaining cases were unknown.

[§]GWAS2 was the newly recruited cohort, consisting of 1,132 PR3+ and 606 MPO+ cases.

*The genotyping for rs6679677 was performed using iPLEX assays on the Sequenom MassARRAY platform. This cohort was part of the replication cohort used in the Lyons et al, of which 1,122 were PR3+, 347 were MPO+ and the ANCA status for the remaining cases were unknown.

[‡]The number of cases and controls provided here were those post QC.

Table 2.2: Breakdown of 1,738 cases and 6,688 controls by country for GWAS2

Country	Cases	Controls
Australia	46	0
Austria	40	0
Czech Republic	134	130
Denmark	76	0
France	228	0
Germany	140	273
Italy	145	266
Poland	60	118
Spain	10	93
Sweden	261	343
United Kingdom and Republic of Ireland	598	5,465 (EPIC)
Total	1,738	6,688

EPIC, the European Prospective Investigation into Cancer consortium (323).

2.1.3 Power calculation

Based on the sample size described in Table 2.1, power calculation was performed using Quanto (software by USC Biostats). For a MAF of 0.05, we had $\geq$ 80% power to detect genome-wide significant ($p < 5 \times 10^{-8}$) SNPs with expected ORs of $\geq$ 1.60 for PR3-AAV (see Figure 2.1A). However, we had less power to detect genome-wide significant SNPs with MAF of 0.01, where the expected OR was increased to $\geq$ 2.45. Similarly, we had $\geq$ 80% power to detect genome-wide significant SNPs with expected ORs $\geq$ 1.83 and 3.0 for SNPs with MAF 0.05 and 0.01 respectively for MPO-AAV (see Figure 2.1B).
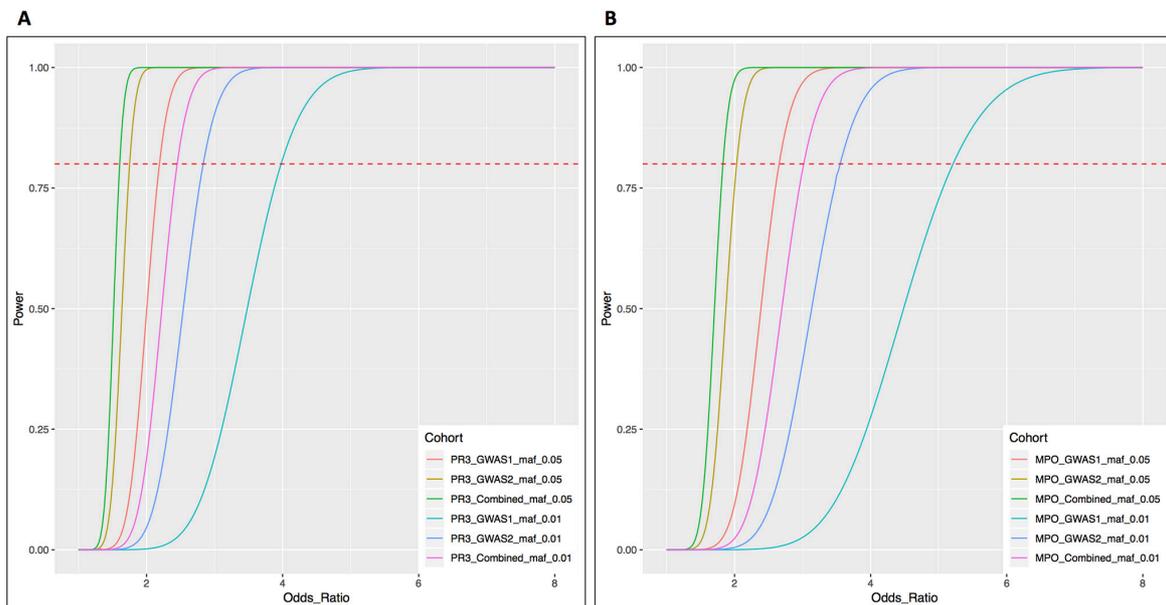
Figure 2.1: Power curves for the individual subset analyses.

Power curves are shown for both (A) PR3-AAV and (B) MPO-AAV cohorts. Power was calculated for various effect sizes assuming a minor allele frequency of 1% and 5% with an alpha value of 5×10^{-8}.

2.1.4 Genotyping and quality control of genotype data

Briefly, for GWAS1, the genotyping, calling and data QC were described in the supplementary appendix by Lyons et al (70). The genotyping was performed by AROS Applied Biotechnology (Aarhus, Denmark) using the Affymetrix SNP6 platform, consisting of 934,968 SNPs, of which 612,676 passed QC analysis. In addition, four SNPs were genotyped using TaqMan SNP

genotyping assays (Applied Biosystems) and these included rs28929474 and rs17580 (*SERPINA1*), rs5000634 (*HLA-DQ*) and rs62132296 (*PRTN3*).

For GWAS2, genomic DNA was extracted from whole blood using magnetic bead technologies at the Centre for Integrated Genomic Medical Research (Manchester, UK) according to the manufacturer's instructions. Both cases and controls were genotyped using the Affymetrix UK Biobank Axiom array according to the manufacturer's protocol. Genotyping of cases and non-UK controls was performed by AROS Applied Biotechnology (Aarhus, Denmark) while the genotyping of UK controls was performed by the EPIC-Norfolk consortium using the same array.

The UK Biobank Affymetrix Axiom array is a customised genotyping array comprising 820,967 SNPs and short insertions/deletions. The array includes the "exome" component, which were designed to capture variants likely to have transcriptional consequences including missense, splice altering and truncating and a good genome-wide coverage in European populations to ensure downstream good quality genome-wide imputation of variants that are common (MAF > 5%) or low-frequency (MAF = 1-5%). The exome component includes over 130,000 predominantly rare variants and was designed using data from three large exome sequencing projects: the NHLBI Exome Sequencing Project (ESP) (324), the Exome Aggregation Consortium (ExAC) (325) and the UK10K non-Finnish project (326). Additional rare variants were included in cardiac disease and cancer predisposition genes, as well as other variants from the Human Gene Mutation Database (HGMD) (327). The genome-wide imputation scaffold was designed by selecting tagging variants from Affymetrix databases using a custom algorithm (328). The mean R^2 between observed and imputed genotypes for common variants was estimated to be 0.92, while for low-frequency variants was 0.79. The remaining content on the array includes markers that are disease specific, including markers related to Alzheimer's, autoimmune and inflammatory, blood phenotypes, cancers, cardiometabolic, neurological disease as well as dense coverage of selected genomic regions such as ApoE, HLA, killer-cell immunoglobulin-like receptor, Y chromosome and mitochondria. Of particular relevance to this study, the array includes 258 variants related to autoimmune and inflammatory diseases and 2,545 variants related to blood phenotypes.

For GWAS2, genotype calling was performed using the Affymetrix Powertools software. Calling was performed in batches of contemporaneously run plates as per the manufacturer's protocol. After genotype calling, SNP data was processed in the following sequence using PLINK v1.9 (329). Samples with proportion of missing SNPs > 5%, sex mismatch or abnormal heterozygosity were removed. SNPs with a genotyping call rate < 98%, deviation from Hardy-Weinberg Equilibrium (p < 1 x 10^{-6}) or monomorphic were removed. The QC process was performed separately for each batch. The post-QC genotype data from each batch was subsequently merged. Duplicated or related samples (identified using Identity by State) were removed and the SNP QC was performed again so that SNPs with a genotype call rate < 98% or deviation from Hardy-Weinberg Equilibrium (p < 1 x 10^{-6}) in the combined data were excluded. SNPs that showed significant differential missingness (p < 1.0 x 10^{-5}) between cases and controls were removed. Finally, principal component analysis (PCA) of the post-QC genotype calls together with the 1000 Genome individuals was performed. The means and standard deviations of PC1 and PC2 were calculated for the EUR subset of the 1000 Genomes samples. Cases or controls which were outside of ± 3 standard deviations from the mean on either PC1 or PC2 were removed. Following these QC steps, 1,738 cases and 6,688 controls remained with a total of 557,075 autosomal SNPs (see Figure 2.2).
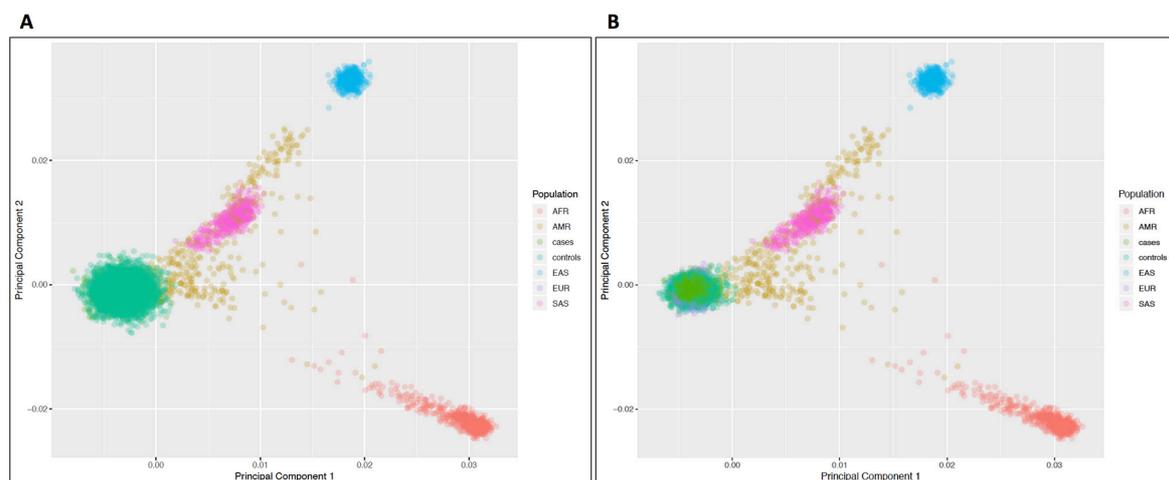
Figure 2.2: Principal component analysis (PCA) of genotype data.

PCA plots show ancestry of AAV cases and controls in (A) GWAS1 and (B) GWAS2 in relation to the 1000 Genomes Project individuals. The legend details the ethnic origin of samples included in the PCA. AFR, African; AMR, ad mixed African; EAS, East Asian; SAS, South East Asian and EUR, European.

2.1.5 Association testing on the directly genotyped data

Case-control association testing on the directly genotyped autosomal SNPs was performed using logistic regression with PLINK v1.9. Principal components (PCs) were calculated using plink --pca containing both cases and controls and included as covariates to adjust for potential confounding and population substructures. Although this traditional genomic control approach (330) carries a small risk of masking true disease associations, it is a more appropriate test given (i) the UK controls were genotyped in a different facility thereby potentially leading to confounding from a batch effect and (ii) French controls were not included in GWAS2. I also tested the effect of using increasing numbers of PCs as covariates on the genomic inflation factor (lambda, λ) which is the ratio of the median of the observed chi-squared statistics to the median of the expected chi-squared statistics under the null hypothesis. There was minimal inflation with a lambda of 1.03 for GWAS1, with no benefit of including further PCs. In GWAS2, the lambda was reduced from 1.31 to 1.15 using the first 20 PCs as covariates.

2.1.6 Association testing using a linear mixed model

As an alternative analysis approach, a linear mixed model (LMM) was employed to perform case-control association testing. This method compares the deviation of each individual from the population mean at each variant in the data set and compares individuals pairwise to establish a value for overall genetic similarity. This value will then be used as a random variable in a mixed linear regression which could capture population variance at a finer scale level than principal component analysis (331). The association testing was carried out using the BOLT-LMM software v2.3.2 (332, 333). Although the mathematical derivations underlying BOLT-LMM are based on a quantitative trait model, it can be applied to analyse case-control traits simply by treating the binary trait as a quantitative trait with the caveat that it can become miscalibrated for unbalanced case-control traits resulting in false positive association at rare SNPs. Moreover, a transformation of the effect size estimates (beta, β) on a quantitative scale to traditional ORs is required, using the formula $\log OR = \beta / (\mu*(1-\mu))$, where μ = case fraction. Similarly, the standard error (SE) of SNP effect sizes should also be divided by $(\mu*(1-\mu))$ when applying the transformation to obtain log ORs.

2.1.7 Pre-phasing and imputation

Pre-phasing is a process of identifying alleles colocalised on the same strand and estimating the haplotypes of GWAS samples. The current recommendation is that GWAS samples should first be pre-phased using the most accurate methods available such that the subsequent imputation step, which involves imputing alleles from one set of haplotypes into another set, will be more accurate and computationally efficient. As new haplotype reference sets become available, imputation can be repeated more rapidly (281). Imputation was performed on the Michigan Imputation Server v1.0.3 that facilitates access to the HRC reference panel (HRC release version 1.1 April 2016, consisting of 64,940 haplotypes, 32,470 samples of predominantly European ancestry) (286, 334). Several steps were executed including variant call format (VCF) file check, QC (where population EUR was selected for the allele frequency check), pre-phasing using Eagle2 (335) and lastly imputation was performed using minimac3.

2.1.8 Post-imputation quality control

Over 39 million variants were generated following imputation and therefore QC is paramount to create high quality data for downstream analyses. QC was performed with BCFtools version 1.2 using HTSlib 1.2.1, which is a program for variant calling and manipulating VCF files. SNPs with MAF < 0.01 (1 in 100), R^2 < 0.3 or which were monomorphic were removed (336). R^2 is a variance ratio, being calculated as the proportion of the observed dosage variance to the expected dosage variance, given observed allele frequency. Higher values of R^2 indicate more accurate genotype imputation. Table 2.3 shows the total counts of SNPs at each chromosome after applying the imputation quality and variant frequency filters for both the GWAS1 and GWAS2 datasets.

2.1.9 Association testing following imputation

Association testing on the imputed data was carried out using the SNPTEST v2.5.2 software with an additive genetic model (option '--frequentist 1'), with the first 20 PCs as covariates to adjust for population stratification. Uncertainty in the imputed genotypes was taken into account in the association testing by using a missing data likelihood score test. In addition, association testing was performed using a LMM with the BOLT-LMM software (332, 333).

Table 2.3: SNP summary for samples from GWAS1 and GWAS2

Chromosome	GWAS1		GWAS2	
	Imputation output	Filter at R^2 0.3	Imputation output	Filter at R^2 0.3
1	3,069,931	587,094	3,069,931	593,584
2	3,392,237	644,689	3,392,237	650,354
3	2,821,894	548,655	2,821,894	551,007
4	2,787,581	559,243	2,787,581	563,826
5	2,588,168	500,249	2,588,168	502,202
6	2,460,111	510,367	2,460,111	515,160
7	2,289,305	447,339	2,289,305	453,673
8	2,242,705	428,952	2,242,705	432,423
9	1,675,898	328,463	1,686,471	333,220
10	1,927,503	390,344	1,927,503	393,635
11	1,936,990	381,881	1,936,990	385,876
12	1,848,117	369,942	1,848,117	372,997
13	1,385,433	286,914	1,385,433	288,391
14	1,266,536	251,658	1,270,436	255,973
15	1,139,215	218,634	1,139,215	220,353
16	1,281,297	233,729	1,281,297	239,615
17	1,090,072	199,069	1,090,072	204,325
18	1,104,755	222,763	1,104,755	223,483
19	868,554	160,528	868,554	174,381
20	884,983	174,108	884,983	175,283
21	531,276	106,992	531,276	108,140
22	524,544	104,963	524,544	106,633
Totals	39,117,105	7,656,576	39,131,578	7,744,534

"Imputation output" lists number of SNPs as result of imputation and "Filter at R^2 0.3" lists number of SNPs passing the filtering criteria: MAF $\geq$ 0.01 and $R^2 \geq$ 0.3.

2.1.10 Meta-analysis

The results of the two cohorts (GWAS1 and GWAS2) were meta-analysed using the META v1.7 software with an inverse-variance method based on a fixed-effects model (option '--method 1'). Figure 2.3 provides an overview of the study design for GWAS of AAV.
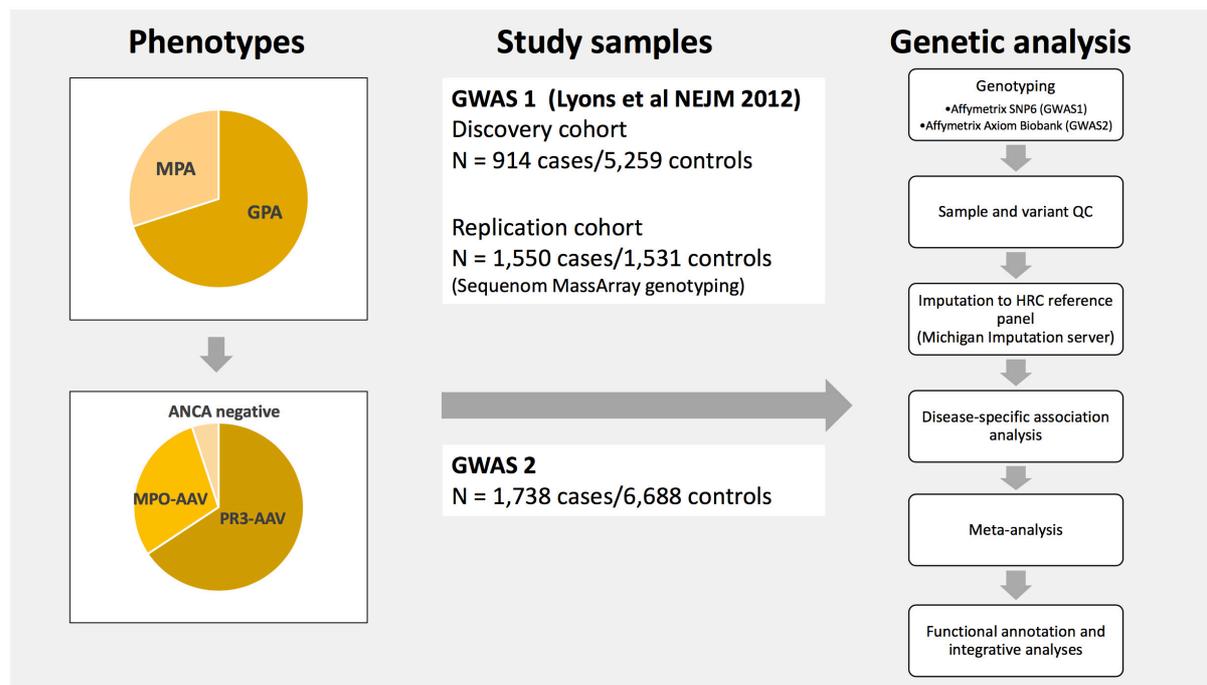
Figure 2.3: An overview of the study design for GWAS of AAV.

The phenotypes and their classification based on ANCA serotype, the study sample size and a summary of the data analysis methods employed to identify the disease-associated risk loci.

2.1.11 HLA imputation and association analysis

The HLA imputation was performed in the GWAS2 dataset. Two thousand seven hundred and seventy-seven SNPs, 343 classical HLA alleles to 2- or 4- digit resolution and 1,092 amino acid variants were imputed at 8 HLA loci (*HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1, HLA-DQB1, HLA-DPA1* and *HLA-DPB1*) from phased genotype data using the HLA*IMP:03 software (https://www.biorxiv.org/content/10.1101/091009v1) (337). Association testing for HLA variants was performed using a LMM with the BOLT-LMM software to adjust for population structure.

51

2.1.12 Heritability explained

A LMM method was applied for estimating the whole-genome SNP-based heritability (h^2_{SNP}) using both common and low-frequency variants, which is implemented in the GCTA software (294). The top 20 ancestry PCs derived from the GRM across the same data set obtained using GCTA (option '--pca 20') were included as covariates in the model when estimating the genetic variance explained by all the SNPs on the observed scale. As the studied phenotypes were dichotomous traits, these results were transformed to the liability scale based on approximately observed disease prevalence from the literature (assuming disease prevalence of 130/million for GPA and 50/million for MPA) (338). Note that only data from cases and controls that were unrelated in the classical sense ('--grm-cutoff 0.05', that is relatedness < 0.05, equivalent to about second-cousins) were included in the analysis. Therefore, the total control sample size utilised varied slightly from the association testing. Given that the UK Axiom Biobank array has a more comprehensive genome coverage than the SNP6 platform (especially for SNPs with MAF range from 0.01 to 0.10) and thus better at tagging the causal variants, the SNP heritability analysis was carried out using the GWAS2 genotype data. Note, GCTA implicitly assumes all classes of SNPs (rare, low-frequency or common) have the same effect on average. However, rare and newer variants tend to have a larger effect as they represent mutation load being purged by negative selection. Rare variants (MAF < 0.01) were not included in the heritability estimation and hence this might bias GCTA estimates towards underestimating heritability.

2.1.13 Genetic correlation between PR3-AAV and MPO-AAV

The proportion of genetic variation as tagged by common genome-wide SNPs that is shared between PR3-AAV and MPO-AAV was estimated using the bivariate linear mixed-effects model implemented in GCTA (339). This method uses genome-wide SNPs to estimate genetic similarities between pairs of individuals, and uses bivariate REML to estimate covariance components (r_G) of the linear mixed model. The genetic correlation analysis was performed using the GWAS2 genotype data for the reasons as discussed in section 2.1.12. Given that the control samples were completely overlapped between PR3-AAV and MPO-AAV, the control samples were randomly allocated to each group in 1:1 ratio and the first 20 PCs were fitted

as covariates. The random allocation process was repeated for n = 100 and the median r_G was obtained.

2.1.14 Haplotype block estimation

Millions of genetic variants were available following genotype and imputation and each variant was tested to see whether an individual's genotype is predictive of the phenotype. A single genetic variant with a causal effect on the phenotype could therefore lead to multiple statistical but non-causal associations at the nearby variants due to LD in the genome. It is important to count the number of independent signals in the genome while accounting for LD. There are few ways to define approximately independent LD blocks including non-overlapping segments of 1 megabase (340) and non-overlapping segments of 5000 SNPs (341). The breakpoints of these segments sometimes fall in regions of strong LD and therefore lead to over-counting of the number of associated variants. A better approximation could be obtained by considering an empirical LD pattern in a reference panel. For the fine-mapping analysis, I employed the approximately independent LD block generated by Berisa and Pickrell using r^2 metric, of which two sets of SNPs were defined as "approximately independent" if the pairwise r^2 between SNPs in different sets is close to zero. The data is available at http://bitbucket.org.nygcresearch/ldetect-data (342).

2.1.15 Fine-mapping

A Bayesian approach was employed to generate the credible sets of causal variants. Credible set construction relies on two underlying assumptions: (i) the causal variant has been tested for association and (ii) there is a single causal variant at a locus. Variants are first ranked according to their Bayes' factor in favour of association (from largest to smallest). Variants are then added to the 99% credible set, by moving down the ranking, until the cumulative posterior probability of association exceeds 0.99 (343). The approximate Bayes factor for each variant was computed using the Wakefield approximation (344), with a prior parameter W = 0.09, indicating the prior expectation of true effect sizes (relative risks) exceed 2 only 1% of the time. The code is available at https://github.com/chr1swallace/lyons-et-al.

2.1.16 Colocalisation

Hundreds and thousands of risk loci have been identified by GWAS affecting different diseases and traits. The integration of multiple association studies could improve our understanding of the molecular basis of the association signals. Visual comparison of overlaps of association signals are insufficient to make inferences about causality. A Bayesian test for colocalisation between pairs of genetic association studies has been developed by using summary statistics to assess whether two association signals are consistent with a shared causal variant (345). Colocalisation analysis was performed using the 'coloc' R package to evaluate whether intermediate cellular phenotypes (e.g. eosinophil, neutrophil and monocyte counts) and AAV share the same causal variant at *BCL2L11* locus. Summary statistics were obtained for all variants within the LD-defined haplotype region for peripheral blood eosinophil, neutrophil and monocyte counts in a population study by Astle et al (346). A total of 5,505 of these variants was available in the AAV dataset. The posterior probabilities of different causal variant configurations were calculated under the assumption of a single causal variant for each trait, including H0 (no causal variant), H1 (causal variant for trait 1 only), H2 (causal variant for trait 2 only), H3 (two independent causal variants) and H4 (one common causal variant). The results of colocalisation analysis include five posterior probabilities (PP0, PP1, PP2, PP3 and PP4) where a large posterior probability for hypothesis 4 (PP4) would indicate that a single variant affects both traits.

2.1.17 Mendelian Randomisation

Mendelian randomisation (MR) is an analytic approach that utilises genetic variants that are robustly associated with a modifiable exposure or risk factor to infer a more reliable causal relationship with a clinically relevant outcome. It is an alternative strategy, especially when randomised controlled trials to examine the causality are not feasible and observational studies provide biased associations because of confounding or reverse causality. MR analysis is an alternative of randomised trial, exploiting the random allocation of alleles at birth to allow causal inference. The choice of a genetic instrumental variable is essential to a successful MR study. In order to achieve an unbiased estimation of the causal effect of the exposure on the outcome, a valid genetic instrumental variable must fulfil three core

assumptions: (i) it must be reproducibly and strongly associated with the exposure, (ii) it must not be associated with confounders between exposure and outcome and (iii) it is only associated with the outcome through the exposure (347).

"Two-sample" MR was performed to assess whether there is a causal effect of eosinophil, neutrophil or monocyte counts (the exposures) on PR3-AAV and MPO-AAV (the outcome). MR analysis was performed using the 'MendelianRandomization' R package. Summary statistics including beta coefficient and standard error of the variants were obtained from a population-based study by Astle et al (346), consisting of 209, 152 and 248 conditionally independent variants associated with peripheral blood eosinophil, neutrophil and monocyte counts respectively. After excluding the HLA region, 147 (eosinophil count), 101 (neutrophil count) and 169 (monocyte count) variants were available in the AAV dataset for MR analysis (see Figure 2.4). The primary analysis was conducted using the inverse-variance weighted (IVW) method and additional analyses were performed using alternative methods including simple median, median-weighted and MR-Egger to test the robustness of our results.
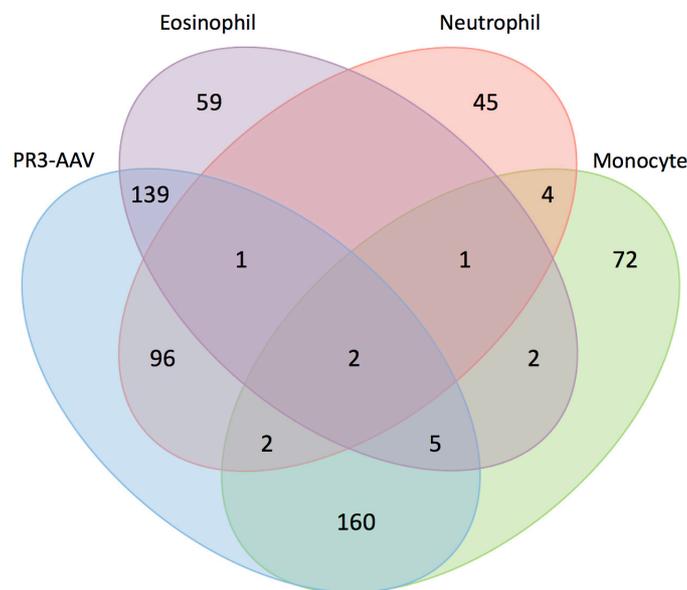
Figure 2.4: Venn diagram depicting the number of SNPs used in MR analysis.

There is a total of 147, 101 and 169 conditionally independent variants associated with eosinophil, neutrophil and monocyte counts respectively that are in common in the AAV dataset. There is minimal overlap between significant SNPs in eosinophil, neutrophil and monocyte counts.

2.1.18 Promoter and DNA regulatory elements interaction

It is well established that DNA regulatory elements such as enhancers or insulators can regulate gene transcription by physically interacting with their target gene promoters via looping of DNA, which are both cell-type and stimulus specific. Long-range interactions between genetic variants associated with AAV and gene promoter and regulatory regions were identified using the promoter capture Hi-C datasets from a wide range of primary cell types such as naïve CD4, activated CD4, naïve CD8, naïve B, monocytes, macrophages, neutrophils, megakaryocytes and endothelial precursors (321) as well as cell lines including CD34 and GM12878 (348) on the CHiCP browser (349).

2.1.19 Functional annotation and data mining

Other diseases and traits associated with AAV-associated loci were identified using PhenoScanner v2 (350). PhenoScanner is a curated database containing over 65 billion association results from multiple sources including the NHGRI-EBI (National Human Genome Research Institute-European Bioinformatics Institute) GWAS Catalog and NHLBI GRASP (Genome-Wide Repository of Associations Between SNPs and Phenotype) Catalog. It accounts for LD between the queried SNPs and those in the catalog of trait-associated variants. For each locus associated with AAV, trait-associated SNPs in LD ($r^2 \geq 0.6$) with the lead AAV SNP and a p-value of $\leq 1 \times 10^{-5}$ or in proximity to the candidate genes were identified. eQTLs at AAV-associated loci were identified from the eQTLGen Consortium (318). It incorporates 37 datasets including the GTEx and BIOS Consortium, with a total of 31,684 individuals. Functional annotation was performed using FUMA (351) – an integrative web-based platform that facilitates functional annotation of GWAS results, gene prioritisation and interactive visualisation using information from 18 biological resources comprising of functional annotations of SNPs (CADD v1.3, RegulomeDB, 15-core chromatin state and GWAS catalog), eQTLs (GTEx v6, Blood eQTL Browser and BIOS QTL browser), HiC, regulatory elements (Roadmap epigenomics project), gene score and gene sets (MsigDB v5.2 and WikiPathways). Regional visualisation of genome-wide association results for each AAV-associated locus was plotted using either LocusZoom or Gviz package.

2.2 Functional validation of significant association signals

2.2.1 Contributions

Patient recruitment to gene expression studies was coordinated by Professor Ken Smith, Dr Paul Lyons and Dr David Jayne as part of a larger immunophenotyping study and physically undertaken by several clinicians and study research nurses, Sister Jane Hollis, Sister Valeria Morrison and Sister Cecilia Matara. The recruitment of the ESKD cohort was undertaken by Dr Elaine Jolly. The initial sample processing was performed at the Smith Laboratory by several clinicians and lab assistants depending on availability and resources. I was primarily involved in DNA/RNA extraction and processing RNA samples for Affymetrix whole transcriptome microarray analysis as well as RNA sequencing for the ESKD cohort. Recruitment for the allele-specific expression (ASE) study and eosinophil apoptosis study together with all the transcriptomic analysis were performed by myself.

2.2.2 Patient recruitment to gene expression studies

2.2.2.1 Cambridge – AAV patients

Patients with AAV were recruited through the Addenbrooke's Vasculitis Service, a quaternary referral centre based in Cambridge, UK between 2004 and 2015 to an ongoing, prospective observational study of gene expression patterns in autoimmune disease. Inpatients and outpatients fulfilling the criteria according to the EMEA or CHCC nomenclature were referred by clinicians of the service and enrolled prior to receiving induction immunosuppression. Initial ethical approval for recruitment and its extension were obtained by Professor Ken Smith from the National Research Ethics Service Committee East of England - Cambridge Central (ref: 08/H0308/176 and 08/H306/21). Patients provided 60-100 ml of peripheral venous blood through a 21-gauge needle at 3 time points: (i) flaring disease: at recruitment, prior to induction therapy; (ii) early remission: 3 months after recruitment usually towards the end of induction therapy; (iii) late remission: 12 months after recruitment, when maintenance therapy was usually well established.

2.2.2.2 Cambridge – ESKD patients

A cohort of 93 ESKD patients who received either living or deceased donor kidney transplants at Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK were recruited between March 2009 and May 2011. A 110 ml venous blood sample was taken prior to renal transplantation and at 3 months following transplantation. Ethical approval for this study was granted by the National Research Ethics Service Committee East of England - Cambridge Central (ref: 08/H0304/121).

2.2.3 Healthy volunteer recruitment to the allele-specific expression study

Eight healthy volunteers who are heterozygous for the variant (rs13405741) were identified and recalled from the database of the ongoing, prospective observational study of gene expression patterns in autoimmune diseases. A total of 100 ml of peripheral venous blood through a 19-gauge needle was taken into a falcon tube containing 3.8% sodium citrate (Martindale Pharma) for neutrophil ASE study. Ethical approval was granted by the National Research Ethics Service Committee East of England - Cambridge Central (ref: 08/H0308/176).

2.2.4 Healthy volunteer recruitment to the eosinophil apoptosis study

Healthy volunteers with the desired genotype for rs13405741 (i.e. homozygous major and minor) were recalled through the Cambridge BioResource. Blood sampling was carried out at the NIHR/Wellcome Trust Clinical Research Facility, Level 5, Addenbrooke's Centre for Clinical Investigation. A total of 100 ml whole blood was collected from each healthy volunteer into a falcon tube containing 3.8% sodium citrate using a 19-gauge needle. Blood samples were processed immediately after collection and handled very gently at all times to avoid immune cell activation. Ethical approval was granted by the National Research Ethics Service Committee East of England - Cambridge Central (ref: 08/H0308/176). A total of 16 individuals were recruited and matched pair experiments were carried out to reduce possible confounding factors. One participant from each of the two genotype groups was recalled simultaneously and were matched for age and gender.

2.2.5　Sample processing

2.2.5.1　Leukocyte separation using immunomagnetic beads

Blood samples were processed immediately to prevent any transcriptional alterations, especially in cells from the myeloid lineage. Individual leukocyte subsets were purified as described (352). The purity of the populations obtained was assessed using the BD FACSCalibur™ flow cytometer. The positively selected cell subsets were resuspended in RLT buffer with 1% β-mercaptoethanol and stored at -80°C for future RNA extraction.

2.2.5.2　Neutrophil isolation

Human peripheral blood neutrophils were purified from peripheral blood using Percoll gradient as previously described (353). Briefly, the collected blood was centrifuged at 300 g for 20 min, acceleration 9, brake 3 at 24°C. Post-centrifugation, two layers were formed: a pellet at the bottom consisting of white and red cells and platelet rich plasma (PRP) as a supernatant. The PRP was removed using a Pasteur pipette into a fresh 50 ml falcon tube. To make autologous serum, 10 ml of PRP was added into a sterile glass vial and followed by 220 µl of 10 mM $CaCl_2$. The glass vial was incubated for 60 min in a CO_2 incubator with the lid of vial loose prior to storing at 4°C. The remaining PRP was centrifuged at 1400 g for 20 min, acceleration 9, brake 9 at 24°C. The supernatant was removed into a fresh 50 ml falcon tube and labelled as platelet poor plasma (PPP).

For the remaining pellet, 6% of dextran solution was added (2.5 ml dextran solution per 10 ml of cell pellet) and followed by warm 0.9% sodium chloride to make up a final volume of 50 ml. The tube was gently rolled several times to resuspend cells and left with cap loosely on at room temperature for 25-40 min until a separation of a pellet at the bottom (red cell) and a supernatant (myeloid cells enriched) was observed. The myeloid cells enriched layer was removed into a fresh 50 ml falcon tube and centrifuged at 260 g for 5 min, acceleration 9, brake 9 at 24°C while the bottom red cell pellet was discarded. Post-centrifugation, the supernatant was discarded and the cell pellet was resuspended by tapping the tube gently. PPP was added to the cell to make up a final volume of 2 ml and transferred into a 15 ml falcon tube. 42% and 51% Percoll gradients were made by diluting 90% Percoll in PPP for a final volume of 2 ml. Subsequently, 2 ml of 42% Percoll was laid under the cells and followed

by 51% Percoll. The resulting gradient was centrifuged at 150 g for 14 min, acceleration 1, brake 0, at 24°C. The resulting bands: top (peripheral blood mononuclear cells (PBMCs) at PPP and 42% interface) and bottom (polymorphonuclear cells (PMNs) at 42% and 51% interface). PMNs were collected into a fresh 50 ml falcon tube containing PPP and centrifuged at 256 g for 5 min, acceleration 9, brake 9 at 24°C. The supernatant was discarded. The cells were consecutively washed twice in 50 ml of phosphate buffered saline (PBS), without Ca^{2+}, Mg^{2+} and followed by PBS with Ca^{2+} and Mg^{2+}, centrifuged at 256 g for 5 min, acceleration 9, brake 9 at 24°C.

Freshly isolated neutrophils were suspended in Iscove modified Dulbecco medium (IMDM) at 5 x 10^6 cells/ml supplemented with 10% autologous serum, 100 U/mL penicillin and 100 µg/mL streptomycin and cultured in a humidified 5% CO_2 atmosphere at 37°C in the presence and absence of predetermined optimal concentrations of granulocyte-macrophage colony-stimulating factor (GM-CSF, 10 ng/ml) for 6 hr and 20 hr. Neutrophils were harvested at different time points and resuspended in RLT buffer with 1% β-mercaptoethanol and stored at -80°C for future RNA extraction. Cell purities were assessed by using air-dried cytocentrifuge preparations fixed in methanol and stained with May-Grünwald-Giemsa (Merck Ltd, Lutterworth, Leicestershire, UK) where at least 300 cells per slide were counted (Figure 2.5A).

2.2.5.3 Eosinophil isolation

Human peripheral blood eosinophils were purified using the RoboSep™ cell separation kit (EasySep™ Human Eosinophil Enrichment Kit and Fully Automated Cell Separator, STEMCELL™ Technologies; catalogue number 17956) which is designed to isolate eosinophils from a polymorphonuclear cell-rich fraction of peripheral blood by negative selection. Briefly, 8 ml of HES 6% solution (hydroxyethyl starch, Grifols) was added into every 40 ml of blood collected. The tubes were gently mixed and left standing at room temperature for at least 30 min to allow sedimentation of the red blood cells. Next, all the top layer was removed and made up to 50 ml with RoboSep™ buffer containing Dulbecco's PBS, 2% fetal bovine serum and 1 mM EDTA. This was centrifuged at 250 g for 10 min, acceleration 9, brake 9 at 24°C. Post-centrifugation, the supernatant was discarded and the cell pellet at the bottom

consisting of PMNs and some red cells. The cell pellet was resuspended in 50 ml of RoboSep™ buffer.

For pre-RoboSep cell counting, 20 µl of 3% acetic acid in PBS was mixed with 20 µl of the cell mixture. Cell counting was performed using a haemocytometer (red blood cells were excluded from the counting). The cell mixture was centrifuged again for 10 min, acceleration 9, brake 9 at 24°C. The cell pellet was then resuspended in RoboSep™ buffer at 50×10^6 cells/ml. Finally, the sample was placed in RoboSep™ Automated Cell Separator and the separation took approximately 30-40 min. Post-RoboSep, cell counting was performed. The cell mixture was centrifuged at 250 g for 10 min, acceleration 9, brake 9 at 24°C. Post-centrifugation, the supernatant was discarded and the cell pellet was resuspended in RPMI supplemented with 100 U/ml streptomycin and 100 U/ml penicillin at 1×10^6 cells/ml and cultured for 24 hr ("serum withdrawal induced apoptosis") (354). Cell purities were assessed by cytospin preparations using May-Grünwald-Giemsa staining (Figure 2.5B). Isolated eosinophils were resuspended in RLT buffer with 1% β-mercaptoethanol and stored at -80°C for future RNA extraction.
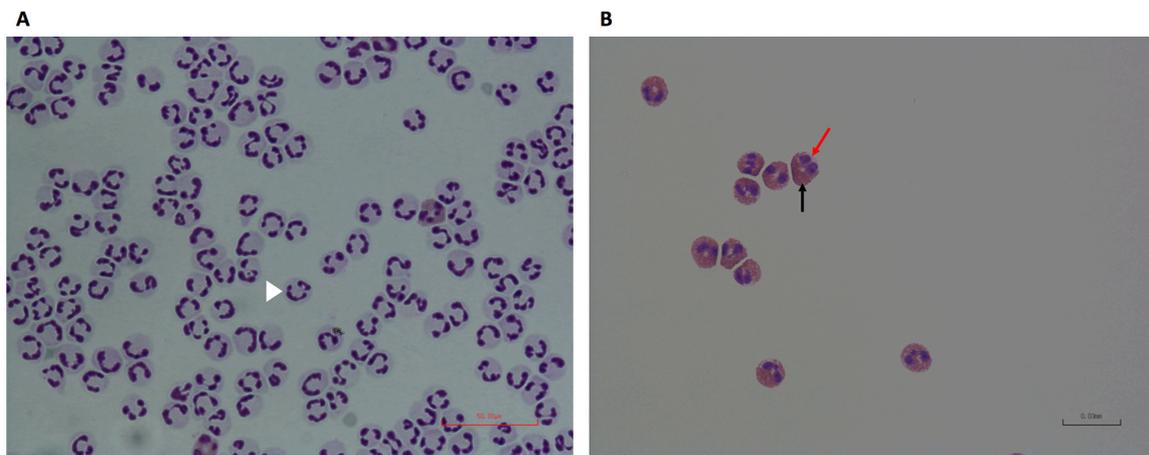
Figure 2.5: Light micrographs (40X) of freshly isolated neutrophils and eosinophils.

(A) Human neutrophils were purified using Percoll gradient while (B) human eosinophils were purified from mixed leukocytes using a Robosep™ negative selection strategy. White arrow head indicates multi-lobulated nucleus in neutrophil, red arrow indicates bi-lobed nucleus in eosinophil and black arrow indicates eosinophilic granules. Both of these isolation methods have been previously demonstrated to induce minimal cell priming and activation (355, 356).

2.2.5.4   Cell culture

THP-1 and MCF cells were cultured in RPMI-1640 supplemented with 10% fetal bovine serum, 1 mM L-glutamine and 1 mM sodium pyruvate. THP-1 is a human leukaemia monocytic cell line, which expresses both PRTN3 isoforms while MCF-7 is a breast cancer cell line that only expresses the alternate isoform (*PRTN3-202*).

2.2.6   Eosinophil apoptosis assay

Eosinophils were cultured in a final volume of 150 µl in flat-bottomed 96-well, ultralow attachment plates (Costar, Hycombe, Bucks, UK) for 24 hours. Following resuspension, the cells were aspirated, cytocentrifuged, fixed and stained as described in section 2.2.5.3. Morphologic analysis of eosinophil apoptosis was assessed under oil immersion light microscopy with the observer masked to the experimental conditions. Apoptotic eosinophils were defined as cells with darkly stained, condensed nuclei (Figure 2.6). A more objective assessment of apoptosis was performed using Annexin V/propidium iodide (PI) flow cytometry (Dead Cell Apoptosis Kit with Annexin V FITC and PI, Invitrogen™). Eosinophils were aspirated and centrifuged at 256 g for 5 min at 4°C. The cell pellet was resuspended in 100 µl master mix containing 98 µl of 1X Annexin-binding buffer, 1 µl of FITC Annexin V and 1 µl of the 100 µg/ml PI. Samples were incubated for 20 minutes at 4°C in the dark and the volume was increased to 500 µl with flow cytometry staining buffer (FACS buffer) immediately before analysis using the BD FACSCanto™ II flow cytometry.
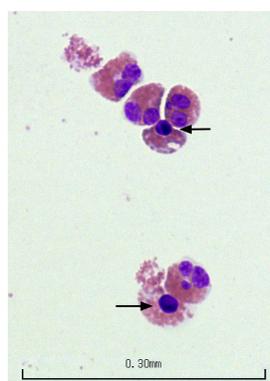
Figure 2.6: Apoptotic eosinophils.

Representative image of apoptotic eosinophils from a healthy volunteer cultured in medium with no serum for 24 hours. The black arrows indicate darkly stained, condensed nuclei.

2.2.7 gDNA and RNA extraction, quantification and quality assurance

Prior to mid-2010, RNA was extracted from thawed lysates using the RNeasy mini kit (Qiagen, cat 74104) according to the manufacturer's instructions. Since then, RNA was extracted using the AllPrep mini kit (Qiagen, cat 80204) which also yielded cell subset specific genomic DNA (gDNA). For the eosinophil apoptosis study, RNA was extracted using the RNEasy micro kit (Qiagen, cat 74004) given the lower cell numbers. The RNA yield and quality were assessed using both an Agilent® ND-1000 spectrophotometer and Agilent® 2100 Bioanalyser (Agilent RNA 6000 Nano kit or Agilent RNA 6000 Pico Kit), depending on the RNA input.

2.2.8 SNP Genotyping analysis

SNP genotyping was performed using TaqMan® SNP Genotyping Assays (Thermo Fisher Scientific) according to the manufacturer's instructions. Wet DNA delivery method was used. Briefly, a 2.75 µl reaction mix consisting of 2X TaqMan® Genotyping Master Mix (2.5 µl per sample) and 20X SNP Genotyping Assay Mix (0.25 µl per sample, Table 2.4) was aliquoted to a 384-well MicroAmp™ optical reaction plate. Next, 2.25 µl (5-10 ng) of purified gDNA was delivered to the final reaction mix including a No-Template-Control (NTC). The reaction plate was sealed with MicroAmp™ optical adhesive film and centrifuged briefly to spin down the contents and to eliminate any air bubbles. The reaction plate was then transferred to LightCycler®480 Real-Time PCR Instrument for polymerase chain reaction (PCR). The thermal cycling condition was as follows: 10 min at 95°C (hold), 15 s at 92°C (denature), 1 min at 60°C (annealing and extension), repeated for a total of 35-40 cycles. After PCR amplification, an endpoint plate read was performed to plot fluorescence (Rn) values based on the signals from each well. Automatic allele calls were generated, reviewed and converted to genotypes.

Table 2.4: A list of the TaqMan® SNP Genotyping Assays

SNP ID (gene)	Assay ID	Context sequence [VIC™ dye/FAM™ dye]
rs62132279 (*PRTN3*)	C_89462672_10	TTGAGATGTCCTTCACACACACCAA[A/G]CAATTCACCCACTTAACAATTCGGT
rs62132296 (*PRTN3*)	C_25654080_10	TGAGCCACGTGCCCATCCATCCAGC[C/T]TCCAGGCCCCGGTGGATTGTGGGGA
rs13405741 (*BCL2L11*)	C_30929894_10	TTGGGTGGGAGTCTTGACCATGGGC[C/T]GGGGCTGCAGTCAGAAGCCCTTAGA

The SNP alleles are included in the square brackets and highlighted in blue. The reporter dye information is represented in the assay context sequence which is provided in the (+) genome strand orientation relative to the NCBI reference genome.

2.2.9 cDNA synthesis and gene expression analysis

Complementary DNA (cDNA) was generated using the SuperScript™ VILO™ cDNA Synthesis Kit (Cat no: 11754050) according to the manufacturer's instructions. Briefly, 20 µl reaction mix consisting of 5X VILO reaction mix (4 µl per sample), 10X SuperScript enzyme (2 µl per sample) and RNA input amounts ranging from 5 ng to 150 ng (16 µl per sample) was aliquoted to a 0.2 ml 8-Strip PCR tube and gently mixed. The PCR tube was transferred to Eppendorf Mastercycler® Nexus for reverse transcription (RT) with the following thermal cycling condition: 10 min at 25°C, 60 min at 42°C and 5 min at 85°C. cDNA was diluted prior to quantitative PCR (qPCR).

qPCR was carried out using TaqMan® Gene Expression Assays (Thermo Fisher Scientific, see Table 2.5) according to the manufacturer's instructions. Briefly, a 20 µl reaction mix consisting of 20X TaqMan Gene Expression Assay (1 µl per sample), 2X Master mix (10 µl per sample), RNase-free water (5 µl per sample) and diluted cDNA template (4 µl per sample) was aliquoted to a 384-well MicroAmp™ optical reaction plate. The reaction plate was sealed with MicroAmp™ optical adhesive film and centrifuged briefly to spin down the contents and to eliminate any air bubbles. The reaction plate was then transferred to LightCycler®480 Real-Time PCR Instrument for PCR. The thermal cycling condition was as follows: 2 min at 50°C (hold), 10 min at 95°C (hold), 15 s at 95°C and 1 min at 60°C (PCR) repeated for a total of 40 cycles. The amplification plots for the entire plate was viewed and comparative cycle threshold (C_T) values were generated for relative quantification of gene expression analysis, which was performed using the ΔC_T method using a reference gene (see formula below). This is a variation of the Livak method that is simpler to perform (357) and gives essentially the same results as those obtained with the $2^{-\Delta\Delta CT}$ method. This method uses the difference between reference and target C_T values for each sample.

$$\text{Ratio (reference/target)} = 2^{CT(reference) - CT(target)}$$

In order to detect the expression of *PRTN3-001* mRNA versus *PRTN3-002* mRNA, I designed two primer sets (Table 2.6) for PCR, which was carried out using the Applied Biosystems AmpliTaq Gold DNA polymerases protocol with the following thermocycler conditions: 95°C

3 min (holding stage), 35 cycles of PCR – denaturation 95°C 15 s, annealing 58°C 30 s and extension 72°C 1 min and final holding stage 72°C 7 min (final extension). After agarose gel electrophoresis of the resulting PCR products, the desired amplicons were gel-purified and sequenced using the Sanger method to confirm their sequences.

Table 2.5: A list of the TaqMan® Gene Expression Assays

Gene	Assay ID	Details (amplicon length)
PRTN3	Hs00160521_m1	Assay probe spans exon 1-2, detect only *PRTN3-201* isoform (100bp)
PRTN3	Hs01553330_m1	Assay probe spans exon 3-4, detect both *PRTN3-201* and *PRTN3-202* isoforms (72bp)
BCL2L11	Hs01076940_m1	Assay probe spans exon 1-2 (68bp)
MIR4435-2HG	Hs03680374_m1	Assay probe spans exon 1-2 (75bp)
B2M	Hs00984230_m1	Assay probe spans exons 3-4 (81bp)
18S	Hs99999901_s1	Both primers and probe map within exon 1 (187bp)
ACTB	Hs99999903_m1	The probe and one of the primers sit within exon 1 (171bp)

18S, Eukaryotic 18S rRNA; *ACTB*, Actin, Beta, cytoplasmic; *B2M*, Beta-2-microglobulin; *BCL2L11*, BCL2 Like 11 (a.k.a BIM); bp, base pairs; *MIR4435-2HG*, MIR4435-2 Host Gene (a.k.a MORRBID); *PRTN3*, proteinase 3.

Table 2.6: A description of the primer sets used to detect different *PRTN3* isoforms

Primer Name	Sequence (amplicon length)
PRTN3-201	FOR 5'-TGC TGC TGG CCT TGC TGC TGA G-3'
	REV 5'-CCG AGA AGT GCT GCT GGG TGG G-3' (267 bp)
PRTN3-202	FOR 5'-GGG AGA CGG AGG CTC GGA GAG G-3'
	REV 5'-AGA AGT GCT GCT GGG TGG GCT C-3' (337 bp)
ACTB	FOR 5'-GAG CAT CCC CCA AAG TTC A-3'
	REV 5'-AGA GAA GTG GGG TGG CTT TT-3' (133 bp)

2.2.10 Western blot of eosinophil lysates

1×10^5 isolated eosinophils aliquots were lysed in 15 µl 1X LDS buffer (Invitrogen™) containing 10% β-mercaptoethanol. Samples were boiled at 70°C for 10 min before loading into pre-cast NuPAGE Novex 4-12% Bis-Tris protein gels (Invitrogen™). Spectra™ Multicolour Broad Range Protein ladder (Invitrogen™) was used as a size ladder for each gel. Proteins were electrophoresed in electrode buffer (1X MOPS - SDS Running Buffer, FORMEDIUM™) at 100 V. Proteins were electrotransferred onto PVDF transfer membrane (Millipore) at 350 mA for 1 hour. Non-specific binding sites on the membranes were blocked with 5% non-fat dry milk power (Marvel, UK) dissolved in TBST buffer (1X Tris buffered saline with 0.1% (vol/vol) Tween 20) at room temperature for 1 hour. Membranes were then incubated with diluted primary antibodies overnight at 4°C (anti-BCL2L11 from Cell Signaling Technology® at 1:1,000 dilution and anti-beta Actin from Abcam at 1:3,000 dilution). Following primary antibody incubation, membranes were washed 3 times in TBST for 10 min on a rocker prior to incubation with species-specific HRP-conjugated secondary antibody (Cell Signaling Technology®) at 1:10,000 dilution for 45 min at room temperature. Membranes were then washed again for 3 times in TBST and once in PBS before being developed with SuperSignal™ West Pico PLUS Chemiluminescent Substrate or SuperSignal™ West Femto Maximum Sensitivity Substrate as per manufacturer's protocol and exposed to photographic film. Semi-quantitative analysis was carried out using the Image Lab software.

2.2.11 cDNA synthesis and hybridisation to Affymetrix arrays

cDNA was synthesized and amplified from total RNA using the Ambion® whole transcript expression kit (Applied Biosystems), fragmented and labelled using the Affymetrix® GeneChip® whole transcript terminal labelling kit before hybridisation to Affymetrix® Gene ST 1.1 arrays according to manufacturer's instructions. Briefly, 200 ng of total RNA was used to synthesise the first-strand cDNA in the initial RT reaction. The whole transcript expression kit uses a priming method that specifically primes non-ribosomal RNA, including both poly(A) and non-poly(A) mRNA. Second-strand cDNA was then synthesised from the first-strand cDNA using DNA polymerase and RNase H to simultaneously degrade the RNA, followed by in vitro transcription to generate synthesise anti-sense cDNA. A purification step was then performed

to improve the stability of cDNA prior to second cycle RT to generate sense-strand cDNA. During this reaction, deoxyuridine triphosphate (dUTP) was incorporated into the DNA, which was then recognized by uracil DNA glycosylase (UDG) and apurinic/apyrimidinic endonuclease 1 (APE1) that cleaved the DNA strand at these sites, a process known as "fragmentation". Fragmented DNA was labelled by terminal deoxynucleotidyl transferase (TdT) covalently linked to biotin. The terminally labelled fragmented cDNA was then hybridised onto Affymetrix Human Gene ST 1.1 arrays and finally scanned using an Affymetrix GeneTitan™ multichannel instrument.

2.2.12 RNA-Seq library preparation

To investigate the differential expression of different isoforms of *PRTN3*, RNA-Seq libraries were prepared with the Illumina® TruSeq® Stranded Total RNA Sample Preparation kit using 150 ng RNA input and the low sample protocol. Briefly, ribosomal RNA (rRNA) was removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. Following purification, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA polymerase I and RNase H. These cDNA fragments had the addition of a single "A" base for subsequent adapter ligation. The products were purified and enriched with PCR to create the final cDNA library. Next, samples were assessed for quality and fragment size using the Bioanalyzer DNA high sensitivity kit (Agilent) prior to library quantification using the KAPA Illumina Sequencing Library qPCR Quantification kit. All 19 libraries were pooled to a final concentration of 4nM. Sequencing of the prepared libraries was conducted at the Cambridge Genomic Services, Department of Pathology, University of Cambridge using the Illumina NextSeq 550 System paired-end 75 bp run. Raw data files were generated in the FASTQ format.

2.2.13 Allele-Specific Expression assay

Gene variants that contribute to common disease susceptibility often alter gene expression only modestly. Small expression changes can be technically challenging to measure robustly, especially since biological variation usually contributes negatively to this goal. A clone-based ASE assay which enables quantification of genotype-determined expression differences, such as those where the susceptibility allele is transcribed 2-fold less than the resistance allele as previously shown by Rainbow et al (358), was performed to analyse *BCL2L11* transcription in neutrophils from heterozygous individuals (rs13405741; C/T). This facilitated a direct comparison between the amount of *BCL2L11* pre-mRNA that was transcribed from each allele or haplotype and ensured that external factors affected both alleles equally. The workflow of a cloned-based ASE assay is summarised in Figure 2.7.
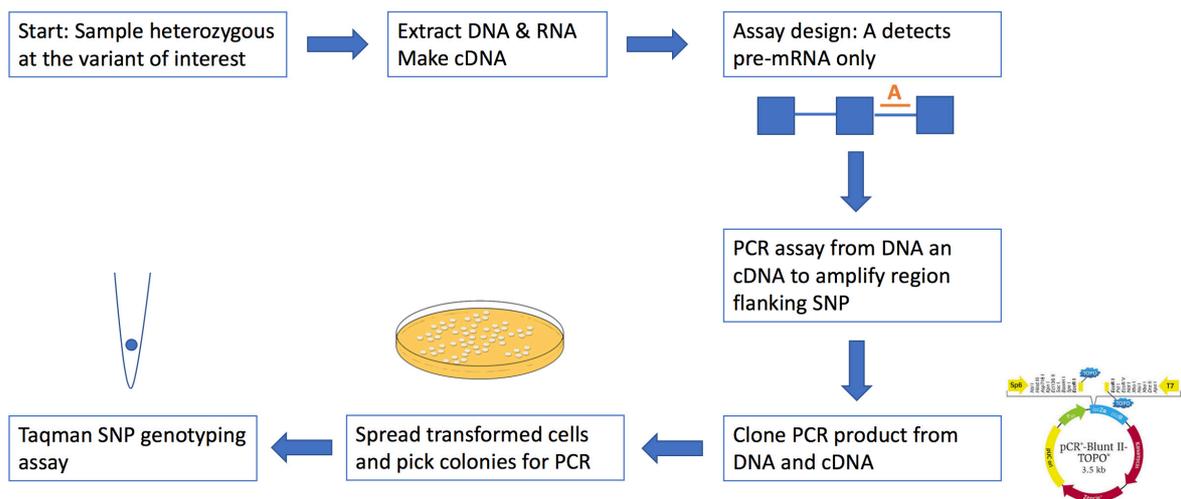
Figure 2.7: Workflow of a clone-based allele-specific expression assay.

Human neutrophils were purified from the peripheral blood of heterozygous healthy volunteers as described in sections 2.2.5.2. Freshly isolated neutrophils were routinely suspended at 5×10^6/mL in IMDM supplemented with 10% autologous serum, 100 U/mL penicillin and 100 µg/mL streptomycin. They were cultured in a humidified 5% CO_2 atmosphere at 37°C. To examine the effect of GM-CSF on the *BCL2L11* transcription, cells were incubated in the presence or absence of predetermined optimal concentrations of GM-CSF (10 ng/mL). Neutrophils were harvested at 0, 6, and 20 hours respectively and lysed with

RLT plus buffer. gDNA and RNA were extracted from the lysates using Qiagen$^®$ AllPrep DNA/RNA Mini kit. RNA was treated with TURBO™ DNase (Thermo Fisher Scientific) to ensure complete removal of gDNA from the samples and reversed transcribed to cDNA using SuperScript$^®$ VILO™ cDNA Synthesis Kit. A nested PCR using two sets of primers was performed (see Table 2.7) with Phusion$^®$ High-Fidelity DNA Polymerase to amplify the region of interest at rs13405741 from both gDNA and cDNA.

After agarose gel electrophoresis of the resulting PCR products, the desired amplicons (301 bp) were gel-purified and sequenced using the Sanger method to confirm the sequence variant. The amplicons were then cloned into pCR™-Blunt II-TOPO$^®$ vectors containing a kanamycin-resistance cassette (Zero Blunt$^®$ TOPO$^®$ PCR Cloning Kit, Invitrogen™) and transfected into chemically competent *E. Coli*. Two different volumes (100 μl and 200 μl) of each transformation were spread on a pre-warmed kanamycin-containing Luria-Bertani agar plates and incubated overnight at 37°C. Next, 96 colonies per condition per individual were picked and added into a 96-well plate containing 20 μl nuclease free water, further diluted with 80 μl nuclease free water and heated at 85°C for 10 minutes for thermal lysis. The reaction plate was centrifuged briefly to spin down the contents and the supernatant was used as DNA template for TaqMan SNP genotyping assay to determine the frequency of each allele.

Table 2.7: Primers and thermocycler conditions for ASE assay

Primer set	Detailed PCR protocol
First set FOR 5'-CCT GAG TGG TGC AGT GTT TG-3' REV 5'-TCA CTG TTT CTT GTG ATC AAA CAA-3'	Initial denaturation 98°C 30 s Denaturation 98°C 10 s, annealing 66°C 30 s and extension 72°C 20 s (x30 cycles) Final extension 72°C 5 min
Second set FOR 5'-TGT ACC GGT TCA TCT GCT GG-3' REV 5'-TGG AGT GAC TTG AGC TGG AC-3'	Initial denaturation 98°C 30 s Denaturation 98°C 10 s, annealing 66°C 30 s and extension 72°C 20 s (x30 cycles) Final extension 72°C 5 min

FOR, forward; min, minute; PCR, polymerase chain reaction; REV, reverse; s, second.

2.2.14 Microarray data analysis

The output .CEL files were processed and analysed using the R statistical software (v 3.2.5). Raw microarray data were background adjusted, quantile normalised and summarised using the robust multichip average method. QC was performed using the arrayQualityMetrics package. The parameters that were assessed as part of this QC process included array intensity distributions, between array distance comparisons, individual array quality and standard deviation versus rank of the mean. An object ("exprsFile") was generated containing QC and normalised microarray data of x number of patients. Following the initial pre-processing steps, an *ExpressionSet* was created using the Biobase package, consisting of expression data from microarray experiments (assayData, a matrix of "expression" values which has F rows and S columns, where F is the number of features on the chip and S is the number of samples), phenotypic data describing samples in the experiment (phenoData), annotations and meta-data about the features on the chip used for the experiment (featureData, Affymetrix human gene 1.1 ST array annotations which contain information such as gene name, symbol, Entrez ID or chromosomal location). The *ExpressionSet* class is designed to combine several different sources of information into a single convenient structure that allows easy manipulation of data for downstream analysis. There was a total of 19,783 probesets remaining after filtering out the control probesets and those that did not have assigned annotations.

The following code was used to create a normalised *ExpressionSet*:

```
# Loading the library
> library(hugene11sttranscriptcluster.db)
> library(Biobase)
> library(arrayQualityMetrics)
> library(oligo)
> library(annotate)


# Identify microarray CEL files and read in the CEL files in the directory
> cels <- list.files(pattern=".*.CEL")
> rawData <- read.celfiles(cels)
```

```
# Normalise the data and perform QC
> exprsFile <- rma(rawData)
> arrayQualityMetrics(exprsFile)


# Create an ExpressionSet
> phenoData <- new("AnnotatedDataFrame", data=pData)
> phenoData(exprsFile) = phenoData
> ID=featureNames(exprsFile)
> Symbol=getSYMBOL(ID,"hugene11sttranscriptcluster.db")
> fData(exprsFile)=data.frame(Symbol=Symbol)
> eSet=exprsFile[!is.na(fData(exprsFile)),]


# Compute principal components
> prData=prcomp(t(exprs(eSet)))
> pdf("PCA Batch Effect.pdf",width=7, height=7, paper="special")
> factor = as.factor(eSet$info_batch_name)
> factors = sort(unique(factor))
> cols = rainbow(length(factors), alpha = 0.8)
> label = eSet$Study_ID
> plot(prData$x[,1],prData$x[,2],pch=21,bg=cols[match(factor,factors)],xlab='PCA1',
  ylab='PCA2',main='Principal Components Plot')
> text(prData$x, labels = label, cex= 0.5,adj=(-.3))
> legend("topright",legend=factors,pch= 21,bty="n",cex=0.9,title="",pt.bg=cols,pt.cex = 1)
> dev.off()


# Batch correction
> pheno = pData(eSet)
> edata = exprs(eSet)
> batch = eSet$info_batch_name
> mod = model.matrix(~as.factor(pheno$rejection_6mth) + as.factor(pheno$rejection) +
as.factor(pheno$infection) + as.factor(pheno$DGF) + as.factor(pheno$gender), data=pheno)
```

71

```
> combat_edata=ComBat(dat=edata, batch=batch, mod=mod, par.prior=TRUE,
prior.plots=FALSE)
> exprs(eSet) <- combat_edata
```

2.2.15 RNA-Seq data analysis

To begin with raw reads that are stored in FASTQ format, the overall quality of the sequenced reads was checked using FastQC software (Babraham Institute, Cambridge). For adapter trimming, the first 13 bp of Illumina standard adapters ('AGATCGGAAGAGC') were removed using Trim Galore with the following specifications --paired --illumina --phred 24 --length 20 --stringency 13. Ribosomal reads were removed in silico using BBSplit prior to read alignment with HISAT2 using the Genome Reference Consortium Human Build 38 (GRCh38). QC of aligned reads were carried out using the Quality of RNA-Seq Toolset (QoRTs). In order to achieve approximately 30 million reads per sample, two sequencing runs were performed and the aligned reads were then merged into a single bam file for each sample. A summary of the mapped sequencing reads to genomic features was obtained using featureCounts (= raw read counts) and normalised for sequencing depth differences with edgeR. The read counts were then transformed to the log scale for downstream analyses. Quantitative visualisation of alternative isoform expression was performed using Sashimi plot.
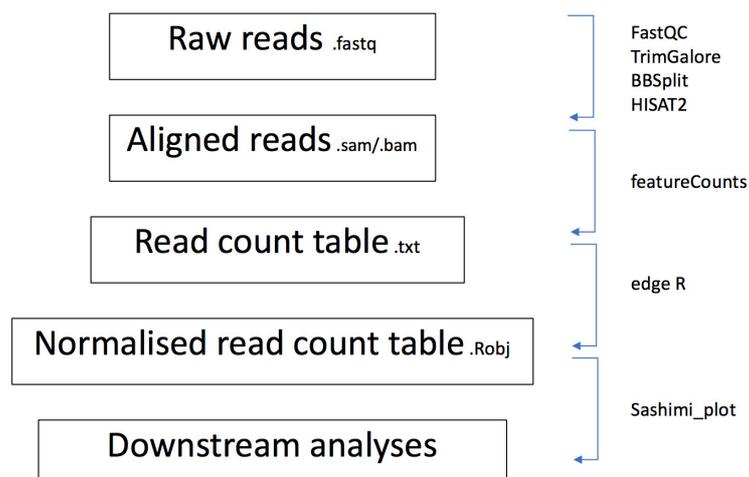
Figure 2.8: Workflow for RNA-Seq data analysis.

The most commonly used file formats to store the results are indicated in smaller font sizes with the most commonly used tools shown on the right.

72
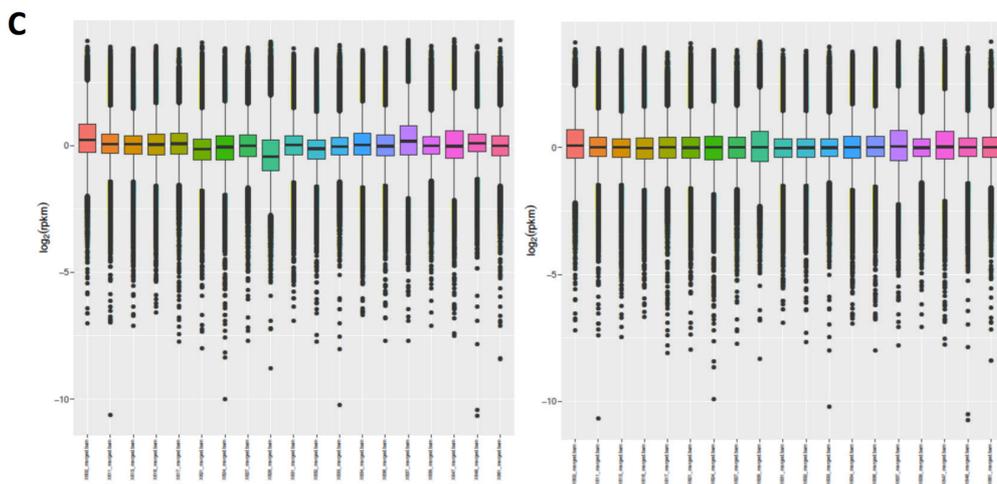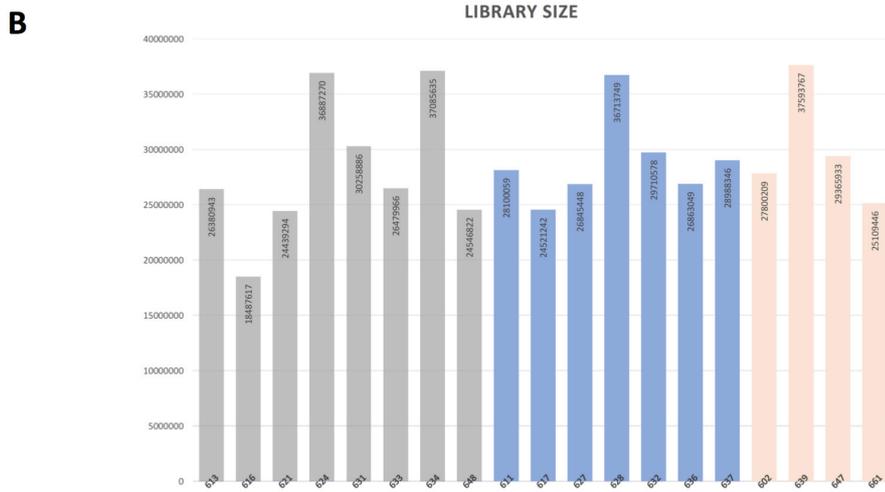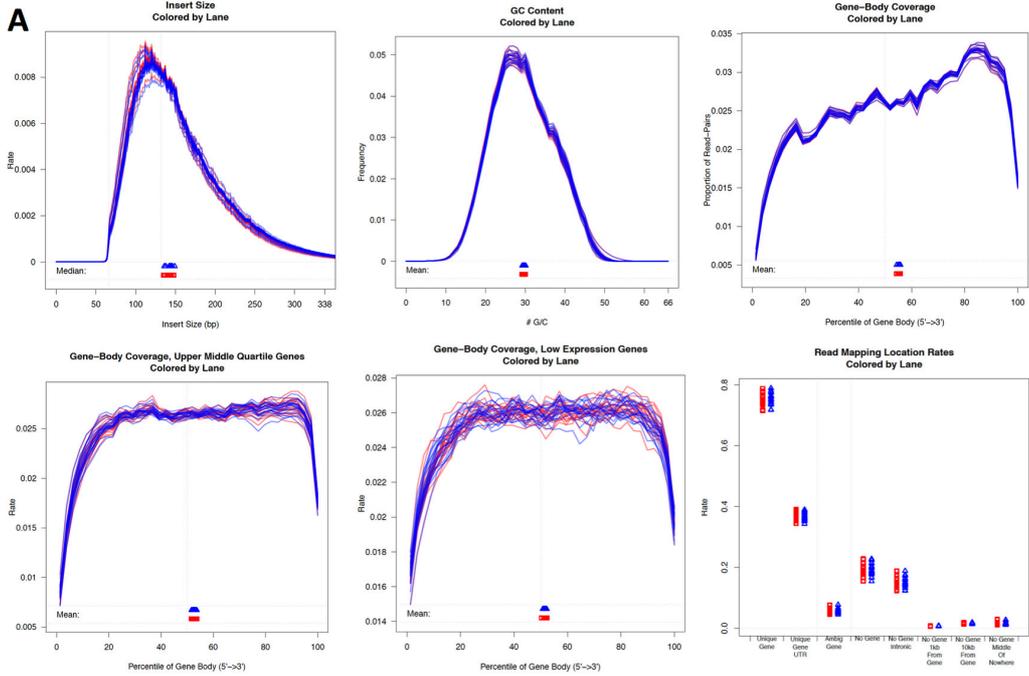
Figure 2.9: Quality control and data processing of RNA-Seq experiment.

(A) Quality control summary plots generated using Quality of RNA-Seq Toolset (QoRTs).

(B) Library size per sample. Y-axis represents read count.

(C) RNA-Seq data normalisation using the edgeR package. Boxplots of log-RPKM (Reads Per Kilobase of transcript, per Million mapped reads) values showing expression distributions for unnormalised data (left) and normalised data (right) for each sample in the CKD datasets.

2.2.16 General statistical analyses

Analyses were performed using GraphPad Prism v6 (GraphPad Software, San Diego, CA) and R statistical software. Unless otherwise specified, comparisons of a single parameter across two groups were performed using the unpaired, non-parametric, two-tailed Mann-Whitney test for continuous variables. Comparisons of a single parameter across more than two groups were performed using the Kruskal-Wallis rank-sum test. Where the comparison was significant, pairwise two-sample Wilcoxon tests were performed to determine which individual comparisons contributed to the significant result. The resulting p-values were adjusted for the number of comparisons made using the false discovery rate method unless otherwise stated. An alpha value < 0.05 was considered significant. Correlation testing was performed with either Pearson or Spearmann correlations, depending on the context.

Table 2.8: Key resources table

Resource	Identifier
Affymetrix Hugene 11 annotation data	https://bioconductor.org/packages/release/data/annotation/html/hugene11sttranscriptcluster.db.html
Annotate	https://www.bioconductor.org/packages/release/bioc/html/annotate.html
ArrayQualityMetrics	https://bioconductor.org/packages/release/bioc/html/arrayQualityMetrics.html
BCFtools	https://samtools.github.io/bcftools/bcftools.html/
Biobase	https://bioconductor.org/packages/release/bioc/html/Biobase.html
BLUEPRINT	http://blueprint-data.bsc.es/release_2016-08/#!/
BOLT-LMM	https://data.broadinstitute.org/alkesgroup/BOLT-LMM/
CHiCP	https://www.chicp.org/
Colocalisation	https://cran.r-project.org/web/packages/coloc/coloc.pdf
edgeR	https://www.bioconductor.org/packages/devel/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf
eQTLgen	http://www.eqtlgen.org/cis-eqtls.html
FastQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FUMA	http://fuma.ctglab.nl/
GCTA	http://cnsgenomics.com/software/gcta
ggplot2	https://cran.r-project.org/web/packages/ggplot2/index.html
GraphPad	https://www.graphpad.com/scientific-software/prism/
Gviz	https://bioconductor.org/packages/release/bioc/html/Gviz.html
Haplotype block	http://bitbucket.org.nygcresearch/ldetect-data
HISAT2	https://ccb.jhu.edu/software/hist2/
Integrative Genomics Viewer	https://software.broadinstitute.org/software/igv/
LocusZoom	http://locuszoom.org/
META v1.7	https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html
Michigan Server	https://imputationserver.sph.umich.edu/index.html#!
MISO	https://miso.readthedocs.io/en/fastmiso/sashimi.html
oligo	https://bioconductor.org/packages/release/bioc/html/oligo.html
PhenoScanner	http://www.phenoscanner.medschl.cam.ac.uk/
PLINK v1.9	https://www.cog-genomics.org/plink/1.9/
QoRTS	https://hartleys.github.io/QoRTs/
Quanto	http://biostats.usc.edu/software.html/
R v3.4.4	https://www.r-project.org/
Rsubread	https://bioconductor.org/packages/release/bioc/html/Rsubread.html
SNPTEST v2.5.2	https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html
TrimGalore	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Two sample MR	https://cran.r-project.org/web/packages/MendelianRandomization/MendelianRandomization.pdf
UCSC	https://genome.ucsc.edu/index.html

3 Discovery of PR3-AAV risk loci

3.1 Introduction

PR3-AAV is more frequent in northwestern Europe and North America than southern Europe and is uncommon in Asian populations. The incidence of PR3-AAV in the United Kingdom has been reported to be 10.6 per million per year while in Spain, the incidence is approximately 4.9 per million per year (32). PR3-AAV patients are more likely to have upper and lower respiratory tract as well as renal involvement as opposed to vasculitis limited to the kidneys in MPO-AAV. The nature of pulmonary involvement differs between ANCA serotypes where cavitating lung nodules are often observed in PR3-AAV while fibrosing lesions are more frequently seen in MPO-AAV. Furthermore, the clinical outcome differs according to ANCA serotype with PR3-AAV predisposing to disease relapse (hazard ratio 1.6-3.2) as compared to MPO-AAV (359, 360).

Over and above the strong association within the HLA region (*HLA-DP*), previous GWAS have identified two genome-wide significant loci at *SERPINA1* (chr14q32.13) and *PRTN3* (chr19p13.3) to be associated with PR3-AAV. In order to identify additional risk loci associated with PR3-AAV, a total of 1,132 PR3-AAV cases from Europe, along with 6,688 population matched controls, were genotyped on the UK Biobank Axiom array. A meta-analysis was subsequently performed with the recent AAV GWAS by Lyons et al (70).

In this chapter, I will discuss the identification of seven genome-wide significant loci associated with PR3-AAV including the HLA region, four of which are implicated in PR3-AAV risk for the first time. Within the HLA region, imputation of HLA classical alleles revealed four independent associations and the strongest association was observed at *HLA-DPB1*04:01*. Furthermore, I performed fine-mapping of the newly identified loci by calculating Bayesian posterior probability of causality for all variants to generate the 99% credible sets of most likely causal SNPs, followed by prioritisation of variants using functional annotation to reduce the size of the associated regions and the number of plausible target genes.

3.2   Samples and genotype data

Following sample and SNP QC, a total of 397,205 SNPs from GWAS1 and 558,875 SNPs from GWAS 2 respectively were available for analysis. The breakdown of the number of patients included in the meta-analysis of PR3-AAV is summarised in Table 3.1.

Table 3.1: The number of cases and controls included in the meta-analysis of PR3-AAV

Cohort	No of Controls	No of Cases
GWAS1	5,259	478
GWAS2	6,688	1,132
[†]Sequenom Genotype (rs6679677)	1,531	1,122
Total	13,478	2,732

[†]For rs6679677, a further cohort was available for meta-analysis where the genotyping was performed using the Sequenom MassARRAY platform in the replication cohort of Lyons et al 2012.

3.3   Association testing with genotype data

I performed association testing of the genotype data using two analytic methods - logistic regression and LMM as implemented in BOLT-LMM. A comparison between the two models is highlighted in Table 3.2. There was minimal genomic inflation observed in GWAS1 (lambda gc, λgc = 1.03) as all the cases and controls were from the UK. In contrast, the λgc for GWAS2 remained over 1.1 despite fitting an increasing number of PCs as covariates in a linear regression model. Not to our surprise, the LMM method that has been shown to better control for population stratification than correction for PCs, reduced the λgc from 1.308 to 1.087, supporting that population stratification in GWAS2 was well controlled (Figure 3.1).

Table 3.2: A comparison of logistic regression and linear mixed model analyses

Genotyped data	Genomic inflation factor (lambda, λ)	
	GWAS1	GWAS2
No covariates	1.033	1.308
3 PCs	1.034	1.156
6 PCs	1.036	1.148
20 PCs	1.039	1.135
BOLT-LMM	1.031	1.087
*BOLT-LMM (imputed data)	1.029	1.065

LMM, linear mixed model; PCs, principal components.

Lambda, λ is the ratio of the median of the observed chi-squared statistics to the median of the expected chi-squared statistics under the null hypothesis.

* indicates genomic inflation factor calculated using the imputed data.
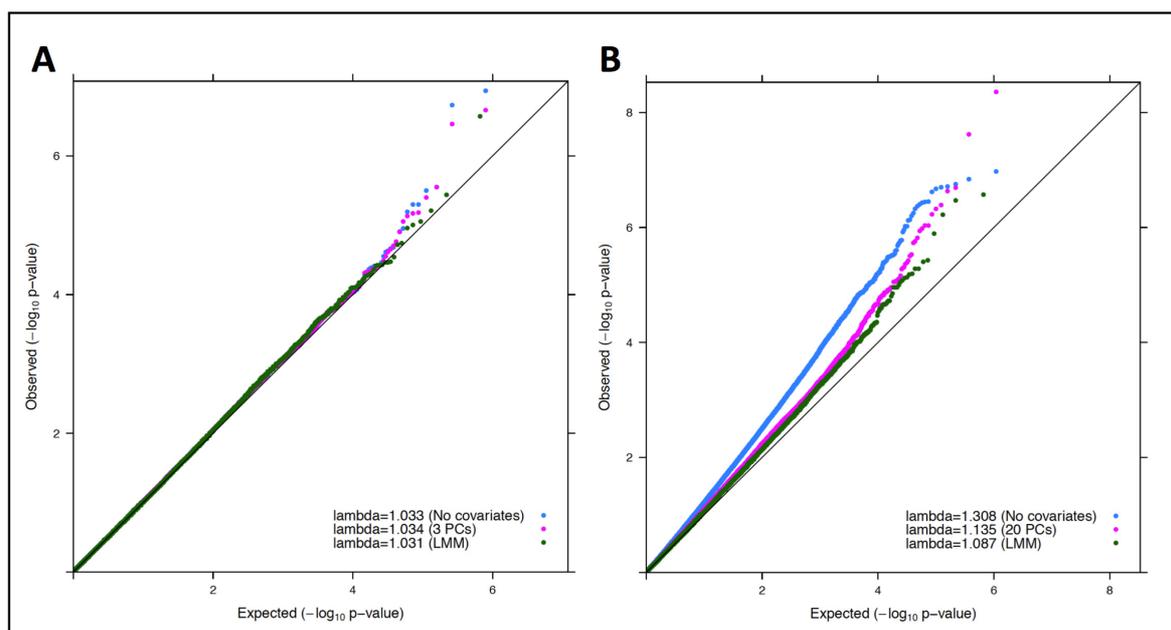
Figure 3.1: Quantile-quantile plots and genomic inflation factors.

Association tests were performed on genotyped data and comparison between logistic regression with no covariates, varying numbers of principal components as covariates and LMM for (A) PR3-AAV in GWAS 1 and (B) GWAS2 respectively. The HLA region and SNPs with genome-wide significant associations were not included in the QQ plots. The black line is y = x.

3.4 Association testing of the imputed data and meta-analysis

Following QC of the imputed data, a total of 7,656,576 and 7,744,534 autosomal markers in GWAS 1 and GWAS2 respectively with $r^2 > 0.3$ were available for analysis. Of these, 7,460,066 SNPs were in common to both cohorts for meta-analysis. Case-control association testing was performed using a LMM with BOLT-LMM software to minimise the effect of population stratification and meta-analysis was performed using an inverse-variance method based on a fixed-effects model.

A total of seven PR3-AAV susceptibility loci were identified (Table 3.3). In five of the seven risk loci identified, the most significant SNP within each locus was a common variant (MAF > 0.05) with the ORs ranging between 1.23-1.68, except for the HLA region where the OR was 2.58 (95% CI 2.41-2.75) with the strongest association ($p = 3.59 \times 10^{-173}$). The most significant associations in the remaining two loci were low-frequency variants (MAF = 0.01-0.05) with ORs of 2.39 and 3.43 respectively. Stepwise conditional regression analysis within each non-HLA locus did not identify additional independent genome-wide significant signals.

Two previously reported genome-wide significant non-HLA risk loci were replicated, rs112635299 at chr14q32.13 (*SERPINA1*) and rs55952682 at chr19p13.3 (*PRTN3*). Although not the same lead SNP, they were in moderate-to-high LD with the lead SNPs from the previous GWAS including rs7151526 (r2 = 0.44) and rs28929474 (r2 = 1.0) at *SERPINA1* and rs62132295 (r2 = 0.66) and rs62132293 (r2 = 0.70) at *PRTN3* (70, 71). As *PRTN3* locus was not well represented on the Affymetrix SNP 6.0 Platform, the quality of imputation in this region was very poor in GWAS1 and majority of the SNPs were filtered out due to low r^2. Therefore, only the p-value of GWAS2 was reported for this locus.

3.5 Genetic heritability

To quantify the genetic influence on PR3-AAV, the total SNP heritability (h^2_{SNP}) was estimated using the genome-wide genotyped variants (MAF ≥ 0.01) and additively they explained 46% of the total disease liability (se = 0.03), for a population prevalence of 130 per million population (338).

Table 3.3: HLA and non-HLA associations with PR3-AAV

[§]SNP	[¶]Chr	[¶]Position	Gene(s)	[*]Risk allele	RAF	GWAS1 (478 cases vs 5,259 controls)		GWAS2 (1,132 cases vs 6,688 controls)		Meta-analysis (1,610 cases vs 11,947 controls)	
						[†]OR (95% CI)	[Ψ]P	[†]OR (95% CI)	[Ψ]P	[†]OR (95% CI)	[Δ]P
rs115670873	2p13.1	74,171,061	*DGUOK, TET3*	A	0.93	1.11 (0.70-1.78)	6.50×10^{-1}	1.72 (1.48-2.00)	8.60×10^{-12}	1.68 (1.44-1.97)	1.22×10^{-10}
rs13405741	2q13	111,913,056	*BCL2L11, MIR4435-2HG*	C	0.11	1.55 (1.20-1.99)	6.60×10^{-4}	1.34 (1.17-1.53)	1.60×10^{-5}	1.40 (1.24-1.58)	3.76×10^{-8}
rs3130216	6p21.32	33,077,271	*HLA-DP*	G	0.48	3.06 (2.68-3.49)	1.50×10^{-61}	2.46 (2.28-2.65)	4.30×10^{-114}	2.58 (2.41-2.75)	3.59×10^{-173}
rs145007430	10q26.3	131,690,291	*EBF3, MGMT*	A	0.02	4.75 (2.50-9.04)	2.00×10^{-6}	1.77 (1.26-2.49)	5.70×10^{-4}	2.39 (1.76-3.25)	2.38×10^{-8}
rs112635299	14q32.13	94,838,142	*SERPINA1*	T	0.02	6.51 (4.21-10.07)	4.20×10^{-17}	2.56 (1.99-3.30)	8.90×10^{-13}	3.43 (2.75-4.27)	2.37×10^{-28}
rs11845244	14q32.33	107,170,077	*IGHV1-69*	T	0.37	1.29 (1.12-1.49)	4.70×10^{-4}	1.21 (1.11-1.32)	2.00×10^{-5}	1.23 (1.15-1.33)	2.16×10^{-8}
[‡]rs55952682	19p13.3	835,908	*PRTN3*	T	0.26			1.39 (1.26-1.54)	1.90×10^{-11}	1.39 (1.26-1.54)	1.90×10^{-11}

[§] SNPs showing the most significant level of association after meta-analysis of GWAS1 and GWAS2.

[¶] Chr, chromosome; Genomic coordinates from Genome Reference Consortium Human Build 37 (GRCh37/hg19).

* The risk allele refers to the effect allele in the overall meta-analysis; RAF, risk allele frequency.

† The odds ratio (OR) is with respect to the risk allele; CI, confidence interval.

‡ The SNP6 platform that was used in the GWAS1 has a poor coverage for PRTN3 locus and imputed SNPs were removed due to poor quality ($R^2 < 0.3$).

Ψ The p-values were calculated using a linear mixed model (BOLT-LMM).

Δ The p-values were calculated using inverse-variance method based on a fixed-effects model.

3.6 Candidate gene prioritisation (non-HLA loci)

To prioritise candidate genes, I first performed fine mapping of all the risk loci identified using the Bayesian approach to generate credible sets of SNPs. Next, I searched for nonsynonymous coding and known eQTL among the SNPs in high LD (r2 > 0.6) with the most associated SNPs. Risk loci were functionally annotated using publicly available data from a number of resources using FUMA while long-range interactions between the PR3-AAV associated SNPs and gene promoters and regulatory regions were identified in promoter capture Hi-C datasets using the CHiCP browser. In addition, I cross-referenced the lead associated variant at each locus with disease-associated variants that are in high LD from the Phenoscanner. I also sought external evidence from experimental literature and other genomic databases to provide further evidence to implicate candidate genes in the pathogenesis of PR3-AAV. The results are summarised in Table 3.4.

3.6.1 DGUOK/TET3

The top variant found in chr2p13.1, rs115670873 (p_{meta} = 1.22 x 10^{-10}), is located in the intron of deoxyguanosine kinase (*DGUOK*). It is an imputed SNP with info scores of 0.53 and 0.68 for GWAS1 and GWAS2 respectively. It does not have any LD buddies (or "proxy" SNPs; defined by r2 $\geq$ 0.4). The cluster plots of the genotyped SNPs in this region ($\pm$ 500 kb from the lead variant) were satisfactory. Although this SNP has not been reported to be associated with any trait, this locus has been reported to be associated with SLE, eosinophil count, seropositive RA and dialysis-related mortality. Promoter capture Hi-C (PCHi-C) data showed interactions of this SNP with the promoter region of ten-eleven translocation methylcytosine dioxygenase 3 (*TET3*), DGUOK antisense RNA 1 (*DGUOK-AS1*) and signal transducing adaptor molecule-binding protein (*STAMBP*) in primary immune cell subsets and GM12878 cell line. Other PCHi-C prioritised plausible candidate genes in this region are summarised in Table 3.4. *DGUOK* encodes a protein that is responsible for phosphorylation of purine deoxyribonucleosides in the mitochrondrial matrix and is highly expressed in common lymphoid progenitor, common myeloid progenitor and haematopoietic stem cells (BLUEPRINT data release 2016-08). Mutations in this gene has been associated with mitochondrial DNA (mtDNA) depletion syndrome (OMIM:251880) where affected individuals suffer from progressive liver failure,

neurological abnormalities, hypoglycaemia and increased lactate in body fluids. However, immune phenotypes have not been reported. Affected tissues were previously shown to have decreased activity of the mtDNA-encoded respiratory chain complexes (I, III, IV, V) and mtDNA depletion, suggesting that this enzyme might be involved in the maintenance of balanced mitochondrial dNTPs pools (361).

The ten-eleven translocation (TET) family of proteins consisting of TET1, TET2 and TET3 can catalyse the oxidation of the methyl group on cytosine and thus play an important role in DNA demethylation. They have been shown to participate in numerous developmental processes including haematopoiesis (362). TET2 plays a role in the regulation of FoxP3 expression and influences the differentiation and homeostasis of CD4$^+$FoxP3$^+$ regulatory T cells (363) while in myeloid cells, TET2 is involved in IL-6 production (364). *TET3* is ubiquitously expressed, but is most abundantly in the myeloid and monocytic haematopoietic cells, and it has been suggested that it might play a role in the development of these immune cells (365). More recently, Xue et al and colleagues investigated the role of TET3 in the production of type I IFN by immune cells and demonstrated that TET3 is a negative regulator of IFN-β in response to viral infection and by deleting *TET3*, the expression of IFN-β and IFN-stimulated genes were increased resulting in a heightened antiviral response. They further showed that this effect was independent of DNA demethylation, instead TET3 recruited histone deacetylase 1 (HDAC1) to the promoter of *IFNB1* (366). In addition, global hypomethylation has been observed in CD4$^+$ T cells in individuals with SLE and it has been postulated that demethylation induced by TET3 might play a role in lymphocyte development and induction of tolerance (367).

STAMBP plays a critical role in cytokine-mediated signalling for MYC induction and cell cycle progression. It binds to the SH3 domain of the signal-transducing adaptor molecule in the JAK-STAT signalling cascade. It is most highly expressed in the common lymphoid and myeloid progenitor cells. In the murine model, complete knockout of *STAMBP* leads to preweaning lethality while heterozygous mice develop an abnormal haematopoietic system affecting the monocyte and lymphocyte numbers. A recent publication by Bednash et al which discovered the role of STAMBP in deubiquitinating NALP7 (NACHT, LRR and PYD domains-containing

protein 7) which is one of the inflammasome constituents, to impede NALP7 trafficking to lysosomes, leading to an increase NALP7 abundance and inflammasome-dependent IL-1β cleavage and release. A small-molecule inhibitor of STAMBP deubiquitinase activity, decreases NALP7 protein levels and suppressed IL-1β release after TLR agonism, suggesting it could be a potential therapeutic target to reduce pro-inflammatory stress (368). These relevant functional associations involving the immune system suggest that *TET3*, *DGUOK* and *STAMBP* are more likely to be the causal gene at this locus as compared to other candidate genes.
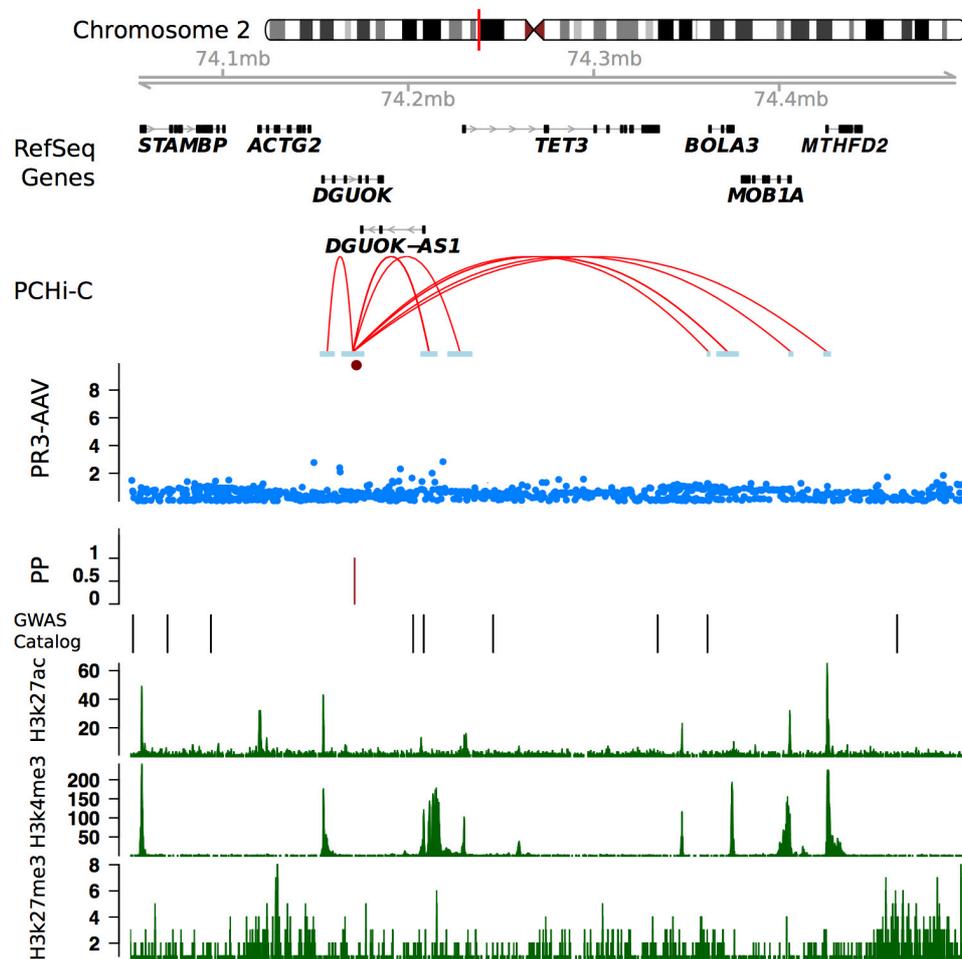
Figure 3.2: Genomic feature at chr2p13.1 (*DGUOK* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq genes, promoter capture Hi-C (PCHi-C) data from Javierre et al (primary immune cell subsets), SNPs association with PR3-AAV with filled red circle indicates the lead SNP (rs115670873), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data.

Table 3.4: Candidate genes at PR3-AAV associated loci

SNP	Most likely causal gene(s)	Type of variants	Diseases or traits associated with the lead SNP or in high LD with lead SNP (r2 > 0.6) or in proximity to candidate gene(s)		*Cis-eQTL (blood/relevant tissues)	‡Immune phenotype in murine model	PCHi-C prioritised protein coding gene(s)	PCHi-C prioritised non-protein coding transcript(s)
			Disease or trait (SNP)	Relationship				
rs115670873	*DGUOK, TET3*	Intronic	SLE (158, 369)	In proximity	n/a	*STAMBP*	*TET3, BOLA3, MOB1A, MTHFD2, ACTG2, STAMBP, ALMS1, SEMA4F, M1AP, INO80B, WBP1, DCTN1, SCL4A5*	*DGUOK-AS1, BOLA-AS1, AC073046.1-201, DCTN1-AS1*
			Eosinophil counts (370)					
			†Systemic sclerosis					
			†Seropositive RA					
			Dialysis-related mortality (371)					
			†Crohn's disease					
rs13405741	*BCL2L11, MIR4435-2HG*	Intronic	¶mQTL (cg04202892) (372)	Same	*BCL2L11* (blood), *MIR4435-2HG* (brain)	*BCL2L11, MIR4435-2HG*	*BCL2L11, ACOXL*	*MIR4435-2HG*
			EGPA (rs72836352)	r2 = 1.0				
			Eosinophil count, neutrophil % of granulocytes (rs17484848) (346)	r2 = 0.91				
			PSC (373), †Asthma (rs72837826)	r2 = 0.89				
			Monocyte % of white cells (rs72836346) (346)	r2 = 0.65				
			Vitiligo (rs4308124) (374)	r2 = 0.17				
			CLL (375), monocyte count (346), atopic dermatitis (376)	In proximity				
rs145007430	*EBF3, MGMT*	Intronic	Atopic eczema, psoriasis (377)	In proximity	n/a	n/a	*MGMT, PTPRE*	n/a
			†RA					
rs112635299	*SERPINA1*	Intergenic	§pQTL (SNAP25, C1QTNF1)	Same	*SERPINA1, IFI27L2, IFI27L1, IFI27* (blood)	*SERPINA1*	*SERPINA9, SERPINA12, SERPINA4, SERPINA5*	*RP11-349I1.2, RP11-986E7.7*
			§pQTL (MRPL33, ZNF175, ACP2, NCF2, WISP3, PIM1, ARFIP2, SLC14A2, TXNDC5, DNAJB9, GPC6) (rs28929474)	r2 = 1.0				

SNP	Most likely causal gene(s)	Type of variants	Diseases or traits associated with the lead SNP or in high LD with lead SNP (r2>0.6) or in proximity to candidate gene(s)		*Cis-eQTL (blood/relevant tissues)	[‡]Immune phenotype in murine model	PCHi-C prioritised protein coding gene(s)	PCHi-C prioritised non-protein coding transcript(s)
			Disease or trait	Relationship				
rs11845244	*IGHV1-69*	Missense	[¶]mQTL	Same	*IGHV1-69, IGHV3-66, IGHV3-64, IGHV4-61, IGHV4-59, IGHV3-53, IGHV4-39* (blood)	n/a	n/a	n/a
			[§]pQTL (DEF119, LAP, EPHB3)	Same				
			[§]pQTL (ACP2) (rs10139058)	r2 = 0.79				
			[§]pQTL (STIM1) (rs3814922)	r2 = 0.78				
rs55952682	*PRTN3*	Intergenic	[§]pQTL (PRTN3)	Same	*PRTN3, CFD, MED16*	*PRTN3, ELANE, CFD, KISS1R, ARID3A, CNN2*	*PRTN3, ELANE, CFD, AZU1, PLPPR3, PTBP1, MISP, PALM, PRSS57*	*CTB-31O20.4, CTB-31O20.3, MIR1909, CTB-25B13.6, RPS15P9, CIRBP-AS1, AC004221.2*
			[¶]mQTL, Basophil % of white cells (346)	Same				
			Monocyte count	Same				
			Monocyte % of white cells (rs138303849) (346)	r2 = 0.99				
			White blood cell count (rs56757486) (346)	r2 = 0.99				
			[§]pQTL (PRTN3) (rs7254911)	r2 = 0.97				

CLL, chronic lymphocytic leukaemia; EGPA, eosinophilic granulomatosis with polyangiitis; eQTL, expression quantitative trait loci; mQTL, methylation quantitative trait loci; PCHi-C, promoter capture Hi-C; pQTL, protein quantitative trait loci; PSC, primary sclerosing cholangitis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; %, percentage.

[†]Results obtained from UK Biobank study (378) using Phenoscanner database (http://www.phenoscanner.medschl.cam.ac.uk/).

*The lead SNP or SNPs in high LD with the lead SNP in the locus that are known eQTLs (obtained from http://www.eqtlgen.org/index.html).

[§]pQTL was obtained Sun et al from NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/) or Phenoscanner (319).

[‡]The immune phenotype designation was taken from http://www.informatics.jax.org/phenotypes.shtml.

[¶]mQTL was obtained from mQTL database http://www.mqtldb.org/ (372).

3.6.2   BCL2L11

The top variant found in chr2q13, rs13405741 (p_{meta} = 3.76 x 10^{-8}), is located in the intron of Bcl-2-like protein 11 (*BCL2L11*), encoding BIM which is a Bcl2 family member that is essential for controlling apoptosis and immune homeostasis. This SNP and several SNPs in strong LD with it (r2 > 0.6) have been reported to be associated with PSC (373), EGPA (Lyons et al 2018 bioRxiv), eosinophil count, neutrophil percentage of white cells, monocyte percentage of white cells (346) and asthma (378). Specifically, the PR3-AAV risk allele (rs13405741, C allele) is associated with a higher eosinophil count, a greater eosinophil percentage of white cells, a decreased neutrophil percentage of granulocytes, a higher monocyte percentage of white cells (346) and an increased risk of PSC (373). This locus has been independently (r2 < 0.1) reported to be associated with CLL (375) and atopic dermatitis (376). PCHi-C data showed interactions of rs17484848 (in high LD with rs13405741, r2 = 0.91) with the putative enhancer region of *MIR4435-2HG* and the promoter region of both *BCL2L11* and Acyl-CoA oxidase like (*ACOXL*) (Figure 3.3). Furthermore, rs13405741 is a known eQTL for *BCL2L11* in the whole blood with p-value of 8.75 x 10^{-36}, where the minor allele (C, risk allele for PR3-AAV) is associated with a higher expression of *BCL2L11*, supporting that the SNP might play a role in regulating the expression of *BCL2L11* (318). In addition, *BCL2L11^{tm1.1Ast}* knockout mice exhibit progressive systemic autoimmune diseases with abnormal accumulation of lymphoid and myeloid cells (379).

MIR4435-2HG (a.k.a *Morrbid*), encodes a long non-coding RNA, was recently discovered to play a critical role in the regulation of *BCL2L11* transcription in the myeloid cells and hence their lifespan, by promoting the enrichment of the polycomb repressive complex 2 at the promoter of *BCL2L11* to keep it in a poised state (380). *Morrbid*-deficient mice had markedly reduced number of eosinophils, neutrophils and Ly6C^{hi} classical monocytes in the blood and tissues but no effects were observed in other immune cell types. Interestingly, the authors also reported an overexpression of *Morrbid* in the eosinophils of patients who suffer from hypereosinophilic syndrome (HES), suggesting its potential role in the pathogenesis of HES. The relevant functional associations suggest that *BCL2L11* and *Morrbid* are more likely to be the causal gene at this locus than *ACOXL*, of which its function is not well defined and it is highly expressed in the lung, testis and prostate.
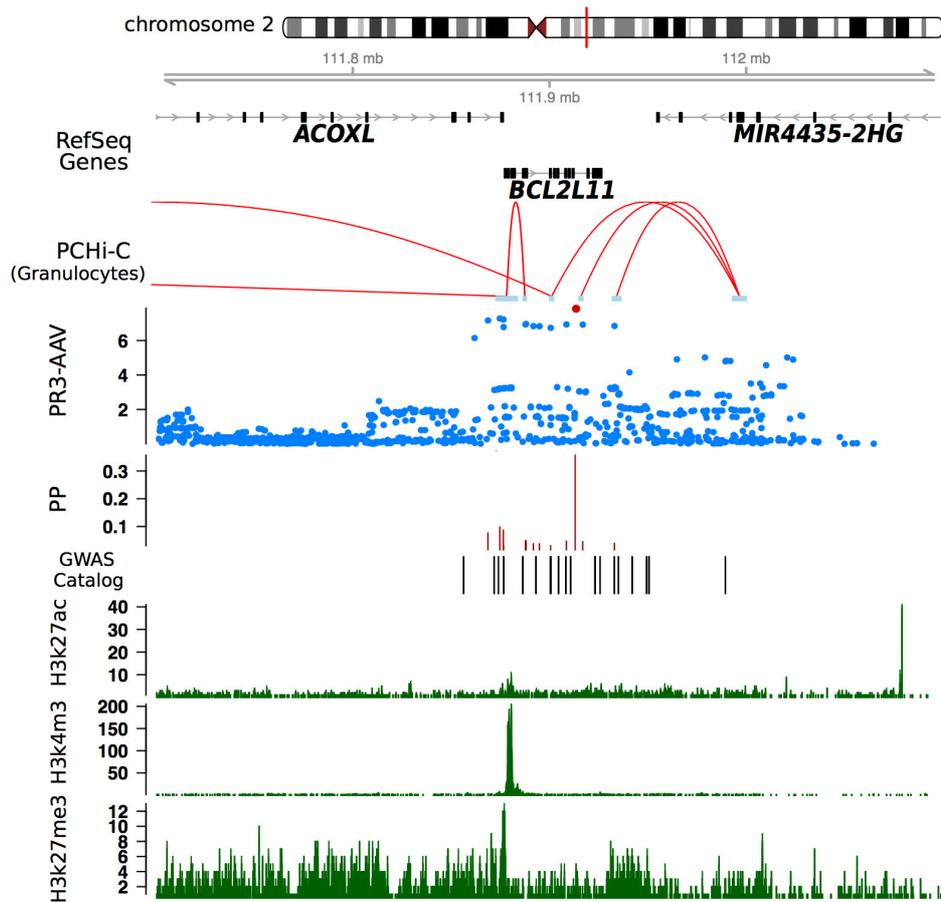
Figure 3.3: Genomic features at chr2q13 (*BCL2L11* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq genes, PCHi-C data (Javierre et al - granulocytes subset), SNPs association with PR3-AAV where the filled red circle indicates the lead SNP (rs13405741), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data.

3.6.3 EBF3

The top variant found in chr10q26.3, rs145007430 (p_{meta} = 2.38x 10^{-8}), is located in the intron of early B-cell factor 3 (*EBF3*), encoding for a member of the EBF family of DNA binding transcription factors. It is an imputed SNP with info scores of 0.63 and 0.83 for GWAS1 and GWAS2 respectively. It does not have any strong LD buddies (r2 $\geq$ 0.4) and the cluster plots of the genotyped SNPs in this region ($\pm$ 500 kb from the lead variant) were satisfactory. EBF proteins are involved in differentiation and maturation of several cell lineages, including B-progenitors lymphoblasts, neuronal precursors and osteoblast progenitors (381). The encoded protein inhibits cell survival through the regulation of genes involved in cell cycle arrest and apoptosis, and aberrant methylation or deletion of this gene have been associated with multiple malignancies such as acute lymphoblastic leukaemia and glioblastoma (382). *EBF3* is expressed in the common lymphoid progenitor and plasma cells, but not in naïve and memory B cells. The PCHi-C dataset showed interactions of rs145007430 with the putative regulatory region and the promoter region of O-6-methylguanine-DNA methyltransferase (*MGMT*) in GM12878 cell line, which encodes a protein that catalyses transfer of methyl groups from O(6)-alkylguanine and other methylated moieties of the DNA to its own molecule. Mechanistically, DNA methylation can modulate gene expression by influencing the binding affinities of transcription factors. It was demonstrated that patients with hypersensitivity pneumonitis had evidence of hypomethylation of *EBF3* as compared to healthy controls with significantly higher expression of *EBF3*, IL-4 and IL-10, indicating that *EBF3* hypomethylation in B cells might play a role in the conversion of Th1 dominant to Th2 dominant response in patients who were in the chronic phase of the disease (383). In contrast, hypermethylation at CpG sites in the *EBF3* gene were observed in RA synovial fibroblasts with a marked decrease in *EBF3* expression as compared with OA patients and healthy controls, suggesting its potential role in the pathogenesis of RA (384). This locus has independently been reported to the associated with inflammatory skin disorders including psoriasis and eczema (377).
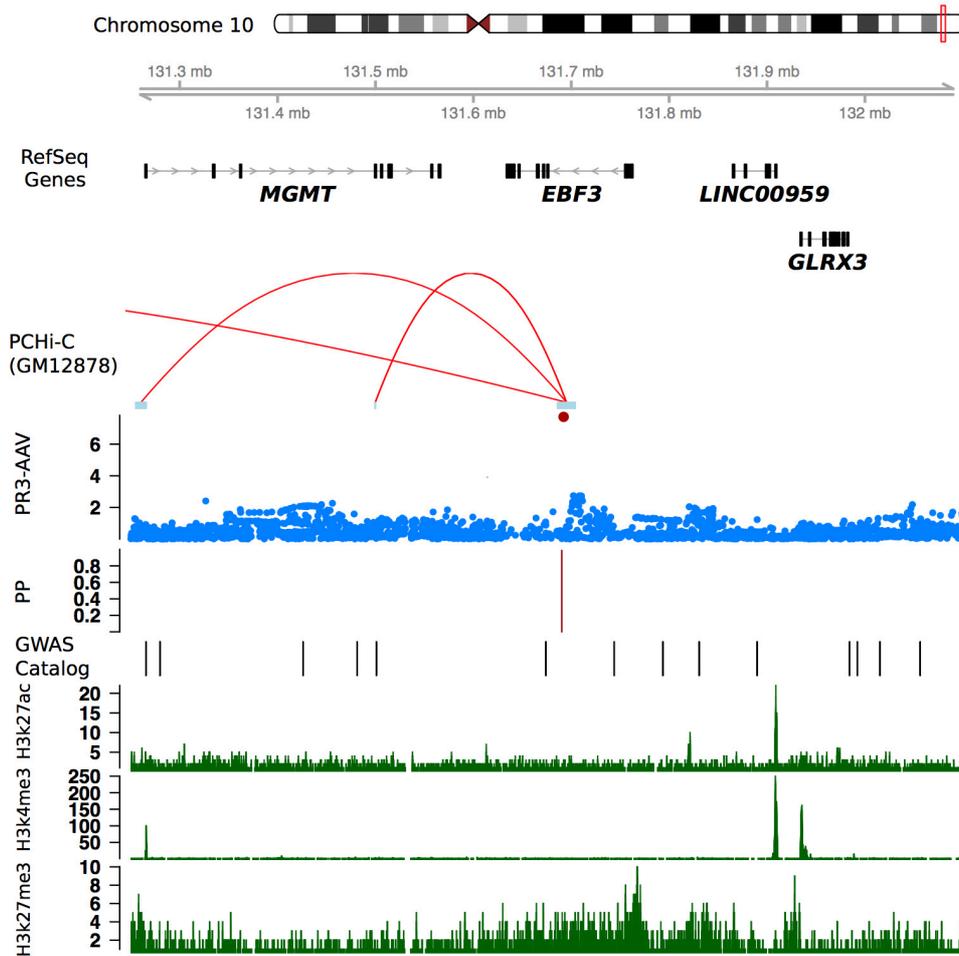
Figure 3.4: Genomic features at chr10q26.3 (*EBF3* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq genes, PCHi-C data (Mifsud et al - GM12878 cell line), SNPs association with PR3-AAV where the filled red circle indicates the lead SNP (rs145007430), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data.

3.6.4   SERPINA1

This locus has been identified in previous GWAS and its potential role in the pathogenesis of PR3-AAV is discussed in section 1.4.2. The top variant identified in chr14q32.13, rs112635299 (p_{meta} = 2.37 x 10^{-28}), is located in the intergenic region with close proximity to *SERPINA1*. It is a known eQTL for *SERPINA1* and *IFI27* (interferon-alpha-inducible protein 27). In addition, it is a pQTL for two proteins including synaptosomal nerve-associated protein 25 (SNAP-25) and complement C1q tumour necrosis factor-related protein 1 (C1QTNF1). The second most significant variant in this locus was rs28929474 (p_{meta} = 9.40 x 10^{-28}) and it is in complete LD

89

with rs112635299 (r2= 1.0). It is a missense variant that leads to substitution of the amino acid glutamine in position 342 in the protein for a lysine amino acid and is also known as the "Z" allele. The PCHi-C dataset showed interactions with the promoters of *SERPINA5*, *SERPINA9, SERPINA11* and *SERPINA12* in GM12878 cell line. *SERPINA5* encodes for protein c inhibitor which interacts with lipid membrane leading to permeabilisation of bacterial pathogen to exert antimicrobial activity while *SERPINA12* encoding vaspin, has been previously shown to reduce expression of proinflammatory factors, improve lipid catabolism and increase insulin sensitivity.
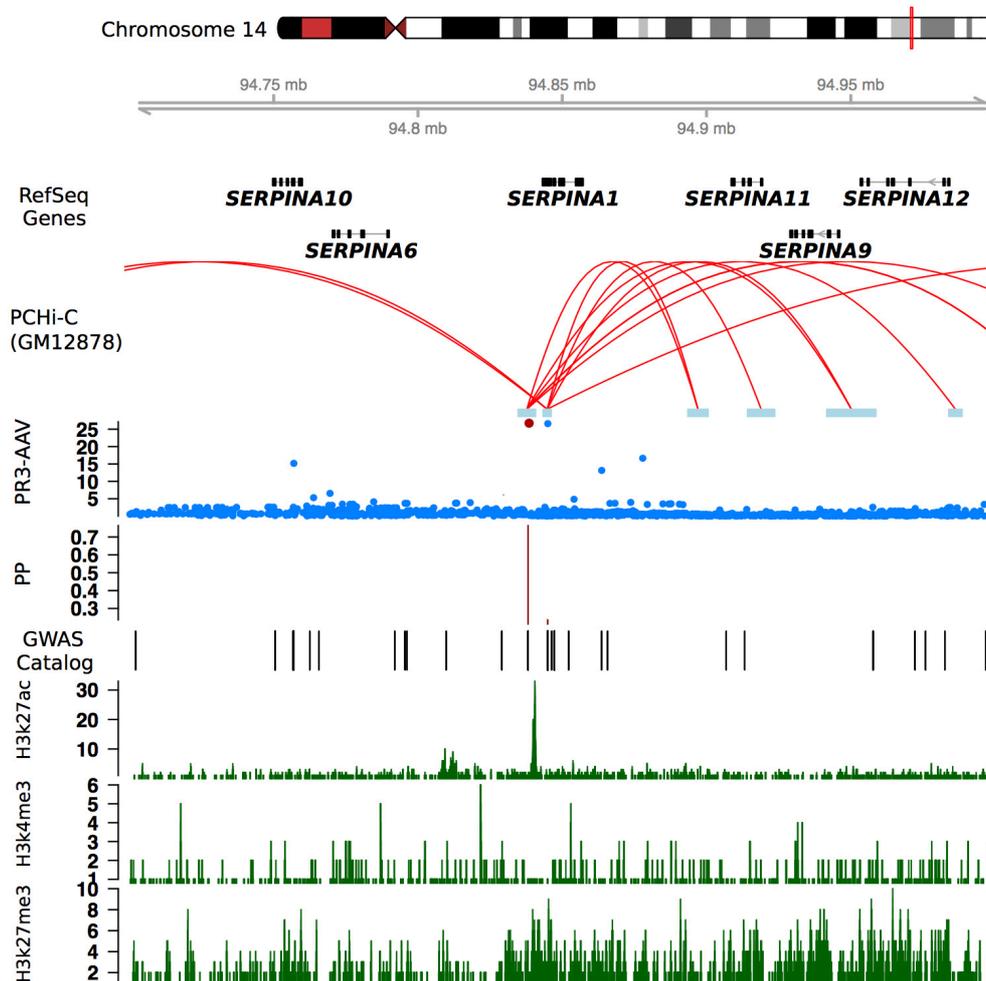
Figure 3.5: Genomic features at chr14q32.13 (*SERPINA1* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq genes, PCHi-C data (Mifsud et al - GM12878 cell line), SNPs association with PR3-AAV where the filled red circle indicates the lead SNP (rs112635299), causal variant mapping expressed as posterior probability (PP), other GWAS significant hits in the region and methylation data.

3.6.5 IGHV1-69

The most significant association, rs11845244 is located in the immunoglobulin heavy variable 1-69 (*IGHV1-69*) gene (Figure 3.6). It is a missense variant that results in a change of amino acid at position 50 from glycine to arginine (G50R based on Kabat numbering system), which is the last amino acid in the framework region 2 and is immediately adjacent to the complementarity determining region 2 (CDR2) as demonstrated in Figure 3.7. It is a trans-pQTL for defensin beta 119 (DEFB119), an antimicrobial peptide that protects tissues and organs from infection by a variety of microorganisms.
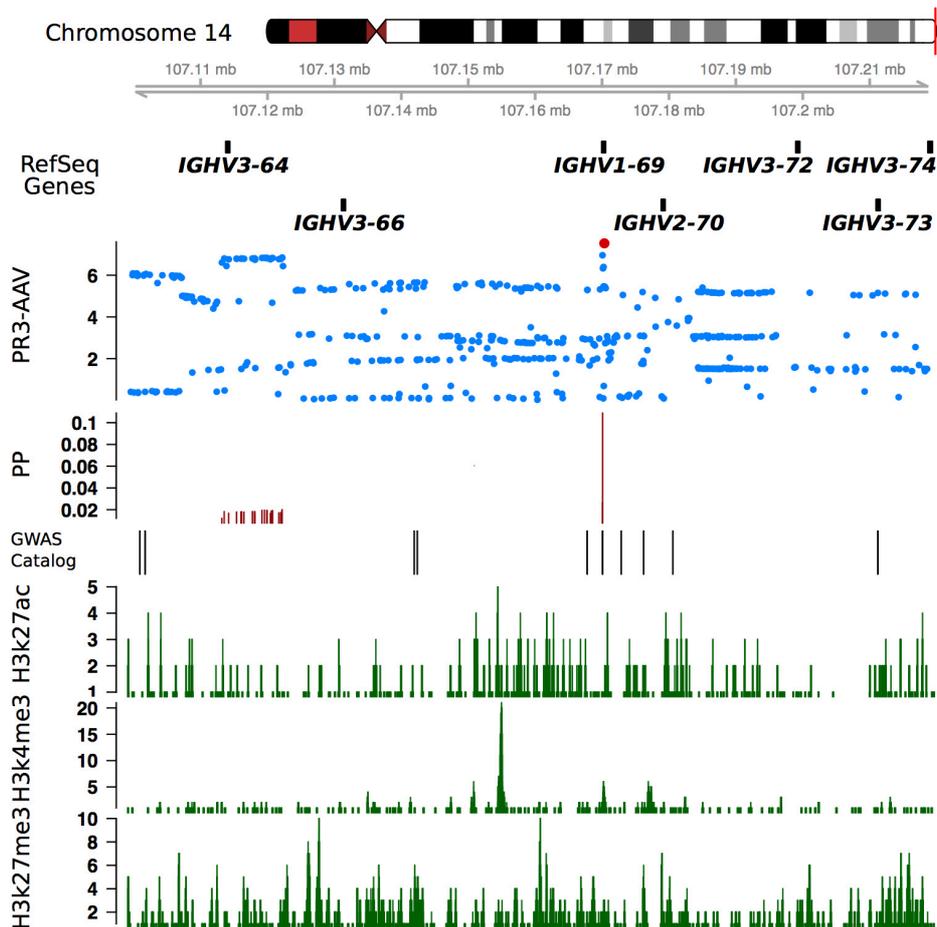
Figure 3.6: Genomic features at chr14q32.33 (*IGHV1-69* gene).

Top to bottoms panels: Genomic positions (hg19), representative RefSeq genes, SNPs association with PR3-AAV where the filled red circle indicates the lead SNP (rs11845244), causal variant mapping expressed as posterior probability (PP), other GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data (H3k27ac, H3k4me3 and H3k27m3).
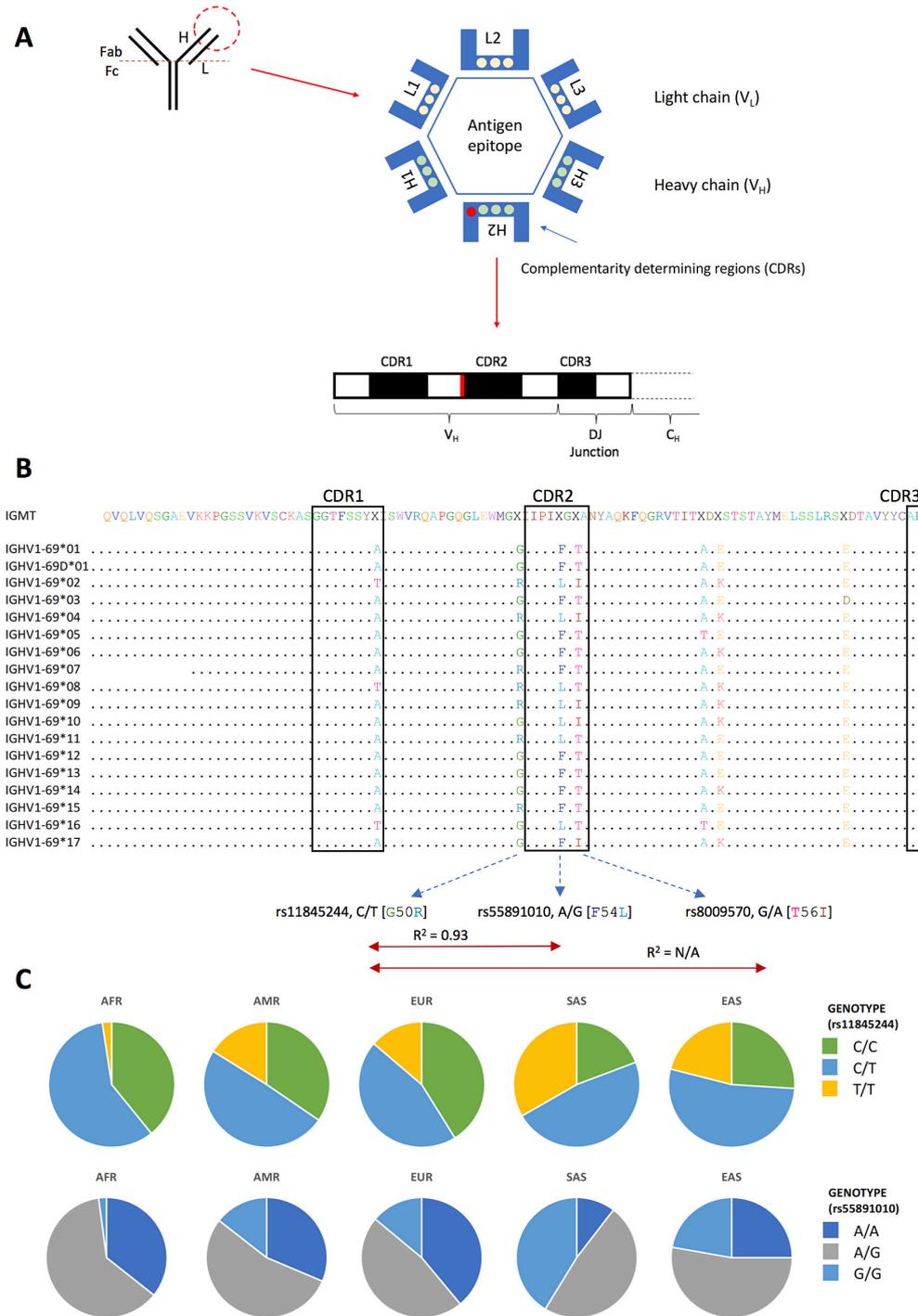
Figure 3.7: *IGHV1-69* germline polymorphisms.

(A) This schematic representation of an immunoglobulin (Ig), which is made up of two heavy chains (V$_H$) and two light chains (V$_L$). The variable region of Ig that binds to the antigen for which the antibody is specific to is highlighted in a red dotted circle. Each heavy and light chain contains three complementarity-determining regions (CDRs) that are folded together to form the antigen binding site, which determine the specificity and the affinity of binding. rs11845244 is a missense variant that results in a change of amino acid at position 50 from glycine to arginine (G50R), which is located at the border of framework region 2 of the heavy chain and is

indicated by a filled red circle. The rearranged genomic V$_{H1-69}$ gene encoding the protein heavy chain is demonstrated.

(B) The amino acid sequences of *IGHV1-69* alleles are aligned and grouped based on the occurrence of phenylalanine or leucine at position 54 (F54L). The black boxes highlight the location of CDR1, CDR2 and partial CDR3 based on the International Immunogenetics Information System (IMGT) database. Positions where the alleles are different among the sequences are depicted with colored amino acids.

(C) The genotype frequency differences at two SNPs in *IGHV1-69* that have been shown to encode functional residues that are critical for neutralizing antibodies against the influenza hemagglutinin stem region (rs55891010, where A/A encodes the functional critical F54 residue) and NEAT2 domain of *S. aureus* (rs11845244, where C/C encodes the primary G50 residue), between five human ethnic groups (Africans (AFR); Central and South Americans (AMR); Europeans (EUR); South Asians (SAS) and East Asians (EAS)), published by the 1000 Genomes project (285).

Antibodies are key components of the adaptive immune response and can selectively recognise and mediate immune responses to foreign antigens. This is accomplished through the vast diversity of the B-cell antibody repertoire, which begins in the bone marrow where B lymphocytes rearrange their immunoglobulin (Ig) variable (V), diversity (D) and joining (J) region gene segments coding for the B-cell receptor for antigen. This specialised process is known as V(D)J recombination that somatically joins various V, D and J segments at Ig heavy-chain locus or V and J segments at Ig Kappa and Ig Lambda chain loci. Ig is made up of two heavy chains and two light chains. It can be divided into two fragments, Fab and Fc fragments. The Fab fragment engages antigen while the Fc fragment mediates effector functions. They differ in *IgV* gene composition where the Fab region contains the antigen-binding site, being composed predominantly of the Ig variable region gene segments (*IGHV*, *IGHD* and *IGHJ*) and the Fc region composed completely by an *IGHC* gene (385). Each heavy and light chain contains three CDRs that are folded together to form the antigen-binding site, which determine the specificity and the affinity of binding. Random rearrangement of the heavy- and light-chain genes further increases genetically encoded diversity to ensure that it will likely contain a naïve antibody with at least weak initial binding against a vast array of antigens. This diversity will then be augmented following antigen stimulation where B cells undergo somatic hypermutation, resulting in thousands of clonally derived affinity maturation variants of the initial antibody to mount a highly specific antigen response (386).

Despite the importance of the unique molecular mechanisms in generating highly diverse functional antibodies, there is a growing appreciation that the Ig genes exhibit a high level of allelic polymorphism. Recent studies have demonstrated that Ig germline variation plays a crucial role in the development of naïve B cell repertoire and downstream implications in the B cell memory compartment and individual's ability to mount antibody response to specific epitopes (387-390). It has been shown in several studies that there is biased use of the heavy-chain gene *IGHV1-69* in generating stem-directed broadly neutralising antibodies against influenza hemagglutinin (HA) (391, 392). More importantly, there are two SNPs in *IGHV1-69* that have been shown to encode functional residues that are critical for broadly neutralising antibodies against influenza HA (F54 and L54 amino acid-associated alleles; rs55891010) (387, 390, 393) and NEAr-iron Transporter 2 (NEAT2) domain of *S. aureus* (G50 and R50 amino acid-associated alleles; rs11845244) (394). Yeung and colleagues demonstrated that it is the CDR-H2 that dominates the binding interaction between neutralising antibodies and NEAT2. More importantly, using enzyme-linked immunosorbent assay, they also showed that allelic variation at position 50 (R50) completely abolished NEAT2 binding, presumably due to steric hindrance (394). Both rs55891010 and rs11845244 are in high LD (r2 = 0.93) and therefore L54 and R50 amino acid tend to co-occur in alleles of *IGHV1-69*. The frequencies of these critical alleles vary across populations where the L54 and R50 amino acids are most commonly seen in South Asians and least frequently in Africans (Figure 3.7). rs55891010 was not available in our imputed data.

3.6.6 PRTN3

This locus has been identified in previous GWAS and its potential role in the pathogenesis of PR3-AAV is discussed in section 1.4.2. The top variant identified in chr19p13.3, rs55952682 (p_{GWAS2} = 1.90 x 10^{-11}), is located in the upstream region of *PRTN3* and downstream of azurocidin 1 (*AZU1*). *AZU1* encodes a protein precursor that is proteolytically processed to generate a mature azurophil granule antibiotic protein with antimicrobial and monocyte chemotactic activity. *AZU1*, neutrophil elastase 2 (*ELANE*) and *PRTN3* are located in a cluster. Their mRNA expressions are coordinately controlled during neutrophil differentiation and their protein products are packaged into azurophil granules. The lead SNP and several SNPs in strong LD with it (r2 > 0.6) have been reported to be associated with monocyte count,

monocyte percentage of white cells, white blood cell count and basophil percentage of white cells. Furthermore, it is a known eQTL for *PRTN3*, complement factor D (*CFD*) and mediator complex subunit 16 (*MED16*) in the whole blood, where the risk allele for PR3-AAV (T) is associated with higher expressions of *PRTN3* (p = 9.69 x 10^{-117}) and *CFD* (p = 4.35 x 10^{-17}) but a lower expression of *MED16* (p = 1.394 x 10^{-5}) (318). It is also a pQTL where the PR3-AAV risk allele is associated with an increased plasma PR3 levels (319). There are four independent cis-pQTL association signals for plasma PR3 levels in this locus and the strongest association is rs10425544, which lies in the promoter region of *PRTN3*. The relationship between PR3-AAV associated variants and PR3 pQTL variants are shown in Table 3.5. Furthermore, a higher frequency of chromatin interactions was observed in this region (see Figure 3.8), suggesting this region could be a topologically associating domain (TAD) (395), where the DNA sequences physically interact with each other more often than with sequences outside the TAD, which has been postulated to play a role in gene regulation and its disruption, leading to disease.
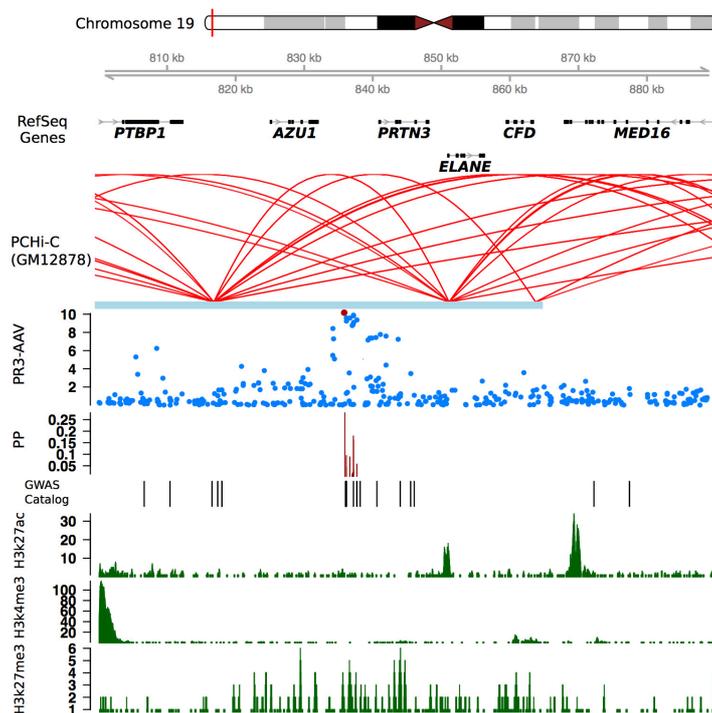
Figure 3.8: Genomic features at chr19p13.3 (*PRTN3* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq genes, PCHi-C data (Mifsud et al - GM12878 cell line), SNPs association with PR3-AAV where the filled red circle indicates the lead SNP (rs55952682), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data.

Table 3.5: LD between PR3-AAV associated variants and PR3 pQTL variants.

	*Lead pQTLs	*Conditionally independent pQTLs			Lyons et al	Merkel et al	Present study
	rs10425544	rs7254911	rs351111	rs6510982	rs62132295	rs62132293	rs55952682
rs7254911	D': 1.0 r2: 0.0984	-	-	-	-	-	-
rs351111	D': 0.86 r2: 0.3079	D': 0.89 r2: 0.1901	-	-	-	-	-
rs6510982	D': 0.97 r2: 0.6068	D': 0.88 r2: 0.0491	D': 1.0 r2: 0.2692	-	-	-	-
rs62132295	D': 0.70 r2: 0.0620	D': 0.90 r2: 0.6317	D': 0.91 r2: 0.2523	D': 0.90 r2: 0.0656	-	-	-
rs62132293	D': 0.72 r2: 0.0709	D': 1.0 r2: 0.7222	D': 0.88 r2: 0.2533	D': 0.81 r2: 0.0577	D': 1.0 r2: 0.9202	-	-
rs55952682	D': 1.0 r2: 0.0952	D': 1.0 r2: 0.9676	D': 0.89 r2: 0.1828	D': 0.87 r2: 0.0471	D': 0.93 r2: 0.6562	D': 1.0 r2: 0.0471	-

*pQTL data was obtained from Sun et al (319).

LD calculations were based on 1000 Genomes phase 3 European ancestry individuals using LDlink (https://ldlink.nci.nih.gov/?tab=home).

3.7   Fine-mapping of the HLA region

The HLA region contained the strongest association signals in the PR3-AAV GWAS, including the *HLA-DQ* and *HLA-DP* loci. The most significant associated variant was rs3130216 (*HLA-DP*; OR = 2.58, p = 3.59 x 10^{-173}) following a meta-analysis of GWAS1 and GWAS2. Given the HLA region is known to show extensive linkage disequilibrium, I performed reciprocal conditional association analysis to dissect the independent signals in this region. As the HLA imputation was performed only on the GWAS2 dataset, I focused on the same.

The lead SNP in the HLA region for GWAS2 PR3-AAV was rs1042169, which is in moderate LD with rs3130216 (r2 = 0.50) and was reported in Merkel et al as one of the top hits in the *HLA-DPB1* region (71). It is a missense variant that results in a change of amino acid at position 113 from glycine to aspartate in the DP beta 1 chain precursor. A conditional analysis of the top variant identified two additional independent variants, rs9277569 (*HLA-DPB1*; OR = 2.49, conditional p = 8.20 x 10^{-39}) and rs28407950 (*HLA-DQ*; OR = 1.50, conditional p = 1.50 x 10^{-19}) (Table 3.6). Lastly, conditioning on all three variants accounted for the entire signal seen in the HLA region (Figure **3.9**).

Table 3.6: Stepwise conditional analysis of the HLA region for PR3-AAV (GWAS2)

SNPs (*Chr:Position)	Gene(s)	Type	†Risk allele (RAF)	Unconditioned ‡p value (OR)	Conditioned on		
					rs1042169	rs9277569	rs1042169 rs9277569
rs1042169 (6:33,048,686)	HLA-DPB1	Missense	G (0.721)	1.2×10^{-117} (2.74)	-	6.2×10^{-140} (3.08)	-
rs9277569 (6:33,058,402)	HLA-DPB1	Intergenic	C (0.892)	1.6×10^{-14} (1.69)	8.2×10^{-39} (2.49)	-	-
rs28407950 (6:32,626,348)	HLA-DQ	Regulatory region	C (0.730)	9.0×10^{-23} (1.54)	5.2×10^{-22} (1.54)	3.1×10^{-21} (1.52)	1.5×10^{-19} (1.50)

* Genome Reference Consortium Human Build 37 (GRCh37).

† The risk allele refers to the effect allele. RAF, risk allele frequency.

‡ The p-values were calculated using linear mixed model (using BOLT-LMM); OR, odds ratio with respect to the risk allele.

Figure 3.9: **The HLA associations with PR3-AAV is localised to the Class II region.**

(A) The HLA association signals in PR3-AAV (GWAS2) and the genomic coordinates are from the hg19 genome build. The HLA association signals following conditioning on (B) rs1042169, (C) rs1042169 and rs9277569 and (D) rs1042169, rs9277569 and rs28407950 where no further genome-wide significant signals were observed.

The classical HLA alleles at 2- or 4-digit resolution and amino acid variants at eight HLA loci were imputed. Univariate LMM analysis identified nine HLA alleles conferring either susceptibility or protection to PR3-AAV. To test association among imputed HLA alleles and amino acid variants, I used a forward stepwise strategy to establish statistically independent signals. Within each variant type, I first identified the most significant associated signal as defined by the lowest p-value and performed forward iterative conditional regression analysis to identify other independent signals.

I identified four independent HLA signals for PR3-AAV, three of which conferred risk including *HLA-DPB1*04:01-HLA-DQB1*06:02-HLA-DRB1*15:01*, *HLA-DPB1*04:02* and *HLA-DPA1*01:03-HLA-DPB1*02:01* while the remaining signal conferred protection by an extended haplotype encoding *HLA-DRB1*01:01-HLA-DQA1*01:01-HLA-DQB1*05:01*. The strongest independent associations with disease susceptibility were observed at *HLA-DPB1*04:01* (OR = 2.52, p = 4.60 x 10^{-116}), *HLA-DPB1*04:02* (OR = 2.47, conditional p = 5.90 x 10^{-38}), *HLA-DQA1*01:01* (OR = 0.71, conditional p = 1.20 x 10^{-10}) and *HLA-DPB1*02:01* (OR = 1.41, conditional p = 3.50 x 10^{-7}) (Table 3.7 and Figure 3.10). In addition, *HLA-DPB1*04:01* was a stronger signal than the top hit SNP, supporting that the main signal in the region is indeed driven by the HLA allele.

Individual amino acid variants in *HLA-DPB1*, *HLA-DQB1* and *HLA-DRB1* were associated with disease. Within *HLA-DPB1*, the HLA classical alleles had stronger associations than the amino acid variants. There were significant residual effects for *HLA-DPB1*04:01* (conditional p = 2.00 x 10^{-7}) and *HLA-DPB1*02:01* (conditional p = 1.70 x 10^{-9}) after conditioning on all the amino acid associations in this region. Within *HLA-DQB1*, the amino acid polymorphism HLA-DQB1 His30 (OR = 0.64, p = 2.00 x 10^{-20}), which is located in the peptide-binding groove, accounted for most of the HLA allele effects (*HLA-DQB1*06:02*, conditional p = 0.00054 and *HLA-DQB1*05:01*, conditional p = 0.23). Within *HLA-DRB1*, there were no significant residual effects for HLA classical alleles after conditioning on the most significant amino acid variants consisting of HLA-DRB1 Ala or Lys71 (OR = 1.34, p = 1.10 x 10^{-11}), HLA-DRB1 Gly or Tyr13 (OR = 1.36, conditional p = 5.50 x 10^{-9}) and HLA-DRB1 Glu58 (OR = 1.54, conditional p = 2.30 x 10^{-8}) - all of which are part of the peptide-binding site of *HLA-DRB1*.

Table 3.7: Association of classical HLA alleles in PR3-AAV (GWAS2)

MHC allele	Unconditioned		HLA-DPB1*04:01		HLA-DPB1*04:01 HLA-DPB1*04:02		HLA-DPB1*04:01 HLA-DPB1*04:02 HLA-DQA1*01:01		HLA-DPB1*04:01 HLA-DPB1*04:02 HLA-DQA1*01:01 HLA-DPB1*02:01	
	OR	P	OR	P	OR	P	OR	P	OR	P
HLA-DPB1*04:01	2.52	4.6×10^{-116}	-	-	-	-	-	-	-	-
HLA-DPA1*01:03	2.36	2.9×10^{-54}	1.58	7.1×10^{-14}	1.25	3.6×10^{-5}	1.27	2.4×10^{-5}	1.12	0.036
HLA-DPA1*02:01	0.42	1.7×10^{-44}	0.65	6.6×10^{-11}	0.82	1.1×10^{-3}	0.82	1.0×10^{-3}	0.93	0.20
HLA-DPB1*03:01	0.43	4.8×10^{-32}	0.66	1.4×10^{-7}	0.83	0.054	0.83	0.045	0.94	0.80
HLA-DPB1*01:01	0.44	8.1×10^{-18}	0.69	1.7×10^{-4}	0.86	0.10	0.83	0.042	0.93	0.38
HLA-DPB1*02:01	0.60	1.8×10^{-15}	1.02	0.54	1.40	9.4×10^{-7}	1.41	3.5×10^{-7}	-	-
HLA-DPA1*02:02	0.49	7.1×10^{-9}	0.71	5.1×10^{-3}	0.84	0.086	0.82	0.065	0.90	0.23
HLA-DQA1*01:01	0.74	8.2×10^{-9}	0.76	1.2×10^{-7}	0.71	1.2×10^{-10}	-	-	-	-
HLA-DQB1*06:02	1.37	1.7×10^{-8}	1.04	0.41	-	-	-	-	-	-
HLA-DRB1*15:01	1.34	9.2×10^{-8}	1.01	0.62	-	-	-	-	-	-
HLA-DQB1*05:01	0.74	2.1×10^{-7}	0.77	8.0×10^{-6}	0.71	2.1×10^{-9}	0.92	0.51	-	-
HLA-DRB1*01:01	0.75	3.3×10^{-6}	0.80	1.6×10^{-4}	0.71	4.0×10^{-8}	0.92	0.36	-	-
HLA-DPB1*04:02	1.29	3.3×10^{-4}	2.47	5.9×10^{-38}	-	-	-	-	-	-

OR, odds ratio; P, p-values were calculated using a linear mixed model with BOLT-LMM software.
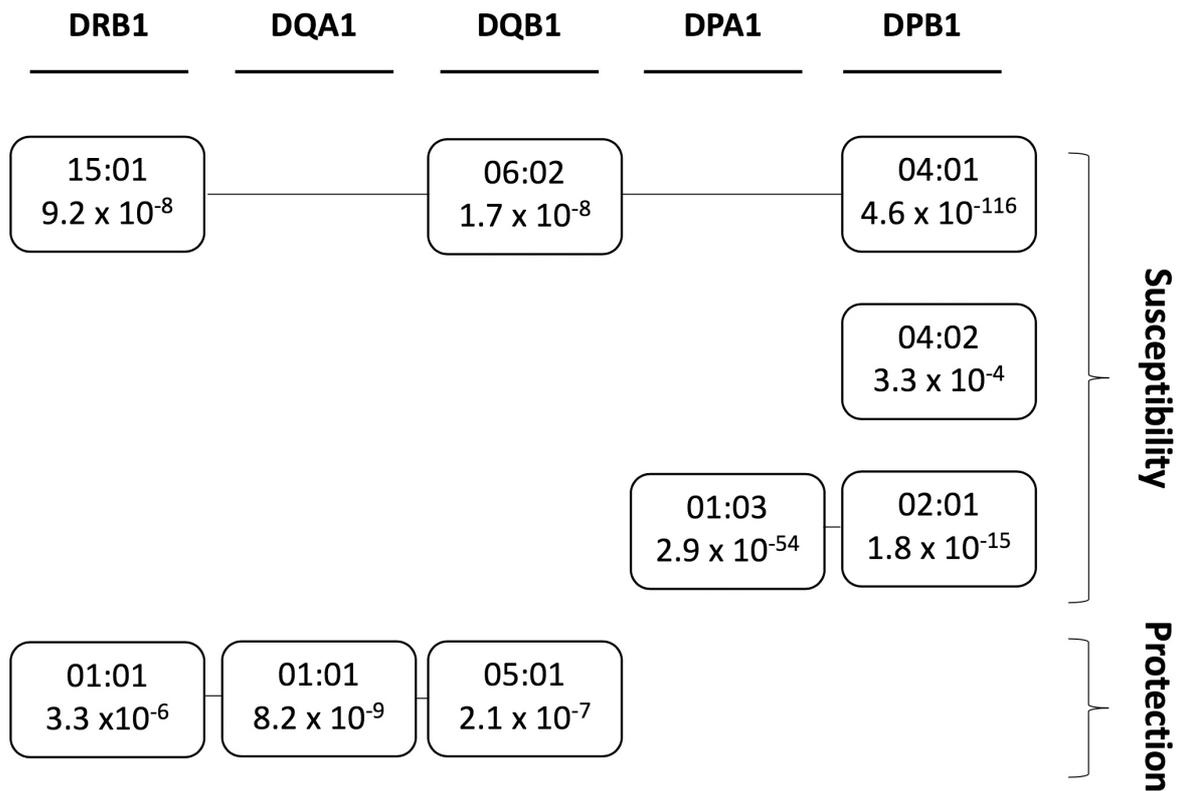
Figure 3.10: Associations of classical alleles with PR3-AAV.

Imputation of classical alleles identified four independent signals to be associated with PR3-AAV, three of which conferred susceptibility while one was protective.

3.8 Discussion

In this study, seven risk loci for PR3-AAV were identified. We replicated three previously identified genome-wide significant risk loci including *HLA-DP*, *SERPINA1* and *PRTN3*. Additional four risk loci were implicated in PR3-AAV for the first time. Within the classical HLA alleles, three independent signals were identified to confer disease risk, including *HLA-DPB1*04:01-HLA-DQB1*06:02-HLA-DRB1*15:01*, *HLA-DPB1*04:02* and *HLA-DPA1*01:03-HLA-DPB1*02:01*. The estimated SNP heritability (h^2_{SNP}) for PR3-AAV was modest (0.46 $\pm$ se 0.03), indicating that 46% of the variability in the trait in a population is due genetic differences among people. This estimate is largely consistent with other population-based heritability estimates such as Crohn's disease (0.55), type 2 diabetes (0.42) and it is lower than those estimated for T1D (0.80) and SLE (0.66) (396). This finding supports that PR3-AAV is a

polygenic disease and with increasing sample sizes, more genetic variants are expected to be identified.

Among the potential candidate susceptibility genes identified, *PRTN3* is well studied and its potential role in autoimmunity is supported by numerous functional studies as discussed in section 1.4.2. This locus was reported in previous GWAS by EVGC and VCRC (albeit with different lead SNPs). The association was limited to the GPA and more importantly, the association was stronger in the PR3-ANCA subset. Together with the eQTL and pQTL association signals where the disease risk allele is associated with a higher expression of *PRTN3* and a higher plasma level of the autoantigen PR3, these findings are consistent with the pathophysiologic relevance of PR3 serine protease to the phenotypes.

In addition, the second most significant risk allele at the *SERPINA1* locus was rs28929474:T (also known as the "Z" allele, which is in complete LD with the lead SNP, rs112635299). rs28929474 is a misssense variant (Glu342Lys) which results in defective secretion and intracellular accumulation of A1AT, leading to deficiency of circulating A1AT and eventually leading to respiratory complications such as emphysema and liver disease. Recent genome-wide plasma proteomic data by Sun et al had identified that the Z allele was associated with lower level of plasma A1AT. Despite using conservative significance threshold (p < 1.5 x 10^{-11}), it was found to be associated with 13 proteins (trans-pQTL), supporting that a single mutation can cause perturbation of a wide range of downstream proteins (319). Interestingly, one of these proteins was neutrophil cytosolic factor 2 (NCF2), which is involved in the neutrophil oxidative burst reactions. Mutations in *NCF2* can lead to chronic granulomatous disease which exhibits granulomatous inflammation, a pathognomonic pathological lesion with a predilection for the upper and lower respiratory tracts in patients with PR3-AAV. Furthermore, there were significant interactions between these SNPs (rs112635299 and rs28929474) and the promoters of several candidate target genes, some of which are involved in anti-bacterial (*SERPINA5*) and anti-inflammatory (*SERPINA12*) responses, altogether suggesting that there might be additional pathways contributing to the pathogenesis of disease.

These data show that the PR3-AAV risk allele at *PRTN3* increases *PRTN3* mRNA expression in the whole blood (318) and total PR3 plasma levels while the *SERPINA1* Z-allele causes a reduction in circulating A1AT levels. These risk alleles lead to an altered availability of PR3 autoantigen, together with HLA associated genetic predisposition (e.g. *HLA-DPB1*04:01*), increase the susceptibility for breaking immune tolerance and the production of anti-PR3 antibodies. Apart from in vitro experiments and animal models as well as clinical studies, the genetic evidence strongly supports that anti-PR3 antibodies are pathogenic rather than an epiphenomenon.

Infection has long been suggested as a trigger for induction and reactivation of different forms of vasculitides. Large-vessel vasculitis has been associated with *Mycobacterium tuberculosis* (Takayasu arteritis) and *Burkholderia bacterium* (GCA) while hepatitis B and C viruses have been related to PAN, which have a different clinical phenotype from non-viral PAN (397). The pathogenesis of GPA, where PR3-ANCA is the predominant serotype, has been linked to *S. aureus* which is a Gram-positive, commensal bacterium. It has consistently been shown that about 60-70% of GPA patients are chronic nasal carriers of *S. aureus* as compared to only 20-30% of general population (42, 398-400). Moreover, *S. aureus* nasal carriage is associated with an increased risk of relapse in GPA patients and prophylactic antimicrobial co-trimoxazole (trimethoprim/sulfamethoxazole) treatment reduces the risk of relapse (42, 231). A recent nasal microbiome study using deep sequencing methods in GPA patients demonstrated that GPA was associated with an altered microbial composition, at both the bacterial and fungal levels (401). However, the authors did not observe a difference in the relative abundance of *S. aureus* between GPA patients (n = 60, comprising 60% PR3-ANCA positive, 25% MPO-ANCA positive and 15% ANCA negative individuals) and healthy controls (n = 41), of which they attributed the discrepancy with prior studies to the differences in methodology including the location within the nasal cavity where the swabbing was performed. In contrast, a recent abstract by Kronbichler et al, where they investigated the changes of the nasal microbiome using culture, bacterial 16S rRNA profiling and shotgun metagenomics sequencing in GPA patients (both active and in remission), disease controls including MPA or EGPA patients as well as healthy controls, observed a different composition of *Staphylococcus spp.* between active GPA patients versus healthy controls and disease controls. They found that patients

with active GPA had a higher abundance of *S. aureus*, while healthy controls had a higher abundance of *S. epidermidis (402)*. However, the exact mechanism by which the pathogen triggers vasculitis remains unclear.

For most microorganisms, iron is an essential nutrient for colonization and pathogenesis. *S. aureus* has evolved sophisticated mechanisms to acquire iron from the host organism, known as the iron-regulated surface determinant (Isd) system, which is activated in response to iron depletion. It comprises a group of proteins that captures (IsdH and IsdB), transports (IsdA, IsdC and transmembrane complex IsdEF-FhuD) and metabolises (IsdD and IsdI) heme molecules to extract the iron atom. IsdB, which is an extracellular receptor, functions as a central component of removing heme from haemoglobin (403, 404). It contains two structurally conserved NEAr-iron Transporter (NEAT) domains, NEAT1 (Hb binding) and NEAT2 (heme binding), that function synergistically (405). The NEAT domains are also found in other Gram-positive human pathogens such as *Bacillus anthracis*, *Streptococcus pyogenes*, *Clostridium perfringens* and *Listeria monocytogenes* (406). It has been demonstrated that *S. aureus* strains lacking IsdB had a reduction in virulence due to inability to bind haemoglobin for iron acquisition in a murine model of abscess formation (407, 408). A recombinant anti-IsdB was isolated from the Cambridge Antibody Technology antibody library over a decade ago and it was shown to confer protection against *S. aureus* infection in a murine lethal sepsis model when dosed prior to challenge (409). Importantly, a high level of serum titre against IsdB had been observed in patients with *S. aureus* infection (410).

Yeung and colleagues recently demonstrated that there was a heavily biased use of two immunoglobulin heavy chain germlines generated antibodies that neutralise NEAT1 (*IGHV4-39*) and NEAT2 (*IGHV1-69*) domains respectively. Notably, they found that the antibody-antigen binding mechanism was nearly identical despite antibody derivation from different donors. Though being a part of the adaptive immune system, the authors postulated that both *IGHV1-69* and *IGHV3-49* might have evolved under selection pressure to encode binding motifs that are capable of recognising and neutralising structurally conserved protein domains, which are involved in iron acquisition across different pathogens. More importantly, the authors demonstrated that allelic variation at position 50 in *IGHV1-69* (G50R, the lead

SNP in chr14q32.33 locus) completely abolished NEAT2 binding presumably from steric hindrance, which led us to posit a potential host genetic factor in PR3-AAV patients that predisposes them to chronic *S. aureus* nasal carriage, where the risk allele results in a change of amino acid in the framework region 2 of the heavy chain leading to a reduction in their ability to neutralise NEAT2-mediated heme-iron acquisition. Furthermore, rs55891010 (F54L) which is in high LD with the lead SNP ($r2 = 0.93$) was previously shown to encode a functionally critical residue for broadly neutralising antibodies against the influenza HA stem (HV1-69-sBnAb) where the L-allele was associated with a weaker binding affinity to HA (411, 412) and individuals carrying L-alleles had consistently been shown to have lower levels of HV1-69-sBnAbs (387, 390, 393). Several case reports have described a temporal association of AAV to influenza vaccination (413, 414), which raises another possibility that influenza antigens might serve as a trigger for AAV induction in genetically predisposed individuals. Although the roles of Ig germ-line polymorphisms have not been extensively explored, there is a growing body of evidence to support that they can influence antibody repertoire and functional variation in protective antibody responses.

3.9   Conclusions

Through meta-analysis of two PR3-AAV European cohorts comprising 13,557 individuals, this study provides conclusive evidence that PR3-AAV is a polygenic disease. A majority of the candidate genes identified are involved in the immune pathway, offering new leads into the pathogenesis of PR3-AAV. Our study also presents a plausible host genetic factor (*IGHV1-69*; rs11845244:C>T; G50R) in determining the susceptibility to infectious disease and as a potential driver of PR3-AAV.

4 Genetic relationship of MPO-AAV and PR3-AAV

4.1 Introduction

The epidemiological and clinical manifestations of AAV differ geographically. The overall annual incidence of AAV was similar between Japan and the UK (36). However, the predominant subtype observed in Japan was MPA, accounting for 83% of patients and more than 80% were MPO-ANCA positive. A similar pattern was observed in China (415, 416) although population-based data is lacking. Previous GWAS in the European population have only identified one genome-wide significant locus within the HLA region (*HLA-DQ)* to be associated with MPO-AAV (70, 71) while genetic data from Asian countries are scarce, primarily candidate gene analyses and most genetic variants are not widely confirmed except the HLA alleles. The presence of *HLA-DRB1*09:01* and *HLA-DRB1*11:01* alleles were associated with MPO-AAV in both Japanese and Chinese populations respectively (68, 69). Although there is a substantial overlap in the clinicopathological phenotypes between MPO-AAV and PR3-AAV, the genetic relationship between the two syndromes has not previously been defined.

In order to discover additional risk loci associated with MPO-AAV, a total of 606 MPO-AAV cases from Europe, along with 6,688 population matched controls, were genotyped on the UK Biobank Axiom array. A meta-analysis was then performed with the AAV GWAS by Lyons et al (70).

In this chapter, I will discuss the identification of five risk loci associated with MPO-AAV including the HLA region. Within the HLA region, two independent signals were identified in a region encompassing the butyrophilin-like protein 2 (*BTNL2*) and *HLA-DQ*. Furthermore, the genome-wide genetic correlation (r_G) between MPO-AAV and PR3-AAV was estimated based on common variants of the genotyped data using a bivariate linear mixed model, confirming a high similarity in the genetic architecture between the two syndromes.

4.2 Samples and genotype data

Following sample and SNP QC, a total of 469,699 SNPs from GWAS1 and 558,830 SNPs from GWAS2 respectively were available for analysis. The breakdown of the number of patients included in the meta-analysis of MPO-AAV is summarised in Table 4.1 .

Table 4.1: The number of cases and controls included in the meta-analysis of MPO-AAV

Cohort	No of Controls	No of Cases
GWAS1	5,259	264
GWAS2	6,688	606
[†]Sequenom Genotype (rs6679677)	1,531	347
Total	13,478	1,217

†For rs6679677, a further cohort was available for meta-analysis where the genotyping was performed using the Sequenom MassARRAY platform in the replication cohort of Lyons et al 2012.

4.3 Association testing with genotype data

I performed association testing of the genotype data using two analytic methods – logistic regression and LMM as implemented in BOLT-LMM. A comparison was made between the two methods and summarised in Table 4.2. There was a minimal genomic inflation observed in GWAS1 (λgc = 1.01) as all the cases and controls were from the UK. In contrast, the λgc for GWAS2 was reduced from 1.242 to 1.094 after fitting 20 PCs as covariates in a linear regression model. Similarly, the LMM method reduced the λgc to 1.094, supporting that population stratification in GWAS2 was well controlled for (Figure 4.1).

Table 4.2: A comparison of logistic regression and linear mixed model analyses

Genotyped data	MPO-AAV (lambda, λ)	
	GWAS1	GWAS2
No covariates	1.014	1.242
3 PCs	1.015	1.121
6 PCs	1.014	1.099
20 PCs	1.012	1.094
LMM	0.999	1.094
*LMM (imputed data)	0.999	1.073

LMM, linear mixed model; PCs, principal components.

Lambda, λ is the ratio of the median of the observed chi-squared statistics to the median of the expected chi-squared statistics under the null hypothesis.

* indicates genomic inflation factor calculated using imputed data.
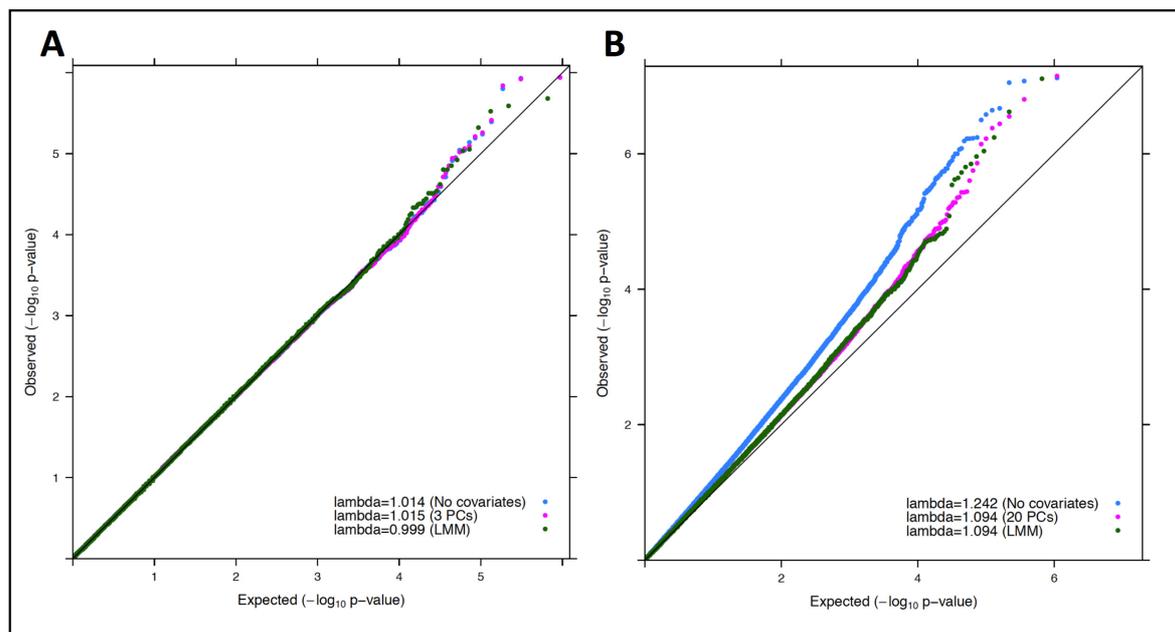
Figure 4.1: Quantile-quantile plots and genomic inflation factors.

Association tests were performed on genotyped data and comparison between logistic regression with no covariates, varying numbers of principal components (PCs) as covariates and LMM for MPO-AAV in (A) GWAS1 and (B) GWAS2 respectively. The HLA region and SNPs with genome-wide significant associations were not included in the QQ plots. The black line is y = x. Additional PCs as covariates did not alter the genomic inflation factors.

4.4 Association testing of the imputed data and meta-analysis

Following QC of the imputed data, a total of 7,656,576 and 7,744,534 autosomal markers in GWAS1 and GWAS2 respectively with MAF ≥ 0.01 and r2 ≥ 0.3 were available for downstream analyses. Of these, 7,576,565 SNPs were in common to both cohorts for meta-analysis. Case-control association testing was performed using the LMM model to minimise the effect of population stratification and meta-analysis was performed using an inverse-variance method based on a fixed effects model.

A total of five MPO-AAV susceptibility loci were identified (Table 4.3). In three of the five risk loci identified, the most significant SNP within each locus was a common variant (MAF > 0.05) of modest effect where ORs ranged between 1.37-1.82. The strongest association was observed at the HLA region (rs116522341, MAF = 0.04) where the OR was 4.80 (95% CI 3.67-6.26, p_{meta} = 1.09 x 10^{-30}). The lead SNP (rs79005509) in the remaining risk locus (chr6q15) was a low- frequency variant (MAF = 0.02) with an OR of 2.90 (95% CI 2.00-4.20, p_{meta} = 1.90 x10^{-8}). Subsequent stepwise conditional regression analyses within each non-HLA locus did not identify additional independent genome-wide significant signals.

For *PTPN22* (rs6679677), a further replication cohort that was previously genotyped using the Sequenom MassARRAY platform was included for combined analysis, consisting of 347 MPO-AAV cases and 1,531 controls. Following the second-stage meta-analysis, rs6679677:A was associated with an increased risk of MPO-AAV with an OR of 1.47 (95% CI 1.29-1.67, p_{meta} = 5.50 x 10^{-9}).

4.5 Genetic heritability

To quantify the genetic influence on MPO-AAV, the total SNP heritability (h^2_{SNP}) was estimated using the genome-wide genotyped variants (MAF ≥ 0.01) and additively, they explained 33% of the total disease liability (se = 0.02), for a population prevalence of 50 per million population (338). This was slightly lower than PR3-AAV (h^2_{SNP} = 0.46, se = 0.03).

Table 4.3: HLA and non-HLA associations with MPO-AAV.

[§]SNP	[¶]Chr	[¶]Position	Gene(s)	[*]Risk allele	RAF	GWAS1 (264 cases vs 5,259 controls)		GWAS2 (606 cases vs 6,688 controls)		Meta-analysis (870 cases vs 11,947 controls)	
						[†]OR (95% CI)	[Ψ]P	[†]OR (95% CI)	[Ψ]P	[†]OR (95% CI)	[Δ]P
[‡]rs2476601	1p13.2	114,377,568	*PTPN22*	A	0.10	1.52 (1.13-2.05)	5.50×10^{-3}	1.52 (1.27-1.83)	1.40×10^{-6}	1.49 (1.28-1.74)	3.13×10^{-7}
[‡]rs6679677	1p13.2	114,303,808	*PTPN22*	A	0.10	1.50 (1.13-2.01)	5.70×10^{-3}	1.52 (1.27-1.83)	1.30×10^{-6}	1.48 (1.27-1.73)	3.51×10^{-7}
rs115670873	2p13.1	74,171,061	*DGUOK, TET3*	A	0.93	1.97 (1.05-3.68)	3.50×10^{-2}	1.69 (1.38-2.08)	8.50×10^{-7}	1.82 (1.46-2.27)	7.60×10^{-8}
rs116522341	6p21.32	32,367,697	*BTNL2, HLA-DQ*	G	0.04	4.80 (2.80-8.23)	1.10×10^{-8}	5.11 (3.76-6.94)	6.00×10^{-25}	4.80 (3.67-6.26)	1.09×10^{-30}
rs79005509	6q15	90,850,730	*BACH2*	T	0.02	7.28 (3.21-16.52)	2.00×10^{-6}	1.98 (1.33-2.95)	8.90×10^{-4}	2.90 (2.00-4.20)	1.90×10^{-8}
rs2965955	16q24.3	89,524,923	*ANKRD11, SPG7*	C	0.69	1.40 (1.15-1.69)	5.80×10^{-4}	1.39 (1.23-1.58)	7.70×10^{-8}	1.37 (1.24-1.51)	9.64×10^{-10}

[§] SNPs showing the most significant level of association after meta-analysis of GWAS1 and GWAS2.

[¶] Genomic coordinates from Genome Reference Consortium Human Build 37 (GRCh37/hg19); Chr, chromosome.

[*] The risk allele refers to the effect in the overall meta-analysis. RAF, risk allele frequency.

[†] The odds ratio (OR) is with respect to the risk allele. CI, confidence interval.

[Ψ] The p-values were calculated using a linear mixed model (BOLT-LMM).

[Δ] The p-values were calculated using inverse-variance method based on a fixed-effects model.

[‡] rs2476601 and rs6679677 are the lead SNPs at chr1p13.2. A further replication cohort was available for rs6679677 for second-stage meta-analysis consisting of 347 MPO-AAV cases and 1,531 controls, resulting in a final $p_{meta} = 5.50 \times 10^{-9}$.

4.6 Candidate gene prioritisation (non-HLA loci)

To prioritise candidate genes, I first performed fine mapping of all the risk loci identified using the Bayesian approach to generate credible sets of SNPs. I then searched for nonsynonymous coding and known eQTL among the lead SNPs or SNPs in high LD with them (r2 > 0.6). Risk loci were functionally annotated using multiple publicly available datasets, similar to PR3-AAV (see section 3.6). The results are summarised in Table 4.4.

4.6.1 PTPN22

The two most significant associations that were observed in chr1p13.2 were rs2476601 and rs6679677. They are in complete LD (r2 = 1.0) and as expected, the effect sizes for these two variants were equivalent. rs2476601 is a missense variant (R620W) which has been studied quite extensively over the last two decades. Its potential role in autoimmunity is supported by several functional studies as discussed in section 1.4.2. It has been reported to be associated with a wide range of autoimmune diseases including RA (90, 417), T1D (91-93), thyroid diseases (95, 96), Crohn's disease (131), alopecia areata (418), vitiligo (374) and SLE (129). Importantly, this variant showed the same direction of effect (i.e. same allele associated with higher risk of MPO-AAV and other autoimmune diseases). It was also reported in the VCRC GWAS (71), although the p-value for the MPO-AAV subset fell just below the genome-wide significance threshold. Furthermore, the PCHi-C data showed interactions of this SNP with multiple gene promoters in total B cell, CD4 and CD8 T cells (see Table 4.4 and Figure 4.2) and notably, three of the new plausible candidate genes (*TRIM33*, *HIPK1* and *NRAS*) might be of biological relevance to the pathogenesis of disease.

TRIM33 (tripartite motif-containing 33), also known as transcriptional intermediary factor 1-γ, which is a modulator of transforming growth factor-β (TGF-β) signalling that associates with SMAD family member 2 (SMAD2), was recently demonstrated to play a critical role in the regulation of proinflammatory function of Th17 cells. *TRIM33* deficient T cells ameliorated autoimmune phenotype in the experimental autoimmune encephalomyelitis mouse model by upregulating the expression of IL-10 and suppressing the production of IL-17 (419). Several studies have indicated that Th17 cells might play a role in AVV pathogenesis. Expansion of

Th17 cells was found in both active and quiescent GPA patients (420). In addition, higher serum levels of IL-17A and IL-23 were observed in acute AAV patients as compared to healthy controls, and levels remained elevated in a proportion of patients despite disease remission (421). IL-17A deficient mice also displayed substantially less renal injury in the experimental autoimmune anti-MPO vasculitis model (422).

HIPK1 (homeodomain-interacting protein kinase-1) is one of the four highly related serine/threonine kinases identified as corepressors for several homeodomain-containing transcription factors and TNF-mediated cellular apoptosis (423). *HIPK1* is ubiquitously expressed and relatively abundant in the naïve B cells, CD8 T cells, neutrophils and monocytes. HIPK1 was shown to play an essential role in splenic B cell homeostasis and function where *HIPK1*$^{-/-}$ mice exhibited a reduction of total B cells in the spleen, especially transitional-1 and follicular B cell populations and the humoral response was impaired despite an expansion of marginal zone B cell population (424).

NRAS (also known as NRAS proto-oncogene, GTPase) encodes N-Ras which is involved in regulating cell division. Mutations in *NRAS* causes autoimmune lymphoproliferative syndrome, which is characterised by an expansion of non-malignant lymphocytes due to defective apoptosis, an increased risk of autoimmunity and haematological malignancies (425). Similar phenotypes are observed in the mutant mouse strain (obtained from mouse genome informatics http://www.informatics.jax.org).

4.6.2 DGUOK/TET3

The lead SNP (rs115670873) in this locus was the same as PR3-AAV, with equivalent effect size and direction, although its p-value was just below the threshold for genome-wide significance (p_{meta} = 7.60 x 10^{-8}). See 3.6.1 for further details about this region.

Table 4.4: Candidate genes at MPO-AAV associated risk loci

SNP	Most likely causal gene(s)	Type of variants	Diseases or traits associated with the lead SNP or in high LD with lead SNP (r2 > 0.6) or in proximity to candidate gene(s)		*Cis-eQTL (blood/relevant tissues)	‡Immune phenotype in murine model	PCHi-C prioritised protein coding gene(s)	PCHi-C prioritised non-protein coding transcript(s)
			Disease or trait (SNP)	Relationship				
rs2476601	*PTPN22 (R620W)*	Missense	RA, RA (ACPA-positive) (90, 417)	Same	*PTPN22, DCLRE1B, PHTF1, AP4B1-AS1, AP4B1*	*PTPN22, NRAS*	*PTPN22, NRAS, CSDE1, TRIM33, HIPK1, AP4B1, DCLRE1B, OLFML3, BCL2L15, RHOC, RHOC*	*RP5-1073O3.7, AP4B1-AS1*
			T1D (91-93)					
			Hypothyroidism, Grave's disease (95, 96)					
			Crohn's disease (131)					
			Alopecia areata (418), vitiligo (374)					
			SLE (129)					
			JIA (130), AS (99) (rs6679677)	r2 = 1.0				
rs79005509	*BACH2*	Intronic	Eosinophil counts (346)	In proximity	*BACH2*	*BACH2*	*BACH2*	*AL391559.1, snoU13*
			T1D (426) , coeliac disease (97)					
			Crohn's disease (427)					
			Vitiligo (374), SLE (369)					
			MS (428)					
rs2965955	*ANKRD11 SPG7*	Intronic	Lymphocyte count (346)	In proximity	*ANKRD11, RPL13, MC1R, SPG7, SPATA2L, DBNDD1, SPIRE2, ZNF276, FANCA, AC137932.6, AC137932.5, TRAPPC2L, VPS9D1, TCF25, CHMP1A* (whole blood); †*SPG7* (renal tubule); *RP11-368I7.4* (testis); *CPNE7* (prostate); *SPG7* (muscles); *CHMP1A* (thyroid); *SPIRE2* (skin); *FANCA* (nerve)	*FANCA*	*ANKRD11, CDK10, SPG7, PIEZO1, MVD*	*RP11-368I7.4*
			CD4 cells (%Th22) (429)					
			Multiple myeloma (IgH translocation) (430)					

ACPA, anti-citrullinated protein antibody; AS, ankylosing spondylitis; eQTL, expression quantitative trait loci; JIA, juvenile idiopathic arthritis; MS, multiple sclerosis; PCHi-C, promoter capture Hi-C data; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; %, percentage.

*The lead SNPs or SNPs in high LD with them that are known eQTLs (obtained from GTEx and http://www.eqtlgen.org/index.html) (317, 318).

‡The immune phenotype designation was taken from http://www.informatics.jax.org/phenotypes.shtml.

†eQTL in renal tubule was obtained from human kidney eQTL atlas (http://susztaklab.com/eqtl) (431).
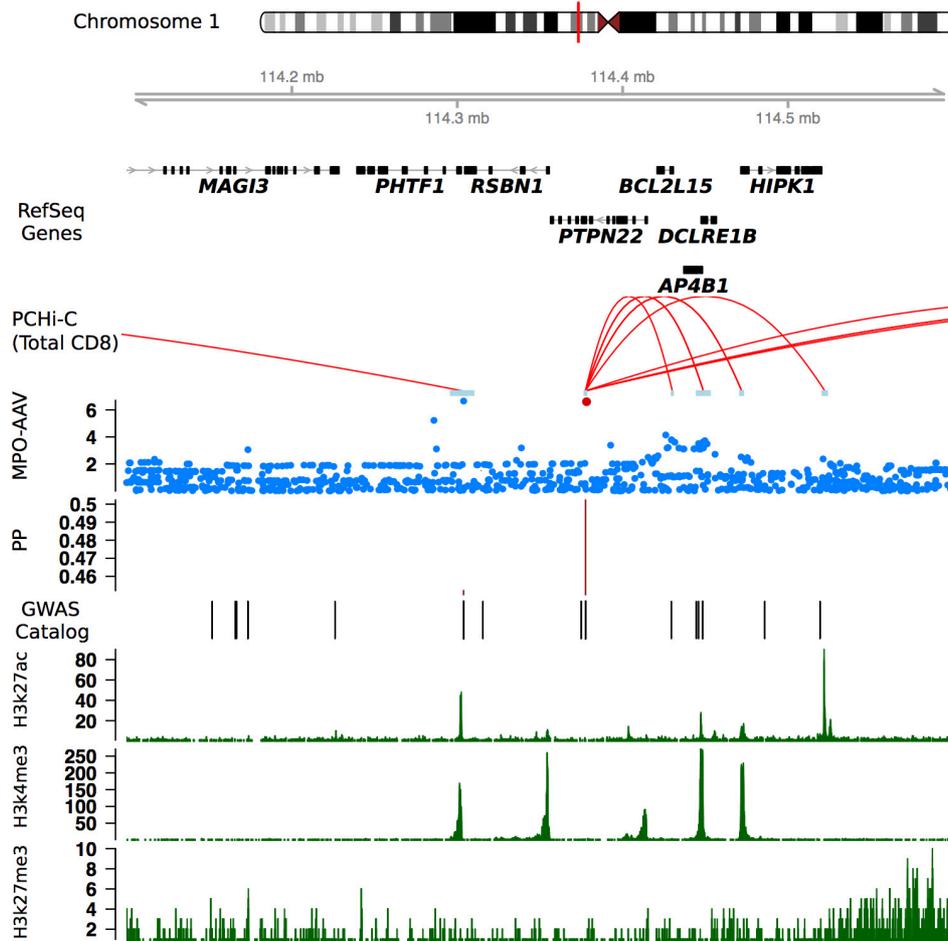
Figure 4.2: Genomic features at chr1p13.2 (*PTPN22* region).

Top to bottom panels: Genomic position (hg19), representative RefSeq genes, PCHi-C data (Javierre et al – total CD8), SNPs association with MPO-AAV where the filled red circle indicates the lead SNP (rs2476601), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI catalog of published GWAS and methylation data.

4.6.3   BACH2

The two most significant associations in chr6q15, rs79005509 and rs78275221 (p_{meta} = 1.90 x 10^{-8} and p_{meta} = 7.87 x 10^{-8} respectively, r2 = 0.88) are located in the intron of BTB domain and CNC homolog 2 (*BACH2*). It encodes a leucine zipper transcription regulator. Although neither of these variants has been reported to be associated with any trait or disease, this locus has independently been linked to multiple inflammatory diseases including T1D (426), coeliac disease (97), Crohn's disease (427), vitiligo (374), SLE (369) and MS (428). In addition, both rs79005509 and rs78275221 are known eQTLs where the risk alleles for MPO-AAV are

associated with reduced expression of *BACH2* in the whole blood (p = 2.59 x 10^{-21} and Z-score = -9.48 for rs79005509; p = 4.44 x 10^{-24} and Z-score = -10.12 for rs78275221) (318). The PCHi-C data showed significant interactions between these SNPs with the promoter of *BACH2* in CD4 and CD8 T cells (Figure 4.3).

BACH2 is most abundantly expressed in the bone marrow and immune system with the highest expression found in germinal centre B cells, followed by naïve B cells, CD4 T cells, class switched memory B cells and CD8 T cells (BLUEPRINT data release 2016-08). It acts as a repressive transcription regulator, maintaining the balance of a network of all other transcription factors which are critical for B and T-cell differentiation and maturation. Its role in B-cell development has been studied greatly where it represses the expression of Blimp-1 (a.k.a. PR domain zinc finger 1) (432), to decelerate plasma cell differentiation and permit antibody class switch recombination (433-435). In T cells, *BACH2* modulates networks of genes that control T cell effector lineages (316), balancing differentiation into effector cells and development of FoxP3+ regulatory T cells (T$_{reg}$) (436). *BACH2*$^{-/-}$ mice exhibit impaired B-cell differentiation in the marginal and follicular zones of the spleen, impaired somatic hypermutation with decreased number of mature B cells and absence of germinal center B cells. These mice also have low number T$_{reg}$ cells and an excess number of effector T cells, leading to autoimmunity and death prematurely between 3-9 months of age (436). Afzali et al recently described a syndrome called BACH2-related immunodeficiency and autoimmunity (BRIDA) resulting from *BACH2* haploinsufficiency, where patients displayed reduced T$_{reg}$ cells, enhanced Th1 differentiation, defective B-cell proliferation and maturation, leading to profound Ig deficiency and colitis requiring intravenous Ig treatment (437).
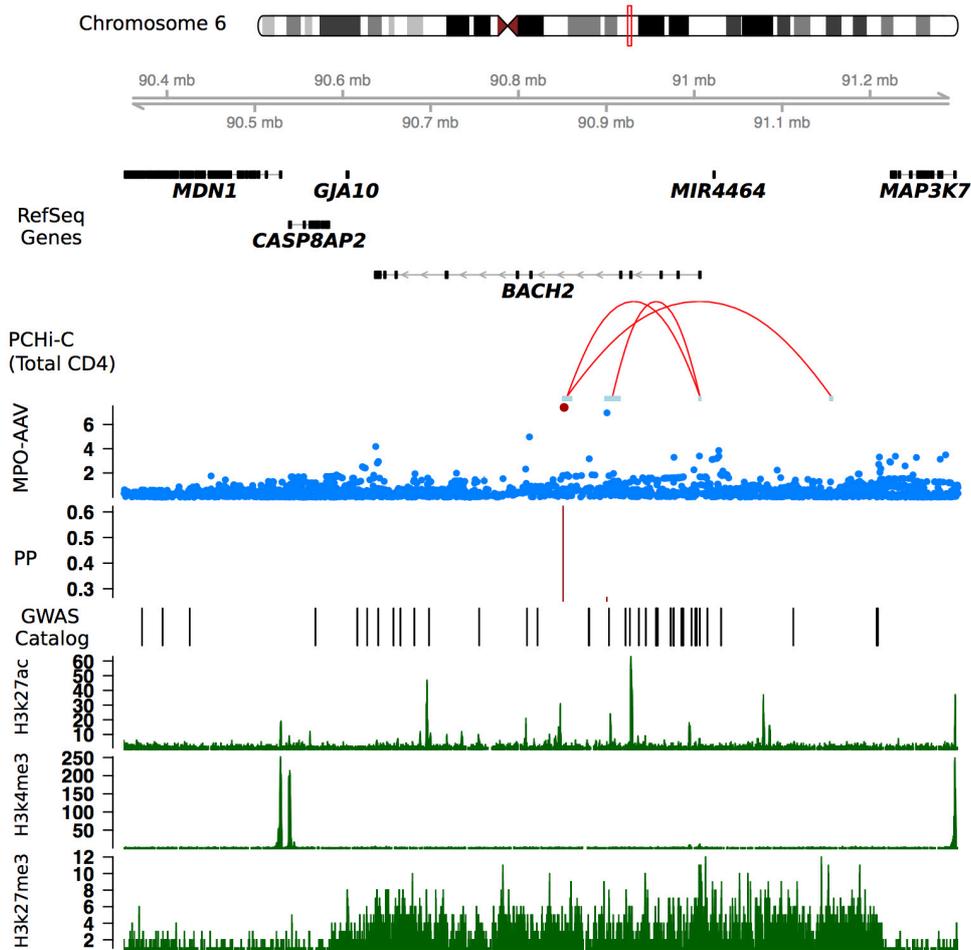
Figure 4.3: Genomic feature at chr6q15 (*BACH2* region).

Top to bottom panels: Genomic position (hg19), representative RefSeq genes, PCHi-C data (Javierre et al – total CD4), SNPs association with MPO-AAV where the filled red circle indicates the lead SNP (rs79005509), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI catalog of published GWAS and methylation data.

4.6.4   ANKRD11/SPG7

The top association signal in chr16q2.3, rs2965955 (p_{meta} = 9.64 x 10^{-10}), is located in the intron of ankyrin repeat domain 11 (*ANKRD11*). More than half of the variants within the credible set are located in the region with evidence of chromatin signatures of active enhancers (enrichment for histone H3K27ac, Figure 4.4). None of these variants has been reported to be associated with other traits or diseases, but this locus has been linked to immune cell frequencies including lymphocyte count (346) and T helper type 22 cells (429) and multiple myeloma (430). rs2965955 is a known eQTL for 16 genes in the whole blood and a wide range

of other tissue types including renal tubules, muscle, skin, nerve, testes, prostate and thyroid (Table 4.4) (317, 318, 431). These genes are enriched for different immune related pathways (Figure 4.5). The PCHi-C data showed evidence of interactions between these potential variants with promoters of several candidate genes in GM12878 cell line.

ANKRD11 encodes a protein that contains multiple ankyrin domains, which allows its interaction with other proteins such as histone deacetylases (chromatin modifying enzymes) to modulate gene expression. For example, it has been shown to inhibit ligand-dependent transactivation by recruiting histone deacetylases to p160 coactivators/nuclear receptor complex (438). Its critical role in regulating transcription might explain the high number of eQTLs observed in this region. It is ubiquitously expressed, being most abundant in the cerebral cortex, skin, bone marrow, spleen, testis and ovary. Mutations in *ANKRD11* causes KBG and 16q24.3 microdeletion syndromes which are characterised by abnormal craniofacial features and intellectual disability with a higher incidence of autism spectrum disorders in the latter, but immune related phenotypes have not been reported to date.

SPG7 encodes for paraplegin, a mitochondrial metalloprotease protein that is a member of the AAA (ATPases associated with diverse cellular activities) protein family, which regulates ribosome assembly and degrades misfolded proteins (439). It is ubiquitously expressed and mutations in *SPG7* causes spastic paraplegia, a neurodegenerative disease characterised by a slowly progressive weakness and spasticity in the lower limbs. Notably, the risk allele for MPO-AAV (rs2965955:C) has opposing directions of effect on *SPG7* expression in whole blood and renal tubules, where a higher expression is found in the former (318) and a lower expression in the latter (431). This observation is not uncommon and eQTLs with contrasting effects have been demonstrated in different immune cell types and in different disease states, and is postulated to have complex downstream implications on immunity (440). Furthermore, *SPG7* has been inferred to be associated with glomerulonephritis and nephrotic syndrome with moderately high confidence scores, using the DISEASES database from the University of Copenhagen (http://diseases.jensenlab.org).
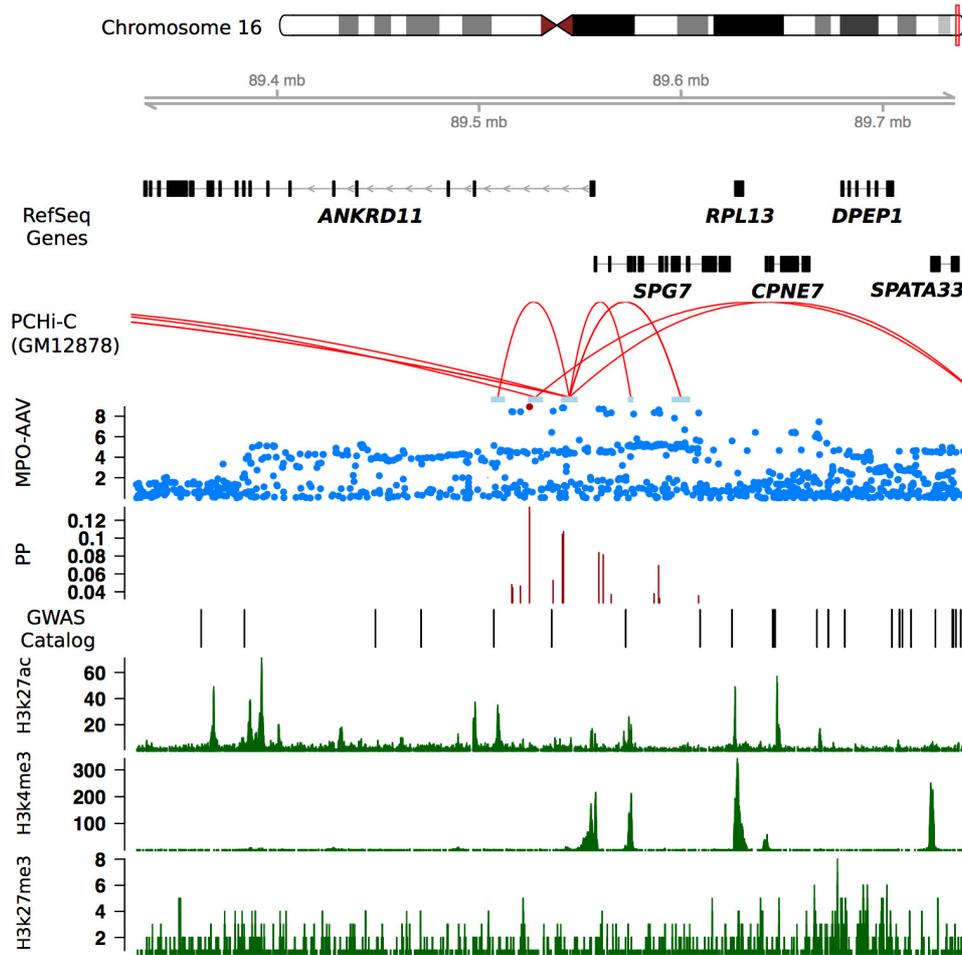
Figure 4.4: Genomic features at chr16q24.3 (*ANKRD11-SPG7* region).

Top to bottom panels: Genomic position (hg19), representative RefSeq genes, PCHi-C data (Mifsud et al – GM12878 cell line), SNPs association with MPO-AAV where the filled red circle indicates the lead SNP (rs2965955), causal variant mapping expressed as posterior probability (PP), GWAS significant hits in the region obtained from the NHGRI-EBI catalog of published GWAS and methylation data.
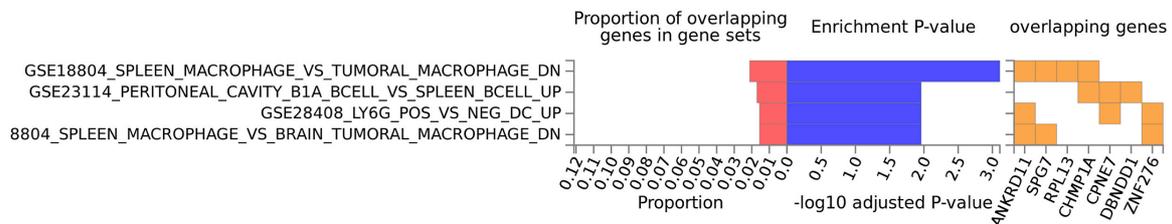
Figure 4.5: Gene-set enrichment in immune related pathways.

The lead SNP and those in high LD with it are known eQTLs for large number of genes, which are significantly enriched in different immune related pathways. The diagram is generated using FUMA.

4.7   Fine-mapping of the HLA region

Following a meta-analysis of GWAS1 and GWAS2, there were two peaks showing the strongest associations in the MHC region, spanning approximately 300 kb region and including *BTNL2*, *HLA-DRA*, *HLA-DRB5*, *HLA-DRB6*, *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* (Figure 4.6). The most significant association for each peak was rs116522341 and rs9275319 respectively. rs116522341 is an intronic SNP at butyrophilin-like protein 2 (*BTNL2*) that is part of the extended region of HLA class III and it has a stronger signal (p_{meta} = 1.09 x 10^{-30}). *BTNL2* encodes a HLA, class II associated, type 1 transmembrane protein which is involved in immune surveillance, acting as a negative T-cell regulator by inhibiting T cell proliferation and TCR activation (441). This variant is implicated in a number of autoimmune diseases including RA, UC, PSC, PBC and asthma (350). rs9275319 is an intergenic SNP at the *HLA-DQ* loci (p_{meta} = 1.78 x 10^{-22}), which is consistent with the previous GWAS finding albeit with a different lead SNP.
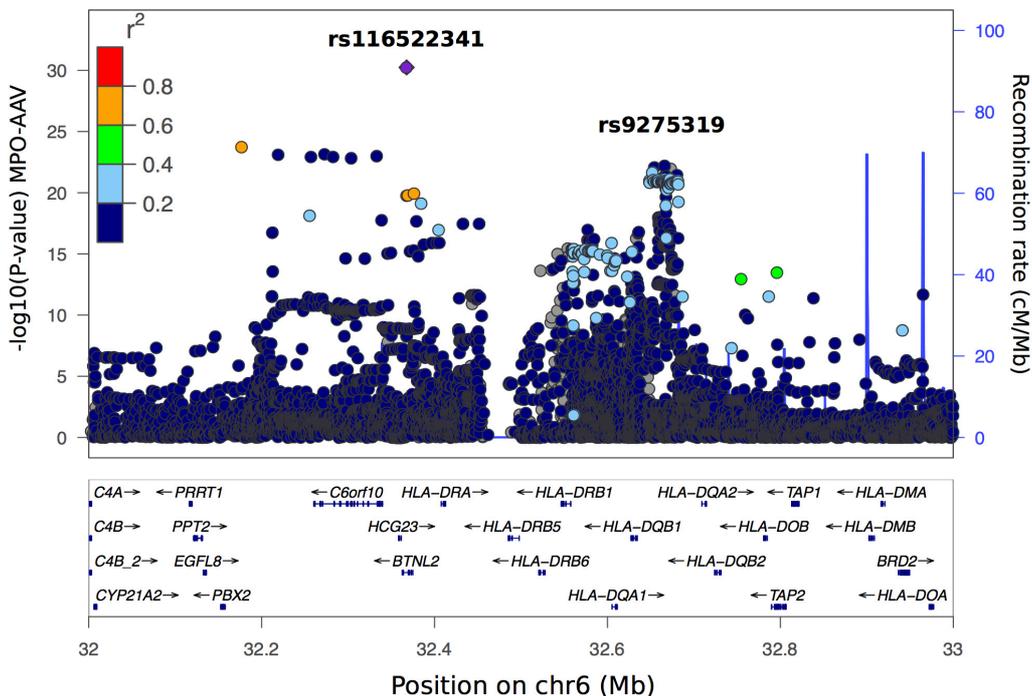
Figure 4.6: Regional plot of the HLA class II/III of chromosome 6.

Meta-analysis of the association analyses of GWAS1 and GWAS2 for MPO-AAV. Two separate peaks were observed at chr6p21.32 where the strongest association was rs116522341, indicated by the purple diamond and the lead SNP for the other peak region was rs9275319.

I performed a stepwise conditional association analysis in this region to confirm the independent signals and since the HLA imputation was performed on the GWAS2 dataset, I focused on the same. The lead SNP for the weaker cluster in GWAS2 MPO-AAV was rs9275318, which is in complete LD with rs9275319 ($r2 = 1.0$). Upon conditioning on rs116522341, there was indeed an additional genome-wide significant association – rs3104415 which was part of the weaker cluster (OR = 1.42, conditional $p = 1.80 \times 10^{-8}$) while rs9275318 did not achieve significant association (OR = 1.45, conditional $p = 8.00 \times 10^{-5}$, Table 4.5). Reciprocal conditional analysis on rs3104415 demonstrated that rs116522341 remained genome-wide significant, confirming that there were two independent signals in the HLA region for MPO-AAV and lastly, conditioning on both rs116522341 and rs3104415 accounted for the entire signal seen in this region.

Table 4.5: Stepwise conditional analysis of the HLA region in MPO-AAV (GWAS2)

SNPs (*Chr:Position)	Gene(s)	Type	†Risk allele (RAF)	Unconditioned ‡p value (OR)	Conditioned on		
					rs116522341	rs3104415	rs116522341 rs3104415
rs116522341 (6:32,367,697)	*BTNL2*	Intronic	G (0.037)	6.00×10^{-25} (5.11)	-	1.00×10^{-17} (4.04)	-
rs9275318 (6:32,666,173)	*HLA-DQ*	Intergenic	G (0.137)	6.60×10^{-17} (2.04)	8.00×10^{-5} (1.45)	6.50×10^{-7} (1.64)	0.14 (1.17)
rs3104415 (6:32,582,577)	*HLA-DQ*	Regulatory region	C (0.355)	1.00×10^{-15} (1.64)	1.80×10^{-8} (1.42)	-	-

* Genome Reference Consortium Human Build 37 (GRCh37).

† The risk allele refers to the effect allele. RAF, risk allele frequency.

‡ The p-values were calculated using a linear mixed model with BOLT-LMM. OR, odds ratio with respect to the risk allele.

The classical HLA alleles at 2- or 4-digit resolution and amino acid variants at eight HLA loci were imputed. Univariate LMM analysis identified three HLA alleles conferring susceptibility to MPO-AAV. Forward stepwise iterative conditional analysis confirmed one signal conferred by an extended haplotype encoding *HLA-DRB1*04:04-HLA-DQA1*03:01-HLA-DQB1*03:02* (Table 4.6), where the strongest independent association with disease risk was observed at *HLA-DRB1*04:04* (OR = 4.79, $p = 3.10 \times 10^{-30}$). Furthermore, *HLA-DRB1*04:04* was a stronger signal than the top hit SNP, supporting that the main signal in the region is indeed driven by the HLA allele. Individual amino acid variants in *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* were

associated with disease. Conditioning on the most significant amino acid variant (absence of Valine at position 180 in *HLA-DRB1*; OR = 1.77, p = 2.20 x 10[-16]) eliminated most of the effects at all other amino acid variants and all the classical HLA alleles except *HLA-DRB1*04:04* (OR = 3.79, conditional p = 2.60 x 10[-18]). Conditioning on *HLA-DRB1*04:04* accounted for the entire signal observed in the HLA region (Figure 4.7).

Table 4.6: Association of classical HLA alleles with MPO-AAV.

MHC Allele	Unconditioned		Conditioned on					
			HLA-DRB1*04:04		*HLA-DQB1*03:02*		*HLA-DQA1*03:01*	
	OR	P	OR	P	OR	p	OR	P
HLA-DRB1*04:04	4.79	3.10 x 10[-30]	-	-	4.00	8.70 x 10[-16]	3.99	1.70 x 10[-19]
HLA-DQB1*03:02	2.14	4.20 x 10[-17]	1.26	0.045	-	-	1.74	1.70 x 10[-5]
HLA-DQA1*03:01	1.71	3.50 x 10[-14]	1.26	0.0036	1.26	0.019	-	-
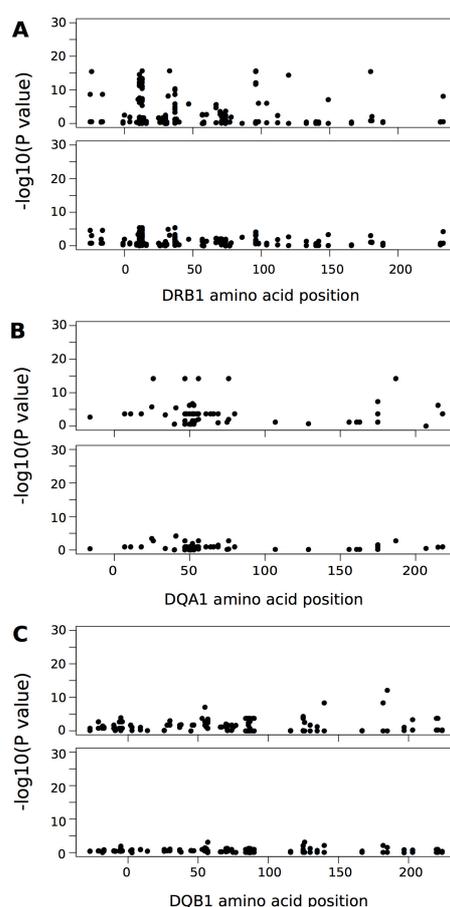
Figure 4.7: Amino acid positions in *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1*.

Amino acid positions in (A) *HLA-DRB1*, (B) *HLA-DQA1* and (C) *HLA-DQB1* associated with susceptibility to MPO-AAV (upper panels) and following conditioning on *HLA-DRB1*04:04* (lower panels).

4.8   The genetic relationship between MPO-AAV and PR3-AAV

To address the genetic relationship between MPO-AAV and PR3-AAV, I performed a combined AAV analysis, including both MPO-AAV and PR3-AAV cases to increase the power to detect potential shared susceptibility loci. Next, a comparison of the strength of the association signals between MPO-AAV and PR3-AAV as well as the direction of effects was made. A signal that appears specific to one group might reflect lack of power to detect it in the other and to formally address this issue, I used a Bayesian test of colocalisation to identify loci with strong evidence of either shared or independent causal variants between MPO-AAV and PR3-AAV across each of the ten non-HLA risk loci. Furthermore, I compared the genotypes of MPO-AAV and PR3-AAV samples directly with one another independent of controls ("within-case" analysis) and computed p-values for the 12 SNPs that declared hits in the primary meta-analysis. A significant difference in allelic frequency between the two traits at a given SNP supports a ANCA specific effect on disease susceptibility.

4.8.1   Combined AAV meta-analysis

The number of patients included in the meta-analysis of AAV is summarised in Table 4.7.

Table 4.7: The number of cases and controls included in the meta-analysis of AAV

Cohort	No of Controls	No of Cases
GWAS1	5,259	914
GWAS2	6,688	1,738
†Sequenom Genotype (rs6679677)	1,531	1,550
Total	13,478	4,202

†For rs6679677, a further cohort was available for meta-analysis where the genotyping was performed using the Sequenom MassARRAY platform in the replication cohort of Lyons et al (70).

Following QC of the imputed data, over 7.4 million SNPs were in common to both GWAS1 and GWAS2 for meta-analysis. Case-control association was performed using a LMM to account for population stratification and meta-analysis was performed using an inverse-variance method based on a fixed-effects model. An additional genome-wide significant association

was identified at chr2q33.2, rs1863800 which is located at the upstream of *CTLA4* and downstream of *CD28* (OR 1.19, 95% CI 1.12-1.26 and p_{meta} = 1.39 x 10^{-9}). This variant and those with high LD it (r2 > 0.6) have been implicated in Graves' disease, RA, T1D, alopecia areata and coeliac disease (350). Furthermore, rs1863800 is an eQTL where the risk allele for AAV is associated with a lower expression of *CTLA4* (p = 2.37 x 10^{-70} and Z-score = -17.73) but a higher expression of inducible T-cell costimulator (*ICOS*) (p = 2.41 x 10^{-8} and Z-score = 5.58) in the whole blood (318). Although it is an intergenic variant, the PCHi-C data showed evidence of interactions between this SNP with *CTLA4*, *CD28* and *RNU6-474P* (Figure 4.8). Genetic association of *CTLA4* with AAV was investigated extensively in the past using primarily candidate gene approach with conflicting results. We report for the first time, its genome-wide significant association with AAV in the present study. Further details about this region are discussed in section 1.4.2.
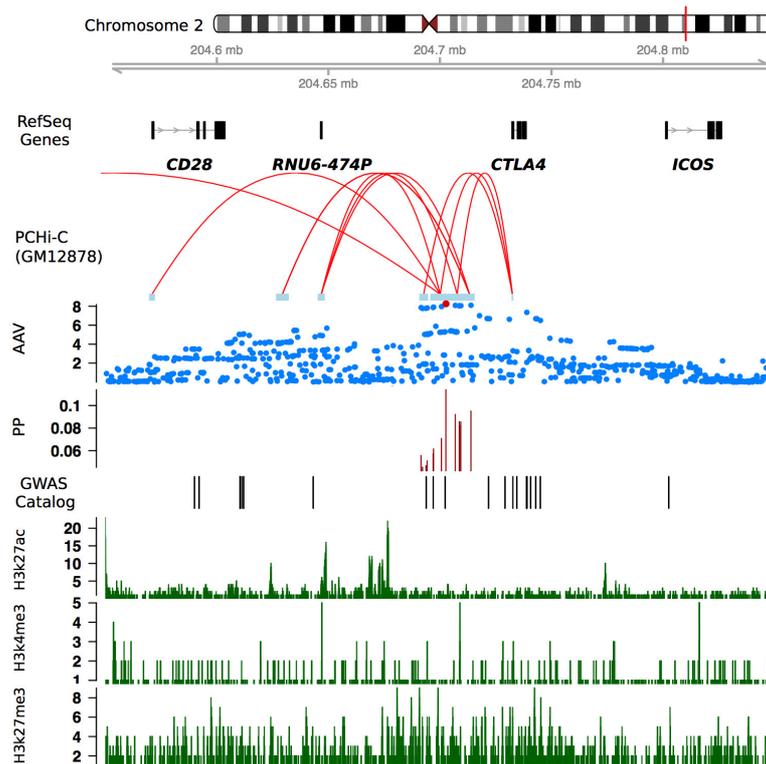
Figure 4.8: Genomic features at chr2q33.2 (*CTLA4* region).

Top to bottom panels: Genomic positions (hg19), representative RefSeq gene, PCHi-C data (Mifsud et al – GM12878 cell line), SNPs association with AAV where the filled red circle indicates the lead SNP (rs1863800), causal variant mapping expressed as posterior probability (PP), other GWAS significant hits in the region obtained from the NHGRI-EBI Catalog of published GWAS and methylation data.

Table 4.8: A comparison between combined AAV meta-analysis with PR3-AAV and MPO-AAV

SNP	Candidate gene(s)	Chr: Position (hg19)	•Risk allele (RAF)	AAV vs Control (2,652 vs 11,947) OR (95% CI)	P	PR3-AAV vs Control (1,610 vs 11,947) OR (95% CI)	P	MPO-AAV vs Control (870 vs 11,947) OR (95% CI)	P	PR3-AAV vs MPO-AAV (1,610 vs 870) OR (95% CI)	§P	¶Bayesian test of Colocalisation PP (%) Shared	PP (%) MPO	PP (%) PR3
‡rs6679677	PTPN22	1:114,303,808	A (0.10)	1.27 (1.18-1.38)	1.25×10^{-9}	1.22 (1.11-1.34)	2.73×10^{-5}	1.47 (1.29-1.67)	5.50×10^{-9}	0.85 (0.74-0.99)	3.03×10^{-2}	75.20	17.00	0.74
rs115670873	DGUOK, TET3	2:74,171,061	A (0.94)	1.68 (1.48-1.91)	7.35×10^{-16}	1.68 (1.44-1.97)	1.22×10^{-10}	1.82 (1.46-2.27)	7.60×10^{-8}	0.97 (0.69-1.37)	8.62×10^{-1}	99.80	0.00	0.03
rs13405741	BCL2L11, MIR4435-2HG	2:111,913,056	C (0.11)	1.25 (1.13-1.37)	7.36×10^{-6}	1.40 (1.24-1.58)	3.76×10^{-8}	1.09 (0.92-1.29)	3.10×10^{-1}	1.31 (1.09-1.57)	3.64×10^{-3}	3.99	0.35	62.40
rs1863800	CTLA4, CD28	2:204,702,660	C (0.58)	1.19 (1.12-1.26)	1.39×10^{-9}	1.17 (1.10-1.26)	5.15×10^{-6}	1.19 (1.09-1.31)	2.33×10^{-4}	0.95 (0.84-1.07)	3.98×10^{-1}	71.80	2.61	8.34
rs116522341	BTNL2, HLA-DQ	6:32,367,697	G (0.04)	1.56 (1.33-1.82)	3.56×10^{-8}	0.83 (0.68-1.02)	6.94×10^{-2}	4.80 (3.67-6.26)	1.09×10^{-30}	0.23 (0.17-0.31)	8.55×10^{-23}	-	-	-
rs3130216	HLA-DPB	6:33,077,271	G (0.49)	1.90 (1.80-2.00)	2.50×10^{-119}	2.58 (2.41-2.75)	3.59×10^{-173}	1.08 (0.98-1.19)	1.10×10^{-1}	2.90 (2.58-3.25)	2.44×10^{-73}	-	-	-
rs79005509	BACH2	6:90,850,730	T (0.02)	1.50 (1.21-1.86)	1.85×10^{-4}	1.07 (0.82-1.40)	6.12×10^{-1}	2.90 (2.00-4.20)	1.90×10^{-8}	0.52 (0.34-0.78)	1.52×10^{-3}	3.13	40.30	9.35
rs145007430	EBF3, MGMT	10:131,690,291	A (0.02)	1.75 (1.37-2.24)	8.36×10^{-6}	2.39 (1.76-3.25)	2.38×10^{-8}	1.23 (0.79-1.90)	3.57×10^{-1}	1.58 (1.04-2.40)	3.10×10^{-2}	1.95	7.40	45.10
rs112635299	SERPINA1	14:94,838,142	T (0.02)	2.21 (1.85-2.64)	3.09×10^{-18}	3.43 (2.75-4.27)	2.37×10^{-28}	1.13 (0.81-1.57)	4.77×10^{-1}	2.16 (1.64-2.86)	5.91×10^{-8}	4.77	0.00	65.10
rs11845244	IGHV1-69	14:107,170,077	T (0.37)	1.13 (1.06-1.20)	6.49×10^{-5}	1.23 (1.15-1.33)	2.16×10^{-8}	0.99 (0.89-1.09)	8.01×10^{-1}	1.28 (1.13-1.45)	1.00×10^{-4}	4.18	0.01	88.00
rs2965955	ANKRD11, SPG7	16:89,524,923	C (0.69)	1.12 (1.06-1.19)	1.64×10^{-4}	1.03 (0.96-1.11)	3.97×10^{-1}	1.37 (1.24-1.51)	9.64×10^{-10}	0.70 (0.61-0.80)	1.04×10^{-7}	2.95	83.50	0.00
rs55952682	PRTN3	19:835,908	T (0.26)	1.22 (1.12-1.32)	2.40×10^{-5}	1.39 (1.26-1.54)	1.90×10^{-11}	0.96 (0.84-1.11)	5.60×10^{-1}	1.51 (1.27-1.78)	1.41×10^{-6}	2.75	0.00	66.10

‡ The values reported for rs6679677 included a further replication cohort, comprising 1,531 controls, 347 MPO-AAV, 1,122 PR3-AAV and 81 unknown ANCA status.

* The risk allele refers to the effect in the overall meta-analysis. RAF, risk allele frequency; OR, odds ratio; CI, confidence interval.

§ The nominal p-value for PR3-AAV versus MPO-AAV was 0.00417. The p-values were calculated using the SNPTEST v2.5.2 software with an additive genetic model.

¶ A Bayesian test of colocalisation was used to identify loci with strong evidence of either shared or independent causal variants between PR3-AAV and MPO-AAV across each of the 10 non-HLA susceptibility loci, expressed as posterior probability in percentage, PP (%).

A comparison of the odds ratios between combined AAV meta-analysis with PR3-AAV and MPO-AAV is summarised in Table 4.8. Three of the 12 loci associated with AAV displayed strong evidence of a shared causal variant between MPO-AAV and PR3-AAV of equivalent effect size in the same direction (*CTLA4, PTPN22, DGUOK-TET3*). Although rs1863800 (*CTLA4*) did not reach the genome-wide significance for both MPO-AAV and PR3-AAV, rs6679677 (*PTPN22*) and rs115670873 (*DGUOK/TET3*) were only genome-wide significant in MPO-AAV and PR3-AAV respectively, these SNPs demonstrated strong evidence for colocalisation with high posterior probability scores and attained genome-wide significance in the combined AAV analysis where the total sample size was larger and thus better powered to detect weaker signals. Furthermore, in the within-case analysis, none of these SNPs achieved nominal significance, supporting that the signals at these loci are not subgroup specific. Notably, the shared risk loci between MPO-AAV and PR3-AAV also shared commonalities with other immune-related diseases (Figure 4.9).

The remaining nine loci displayed strong evidence of independent causal variant for either MPO-AAV or PR3-AAV. Three risk loci were MPO-AAV specific (*BTNL2-HLA-DQ, BACH2* and *ANKRD11-SPG7*) while six risk loci were PR3-AAV specific (*HLA-DP, PRTN3, SERPINA1, BCL2L11-MIR4435-2HG, EBF3-MGMT* and *IGHV1-69*) (Figure 4.10). In the within-case analysis, all variants in these risk loci revealed nominally significant associations with appropriate direction of effects except rs145007430 (*EBF3-MGMT*), which could be due to lack of power or less likely the signal at this locus might not be subgroup specific.

In summary, both colocalisation and within-case analysis provide robust evidence of a differential genetic basis of MPO-AAV and PR3-AAV at the HLA region and seven non-HLA loci including *BACH2, ANKRD11-SPG7, PRTN3, SERPINA1, BCL2L11-MIR4435-2HG, EBF3-MGMT* and *IGHV1-69* as well as common susceptibility loci comprising *PTPN22, CTLA4* and *DGUOK-TET3*.
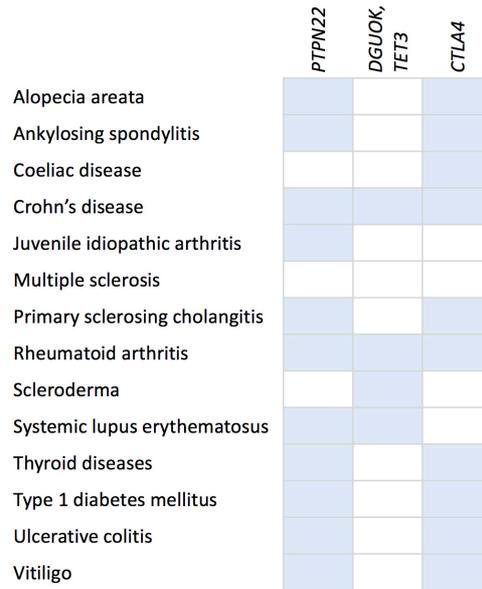
Figure 4.9: Shared genetic architecture with other autoimmune diseases.

Three risk loci are shared between MPO-AAV and PR3-AAV and their associations with other autoimmune diseases in genome-wide association studies obtained from the Phenoscanner and the NHGRI-EBI Catalog of published GWAS. The filled blue boxes indicate genome-wide significant associations for the autoimmune diseases and the risk loci.
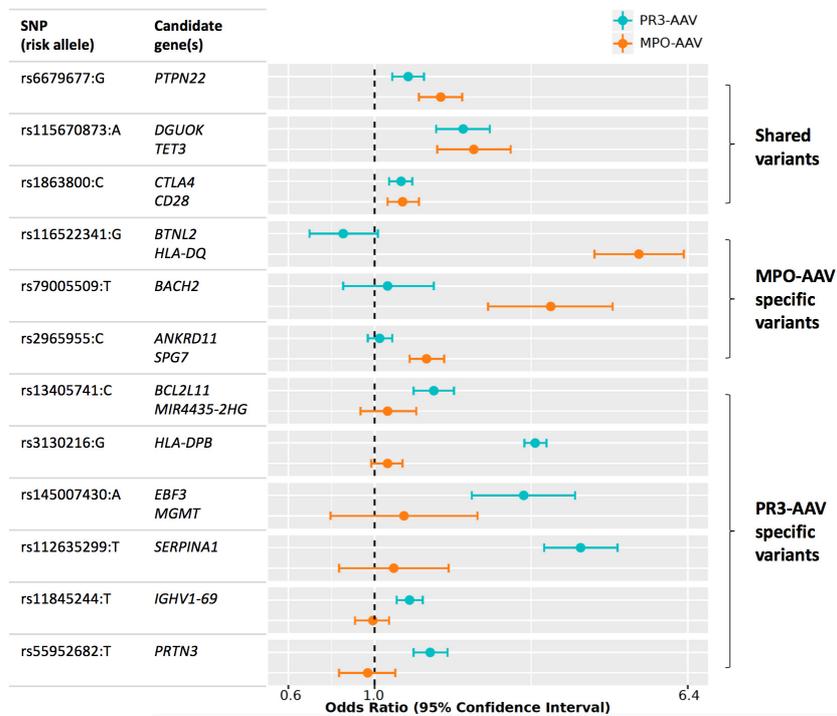
Figure 4.10: Forest plot depicts the odds ratios for MPO- and PR3-AAV risk loci.

The error bars represent 95% confidence intervals and the x-axis is in logarithmic scale.

125

4.8.2 Genetic correlation using GCTA

Using the genome-wide individual level genotype data (MAF ≥ 0.05), I used GCTA to estimate genome-wide genetic correlation (r_G) between MPO-AAV and PR3-AAV using a bivariate LMM. r_G ranges from -1 (negatively correlated/genetically unrelated) to 1 (co-heritable). The r_G between MPO-AAV and PR3-AAV was 0.69 ± se 0.09 (p = 0.0000E+00). This estimate of genetic correlation is similar to that observed between UC and Crohn's disease (r_G = 0.62 ± 0.042) (373, 442) but is greater than that between type 2 diabetes and hypertension (r_G = 0.31 ± 0.14) (339). A comparison of bivariate estimates of genetic correlation across different diseases including those from the paediatric population is shown in Figure 4.11. The r_G was even stronger after exclusion of the MHC region (r_G = 0.84 ± se 0.12).
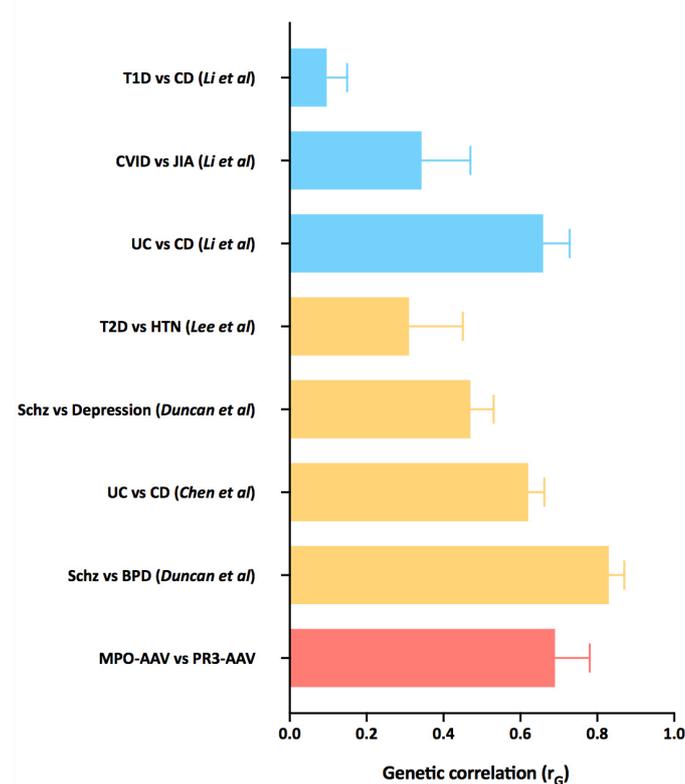
Figure 4.11: Bivariate estimates of genetic correlation across different diseases.

Genetic correlation (r_G) estimated using GCTA across different diseases in paediatric population (indicated by blue bars) – type 1 diabetes and Crohn's disease (CD), common variable immunodeficiency disorder (CVID) and juvenile idiopathic arthritis (JIA), ulcerative colitis (UC) and CD (443); adult population (indicated by yellow bars) – type 2 diabetes and hypertension (339), schizophrenia (Schz) with depression and bipolar disorder (BPD) (444), UC and CD (442); MPO-AAV and PR3-AAV as indicated by the red bar.

4.9   Discussion

We performed the largest genome-wide association study of AAV to date, which was powered to examine both MPO-AAV and PR3-AAV independently. Following a meta-analysis of two European cohorts, a total of 12 regions of the genome were identified to be associated with AAV, including five risk loci from previous studies (*HLA-DP*, *HLA-DQ*, *SERPINA1*, *PRTN3* and *PTPN22*). Apart from *SERPINA1*, the strongest effect size was observed in the HLA region. Three independent HLA signals were confirmed to confer susceptibility for PR3-AAV while there was one for MPO-AAV. Of these 12 loci, three displayed strong evidence for a shared causal variant between MPO-AAV and PR3-AAV (*PTPN22*, *DGUOK-TET3* and *CTLA4*), which are also implicated in a wide range of autoimmune diseases. Three risk loci were MPO-AAV specific and these include *BTNL2-HLA-DQ*, *BACH2* and *ANKRD11-SPG7* while six risk loci were PR3-AAV specific consisting of *HLA-DP*, *BCL2L11-MIR4435-2HG*, *EBF3-MGMT*, *SERPINA1*, *IGHV1-69* and *PRTN3*. These findings have confirmed that AAV is a polygenic disease. In addition, the estimated total SNP heritability (h^2_{SNP}) was slightly higher in PR3-AAV than MPO-AAV, in a similar range to inflammatory bowel disease but was lower than that of T1D and SLE (396, 442), suggesting that environmental factors might play a larger role in AAV, in keeping with previous epidemiological and observational studies demonstrating infections, silica, solvent, asbestos and drug exposures (42-45) to be related to AAV as well as a much later age of diagnosis as compared to T1D, which is commonly diagnosed by or before 23 years of age and SLE, which is classically seen in women of childbearing age.

The shared genetic risk between MPO-AAV and PR3-AAV was also quantified for the first time. The strength of genetic correlation between these two syndromes is similar to those observed between UC and Crohn's disease (373, 443), suggesting that substantial components of the genetic architecture are shared between the two syndromes. Although only three common risk loci have been identified so far, this is likely due to the limitation of a small sample size. For example, in UC and Crohn's disease where a large number of shared risk loci have been identified (167), the sample sizes for most studies were ten to hundred times larger than that in our study. Although a degree of the co-heritability observed between autoimmune diseases could be due to common HLA alleles, Li and colleagues previously estimated genetic correlations across different paediatric autoimmune diseases with and without the HLA

region and they did not observe significant difference between the two methods, suggesting that it is unlikely that genetic sharing of HLA haplotypes can explain the genetic correlation observed (443). In contrary, we found a stronger genetic correlation after exclusion of the HLA region, which is not too surprising, given that MPO-AAV and PR3-AAV have distinct HLA associations and by far have the greatest genetic effect on the risk of developing AAV. Furthermore, in population-based studies, the estimate of genetic effect that is shared between two traits is inevitably confounded by common environmental factors or genetic environment interactions, and therefore the genetic correlation between MPO-AAV and PR3-AAV could be slightly biased to overestimation.

One of the new MPO-AAV specific risk loci is *BACH2* – a classical example of a super-enhancer regulated gene that is associated with many autoimmune diseases (437). Both the PCHi-C and eQTL data support that the lead variant (rs79005509:T), which lies in the distal regulatory region, is likely to regulate the expression of *BACH2* through an interaction with its promoter (321) in CD4$^+$ and CD8$^+$ T cells, potentially influencing the binding affinity of transcription factors. The risks allele reduces *BACH2* mRNA expression (318) and might consequently lead to an imbalance of T_{reg} cell formation and effector cell differentiation, thus promoting immunity over tolerance. T_{reg} cells are an important subpopulation of T cells that modulate immune system to maintain tolerance to self-antigens. The role of T_{reg} in the pathogenesis of AAV has been well demonstrated in the animal models. By depleting the T_{reg} population using anti-CD25 monoclonal antibody, Tan et al showed that the MPO-immunised mice developed more severe segmental necrosis and glomerular cellular proliferation as compared with non-T_{reg}-depleted MPO-immunised mice with an enhanced autoimmune B cell response to MPO (445). Moreover, in the experimental crescentic glomerulonephritis model, mice that lack CXCR3 in T_{reg} which is essential for its migration to site of inflammation, were found to have a reduced number of T_{reg} in the kidneys with an overwhelming Th1 immune response and developed with more severe renal impairment that was reversible with anti-IFNγ treatment (446). Several studies have also showed a reduced suppressor function of T_{reg} in AAV patients as compared to healthy controls (253, 254, 447). Further functional studies will be required to elucidate the biological mechanisms by which the causal variant increases susceptibility to disease.

We also confirmed the association of *CTLA4* with AAV at a genome-wide significance level. CTLA-4 is an important negative regulator of T-cell activation and it is implicated in many organ-specific autoimmune diseases. The lead variant in this locus is a known eQTL where the risk allele (rs1863800:C) is associated with a lower expression of *CTLA4* but a higher expression of *ICOS* in the whole blood (318). ICOS (a.k.a CD278) is a CD28-superfamily costimulatory molecule that is expressed on T cells upon activation. In addition, other reported autoimmunity-susceptibility variants in this region including rs231775, rs3087234 and rs231806, which are in high LD with the lead SNP, have been demonstrated to have functional effects on *CTLA4* transcription (in both CD4$^+$ T cells and CD8$^+$ T cells), splicing, production of the soluble form of CTLA-4 and post-translational modifications (448-451). Taken all together, individuals carrying the risk alleles might have a lower activation threshold of T cells and thus contribute to T-cell activation and autoimmune disease susceptibility. There is an ongoing clinical trial (ABROGATE, ClinicalTrials.gov Identifier: NCT02108860) to evaluate the efficacy of Abatacept to achieve sustained glucocorticoid-free remission in patients with relapsing non-severe GPA. Abatacept is a fusion protein composed of the Fc region of IgG1 fused with the extracellular domain of CTLA-4, and it binds to the CD80 and CD86 molecules to prevent T-cell activation.

4.10 Conclusion

Both MPO-AAV and PR3-AAV are polygenic diseases and they share a substantial component of their genetic architecture. The common risk loci between the two syndromes also share commonalities with other immune-related diseases. Each disease-related region can be fine-mapped to obtain the most likely causal variants. Through eQTL and chromatin interaction datasets, we are able to determine the genes that these SNPs most probably regulate in the disease relevant cell types. This will enable us to better understand the molecular mechanisms underpinning the development of AAV and identify novel therapeutic pathways for drug development.

5 Functional follow-up at *BCL2L11-MIR4435-2HG* and *PRTN3* regions

5.1 Prelude

Statistical and computational analyses provide important information on which SNPs within a trait associated LD block could be a causal variant. However, these analyses do not provide definite proof of causality but rather bring forward strong candidates to undergo functional validation. In addition, cellular or in vivo experiments would be required to prove that the gene(s) regulated by a causal variant do indeed lead to disease relevant phenotypes if perturbed. In this chapter, I will discuss the post-GWAS bioinformatics analyses and functional follow-up of two PR3-AAV susceptibility loci including *BCL2L11-MIR4435-2HG* and *PRTN3*, to understand how the causal variants exert their effects in contributing to the pathogenesis of PR3-AAV.

5.2 BCL2L11-MIR4435-2HG

5.2.1 Introduction

As described in Chapter 3, one of the newly identified PR3-AAV specific risk loci in the meta-analysis was *BCL2L11-MIR4435-2HG*, which is located in the long arm of chromosome 2. The top variant was rs13405741 (risk allele: C, OR = 1.40, p_{meta} = 3.76 x 10^{-8}) which is an intronic variant in *BCL2L11*. Furthermore, PCHi-C data confirmed interactions between the SNPs from the credible set with the promoter region of *BCL2L11* and the putative regulatory region of *MIR4435-2HG* in granulocytes. *BCL2L11* encodes for Bim, which is part of the Bcl2 family and is critical in regulating cell death while *MIR4435-2HG* is a long, non-coding RNA that has been discovered recently to play an essential role in the regulation of *BCL2L11* transcription. To date, no functional data exists to explain the association of this locus with disease.

There are two major mechanisms by which cell death can occur – necrosis or apoptosis. Necrosis is a pathological process that occurs when cells are subjected to severe physical or chemical stress. In contrast, apoptosis is a fundamental physiological process of cell death in multicellular organisms during development and tissue homoeostasis (452). It is also known as programmed cell death and is characterised by cell shrinkage, nuclear and chromatin condensation, DNA fragmentation, membrane blebbing, phosphatidylserine externalisation and formation of membrane-bound apoptotic bodies. These apoptotic bodies are rapidly phagocytosed by either macrophages or adjacent epithelial cells to avoid inflammatory response. In vitro, the apoptotic bodies and the remaining cell fragments ultimately swell and lyse. This process is also known as secondary necrosis. The balance between survival and apoptotic signals is tightly controlled and disruption of this balance leads to diseases such as cancer. In AAV, neutrophils are the primary effector cells in mediating endothelial cell and vascular damage. Several studies have demonstrated that apoptosis of infiltrating neutrophils is disrupted by ANCA activation and removal of apoptotic neutrophils occurs in a proinflammatory manner, resulting in persistent inflammation (453, 454).

There are many players that are known to regulate apoptosis, including caspases, cell death receptors of the TNF family, adaptor proteins, inhibitor of apoptosis proteins and the bcl-2 family. The Bcl-2 family contains at least 20 related proteins. Each member of the family shares one or more Bcl-2 homology (BH) domains and are divided into two groups based on whether they facilitate or inhibit apoptosis. Bim is a BH3-only containing protein and was first discovered over two decades ago as a pro-apoptotic protein (455). Three isoforms of Bim are generated by alternative splicing (Bim_{EL} ~33kDa, Bim_L ~25kDa and Bim_S ~15kDa), all of which induce apoptosis by binding to and antagonising anti-apoptotic members of the Bcl-2 family that ultimately influence the insertion of Bax and Bax-like proteins into the outer mitochondrial membrane, leading to cytochrome C release and activation of caspase9/Apaf-1 (apoptotic protease activating factor 1) apoptosome. Bim_s is the most potent inducer of apoptosis. The lifespans of myeloid and lymphoid cells are substantially increased in Bim-deficient mice, where the numbers of granulocytes, monocytes, B cells and T cells are increased by 2-4 fold (379).

Intriguingly, the transcription of *BCL2L11* was recently discovered to be regulated by its neighbouring long non-coding RNA – *MIR4435-2HG* (also known as *Morrbid*, myeloid RNA regulator of Bim-induced death), specifically in myeloid cells and in a cis manner by promoting the enrichment of the polycomb repressive complex 2 at *BCL2L11* promoter, maintaining the gene in a poised state and enabling rapid control of apoptosis in response to pro-survival cytokines (380). *MIR4435-2HG* is highly conserved across species and is localised in the nucleus, predominantly chromatin bound. It is most abundantly expressed in eosinophils, neutrophils and classical monocytes in both human and mice. These three cell types were greatly reduced in *MIR4435-2HG*-deficient mice, while natural killer and other lymphoid cells remained unaffected. Furthermore, an accelerated apoptosis was observed in these three cell types in bone marrow cultured ex vivo with markedly elevated *BCL2L11* mRNA and protein levels. *MIR4435-2HG*-deficient mice were more prone to bacterial infection but protected from eosinophil-driven inflammatory lung disease. Importantly, the authors demonstrated overexpression of *MIR4435-2HG* in eosinophils isolated from patients with hypereosinophilic syndrome as compared to healthy controls, which was positively correlated with plasma IL-5 levels, supporting the important role of *MIR4435-2HG* in eosinophil-mediated inflammatory disorders (380).

rs13405741 is an eQTL for *BCL2L11* in the whole blood (318). To establish the specific cell types that this variant might have its effect on, I began by looking into the existing microarray whole-genome gene expression datasets for a number of leukocyte lineages in patients with chronic kidney disease (CKD), comprising CD16[+] neutrophils, CD14[+] monocytes, CD8[+] T cells, CD4[+] T cells and CD19[+] B cells that were generated in the Smith lab.

5.2.2 The effect of rs13405741 genotype on *BCL2L11* expression in different immune cells

Firstly, the genotype of each individual in the CKD cohort was determined using the TaqMan[®] SNP genotyping assay. Following QC of the gene expression datasets, a breakdown of the number of patients included in the eQTL analysis for each immune cell subset is summarised in Table 5.1. The allele frequency of the risk allele (rs13405741:C) is 0.11 and not surprisingly, there was only one homozygous minor patient in the CKD cohort. There was a trend towards higher expression of *BCL2L11* in association with the risk allele in neutrophils (one-way

analysis of variance (ANOVA), p = 0.0921; test for linear trend, p = 0.0324) but not in other cell lineages (Figure 5.1A). One major limitation of this analysis was a small sample size, particularly of the statistical power for detecting a difference in minor allele homozygotes.

Table 5.1: The number of CKD patients included in the eQTL analysis

Immune cell subsets	*rs13405741 genotype		
	T/T	T/C	C/C
CD16+ neutrophils	55	21	1
CD14+ monocytes	40	17	0
CD8+ T cells	56	20	1
CD4+ T cells	22	9	1
CD19+ B cells	47	19	1

*rs13405741 (T/T: homozygous major; T/C: heterozygous; C/C: homozygous minor).

5.2.3 Allele-specific expression of *BCL2L11* in neutrophils

To ascertain whether rs13405741:T>C or a variant in strong LD with it, might influence gene expression, a clone-based allele-specific expression assay was used to analyse *BCL2L11* transcription in neutrophils from eight heterozygous healthy individuals. This technique facilitated a direct comparison between the amount of *BCL2L11* pre-mRNA that was transcribed from each allele and ensured that the external factors affected both alleles equally. Under basal conditions (unstimulated, time 0 hr), we observed allele-specific differences in transcription, with ~0.3-fold more *BCL2L11* being transcribed from the DNA strand containing the minor allele (C, the risk allele for PR3-AAV; p = 0.0078) than from the strand containing the major allele (T). There was no difference observed in the genomic DNA which was inherently in a 1:1 ratio, thereby serving as an internal control. At time 6 hr, the same trend was seen in neutrophils that were cultured with standard medium though the effect size was less marked as compared to time 0 hr (p = 0.0391). In contrast, the ratio of major (T) to minor (C) alleles in *BCL2L11* pre-mRNA was no different at time 6 hr when the neutrophils were stimulated with GM-CSF, suggesting that the allele-specific expression might be abrogated by a pro-survival agent (Figure 5.1 B).
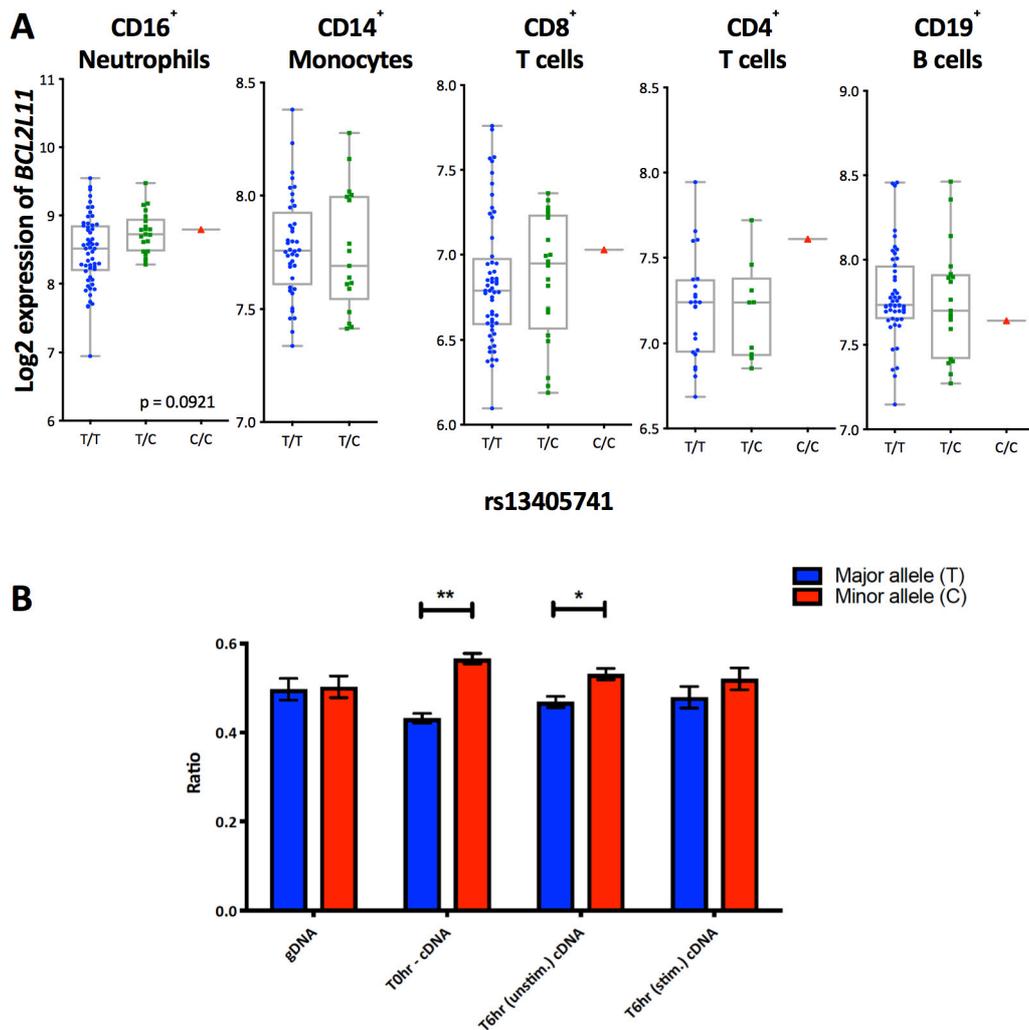
Figure 5.1: rs13405741:C allele is associated with a higher expression of *BCL2L11* in neutrophils.

(A) Microarray whole-genome gene expression of different immune cell subsets for chronic kidney disease patients. There was a trend of higher expression of *BCL2L11* in neutrophils for individuals carrying the PR3-AAV risk alleles (rs13405741:C) but not in other leukocyte lineages. In box plots, each box extends from the 25th to the 75th percentile with a line represents the median and whiskers extend to the minimum and maximum values. (B) Allele-specific expression assay showing the ratio of rs13405741 alleles in genomic DNA (gDNA) and complementary DNA (cDNA; synthesised from pre-mRNA) from neutrophils of heterozygous individuals (n = 8). Ninety-six colonies were genotyped per individuals per condition (either unstimulated [unstim.] or stimulated [stim.] with GM-CSF). Data are represented as mean $\pm$ standard error of mean (SEM), *p < 0.05 and **p < 0.01 (Wilcoxon signed rank test).

5.2.4 PR3-AAV shares the same causal variant as eosinophil count at *BCL2L11-MIR4435-2HG*

Clinical evidence and in vitro studies support the notion that the innate immune system plays an important role in the pathogenesis of AAV (as described in section 1.5.2). Chromosome 2q13 has been reported to be associated with several haematological traits including eosinophil count, eosinophil percentage of granulocytes, neutrophil percentage of white cells and monocyte percentage of white cells in a population-based study (346). Importantly, the candidate genes in this locus – *BCL2L11* and *MIR4435-2HG* are involved in the regulation of myeloid cell lifespan and henceforth, cell numbers. A natural question to ask was whether these haematological traits shared the same causal variant as PR3-AAV at this locus and therefore, also potentially share the same causal mechanism. Visual comparison of overlaps of between the association signals from different traits was a step in the right direction but inadequate to make inferences about causality. The Bayesian statistical test for colocalisation was employed to address this question, using GWAS summary statistics generated by Astle et al (346) for three haematological traits comprising eosinophil, neutrophil and monocyte counts. The results of the colocalisation analysis include five posterior probabilities (PP0, PP1, PP2, PP3 and PP4) where a large posterior probability for hypothesis 3 (PP3) indicates support for two independent causal variants with each trait while a large PP4 supports a single variant affecting both traits (see Section 2.1.16). We demonstrated positive (eosinophil count, PP4 = 98%) and negative (neutrophil count, PP4 < 0.01% and monocyte count, PP4 = 0.07%) colocalisation results (Figure 5.2), supporting that PR3-AAV and eosinophil count share the same causal variant at *BCL2L11-MIR4435-2HG*.
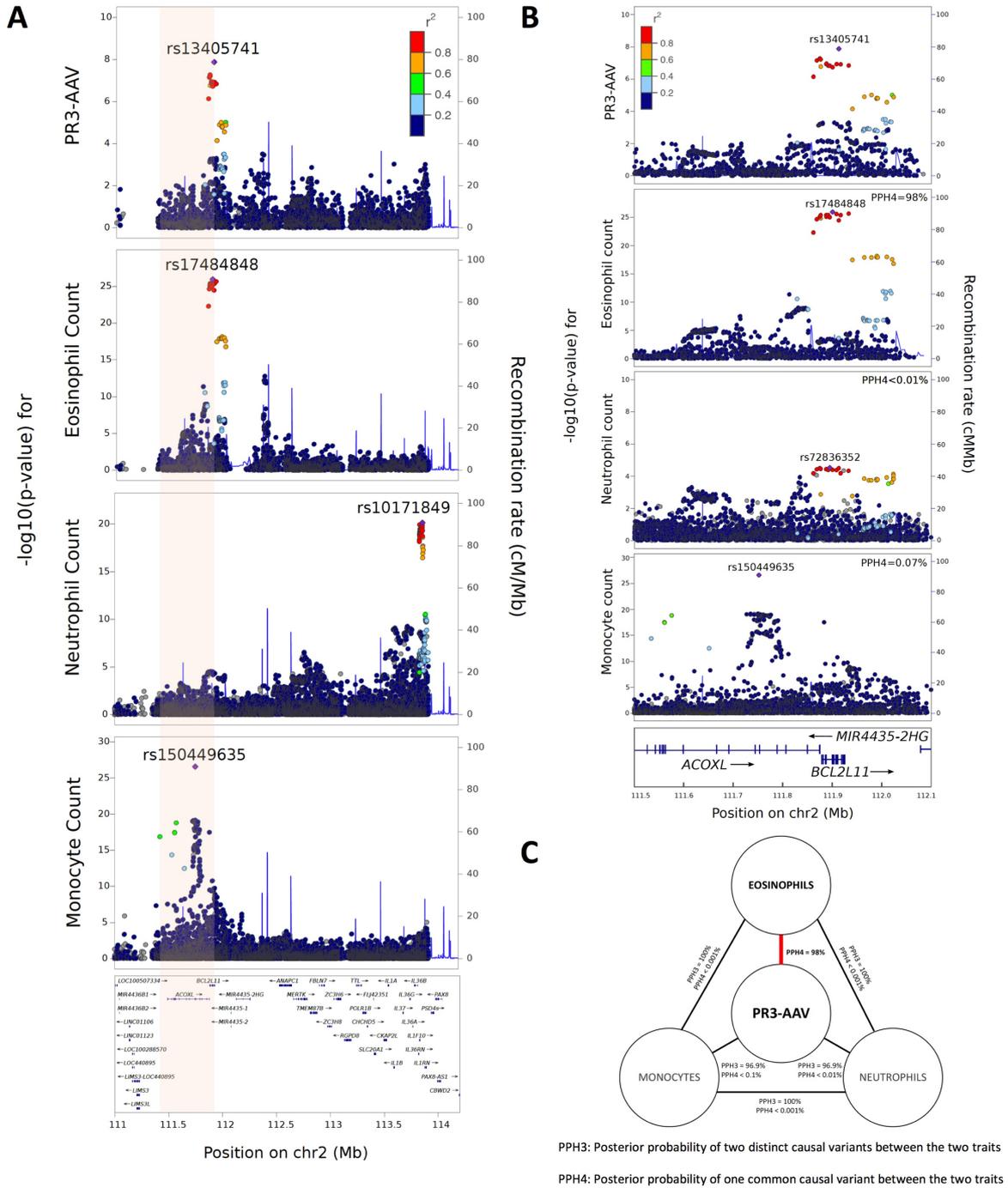
Figure 5.2: Illustration of the colocalisation results.

(A) The 2q13 region containing the candidate genes *BCL2L11* and *MIR4435-2HG*. Top to bottom panels: regional visualisation of genome-wide association results for PR3-AAV (top SNP, rs13405741), eosinophil count (top SNP, rs17484848), neutrophil count (top SNP, rs10171849) and monocyte count (top SNP, rs150449635). Region of interest is highlighted in orange. (B) Magnified view of the region of interest showing positive (PR3-AAV and eosinophil count, PP4 = 98%) and negative (PR3-AAV and neutrophil count, PP4 < 0.01%; PR3-AAV and monocyte count, PP4 = 0.07%) colocalisation results. (C) The relationships between PR3-AAV and 3 other haematological traits including eosinophil, neutrophil and monocyte counts.

5.2.5 A higher eosinophil count is associated with an increased risk of PR3-AAV

Having established that PR3-AAV and eosinophil count have a high probability of sharing the same causal variant at *BCL2L11-MIR4435-2HG* region, we then asked whether eosinophil count was causally related to PR3-AAV. To address this question, we performed Two-sample Mendelian Randomisation analysis using the summary statistics from GWAS of eosinophil counts (346), consisting of 209 conditionally independent genome-wide significant SNPs. After excluding the HLA region, 147 of these variants were available in the PR3-AAV dataset for MR using the IVW method. We found that a higher eosinophil count was strongly associated with an increased likelihood of developing PR3-AAV (causal estimate = 0.249, p = 0.028). The results were similar when different MR methods were used, including simple median MR (causal estimate = 0.593, p = 0.000) and penalised IVW MR (causal estimate = 0.299, p = 0.002), except for MR-Egger approach where the causal estimate was attenuated (causal estimate = 0.086, p = 0.75). The MR-Egger intercept was close to zero, indicating that there was no evidence of pleiotropy (Table 5.2). MR-Egger is a sensitivity analysis which it is generally used to correct for potential violations of assumptions by quantifying the amount of directional pleiotropy. However, its power to detect a causal effect is greatly reduced as compared to the traditional MR methods. In this instance, although MR-Egger analysis did not provide additional evidence for a causal effect, it did not contradict evidence for a causal effect from the conventional MR analysis methods either.

Table 5.2: Estimates from various Mendelian randomisation methods for the association between eosinophil count and PR3-AAV

Method	Causal Estimate	95% CI	P-value
Simple median MR	0.593	0.318 – 0.868	0.000***
Weight median MR	0.243	-0.040 – 0.525	0.092
Penalised weighted median MR	0.245	-0.038 – 0.528	0.090
IVW MR	0.249	0.028 – 0.470	0.028*
Penalised IVW MR	0.299	0.111 – 0.486	0.002**
Robust IVW MR	0.263	0.061 – 0.465	0.011*
Penalised robust IVW MR	0.304	0.117 – 0.491	0.001***
MR Egger estimate	0.086	-0.441 – 0.613	0.750
MR Egger intercept	0.007	-0.015 – 0.030	0.504

IVW, inverse variance weighted; MR, Mendelian randomisation; *Significant at the 0.05 probability level; **Significant at the 0.01 probability level; *** Significant at the 0.001 probability level.

Summary statistics were also obtained for 152 and 248 conditionally independent variants associated with peripheral blood neutrophil and monocyte counts respectively (346). After excluding the HLA region, 101 (neutrophil count) and 169 (monocyte count) variants were available in the PR3-AAV dataset for MR analysis. We did not identify evidence of any association between neutrophil or monocyte counts with susceptibility to PR3-AAV (Figure 5.3).
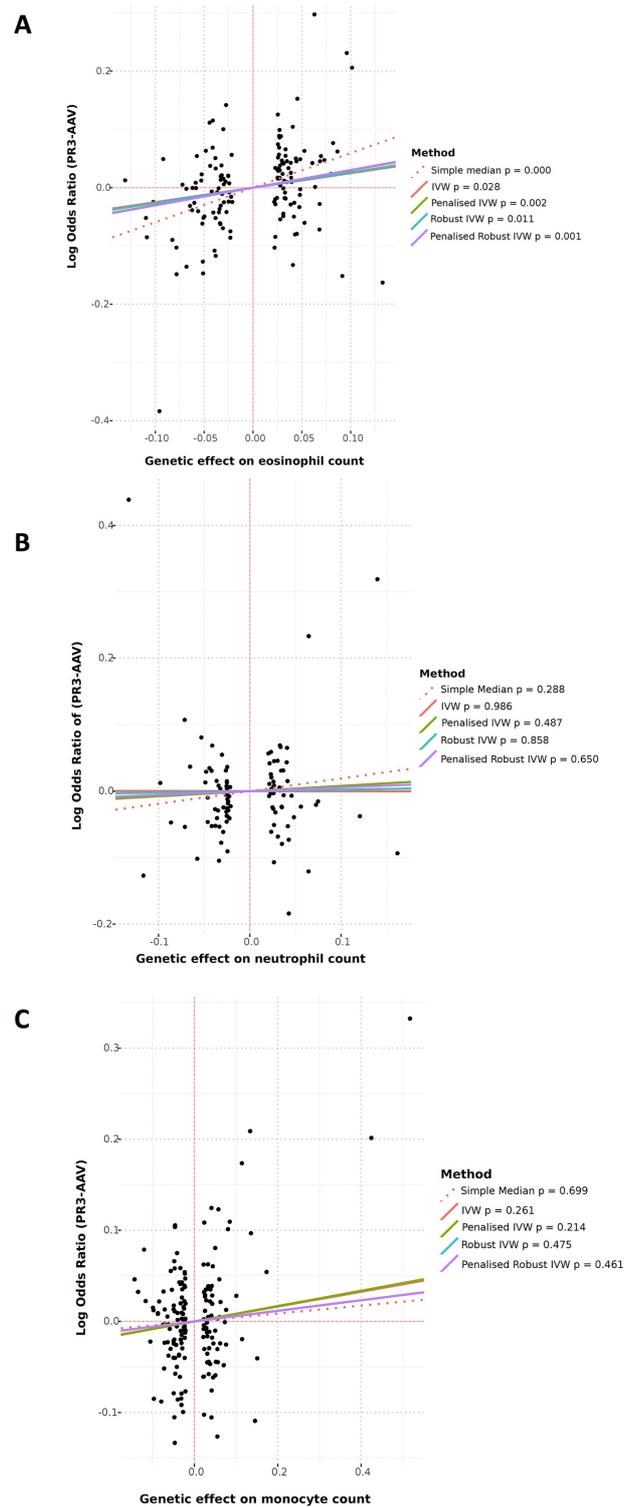
Figure 5.3: A higher eosinophil count is associated with an increased risk of PR3-AAV.

Scatter plots showing correlation of genetic associations of (A) eosinophil count, (B) neutrophil count and (C) monocyte count with genetic associations of PR3-AAV. Points represent genome-wide significant conditionally independent variants associated with each haematological trait in the GWAS by Astle et al and coloured lines represent the slopes (equivalent to the estimated causal effects) of the different MR regression analyses. IVW, inverse-variance weighted.

5.2.6 Genetic variation at rs13405741 modulates the lifespan of eosinophil

Taking all the findings so far and the relevant functional associations of the candidate genes in this region, we next considered whether rs13405741, or a variant in strong LD with it, might influence gene expression in eosinophils and consequently alter eosinophil count through prolongation of cell survival. We recruited eight pairs of healthy individuals via the NIHR BioResource Centre Cambridge, who were homozygous for either the major or minor allele at rs13405741, and were matched for age and gender. Eosinophils were isolated from polymorphonuclear cell-rich fraction of peripheral blood by negative immunomagnetic selection. Eosinophils represent a minor population of leukocytes in the peripheral blood (typically between 2-5%) and therefore cell yield following isolation varied among samples. We first examined the expression of *BCL2L11* and *MIR4435-2HG* to determine whether these differed by genotype. We found that eosinophils from minor (C) allele homozygotes expressed more *MIR4435-2HG* compared to those from major (T) allele homozygotes at time 0 hr (p = 0.080) and 24 hr (p = 0.029) following culture in the absence of serum. However, we did not detect any genotype-specific differences in the transcription of *BCL2L11* at both time points (Figure 5.4A). We then investigated how differences in *MIR4435-2HG* transcription might affect the production of Bim, by performing Western blot of eosinophil lysates recognising the limitations that only three samples from each genotype group had adequate cell yields for this experiment and therefore no finite conclusion could be made. Nevertheless, there was a trend for lower Bim production by eosinophils from minor allele homozygotes (Figure 5.4C). To determine whether differences in *MIR4435-2HG* transcription might have an implication on eosinophil survival, we measured the apoptosis rate in eosinophils following 24 hr culture without serum using flow cytometry (Annexin V and PI) and observed that lower apoptosis occurred in minor allele homozygotes (p = 0.0156), consistent with the earlier discovery of higher *MIR4435-2HG* transcription and lower Bim production in this genotype (Figure 5.4D and E).
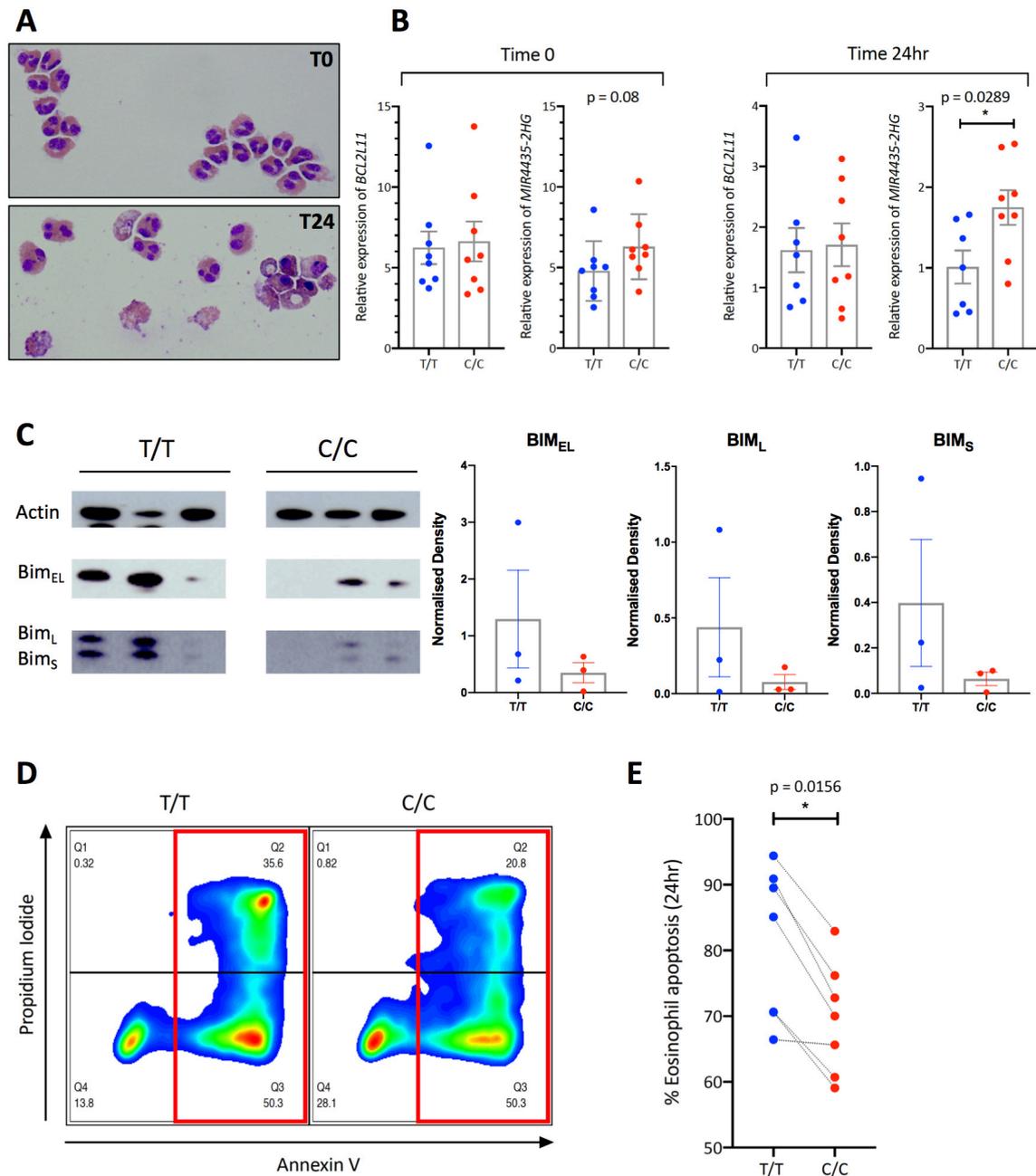
Figure 5.4: rs13405741:C allele is associated with a lower apoptosis rate in eosinophils.

(A) Representative photomicrographs of eosinophils at time 0 hr (top) and 24 hr (bottom) after incubation without serum depicting the morphological features of apoptosis including DNA fragmentation, cell shrinkage, nuclear and chromatin condensation and membrane blebbing. Cytospin preparations using May-Grünwald-Giemsa staining, original magnification 40X.

(B) Evidence of eQTL for *MIR4435-2HG*. Expression of *BCL2L11* and *MIR4435-2HG* were assessed using qPCR at time 0 hr and 24 hr, assayed in triplicate, n = 8 per group except for major allele homozygotes (T/T) at time 24 hr consisting of 7 per group. Mann-Whitney test was performed. Data are represented as mean $\pm$ SEM, *p < 0.05.

(C) Representative western blot images of purified eosinophils lysed in loading buffer following 24 hr culture and blotted for expression of 3 different isoforms of Bim (Bim$_{EL}$, Bim$_L$ and Bim$_S$). β-Actin was used as a loading control. Shown on the right are the densitometries calculated using ImageLab software and expression was normalised to β-Actin (n = 3 per group).

(D) Representative flow cytometry of Annexin V/Propidium iodide-stained eosinophils. Eosinophil apoptosis was measured following 24 hr culture in the RPMI medium in the absence of serum supplement.

(E) Eosinophil apoptosis rate was lower in the minor allele homozygotes. Matched-pairs experiments were carried out and assayed in triplicate. Wilcoxon matched-pairs signed-rank test (*p < 0.05), n = 7 per group was performed and dotted lines link paired samples.

5.3   PRTN3

5.3.1   Introduction

PRTN3 encodes PR3 which is a major antigenic target for AVV. It is most abundantly expressed in the most immature myeloid precursors (myeloblasts and promyelocytes) and subsequently diminishes during granulopoiesis, reaching an almost undetectable level in the circulating mature neutrophils (Figure 5.5A, RNA-seq data obtained from BLUEPRINT datasets). This gene has two transcripts – canonical isoform (*PRTN3-201*) and alternative isoform (*PRTN3-202*) which initiates transcription at a different promoter located in the intron 1. Alternative splicing was first reported by Chow et al in 1977 (456). It is an important biological process that contributes to the diversity of protein isoforms of subtle functional differences, yet translating into significant biological consequences. It has also been implicated in a wide range of diseases, for example in coeliac disease, where patients were found to express high levels of the alternatively spliced FoxP3Δ2 isoform which lacks exon 2, rather than the full-length FoxP3 in the intestinal cells, resulting in failure to achieve adequate downregulation of immune response (457).

As discussed in section 1.4.2, a higher expression of *PRTN3* transcript was demonstrated in AAV patients as compared to healthy controls (118, 119). Furthermore, McInnis and colleagues recently detected the alternative *PRTN3* transcript (*PRTN3-202*) in AAV patients (111), where its expression correlated with disease activity. Notably, its predicted amino acid sequence is identical to that previously reported for myeloblastin (112), which lacks the signal

peptide sequence, dipeptide (alanine-glutamate) and 14 amino acids coded by exon 2 with a predicted size of 24 kDa (Figure 5.5B). Myeloblastin was originally discovered in human leukemic cells and is involved in the proliferation and differentiation of leukaemic cells (112). However, it remains to be confirmed whether is *PRTN3-202* transcript coding for myeloblastin despite identical protein sequence as the published cDNA sequence for myeloblastin differs from the sequence for *PRTN3-202* message. Furthermore, the potential role of the alternative isoform of *PRTN3* in the pathogenesis of AAV is yet to be elucidated.

Three different lead variants at the *PRTN3* locus that have been reported in previous GWAS (rs62132295 and rs62132293) and in the present study (rs55952682) are known eQTLs in the whole blood where the risk alleles for PR3-AAV are associated with a higher expression of *PRTN3* (318). In terms of disease-relevant cell type, there were two studies investigating whether these variants had an influence on *PRTN3* expression in neutrophils with contradictory results (71, 123), while another study failed to detect any quantitative difference in *PRTN3* expression between wild-type and polymorphic variant (122) using the luciferase reporter assay.

To investigate how genetic variation at this locus might influence *PRTN3* expression in neutrophils, we first examined whether rs62132296 (availability of individual level genotype data) was an eQTL using the existing microarray gene expression datasets for AAV and CKD cohorts in the Smith lab. Furthermore, we aimed to determine whether there was evidence of differential expression at the isoform level. See Table 5.3 for LD between rs62132296 with other previously identified lead variants at this locus.

Table 5.3: Relationship of rs62132296 with previously reported lead variants at *PRTN3*

SNP	LD with rs62132296
rs62132295 (Lyons et al)	1.0
rs62132293 (Merkel et al)	0.9202
rs55952682 (present study)	0.6562

A

B

	Signal peptide sequence	Pro	Protein sequence	
PRTN3-201	MAHRPPSPALASVLLALLLSGAARA	AE	IVGGHEAQPHSRPYMASLQMRGNPGSHFCGGTL	60
*PRTN3-202	---		MASLQMRGNPGSHFCGGTL	19
PRTN3-201	IHPSFVLTAAHCLRDIPQRLVNVVLGAHNVRTQEPTQQHFSVAQVFLNNYDAENKLNDVL			120
PRTN3-202	IHPSFVLTAAHCLRDIPQRLVNVVLGAHNVRTQEPTQQHFSVAQVFLNNYDAENKLNDVL			79
PRTN3-201	LIQLSSPANLSASVATVQLPQQDQPVPHGTQCLAMGWGRVGAHDPPAQVLQELNVTVVTF			180
PRTN3-202	LIQLSSPANLSASVATVQLPQQDQPVPHGTQCLAMGWGRVGAHDPPAQVLQELNVTVVTF			139
PRTN3-201	FCRPHNICTFVPRRKAGICFGDSGGPLICDGIIQGIDSFVIWGCATRLFPDFFTRVALYV			240
PRTN3-202	FCRPHNICTFVPRRKAGICFGDSGGPLICDGIIQGIDSFVIWGCATRLFPDFFTRVALYV			199
PRTN3-201	DWIRSTLRRVEAKGRP			256
PRTN3-202	DWIRSTLRRVEAKGRP			215

Figure 5.5: *PRTN3* expression in different stages of granulopoiesis.

(A) Left: There are two *PRTN3* transcripts (*PRTN3-201* and *PRTN3-202*). *PRTN3* gene expression by RNA-seq in different stages of granulopoiesis (BLUEPRINT datasets, auto-scale to the level of expression as depicted on the left axis). Right: The stages of granulopoiesis from myeloblast to the mature circulating neutrophil. During neutrophil maturation, granulocytic cells change shape, acquire primary and specific granules and undergo nuclear condensation (Images were obtained from https://imagebank.hematology.org/).

(B) *In silico translation of *PRTN3-202* transcript indicates that it has an identical amino acid sequence as myeloblastin, which lacks the signal peptide sequence that is present in PR3. Signal peptide is critical in directing newly synthesised protein to the Golgi for post-translational glycosylation. Adapted and modified from McInnis et al.

5.3.2 rs62132296:T allele is associated with a higher expression of *PRTN3* in neutrophils

Firstly, the genotype of each individual in the CKD and AAV cohort was determined using the TaqMan® SNP genotyping assay. Following QC of the CD16+ neutrophil gene expression datasets, the number of patients included in the eQTL analysis is summarised in Table 5.4. To evaluate for allelic effects of rs62312296: C > T on *PRTN3* expression, I first performed whole transcript analysis and showed that in the CKD cohort, *PRTN3* transcript levels were significantly higher in minor (T) allele homozygotes (risk allele for PR3-AAV) than in patients with CC or CT genotypes, with evidence of allelic transcripts dosage effect (ANOVA, p < 0.001; test for linear trend, p < 0.0001; Figure 5.6A). In contrast, I did not detect any genotype-specific differences in *PRTN3* expression in the AAV cohort, which could possibly be attributed to (i) a small sample size or (ii) *PRTN3* expression reaching the saturation limit in the setting of active disease (Figure 5.6B).

To investigate differential isoform transcripts expression, I performed exon-level analysis of the same CKD cohort. There are seven probes targeting *PRTN3* on the Affymetrix GeneChip® microarray – two probes for exons 1 and 5 respectively and one probe for the remaining exons. I detected significantly higher expressions of exons 3 (ANOVA, p = 0.0362; test for linear trend, p = 0.0229), 4 (ANOVA, p = 0.0042; test for linear trend, p = 0.0010) and 5 (ANOVA, p = 0.0002; test for linear trend, p < 0.0001 and ANOVA, p = 0.0286; test for linear trend, p = 0.0272 for each respective probe) in the minor allele homozygotes than those from the major allele homozygotes. These exons are common to both isoforms. However, there were no differences observed in exons 1 and 2, therefore it was not possible to detect any evidence of differential isoform transcripts expression (Figure 5.6C).

Table 5.4: Number of patients included in the eQTL analysis

Cohort	rs62132296: C > T		
	C/C	C/T	T/T
CKD	38	35	4
AAV	7	17	5
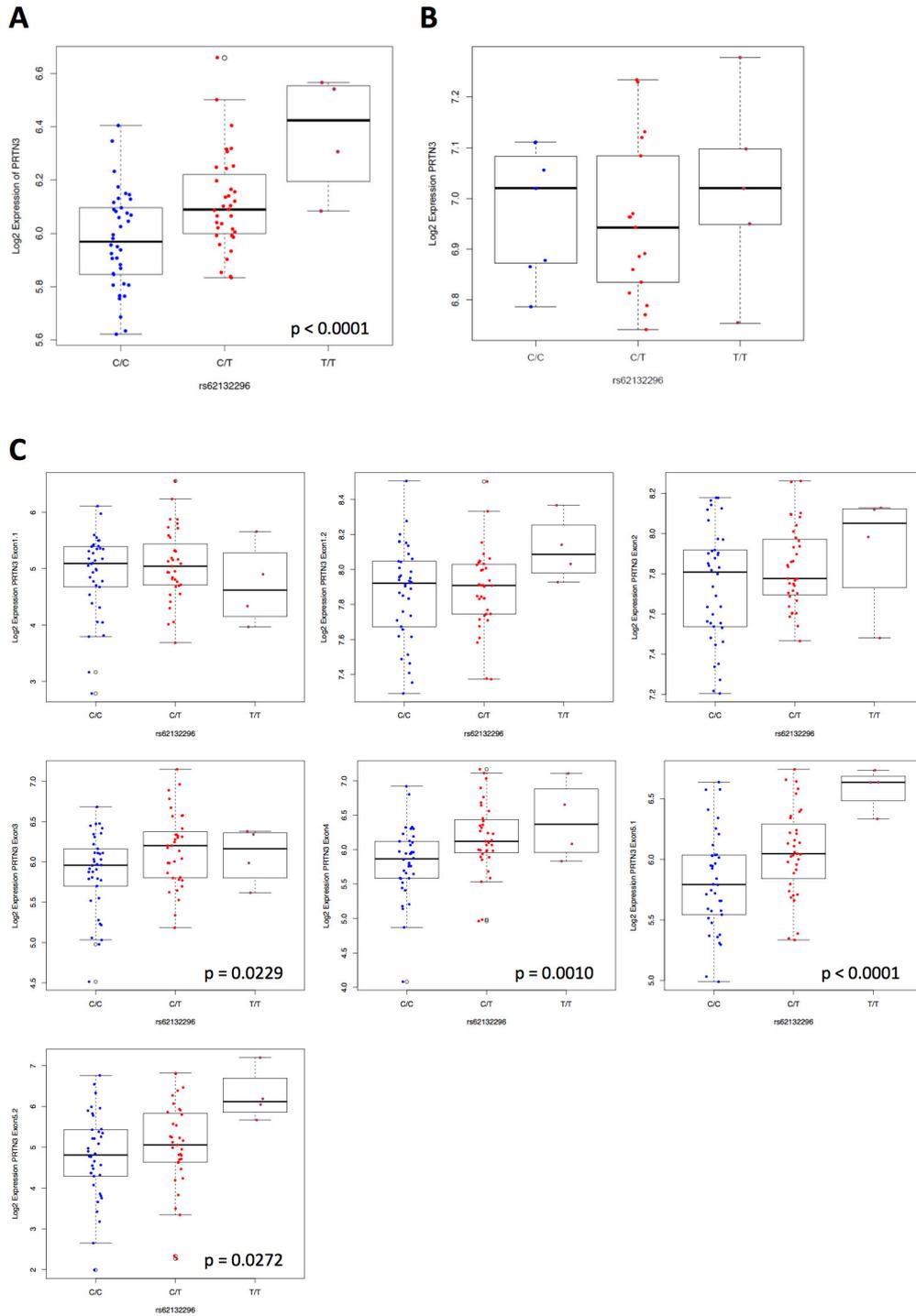
Figure 5.6: *PRTN3* expression in neutrophils

(A) rs62132296 is an eQTL for *PRTN3*. In the CKD cohort, the expression of *PRTN3* was significantly associated with the genotype at rs62132296 by standard ANOVA ($p < 0.0001$) and post-test for linear trend ($p < 0.0001$).

(B) The expression of *PRTN3* in the active AAV cohort was not associated with genotype at rs62132296.

(C) Exon-level expression analysis showing significant differential expressions of exons 3, 4 and 5 by genotype but not exons 1 and 2. The p-values represent test for linear trend.

146

5.3.3   Detection of alternative *PRTN3* transcript (*PRTN3-202*)

To confirm the expression of *PRTN3-202* transcript, I performed PCR with RNA from isolated PMN and CD16⁺ neutrophils (see Table 2.6 for PCR primer sets) and data indicated that healthy controls, active AAV and CKD patients expressed both *PRTN3-201* and *PRTN3-202* transcripts.
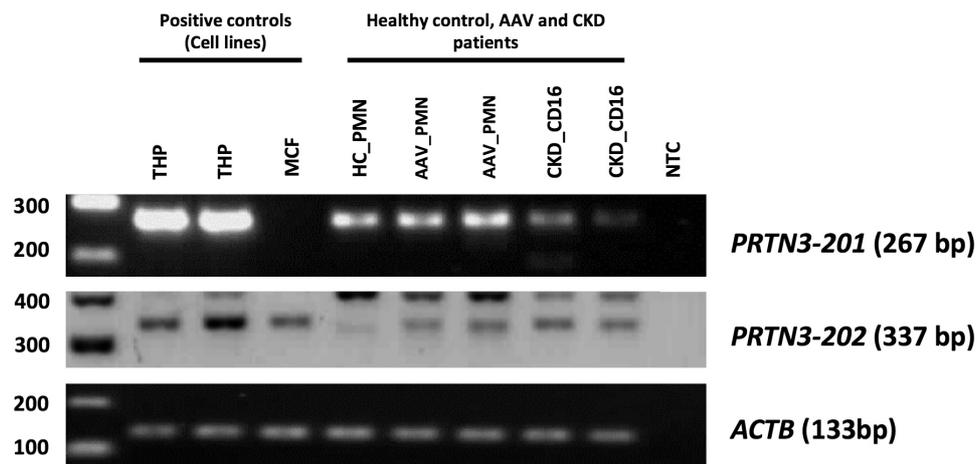
Figure 5.7: Detection of *PRTN3-202* transcript in PMN and CD16⁺ neutrophils

THP-1 cells express both *PRTN3-201* (267 bp) and *PRTN3-202* (337 bp) while MCF cells express only *PRTN3-202*. They were used as the positive controls. AAV, ANCA-associated vasculitis; CKD, chronic kidney disease; HC, healthy controls; NTC, no template control; PMN, polymorphonuclear cells. The target PCR bands were purified prior to Sanger sequencing to confirm the cDNA sequence. The upper bands (~400bp) in the middle panel were non-specific bands.

5.3.4   rs62132296:T allele is associated with a higher expression of *PRTN3-201* in neutrophils

To determine which isoform of *PRTN3* was aberrantly expressed, I performed qPCR with RNA from isolated CD16⁺ neutrophils from CKD patients. Two TaqMan® gene expression assays were used including Hs00160521_m1 where the probe spans exons 1 and 2, detecting only *PRTN3-201* isoform and Hs01553330_m1 where the probe spans exons 3 and 4, allowing the detection both *PRTN3-201* and *PRTN3-202* isoforms. These assays validated the microarray findings (section 5.3.2) where genotype at rs62132296 had an allelic effect on *PRTN3* expression in neutrophils. qPCR analysis using the former probe showed the cellular *PRTN3-*

201 transcript level to be significantly higher in risk allele (T) homozygotes than in individuals with CC or CT genotypes (ANOVA, p = 0.0005; test for linear trend, p = 0.0005). A similar observation was noted with the latter probe (ANOVA, p = 0.0001; test for linear trend, p < 0.0001; Figure 5.8A). However, it was not possible to determine the relative expression of *PRTN3-202* by deducting the C_T of these two probes due to different PCR efficiency. The data imply that the risk allele is associated with a higher expression of *PRTN3-201*, but we could not conclude whether rs62132296 has an allelic effect on *PRTN3-202*. There is a possibility that the differences that we observed using the latter probe could be purely driven by *PRTN3-201* rather than both *PRTN3-201* and *PRTN3-202*.

To further pursue the question of whether rs62132296 or the causal variant in high LD with it had an allelic effect on *PRTN3-202*, I generated RNA-Seq gene expression data with good quality RNA from isolated CD16[+] neutrophils of CKD patients, comprising eight major allele homozygotes, seven heterozygotes and four minor allele homozygotes. Following QC of the raw data, an average of 28.7 million reads were obtained per library and gene expression levels were evaluated by counting the number of transcripts per kilobase million (TPM). Despite having a smaller sample size, a differential *PRTN3* expression was validated (ANOVA, p = 0.0625; test for linear trend, p = 0.0249; Figure 5.8B) and the results were comparable with those obtained using microarray and qPCR analyses.

I then made Sashimi plots, a quantitative visualisation of aligned RNA-seq reads that allows comparison of exon usage across all samples using the Broad Institute Integrative Genomics Viewer, enabling rapid differential isoform usage analysis. Each splice junction track represents one sample and it calls a splicing event when at least a single read splits across two exons in the alignment track and each splice junction is represented by an arc from the beginning to the end of the junction (Figure 5.8C). Despite relatively high sequencing depth, no splicing events between exons 1 and 2 were detected for all except one sample and the overall numbers of reads spliced across any of the other splice junctions were low. These read counts were not able to capture differential isoform usage.
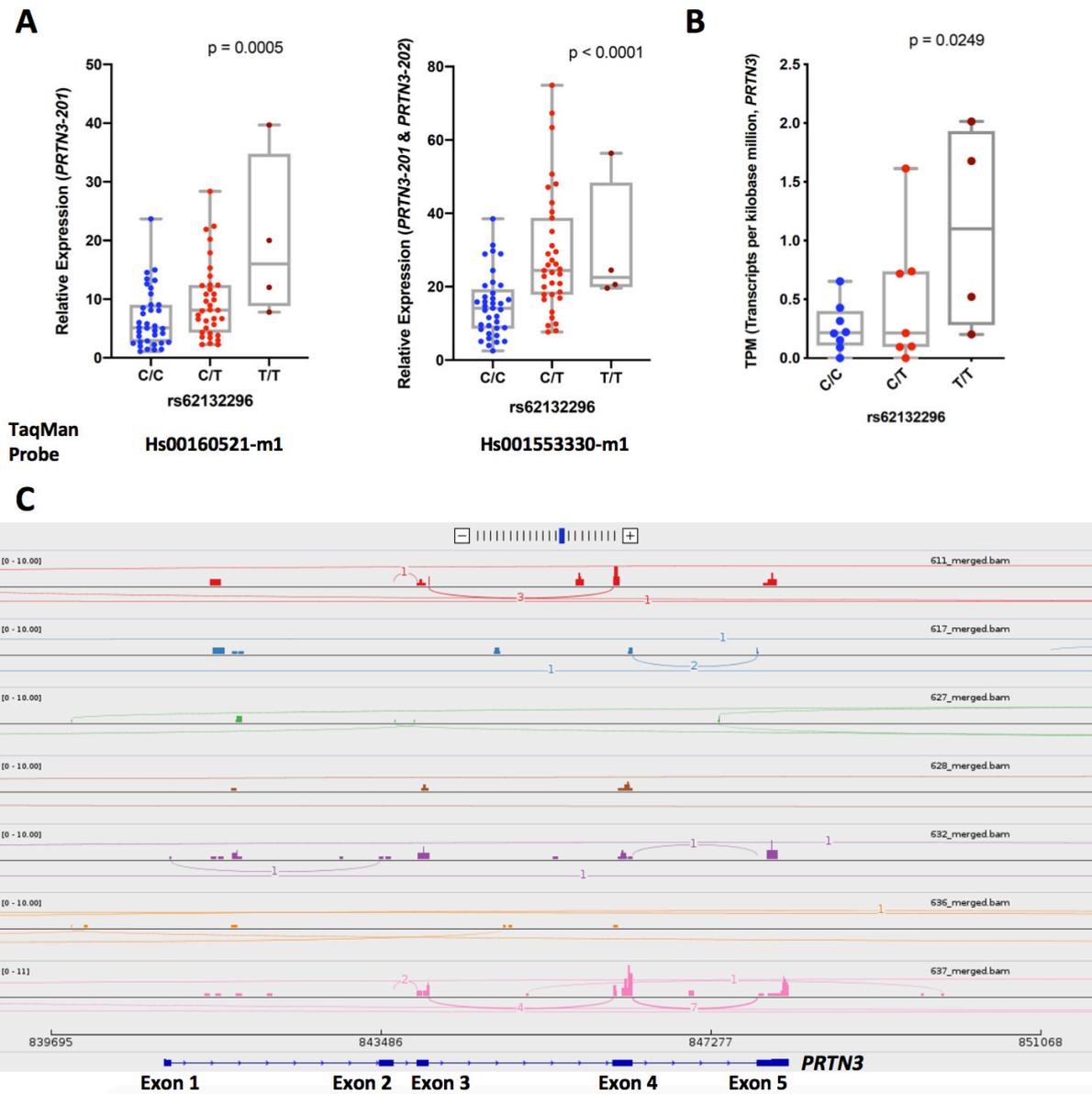
Figure 5.8: rs62132296:T is associated with a higher expression of *PRTN3-201*.

(A) qPCR using the TaqMan genotyping assays with probes targeting different exons showing differential expression of *PRTN3-201* by genotype at rs62132296 (left: test for linear trend, p = 0.0005) and combined expression of *PRTN3-201* and *PRTN3-202* (right: test for linear trend, p < 0.0001).

(B) *PRTN3* expression quantification at gene level using the RNA-seq data of CD16[+] neutrophils showing comparable findings as microarray and qPCR analyses where higher expression of *PRTN3* was observed in patients bearing homozygous risk genotype at rs62132296 (test for linear trend, p = 0.0249).

(C) Representative Sashimi plot consisting 7 samples for *PRTN3* region generated using Broad Institute Integrative Genomics Viewer. Genomic coordinates are plotted on x-axis and read density on y-axis. The raw junction read counts are shown on top of each junction arch.

149

5.4 Discussion

Understanding the fundamental mechanisms of how genetic variants might exert their pathogenic effects is essential for the development of novel diagnostic and treatment avenues. Translating GWAS results into biologically relevant effects remains a major challenge, although these barriers have been gradually overcome with recent advances in technology, statistical and bioinformatics approaches, along with publicly available genomic databases generated by large consortia. We demonstrated that PR3-AAV and eosinophil count shared the same causal variant at *BCL2L11-MIR4435-2HG* locus. Importantly, for the first time, we established a causal association between higher eosinophil count and increased risk of PR3-AAV. We further explored the functional implication of rs13405741 (or the causal variant in high LD with it) in eosinophils by recalling healthy volunteers by genotype and found that the risk allele homozygotes expressed higher levels of *MIR4435-2HG* transcript with an appropriately lower eosinophil apoptosis rate, supporting our hypothesis that the risk variant is associated with a higher expression of *MIR4435-2HG* in eosinophils, which results in downregulation of *BCL2L11* transcription and eventually, reduced apoptosis. Prolongation of cell survival leads to a higher eosinophil count which underpins the pathogenesis of PR3-AAV. Although we did not detect any differences in *BCL2L11* mRNA levels by genotype in our experiments, it does not exclude a possible role of *BCL2L11* for this biological process as the negative results might be attributed to (i) the complex regulatory network for *BCL2L11* expression, (ii) suboptimal harvest time of samples to detect a significant difference if it is present and (iii) insufficient power to detect a small effect size.

According to the 1990 ACR classification criteria, eosinophilia (more than 10% of peripheral white blood cells) and extravascular eosinophilia are two of the six criteria used to distinguish EGPA from other vasculitis (458). However, these two features are not specific to EGPA and have been observed in other types of small and medium-vessel vasculitis, most notably GPA (459). Yousem and Lombard were the first to coin the term "eosinophilic variant of GPA" in 1988 when they described four cases of atypical histological variant of GPA (460) and several studies of similar description have been reported in the literature where tissue eosinophilic infiltration was universally observed in all cases, primarily involving the lungs, nasal, orbital and mastoid regions, with evidence of PR3-ANCA (those who had their ANCA status tested)

and absence of allergy and asthma except for one patient (460-467). Furthermore, Balding and colleagues demonstrated a Th2 dominant cytokine millieu within the nasal mucosa of GPA patients as compared to healthy and disease controls, where higher levels of CD3$^+$ T cells and eosinophils, along with the observation of higher IL-4 but lower IL-2 expression in GPA patients (468). Schnabel et al found higher eosinophil cationic protein levels in the bronchoalveolar lavage fluid from GPA patients but below the detection limit in patients with sarcoidosis, which is a granulomatous disease without vasculitis involvement and healthy controls, suggestive of eosinophil activation and degranulation (469). Nonetheless, it is difficult to ascertain the precise role of eosinophils in disease pathogenesis as these studies are predominantly based on case reports. With the application of MR methodology, our study provides a clear genetic evidence to support the causal role of higher eosinophil count in increasing the risk of PR3-AAV.

Eosinophils were first described by Paul Ehrlich in 1879 due to high affinity of its secondary granules for red dye eosin (470) and more than 90% of eosinophils reside in the tissues, particularly bone marrow, gastrointestinal tract, thymus and the endometrial lining of the uterus (471). Eosinophil differentiation and expansion occur in the bone marrow under the influence of a number of transcription factors (GATA-binding factor 1, PU.1 and CCAAT-enhancer-binding proteins) and several cytokines (IL-3, IL-5 and GM-CSF) (459). The secondary granules contain cationic proteins including major basic protein, eosinophilic cationic protein, eosinophil peroxidase and eosinophil derived neurotoxin that mediate cell death, as well as a vast range of preformed proinflammatory cytokines, growth factors and enzymes, which are released upon cell activation. Well known for their important roles in host defence against helminths and allergic inflammation such as asthma, Lee and colleagues recently hypothesised that resident tissue eosinophils might be essential for local immunity, remodelling and repair in both health and disease (471), as accumulating data supporting their immunomodulatory abilities through expression of cytokines and growth factors characterised as Th1/Th17, Th2 as well as acute phase pro-inflammatory and anti-inflammatory (472). In addition, eosinophils are critical in maintaining the plasma cell survival by secreting APRIL (A PRoliferation-Inducing Ligand) and IL-6. Low plasma cell numbers had

been observed in eosinophil-deficient mice both at steady state and after immunisation, which improved momentarily following eosinophil reconstitution (473).

The pathogenic role of eosinophils in vasculitis is best characterised in EGPA, primarily inferred from clinical studies given the lack of widely accepted animal model of eosinophilic vasculitis. Its role remains underexplored in other forms of vasculitis as peripheral and tissue eosinophilia are not frequently present in these conditions. Hellmark and colleagues recently investigated the functional and activation status of eosinophils in a cohort of 120 healthy controls and 98 GPA/MPA patients, of whom 88% were on active treatment with immunosuppression and 78% in disease remission. The authors reported a lower percentage of eosinophils in the peripheral blood of AAV patients with altered expression of cell surface markers including CD16, CD64, CCR3, CD35, CD11c and CD62L as compared to healthy controls. Although the production of ROS was noted to be lower in the AAV eosinophils, they demonstrated an increased in eosinophil extracellular DNA traps (EETs) formation following ANCA stimulation, postulating that it could lead to deposition of intact eosinophilic granules in tissues and contribute to inflammation (474). However, one must be cautious in interpreting the results as they could potentially be confounded by immunosuppressive treatment particularly glucocorticoids, which are known to induce eosinophil apoptosis by inhibiting survival prolonging signals and influence the levels of NADPH oxidase subunits and ROS (475, 476). Notably, eosinophils have been shown to coexpress eosinophil specific granule proteins and neutrophil primary granule proteins including cathepsin G, azurocidin, lysozyme, neutrophil elastase, MPO and PR3 during early differentiation stages in the bone marrow, which diminish before full maturation (477). *PRTN3* transcript is also detectable in mature eosinophils (BLUEPRINT datasets). Future studies are required to establish the role of eosinophils in the pathogenesis of PR3-AAV though it is tempting to speculate that eosinophil infiltration of inflamed mucosal tissue may support a Th2-mediated inflammatory process and provide a specific microenvironment to prolong plasma cell survival, further enhancing the autoreactive immune response. In addition, eosinophils could increase the availability of PR3 autoantigen via EETs. Eosinophil directed therapy might be a new potential therapeutic approach in AAV, especially in those resistant to currently available treatment regimens along with histopathological evidence of tissue eosinophilic infiltration.

Moreover, we investigated the functional implication of the causal variant at *PRTN3* locus. As we were conducting experiments to assess whether the identified variant exerted its influence by modulating gene expression of *PRTN3*, Merkel and colleagues reported a neutrophil-specific eQTL for *PRTN3* (71) at rs62132293: C > G and demonstrated that cellular *PRTN3* transcript levels were higher in those homozygous for the risk (G) allele than in donors with CC or CG genotypes. Not only did our findings confirm this observation, but we were also able to demonstrate that the canonical transcript (*PRTN3-201*) was differentially expressed by genotype at rs62132296 in neutrophils. These data agree with the proteomic analysis where higher plasma levels of PR3 were observed in individuals carrying the risk alleles (319). Although we detected the expression of the alternative transcript (*PRTN3-202*) in PMN of healthy controls and AAV patients, we were not able to determine differential isoform expression due to low read counts for *PRTN3* in mature neutrophils. *PRTN3-202* transcript has been demonstrated to correlate with disease activity in AAV patients (111) and might potentially be antigenic. Further investigation will be necessary to confirm its role in disease pathogenesis and to address whether the causal variant has an impact on its expression, one will have to consider using bone marrow samples, of which the expression of *PRTN3* in neutrophil precursors can be 100-1000 times higher than in mature neutrophils, facilitating differential isoform analysis.

5.5   Conclusion

This is the first demonstration of genetic evidence supporting a causal role eosinophil number in PR3-AAV and this effect might partly be due to dysregulation of *MIR4435-2HG*, thereby prolonging eosinophil survival. The current paradigm of the pathogenesis of AAV focuses primarily on neutrophils, monocytes and lymphocyte-dependent mechanisms. This study is an important step towards understanding the role of eosinophils in vasculitis other than EGPA. We also validated that the genetic variant at *PRTN3* locus is associated with dysregulation of autoantigen expression.

6 GENERAL DISCUSSION

6.1 Summary and implication of findings

AAV is a rare autoimmune disorder that is associated with substantial morbidity and mortality. The aim of this study was to unravel the genetic basis of PR3-AAV and MPO-AAV independently in order to improve our understanding about the pathogenesis of disease and to pave the way for the development of new diagnostic and therapeutic approaches. We conducted a meta-analysis of two AAV GWAS datasets, comprising 2,652 cases and 11,947 controls of European ancestry. We replicated previously identified genome-wide significant risk loci and discovered seven additional risk loci (Figure 6.1). In addition, fine-mapping of the classical HLA alleles identified three independent signals conferring disease susceptibility for PR3-AAV, including *HLA-DPB1*04:01-HLA-DQB1*06:02-HLA-DRB1*15:01*, *HLA-DPB1*04:02* and *HLA-DPA1*01:03-HLA-DPB1*02:01*, which were distinct from MPO-AAV, conferred by an extended haplotype encoding *HLA-DRB1*04:04-HLA-DQA1*03:01-HLA-DQB1*03:02*. Using Bayesian methods to define credible sets from the GWAS results, together with integration of multi-sourced genomic datasets consisting of functional annotation, gene expression, methylation and PCHi-C data, we were able to identify candidate genes that provide potentially important biological insights into disease pathogenesis.

Further analyses based on genotyped common variants suggested a significant positive genetic correlation between PR3-AAV and MPO-AAV, comparable to that observed between UC and Crohn's disease, supporting that both conditions share a strong genetic architecture. The three risk loci that are common to both PR3-AAV and MPO-AAV (*PTPN22*, *CTLA4* and *DGUOK-TET3*) also share commonalities with a range of other autoimmune diseases and have been implicated in several key immunological pathways. This led us to speculate that the clinical similarities between PR3-AAV and MPO-AAV could be explained by common genetic predisposition in addition to ANCA itself, which, through targeting the neutrophil albeit via different granules antigens, drives a similar pathology.
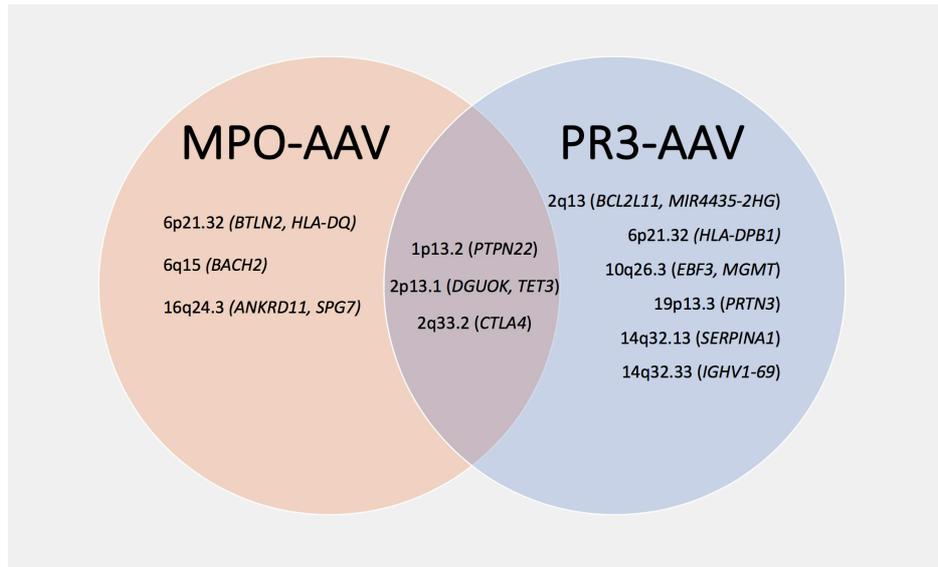
Figure 6.1: The relationship between MPO-AAV and PR3-AAV.

By performing colocalisation analysis, we found that PR3-AAV and eosinophil count had a greater than 98% probability of sharing the same causal variant at the *BCL2L11-MIR4435-2HG* locus. Several studies have consistently reported an atypical histological variant of PR3-AAV associated with tissue eosinophilia, however disentangling causality is challenging as neither eosinophilia nor extravascular eosinophilia are universal clinical features of PR3-AAV. We confirmed a causal association between higher eosinophil numbers and increased risk of PR3-AAV but not MPO-AAV, using two-sample MR and publicly available GWAS summary statistics. This effect could partly be modulated by *MIR4435-2HG* through prolongation of eosinophil survival. Very little is known about the role of eosinophils in the pathogenesis of PR3-AAV. This is an area worthy of further research efforts with potential for becoming a target for new therapeutics.

Finally, we identified an association between a missense variant at the *IGHV1-69* locus (rs11845244) with PR3-AAV. This single point of mutation in the V gene results in a change of amino acid in the framework region 2 (immediately adjacent to the CDR2) of the heavy chain that has previously been shown to completely abolish NEAT2 domain binding which is critical for iron acquisition in *S. aureus* during infection. Our study presents a plausible host genetic factor to support the association between chronic *S. aureus* nasal carriage and PR3-AAV.

6.2 Lessons learnt

The strengths of the current study include the following aspects. This is the largest GWAS of AAV to date and the samples were well phenotyped. We utilised the latest imputation software and a large imputation reference panel, which led to better imputation accuracy, thereby facilitating further discoveries that would have otherwise been challenging to detect. As we recruited patients and controls from a number of European countries, we employed the LMM model for genetic association testing, which has been shown to be better in accounting for population substructure than traditional genomic control using principal component analysis to reduce false positive signals. Through this project, I have developed skills and proficiency in running imputation, genome-wide association tests, visualisation of the GWAS summary statistics and work-flow for loci prioritisation. With the support from the NIHR BioResource Centre Cambridge, we were able to recall age and gender-matched healthy volunteers by genotype to perform functional experiments to uncover the potential mechanisms by which the genetic variants exert their effects.

However, there are some limitations in this study that are worth noting. Firstly, the sample size was still relatively limited given the rare nature of AAV and it was not adequately powered to capture low-frequency and rare variants with small effects sizes. Nonetheless, we have discovered common and a few low-frequency variants with high penetrance, the majority of which are ANCA status specific. Having discovered that both PR3-AAV and MPO-AAV share a strong genetic architecture, a larger sample size would enable discovery of more variants that are common to both. Secondly, the overall eosinophil yield was low during in vitro experiments despite the use of the immunomagnetic purification procedure, which is a highly reproducible method for isolating eosinophils. Therefore, I did not have adequate cells for time-course experiments to investigate time-dependent transcriptional responses under different conditions. As eosinophils represent only a minority of the circulating leukocyte population, this problem could be addressed by sampling a larger volume of whole blood from healthy volunteers. Thirdly, we did not detect differential *PRTN3* isoform expression using RNA-Seq due to low read counts despite acquiring approximately 30 million reads per sample. This is unsurprising as mature neutrophils are generally considered to be transcriptionally inactive. Bone marrow samples will be an ideal research substrate to resolve

this question as *PRTN3* is highly expressed in the myeloid precursor cells. However, it will be immensely challenging to recruit sufficient research participants.

6.3 Future research in the field

GWAS have been tremendously successful in deciphering the genetic component of common complex diseases. Advances in genomic technologies, larger sample sizes and collaborative research generating multi-omics datasets have driven the evolution of bioinformatics, bridging the gap between genetic association and biological function and eventually, translating toward new therapeutics. The data presented here highlight several areas for further research.

6.3.1 Discovering additional genetic variants

The first step is to increase the number of individuals within a sample because a larger sample size will provide the greatest effect gain on power. Although replacing high density SNP chips with whole genome sequencing will tag low-frequency loci, this alone will not be sufficient to capture effects of rare variants as these variants will be present at such low numbers that very large datasets will be required for their detection. Given the rarity of the disease, it is unlikely that we will reach a sufficiently large power, contrary to other common complex diseases such as hypertension and type 2 diabetes, to allow low-frequency and rare variants to be dissected to a far greater extent. An alternative approach is to perform whole genome sequencing in densely affected families, which will enable identification of additional rare variants that can subsequently be validated with a combination of genotyping and deep sequencing of variants in large number of cases and controls.

The differences in the geographical and ethnic distribution of AAV are well reported. To date, GWAS in AAV have only focused on populations of European ancestry. Notably, some of the identified risk variants are monomorphic in the East Asian populations. Therefore, conducting GWAS in different ethnic groups will provide valuable information including identifying genetic determinants that are shared between racial groups or specific to a subpopulation and localisation of functional variant (or variants) by harnessing different LD patterns in

different ethnic groups. In addition, alternative approach (e.g. conditional false discovery rate (cFDR)) could be employed to address the question of genetic overlap as this statistical framework increases the power for discovery of risk loci by leveraging overlapping signals across GWAS.

6.3.2   From association to function

Individual effect sizes at common loci are modest and each risk allele only explains a small fraction of the phenotypic variance of a trait. Computational approaches help to pinpoint the causal variant underlying an association signal but do not provide definitive proof of causality. Rather, it brings forward a reasonable number of strong candidates to undergo functional validation and exploratory experiments in primary cells or animal models to reveal the underlying biological basis of a GWAS signal. This can be facilitated by cutting-edge experimental techniques such as massively parallel reporter assays (MPRA) and various genome-editing tools.

Reporter assays such as luciferase are widely used as a tool to study gene expression at transcriptional level. A DNA region of interest (containing SNP or SNPs) can be cloned either upstream or downstream of a promoter reporter gene cassette in a plasmid, which is then transfected into a cell type of interest. The protein product of the reporter gene is measured to determine allelic effects of SNP on gene regulation. However, this is not an efficient method to screen a large number of candidate causal variants. MPRA is a high-throughput platform that allows the interrogation of the transcriptional regulatory activity of hundreds of sequences simultaneously, by synthesising a large number of oligonucleotides containing a short genomic DNA sequence of interest in parallel. Briefly, each oligonucleotide is coupled to a unique barcode tag and subcloned into a fixed reporter plasmid backbone. The pool of reporter plasmids is then introduced into cells or tissues and incubated for a sufficient period of time to allow the reporters to be fully expressed, followed by high-throughput RNA sequencing and counting of the barcode tags to determine the relative regulatory activities of all the genomic DNA sequences (478). Similar to the individual reporter assay experiments, MPRA lacks the complete chromatin structure of endogenous genomic DNA and is therefore

unable to capture short-range or long-range chromatin interactions that might significantly influence gene expression.

Advances in genome editing technologies have greatly facilitated research progress in studying the biological mechanisms in health and disease. One good example is the clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9) system that has become widely applied in basic science research due its efficacy and ease of use, allowing investigators to introduce, remove or alter specific DNA sequences with a high degree of target specificity within a controlled genetic background. The CRISPR-Cas9 system was adapted from naturally occurring bacterial adaptive immune systems, comprising both protein and RNA components to destroy invading viral and plasmids DNA by generating a double-strand break in the DNA (479). To explore the direct effect of genotype on cellular phenotype, one could use CRISPR-Cas9 system to knock in alternate SNP alleles into cellular or animal models to generate variant and wild-type models that are isogenic. This method, however can be very inefficient and an alternative approach is to knock out the SNP site. It is important to note that any differences between the wild-type and variant models cannot be attributed to the SNP itself, rather the sequence containing the SNP allele. Furthermore, CRISPR-Cas9 technology has been adapted to study transcriptional regulation where the Cas9 protein is altered to be catalytically dead but still forming a protein RNA-DNA complex with specifically engineered guide RNAs and an additional transcriptional activator (CRISPR activation) (480) or repressor (CRISPR interference) (481) domain to alter the target gene expression. For example, by positioning Cas9 at the site of a SNP with or without additional domain, the regulatory activity can be inhibited by steric interference due to reduced binding sites for endogenous regulatory factors. These latest technologies hold promise in expediting future efforts to understand the genetic basis of human disease and clinical translation.

6.3.3   The role of eosinophils in disease pathogenesis of PR3-AAV

Our data support a causal role of eosinophil count in the development of PR3-AAV. Eosinophils have long been perceived as non-specific destructive and cytotoxic cells with prominent roles in asthmatic lung disease and host defences against helminth infections. However, there is emerging evidence to support their diverse functions, including the ability

to polarise T-cell function (through expression of Th1/2 cytokines), to modulate humoral response and to promote immunosuppressive activity as well as tissue remodelling and repair, all of which have been underappreciated in health and disease (482). The functional role of eosinophils in PR3-AAV is a largely unexplored area and greater research endeavours in this area may provide further insight into new therapeutic opportunities. Newer techniques such as single cell RNA-Seq and mass cytometry by time of flight (CyTOF) have the ability to profile individual cells of a heterogeneous population and measure more than 40 cell surface markers on individual cells, allowing interrogation of cellular phenotype, function and signalling status simultaneously. Furthermore, the development of monoclonal antibodies recognising anti-eosinophil peroxidase has enabled the detection and quantification of tissue infiltrating eosinophils, as well as evidence of degranulation, which is far more specific and rigorous than conventional evaluation using haematoxylin and eosin-stained slides and assays detecting eosinophil cationic protein or eosinophil-derived neurotoxin which are also present in other leukocytes such as neutrophils (483).

We further explored the functional implication of the risk locus at chr2p13 and three genes including *ACOXL*, *BCL2L11* and *MIR4435-2HG* were found to interact with credible SNPs in the granulocyte PCHi-C data. We demonstrated that the genotype at rs13405741 was associated with dysregulation of *MIR4435-2HG* which led to prolongation of eosinophil survival. Although no difference in *BCL2L11* expression between groups was present, our results have been encouraging to merit further investigation to understand how the most likely causal variant and potentially, DNA elements within this locus regulate the expression of *MIR4435-2HG* and *BCL2L11*.

In conclusion, the aims of this thesis have been achieved. We have established a robust set of causal variants in AAV and have laid the groundwork for future research to relate genetic variation to the aetiopathogenesis of AAV.

REFERENCES

1. Rokitansky K. On some of the most important diseases of arteries. Denkschr Kais Akad Wissensch 1852;4:1-721.

2. Kussmaul A, Maier R. About a not previously described peculiar arterial disease (periarteritis nodosa), that coincides with morbus brightii and rapidly progressive muscular palsy. Deutsch Arch Klin Med 1866;1:484-518.

3. Ferrari E. About polyarteritis nodosa (so-called periarteriitis nodosa) and its relations to polymyositis and acute polyneuritis. Beitr Pathol Anat 1903;34:350-386.

4. Wohlwill FJ. About a form of periarteriitis nodosa that can only be detected histologically. Virchows Arch Pathol Anat Physiol Klin Med 1923;246:377-411.

5. Klinger H. Borderline variants of periarteriitis nodosa. Frankf Z Pathol 1931;42:455-480.

6. Wegener F. About a perculiar rhinogenic granulomatosis with marked involvement of the arterial system and kidneys. Beitr Pathol Anat 1939;102:30-68.

7. Godman GC, Churg J. Wegener's granulomatosis: Pathology and review of the literature. AMA Arch Pathol 1954;58(6):533-553.

8. Woywodt A, Haubitz M, Haller H et al. Wegener's granulomatosis. Lancet 2006;367(9519):1362-1366.

9. Davies DJ, Moran JE, Niall JF et al. Segmental necrotising glomerulonephritis with antineutrophil antibody: Possible arbovirus aetiology? Br Med J (Clin Res Ed) 1982;285(6342):606.

10. Hall JB, Wadham BM, Wood CJ et al. Vasculitis and glomerulonephritis: A subgroup with an antineutrophil cytoplasmic antibody. Aust N Z J Med 1984;14(3):277-278.

11. van der Woude FJ, Rasmussen N, Lobatto S et al. Autoantibodies against neutrophils and monocytes: Tool for diagnosis and marker of disease activity in wegener's granulomatosis. Lancet 1985;1(8426):425-429.

12. Falk RJ, Jennette JC. Anti-neutrophil cytoplasmic autoantibodies with specificity for myeloperoxidase in patients with systemic vasculitis and idiopathic necrotizing and crescentic glomerulonephritis. N Engl J Med 1988;318(25):1651-1657.

13.   Goldschmeding R, van der Schoot CE, ten Bokkel Huinink D et al. Wegener's granulomatosis autoantibodies identify a novel diisopropylfluorophosphate-binding protein in the lysosomes of normal human neutrophils. J Clin Invest 1989;84(5):1577-1587.

14.   Tervaert JW, Elema JD, Kallenberg CG. Clinical and histopathological association of 29kd-anca and mpo-anca. APMIS Suppl 1990;19:35.

15.   Guillevin L, Visser H, Noel LH et al. Antineutrophil cytoplasm antibodies in systemic polyarteritis nodosa with and without hepatitis b virus infection and churg-strauss syndrome--62 patients. J Rheumatol 1993;20(8):1345-1349.

16.   Schönlein JL. Allegemeine und specielle pathologie und therapie. Literatur-Comptoir 1837;2:48.

17.   Henoch E. Uber den zusammenhang von purpura und intestinal-stoerungen. Berl Klin Wochenschur 1868;5:517-519.

18.   Henoch E. Lectures on diseases of children: A handbook for physicians and students. W Wood & Co 1882.

19.   Osler W. On the visceral complications of erythema eudativum multiforme. Am J Med Sci 1895;2:629-647.

20.   Osler W. The visceral lesions of purpura and allied conditions. Br Med J 1914;1:517-525.

21.   Faille-Kuyber EH, Kater L, Kooiker CJ et al. Iga-deposits in cutaneous blood-vessel walls and mesangium in henoch-schonlein syndrome. Lancet 1973;1(7808):892-893.

22.   Jennette JC. L17. What can we expect from the revised chapel hill consensus conference nomenclature of vasculitis? Presse Med 2013;42(4 Pt 2):550-555.

23.   Hunder GG, Arend WP, Bloch DA et al. The american college of rheumatology 1990 criteria for the classification of vasculitis. Introduction. Arthritis Rheum 1990;33(8):1065-1067.

24.   Seeliger B, Sznajd J, Robson JC et al. Are the 1990 american college of rheumatology vasculitis classification criteria still valid? Rheumatology (Oxford) 2017.

25.   Jennette JC, Falk RJ, Bacon PA et al. 2012 revised international chapel hill consensus conference nomenclature of vasculitides. Arthritis Rheum 2013;65(1):1-11.

26.   Watts R, Lane S, Hanslik T et al. Development and validation of a consensus methodology for the classification of the anca-associated vasculitides and polyarteritis nodosa for epidemiological studies. Ann Rheum Dis 2007;66(2):222-227.

27.   Liu LJ, Chen M, Yu F et al. Evaluation of a new algorithm in classification of systemic vasculitis. Rheumatology (Oxford) 2008;47(5):708-712.

28.   Mahr A, Katsahian S, Varet H et al. Revisiting the classification of clinical phenotypes of anti-neutrophil cytoplasmic antibody-associated vasculitis: A cluster analysis. Ann Rheum Dis 2013;72(6):1003-1010.

29.   Luqmani RA, Suppiah R, Grayson PC et al. Nomenclature and classification of vasculitis - update on the acr/eular diagnosis and classification of vasculitis study (dcvas). Clin Exp Immunol 2011;164 Suppl 1:11-13.

30.   Haugeberg G, Bie R, Bendvold A et al. Primary vasculitis in a norwegian community hospital: A retrospective study. Clin Rheumatol 1998;17(5):364-368.

31.   Gonzalez-Gay MA, Garcia-Porrua C, Guerrero J et al. The epidemiology of the primary systemic vasculitides in northwest spain: Implications of the chapel hill consensus conference definitions. Arthritis Rheum 2003;49(3):388-393.

32.   Watts RA, Gonzalez-Gay MA, Lane SE et al. Geoepidemiology of systemic vasculitis: Comparison of the incidence in two regions of europe. Ann Rheum Dis 2001;60(2):170-172.

33.   el-Reshaid K, Kapoor MM, el-Reshaid W et al. The spectrum of renal disease associated with microscopic polyangiitis and classic polyarteritis nodosa in kuwait. Nephrol Dial Transplant 1997;12(9):1874-1882.

34.   Oh JS, Lee CK, Kim YG et al. Clinical features and outcomes of microscopic polyangiitis in korea. J Korean Med Sci 2009;24(2):269-274.

35.   Chen M, Yu F, Zhang Y et al. Clinical [corrected] and pathological characteristics of chinese patients with antineutrophil cytoplasmic autoantibody associated systemic vasculitides: A study of 426 patients from a single centre. Postgrad Med J 2005;81(961):723-727.

36.   Fujimoto S, Watts RA, Kobayashi S et al. Comparison of the epidemiology of anti-neutrophil cytoplasmic antibody-associated vasculitis between japan and the u.K. Rheumatology (Oxford) 2011;50(10):1916-1920.

37.   O'Donnell JL, Stevanovic VR, Frampton C et al. Wegener's granulomatosis in new zealand: Evidence for a latitude-dependent incidence gradient. Intern Med J 2007;37(4):242-246.

38.   Mahr A, Guillevin L, Poissonnet M et al. Prevalences of polyarteritis nodosa, microscopic polyangiitis, wegener's granulomatosis, and churg-strauss syndrome in a french urban multiethnic population in 2000: A capture-recapture estimate. Arthritis Rheum 2004;51(1):92-99.

39.   Hogan SL, Falk RJ, Chin H et al. Predictors of relapse and treatment resistance in antineutrophil cytoplasmic antibody-associated small-vessel vasculitis. Ann Intern Med 2005;143(9):621-631.

40.   Katsuyama T, Sada KE, Makino H. Current concept and epidemiology of systemic vasculitides. Allergol Int 2014;63(4):505-513.

41.   Cantorna MT, Mahon BD. Mounting evidence for vitamin d as an environmental factor affecting autoimmune disease prevalence. Exp Biol Med (Maywood) 2004;229(11):1136-1142.

42.   Stegeman CA, Tervaert JW, Sluiter WJ et al. Association of chronic nasal carriage of staphylococcus aureus and higher relapse rates in wegener granulomatosis. Ann Intern Med 1994;120(1):12-17.

43.   Kain R, Exner M, Brandes R et al. Molecular mimicry in pauci-immune focal necrotizing glomerulonephritis. Nat Med 2008;14(10):1088-1096.

44.   Lane SE, Watts RA, Bentham G et al. Are environmental factors important in primary systemic vasculitis? A case-control study. Arthritis Rheum 2003;48(3):814-823.

45.   Pelclova D, Bartunkova J, Fenclova Z et al. Asbestos exposure and antineutrophil cytoplasmic antibody (anca) positivity. Arch Environ Health 2003;58(10):662-668.

46.   Aloush V, Litinsky I, Caspi D et al. Propylthiouracil-induced autoimmune syndromes: Two distinct clinical presentations with different course and management. Semin Arthritis Rheum 2006;36(1):4-9.

47.   Pillinger MH, Staud R. Propylthiouracil and antineutrophil cytoplasmic antibody associated vasculitis: The detective finds a clue. Semin Arthritis Rheum 2006;36(1):1-3.

48.   Almroth G, Enestrom S, Hed J et al. Autoantibodies to leucocyte antigens in hydralazine-associated nephritis. J Intern Med 1992;231(1):37-42.

49. Flendrie M, Vissers WH, Creemers MC et al. Dermatological conditions during tnf-alpha-blocking therapy in patients with rheumatoid arthritis: A prospective study. Arthritis Res Ther 2005;7(3):R666-676.

50. Graf J, Lynch K, Yeh CL et al. Purpura, cutaneous necrosis, and antineutrophil cytoplasmic antibodies associated with levamisole-adulterated cocaine. Arthritis Rheum 2011;63(12):3998-4001.

51. Muniain MA, Moreno JC, Gonzalez Campora R. Wegener's granulomatosis in two sisters. Ann Rheum Dis 1986;45(5):417-421.

52. Hay EM, Beaman M, Ralston AJ et al. Wegener's granulomatosis occurring in siblings. Br J Rheumatol 1991;30(2):144-145.

53. Franssen CF, ter Maaten JC, Hoorntje SJ. Brother and sister with myeloperoxidase associated autoimmune disease. Ann Rheum Dis 1994;53(3):213.

54. Nowack R, Lehmann H, Flores-Suarez LF et al. Familial occurrence of systemic vasculitis and rapidly progressive glomerulonephritis. Am J Kidney Dis 1999;34(2):364-373.

55. Hull CM, Couser WG, Knostman JD. A familial case of p-anca glomerulonephritis presenting in a father and daughter. Am J Kidney Dis 2000;35(5):E23.

56. Sollid LM, Pos W, Wucherpfennig KW. Molecular mechanisms for contribution of mhc molecules to autoimmune diseases. Curr Opin Immunol 2014;31:24-30.

57. Chapman SJ, Hill AV. Human genetic susceptibility to infectious disease. Nat Rev Genet 2012;13(3):175-188.

58. Evans DM, Spencer CC, Pointon JJ et al. Interaction between erap1 and hla-b27 in ankylosing spondylitis implicates peptide handling in the mechanism for hla-b27 in disease susceptibility. Nat Genet 2011;43(8):761-767.

59. Grumet FC, Coukell A, Bodmer JG et al. Histocompatibility (hl-a) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease. N Engl J Med 1971;285(4):193-196.

60. Cortes A, Pulit SL, Leo PJ et al. Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with erap1. Nat Commun 2015;6:7146.

61. Haroon N, Inman RD. Endoplasmic reticulum aminopeptidases: Biology and pathogenic potential. Nat Rev Rheumatol 2010;6(8):461-467.

62.   Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annu Rev Genomics Hum Genet 2013;14:301-323.

63.   Wieczorek M, Abualrous ET, Sticht J et al. Major histocompatibility complex (mhc) class i and mhc class ii proteins: Conformational plasticity in antigen presentation. Front Immunol 2017;8:292.

64.   Gencik M, Borgmann S, Zahn R et al. Immunogenetic risk factors for anti-neutrophil cytoplasmic antibody (anca)-associated systemic vasculitis. Clin Exp Immunol 1999;117(2):412-417.

65.   Hagen EC, Stegeman CA, D'Amaro J et al. Decreased frequency of hla-dr13dr6 in wegener's granulomatosis. Kidney Int 1995;48(3):801-805.

66.   Heckmann M, Holle JU, Arning L et al. The wegener's granulomatosis quantitative trait locus on chromosome 6p21.3 as characterised by tagsnp genotyping. Ann Rheum Dis 2008;67(7):972-979.

67.   Jagiello P, Gencik M, Arning L et al. New genomic region for wegener's granulomatosis as revealed by an extended association screen with 202 apoptosis-related genes. Hum Genet 2004;114(5):468-477.

68.   Kawasaki A, Hasebe N, Hidaka M et al. Protective role of hla-drb1*13:02 against microscopic polyangiitis and mpo-anca-positive vasculitides in a japanese population: A case-control study. PLoS One 2016;11(5):e0154393.

69.   Luo H, Chen M, Yang R et al. The association of hla-drb1 alleles with antineutrophil cytoplasmic antibody-associated systemic vasculitis in chinese patients. Hum Immunol 2011;72(5):422-425.

70.   Lyons PA, Rayner TF, Trivedi S et al. Genetically distinct subsets within anca-associated vasculitis. N Engl J Med 2012;367(3):214-223.

71.   Merkel PA, Xie G, Monach PA et al. Identification of functional and expression polymorphisms associated with risk for anti-neutrophil cytoplasmic autoantibody-associated vasculitis. Arthritis Rheumatol 2016.

72.   Papiha SS, Murty GE, Ad'Hia A et al. Association of wegener's granulomatosis with hla antigens and other genetic markers. Ann Rheum Dis 1992;51(2):246-248.

73.   Spencer SJ, Burns A, Gaskin G et al. Hla class ii specificities in vasculitis with antibodies to neutrophil cytoplasmic antigens. Kidney Int 1992;41(4):1059-1063.

74.	Stassen PM, Cohen-Tervaert JW, Lems SP et al. Hla-dr4, dr13(6) and the ancestral haplotype a1b8dr3 are associated with anca-associated vasculitis and wegener's granulomatosis. Rheumatology (Oxford) 2009;48(6):622-625.

75.	Tsuchiya N, Kobayashi S, Hashimoto H et al. Association of hla-drb1*0901-dqb1*0303 haplotype with microscopic polyangiitis in japanese. Genes Immun 2006;7(1):81-84.

76.	Tsuchiya N, Kobayashi S, Kawasaki A et al. Genetic background of japanese patients with antineutrophil cytoplasmic antibody-associated vasculitis: Association of hla-drb1*0901 with microscopic polyangiitis. J Rheumatol 2003;30(7):1534-1540.

77.	Wu Z, Wu Q, Xu J et al. Hla-dpb1 variant rs3117242 is associated with anti-neutrophil cytoplasmic antibody-associated vasculitides in a han chinese population. Int J Rheum Dis 2015.

78.	Xie G, Roshandel D, Sherva R et al. Association of granulomatosis with polyangiitis (wegener's) with hla-dpb1*04 and sema6a gene variants: Evidence from genome-wide analysis. Arthritis Rheum 2013;65(9):2457-2468.

79.	Tsuchiya N. Genetics of anca-associated vasculitis in japan: A role for hla-drb1*09:01 haplotype. Clin Exp Nephrol 2013;17(5):628-630.

80.	Borgmann S, Endisch G, Urban S et al. A linkage disequilibrium between genes at the serine protease inhibitor gene cluster on chromosome 14q32.1 is associated with wegener's granulomatosis. Clin Immunol 2001;98(2):244-248.

81.	Elzouki AN, Segelmark M, Wieslander J et al. Strong link between the alpha 1-antitrypsin piz allele and wegener's granulomatosis. J Intern Med 1994;236(5):543-548.

82.	Mahr AD, Edberg JC, Stone JH et al. Alpha(1)-antitrypsin deficiency-related alleles z and s and the risk of wegener's granulomatosis. Arthritis Rheum 2010;62(12):3760-3767.

83.	Pervakova MY, Emanuel VL, Titova ON et al. The diagnostic value of alpha-1-antitrypsin phenotype in patients with granulomatosis with polyangiitis. Int J Rheumatol 2016;2016:7831410.

84.	Hernandez Perez JM, Fumero Garcia S, Alvarez Pio A. Successful alpha1-antitrypsin replacement therapy in a patient with alpha1-antitrypsin deficiency and granulomatosis with polyangiitis. Rheumatology (Oxford) 2013;52(4):755-757.

85.   Harper K, Balzano C, Rouvier E et al. Ctla-4 and cd28 activated lymphocyte molecules are closely related in both mouse and human as to sequence, message expression, gene structure, and chromosomal location. J Immunol 1991;147(3):1037-1044.

86.   Tivol EA, Borriello F, Schweitzer AN et al. Loss of ctla-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of ctla-4. Immunity 1995;3(5):541-547.

87.   Abdel-Wahab N, Shah M, Suarez-Almazor ME. Adverse events associated with immune checkpoint blockade in patients with cancer: A systematic review of case reports. PLoS One 2016;11(7):e0160221.

88.   Eyre S, Bowes J, Diogo D et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44(12):1336-1340.

89.   Gregersen PK, Amos CI, Lee AT 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):820-823.

90.   Okada Y, Wu D, Trynka G et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506(7488):376-381.

91.   Barrett JC, Clayton DG, Concannon P et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41(6):703-707.

92.   Onengut-Gumuscu S, Chen WM, Burren O et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 2015;47(4):381-386.

93.   Plagnol V, Howson JM, Smyth DJ et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. PLoS Genet 2011;7(8):e1002216.

94.   Chu X, Pan CM, Zhao SX et al. A genome-wide association study identifies two new risk loci for graves' disease. Nat Genet 2011;43(9):897-901.

95.   Cooper JD, Simmonds MJ, Walker NM et al. Seven newly identified loci for autoimmune thyroid disease. Hum Mol Genet 2012;21(23):5202-5208.

96.   Eriksson N, Tung JY, Kiefer AK et al. Novel associations for hypothyroidism include known autoimmune risk loci. PLoS One 2012;7(4):e34442.

97.   Dubois PC, Trynka G, Franke L et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 2010;42(4):295-302.

98. Petukhova L, Duvic M, Hordinsky M et al. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature 2010;466(7302):113-117.

99. Ellinghaus D, Jostins L, Spain SL et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 2016;48(5):510-518.

100. Carr EJ, Niederer HA, Williams J et al. Confirmation of the genetic association of ctla4 and ptpn22 with anca-associated vasculitis. BMC Med Genet 2009;10:121.

101. Chung SA, Xie G, Roshandel D et al. Meta-analysis of genetic polymorphisms in granulomatosis with polyangiitis (wegener's) reveals shared susceptibility loci with rheumatoid arthritis. Arthritis Rheum 2012;64(10):3463-3471.

102. Giscombe R, Wang X, Huang D et al. Coding sequence 1 and promoter single nucleotide polymorphisms in the ctla-4 gene in wegener's granulomatosis. J Rheumatol 2002;29(5):950-953.

103. Huang D, Giscombe R, Zhou Y et al. Polymorphisms in ctla-4 but not tumor necrosis factor-alpha or interleukin 1beta genes are associated with wegener's granulomatosis. J Rheumatol 2000;27(2):397-401.

104. Kamesh L, Heward JM, Williams JM et al. Ct60 and +49 polymorphisms of ctla 4 are associated with anca-positive small vessel vasculitis. Rheumatology (Oxford) 2009;48(12):1502-1505.

105. Slot MC, Sokolowska MG, Savelkouls KG et al. Immunoregulatory gene polymorphisms are associated with anca-related vasculitis. Clin Immunol 2008;128(1):39-45.

106. Spriewald BM, Witzke O, Wassmuth R et al. Distinct tumour necrosis factor alpha, interferon gamma, interleukin 10, and cytotoxic t cell antigen 4 gene polymorphisms in disease occurrence and end stage renal disease in wegener's granulomatosis. Ann Rheum Dis 2005;64(3):457-461.

107. Wu Z, Wu Q, Xu J et al. Hla-dpb1 variant rs3117242 is associated with anti-neutrophil cytoplasmic antibody-associated vasculitides in a han chinese population. Int J Rheum Dis 2017;20(8):1009-1015.

108. Zhou Y, Huang D, Paris PL et al. An analysis of ctla-4 and proinflammatory cytokine genes in wegener's granulomatosis. Arthritis Rheum 2004;50(8):2645-2650.

109.   Steiner K, Moosig F, Csernok E et al. Increased expression of ctla-4 (cd152) by t and b lymphocytes in wegener's granulomatosis. Clin Exp Immunol 2001;126(1):143-150.

110.   Gabay JE, Almeida RP. Antibiotic peptides and serine protease homologs in human polymorphonuclear leukocytes: Defensins and azurocidin. Curr Opin Immunol 1993;5(1):97-102.

111.   McInnis EA, Badhwar AK, Muthigi A et al. Dysregulation of autoantigen genes in anca-associated vasculitis involves alternative transcripts and new protein synthesis. J Am Soc Nephrol 2015;26(2):390-399.

112.   Bories D, Raynal MC, Solomon DH et al. Down-regulation of a serine protease, myeloblastin, causes growth arrest and differentiation of promyelocytic leukemia cells. Cell 1989;59(6):959-968.

113.   Falk RJ, Terrell RS, Charles LA et al. Anti-neutrophil cytoplasmic autoantibodies induce neutrophils to degranulate and produce oxygen radicals in vitro. Proc Natl Acad Sci U S A 1990;87(11):4115-4119.

114.   Csernok E, Ernst M, Schmitt W et al. Activated neutrophils express proteinase 3 on their plasma membrane in vitro and in vivo. Clin Exp Immunol 1994;95(2):244-250.

115.   Witko-Sarsat V, Lesavre P, Lopez S et al. A large subset of neutrophils expressing membrane proteinase 3 is a risk factor for vasculitis and rheumatoid arthritis. J Am Soc Nephrol 1999;10(6):1224-1233.

116.   Halbwachs-Mecarelli L, Bessou G, Lesavre P et al. Bimodal distribution of proteinase 3 (pr3) surface expression reflects a constitutive heterogeneity in the polymorphonuclear neutrophil pool. FEBS Lett 1995;374(1):29-33.

117.   Schreiber A, Busjahn A, Luft FC et al. Membrane expression of proteinase 3 is genetically determined. J Am Soc Nephrol 2003;14(1):68-75.

118.   Yang JJ, Pendergraft WF, Alcorta DA et al. Circumvention of normal constraints on granule protein gene expression in peripheral blood neutrophils and monocytes of patients with antineutrophil cytoplasmic autoantibody-associated glomerulonephritis. J Am Soc Nephrol 2004;15(8):2103-2114.

119.   Ohlsson S, Hellmark T, Pieters K et al. Increased monocyte transcription of the proteinase 3 gene in small vessel vasculitis. Clin Exp Immunol 2005;141(1):174-182.

120.   Ohlsson S, Wieslander J, Segelmark M. Increased circulating levels of proteinase 3 in patients with anti-neutrophilic cytoplasmic autoantibodies-associated systemic vasculitis in remission. Clin Exp Immunol 2003;131(3):528-535.

121.   Gencik M, Meller S, Borgmann S et al. Proteinase 3 gene polymorphisms and wegener's granulomatosis. Kidney Int 2000;58(6):2473-2477.

122.   Pieters K, Pettersson A, Gullberg U et al. The - 564 a/g polymorphism in the promoter region of the proteinase 3 gene associated with wegener's granulomatosis does not increase the promoter activity. Clin Exp Immunol 2004;138(2):266-270.

123.   Abdgawad M, Hellmark T, Gunnarsson L et al. Increased neutrophil membrane expression and plasma level of proteinase 3 in systemic vasculitis are not a consequence of the - 564 a/g promotor polymorphism. Clin Exp Immunol 2006;145(1):63-70.

124.   Bottini N, Musumeci L, Alonso A et al. A functional variant of lymphoid tyrosine phosphatase is associated with type i diabetes. Nat Genet 2004;36(4):337-338.

125.   Todd JA, Walker NM, Cooper JD et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39(7):857-864.

126.   Begovich AB, Carlton VE, Honigberg LA 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):330-337.

127.   Stahl EA, Raychaudhuri S, Remmers EF et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42(6):508-514.

128.   Kyogoku C, Langefeld CD, Ortmann WA et al. Genetic association of the r620w polymorphism of protein tyrosine phosphatase ptpn22 with human sle. Am J Hum Genet 2004;75(3):504-507.

129.   Gateva V, Sandling JK, Hom G et al. A large-scale replication study identifies tnip1, prdm1, jazf1, uhrf1bp1 and il10 as risk loci for systemic lupus erythematosus. Nat Genet 2009;41(11):1228-1233.

130.   Hinks A, Cobb J, Marion MC et al. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat Genet 2013;45(6):664-669.

131.   Barrett JC, Hansoul S, Nicolae DL et al. Genome-wide association defines more than 30 distinct susceptibility loci for crohn's disease. Nat Genet 2008;40(8):955-962.

132.	Vang T, Miletic AV, Bottini N et al. Protein tyrosine phosphatase ptpn22 in human autoimmunity. Autoimmunity 2007;40(6):453-461.

133.	Cao Y, Yang J, Colby K et al. High basal activity of the ptpn22 gain-of-function variant blunts leukocyte responsiveness negatively affecting il-10 production in anca vasculitis. PLoS One 2012;7(8):e42783.

134.	Jagiello P, Aries P, Arning L et al. The ptpn22 620w allele is a risk factor for wegener's granulomatosis. Arthritis Rheum 2005;52(12):4039-4043.

135.	Martorana D, Maritati F, Malerba G et al. Ptpn22 r620w polymorphism in the anca-associated vasculitides. Rheumatology (Oxford) 2012;51(5):805-812.

136.	Rhee I, Veillette A. Protein tyrosine phosphatases in lymphocyte activation and autoimmunity. Nat Immunol 2012;13(5):439-447.

137.	Wang Y, Shaked I, Stanford SM et al. The autoimmunity-associated gene ptpn22 potentiates toll-like receptor-driven, type 1 interferon-dependent immunity. Immunity 2013;39(1):111-122.

138.	Stanford SM, Rapini N, Bottini N. Regulation of tcr signalling by tyrosine phosphatases: From immune homeostasis to autoimmunity. Immunology 2012;137(1):1-19.

139.	Cloutier JF, Veillette A. Association of inhibitory tyrosine protein kinase p50csk with protein tyrosine phosphatase pep in t cells and other hemopoietic cells. EMBO J 1996;15(18):4909-4918.

140.	de la Puerta ML, Trinidad AG, Rodriguez Mdel C et al. The autoimmunity risk variant lyp-w620 cooperates with csk in the regulation of tcr signaling. PLoS One 2013;8(1):e54569.

141.	Vang T, Liu WH, Delacroix L et al. Lyp inhibits t-cell activation when dissociated from csk. Nat Chem Biol 2012;8(5):437-446.

142.	Zikherman J, Hermiston M, Steiner D et al. Ptpn22 deficiency cooperates with the cd45 e613r allele to break tolerance on a non-autoimmune background. J Immunol 2009;182(7):4093-4106.

143.	Vang T, Congia M, Macis MD et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet 2005;37(12):1317-1319.

144.	Rawlings DJ, Dai X, Buckner JH. The role of ptpn22 risk variant in the development of autoimmunity: Finding common ground between mouse and human. J Immunol 2015;194(7):2977-2984.

145.   Aarnisalo J, Treszl A, Svec P et al. Reduced cd4+t cell activation in children with type 1 diabetes carrying the ptpn22/lyp 620trp variant. J Autoimmun 2008;31(1):13-21.

146.   Rieck M, Arechiga A, Onengut-Gumuscu S et al. Genetic variation in ptpn22 corresponds to altered function of t and b lymphocytes. J Immunol 2007;179(7):4704-4710.

147.   Vang T, Landskron J, Viken MK et al. The autoimmune-predisposing variant of lymphoid tyrosine phosphatase favors t helper 1 responses. Hum Immunol 2013;74(5):574-585.

148.   Ferreira RC, Castro Dopico X, Oliveira JJ et al. Chronic immune activation in systemic lupus erythematosus and the autoimmune ptpn22 trp(620) risk allele drive the expansion of foxp3(+) regulatory t cells and pd-1 expression. Front Immunol 2019;10:2606.

149.   Habib T, Funk A, Rieck M et al. Altered b cell homeostasis is associated with type i diabetes and carriers of the ptpn22 allelic variant. J Immunol 2012;188(1):487-496.

150.   Menard L, Saadoun D, Isnardi I et al. The ptpn22 allele encoding an r620w variant interferes with the removal of developing autoreactive b cells in humans. J Clin Invest 2011;121(9):3635-3644.

151.   Zhang J, Zahir N, Jiang Q 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):902-907.

152.   Dai X, James RG, Habib T et al. A disease-associated ptpn22 variant promotes systemic autoimmunity in murine models. J Clin Invest 2013;123(5):2024-2036.

153.   Krausgruber T, Blazek K, Smallie T et al. Irf5 promotes inflammatory macrophage polarization and th1-th17 responses. Nat Immunol 2011;12(3):231-238.

154.   Sweeney SE. Targeting interferon regulatory factors to inhibit activation of the type i ifn response: Implications for treatment of autoimmune disorders. Cell Immunol 2011;271(2):342-349.

155.   Baechler EC, Batliwalla FM, Karypis G et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A 2003;100(5):2610-2615.

156.   Bennett L, Palucka AK, Arce E et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 2003;197(6):711-723.

157.   Kirou KA, Lee C, George S et al. Coordinate overexpression of interferon-alpha-induced genes in systemic lupus erythematosus. Arthritis Rheum 2004;50(12):3958-3967.

158.   Morris DL, Sheng Y, Zhang Y et al. Genome-wide association meta-analysis in chinese and european individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet 2016;48(8):940-946.

159.   Bentham J, Morris DL, Graham DSC et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 2015;47(12):1457-1464.

160.   Langefeld CD, Ainsworth HC, Cunninghame Graham DS et al. Transancestral mapping and genetic load in systemic lupus erythematosus. Nat Commun 2017;8:16021.

161.   Alarcon-Riquelme ME, Ziegler JT, Molineros J et al. Genome-wide association study in an amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of european admixture. Arthritis Rheumatol 2016;68(4):932-943.

162.   Cordell HJ, Han Y, Mells GF et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat Commun 2015;6:8019.

163.   Liu JZ, Almarri MA, Gaffney DJ et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet 2012;44(10):1137-1141.

164.   Gorlova OY, Li Y, Gorlov I et al. Gene-level association analysis of systemic sclerosis: A comparison of african-americans and white populations. PLoS One 2018;13(1):e0189498.

165.   Radstake TR, Gorlova O, Rueda B et al. Genome-wide association study of systemic sclerosis identifies cd247 as a new susceptibility locus. Nat Genet 2010;42(5):426-429.

166.   Lessard CJ, Li H, Adrianto I et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with sjogren's syndrome. Nat Genet 2013;45(11):1284-1292.

167.   Jostins L, Ripke S, Weersma RK et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491(7422):119-124.

168.   Graham RR, Kozyrev SV, Baechler EC et al. A common haplotype of interferon regulatory factor 5 (irf5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 2006;38(5):550-555.

169.   Graham RR, Kyogoku C, Sigurdsson S et al. Three functional variants of ifn regulatory factor 5 (irf5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci U S A 2007;104(16):6758-6763.

170.   Wieczorek S, Holle JU, Muller S et al. A functionally relevant irf5 haplotype is associated with reduced risk to wegener's granulomatosis. J Mol Med (Berl) 2010;88(4):413-421.

171.   Kawasaki A, Inoue N, Ajimi C et al. Association of irf5 polymorphism with mpo-anca-positive vasculitis in a japanese population. Genes Immun 2013;14(8):527-529.

172.   Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med 2010;61:437-455.

173.   Nimmerjahn F, Ravetch JV. Fc-receptors as regulators of immunity. Adv Immunol 2007;96:179-204.

174.   McCarroll SA, Kuruvilla FG, Korn JM et al. Integrated detection and population-genetic analysis of snps and copy number variation. Nat Genet 2008;40(10):1166-1174.

175.   Ravetch JV, Perussia B. Alternative membrane forms of fc gamma riii(cd16) on human natural killer cells and neutrophils. Cell type-specific expression of two genes that differ in single nucleotide substitutions. J Exp Med 1989;170(2):481-497.

176.   Siriboonrit U, Tsuchiya N, Sirikong M et al. Association of fcgamma receptor iib and iiib polymorphisms with susceptibility to systemic lupus erythematosus in thais. Tissue Antigens 2003;61(5):374-383.

177.   Tse WY, Abadeh S, Jefferis R et al. Neutrophil fcgammariiib allelic polymorphism in anti-neutrophil cytoplasmic antibody (anca)-positive systemic vasculitis. Clin Exp Immunol 2000;119(3):574-577.

178.   Willcocks LC, Lyons PA, Clatworthy MR et al. Copy number of fcgr3b, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. J Exp Med 2008;205(7):1573-1582.

179.   Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. Nat Immunol 2010;11(5):373-384.

180.   Husmann CA, Holle JU, Moosig F et al. Genetics of toll like receptor 9 in anca associated vasculitides. Ann Rheum Dis 2014;73(5):890-896.

181.   Zappia M, Manna I, Serra P et al. Increased risk for alzheimer disease with the interaction of mpo and a2m polymorphisms. Arch Neurol 2004;61(3):341-344.

182.   Reynolds WF, Hiltunen M, Pirskanen M et al. Mpo and apoeepsilon4 polymorphisms interact to increase risk for ad in finnish males. Neurology 2000;55(9):1284-1290.

183.   Bouali H, Nietert P, Nowling TM et al. Association of the g-463a myeloperoxidase gene polymorphism with renal disease in african americans with systemic lupus erythematosus. J Rheumatol 2007;34(10):2028-2034.

184.   Taioli E, Benhamou S, Bouchardy C et al. Myeloperoxidase g-463a polymorphism and lung cancer: A huge genetic susceptibility to environmental carcinogens pooled analysis. Genet Med 2007;9(2):67-73.

185.   Hsu PI, Jwo JJ, Yang CL et al. Association of the myeloperoxidase polymorphism with the risk of gastric cancer. Anticancer Res 2008;28(2B):1317-1323.

186.   Piedrafita FJ, Molander RB, Vansant G et al. An alu element in the myeloperoxidase promoter contains a composite sp1-thyroid hormone-retinoic acid response element. J Biol Chem 1996;271(24):14412-14420.

187.   Reynolds WF, Stegeman CA, Tervaert JW. -463 g/a myeloperoxidase promoter polymorphism is associated with clinical manifestations and the course of disease in mpo-anca-associated vasculitis. Clin Immunol 2002;103(2):154-160.

188.   Kabutomori O, Yanagihara T, Iwatani Y et al. Sex difference in myeloperoxidase activity of neutrophils. Am J Hematol 1999;60(4):312-313.

189.   Verbsky JW, Routes JR. Recurrent fever, infections, immune disorders, and autoinflammatory diseases. Nelson pediatric symptom-based diagnosis: Elsevier;2018: p. 767.

190.   Gencik M, Meller S, Borgmann S et al. The association of cd18 alleles with anti-myeloperoxidase subtypes of anca-associated systemic vasculitides. Clin Immunol 2000;94(1):9-12.

191.   Meller S, Jagiello P, Borgmann S et al. Novel snps in the cd18 gene validate the association with mpo-anca+ vasculitis. Genes Immun 2001;2(5):269-272.

192.   Carr EJ, Clatworthy MR, Lowe CE et al. Contrasting genetic association of il2ra with sle and anca-associated vasculitis. BMC Med Genet 2009;10:22.

193. Mamegano K, Kuroki K, Miyashita R et al. Association of lilra2 (ilt1, lir7) splice site polymorphism with systemic lupus erythematosus and microscopic polyangiitis. Genes Immun 2008;9(3):214-223.

194. Murakozy G, Gaede KI, Ruprecht B et al. Gene polymorphisms of immunoregulatory cytokines and angiotensin-converting enzyme in wegener's granulomatosis. J Mol Med (Berl) 2001;79(11):665-670.

195. Wu Z, Xu J, Sun F et al. Single nucleotide polymorphisms in the toll-like receptor 2 (tlr2) gene are associated with microscopic polyangiitis in the northern han chinese population. Mod Rheumatol 2015;25(2):224-229.

196. Zhou XJ, Cheng FJ, Lv JC et al. Higher defb4 genomic copy number in sle and anca-associated small vasculitis. Rheumatology (Oxford) 2012;51(6):992-995.

197. Zhou Y, Giscombe R, Huang D et al. Novel genetic association of wegener's granulomatosis with the interleukin 10 gene. J Rheumatol 2002;29(2):317-320.

198. Miyashita R, Tsuchiya N, Yabe T et al. Association of killer cell immunoglobulin-like receptor genotypes with microscopic polyangiitis. Arthritis Rheum 2006;54(3):992-997.

199. Wieczorek S, Hoffjan S, Chan A et al. Novel association of the cd226 (dnam-1) gly307ser polymorphism in wegener's granulomatosis and confirmation for multiple sclerosis in german patients. Genes Immun 2009;10(6):591-595.

200. Wieczorek S, Holle JU, Bremer JP et al. Contrasting association of a non-synonymous leptin receptor gene polymorphism with wegener's granulomatosis and churg-strauss syndrome. Rheumatology (Oxford) 2010;49(5):907-914.

201. Finn JE, Zhang L, Agrawal S et al. Molecular analysis of c3 allotypes in patients with systemic vasculitis. Nephrol Dial Transplant 1994;9(11):1564-1567.

202. Persson U, Truedsson L, Westman KW et al. C3 and c4 allotypes in anti-neutrophil cytoplasmic autoantibody (anca)-positive vasculitis. Clin Exp Immunol 1999;116(2):379-382.

203. Wiik A. What you should know about pr3-anca. An introduction. Arthritis Res 2000;2(4):252-254.

204. Eisenberger U, Fakhouri F, Vanhille P et al. Anca-negative pauci-immune renal vasculitis: Histology and outcome. Nephrol Dial Transplant 2005;20(7):1392-1399.

205.   Selga D, Segelmark M, Gunnarsson L et al. Epitope shift of proteinase-3 anti-neutrophil cytoplasmic antibodies in patients with small vessel vasculitis. Clin Exp Immunol 2010;160(3):318-324.

206.   Silva F, Hummel AM, Jenne DE et al. Discrimination and variable impact of anca binding to different surface epitopes on proteinase 3, the wegener's autoantigen. J Autoimmun 2010;35(4):299-308.

207.   Roth AJ, Ooi JD, Hess JJ et al. Epitope specificity determines pathogenicity and detectability in anca-associated vasculitis. J Clin Invest 2013;123(4):1773-1783.

208.   Bollee G, Flamant M, Schordan S et al. Epidermal growth factor receptor promotes glomerular injury and renal failure in rapidly progressive crescentic glomerulonephritis. Nat Med 2011;17(10):1242-1250.

209.   Ding M, Cui S, Li C et al. Loss of the tumor suppressor vhlh leads to upregulation of cxcr4 and rapidly progressive glomerulonephritis in mice. Nat Med 2006;12(9):1081-1087.

210.   Peschel A, Basu N, Benharkou A et al. Autoantibodies to hlamp-2 in anca-negative pauci-immune focal necrotizing gn. J Am Soc Nephrol 2014;25(3):455-463.

211.   Tomasson G, Grayson PC, Mahr AD et al. Value of anca measurements during remission to predict a relapse of anca-associated vasculitis--a meta-analysis. Rheumatology (Oxford) 2012;51(1):100-109.

212.   Bansal PJ, Tobin MC. Neonatal microscopic polyangiitis secondary to transfer of maternal myeloperoxidase-antineutrophil cytoplasmic antibody resulting in neonatal pulmonary hemorrhage and renal involvement. Ann Allergy Asthma Immunol 2004;93(4):398-401.

213.   Silva F, Specks U, Sethi S et al. Successful pregnancy and delivery of a healthy newborn despite transplacental transfer of antimyeloperoxidase antibodies from a mother with microscopic polyangiitis. Am J Kidney Dis 2009;54(3):542-545.

214.   ten Holder SM, Joy MS, Falk RJ. Cutaneous and systemic manifestations of drug-induced vasculitis. Ann Pharmacother 2002;36(1):130-147.

215.   Stone JH, Merkel PA, Spiera R et al. Rituximab versus cyclophosphamide for anca-associated vasculitis. N Engl J Med 2010;363(3):221-232.

216.   Jones RB, Tervaert JW, Hauser T et al. Rituximab versus cyclophosphamide in anca-associated renal vasculitis. N Engl J Med 2010;363(3):211-220.

217.   Neumann I, Birck R, Newman M et al. Scg/kinjoh mice: A model of anca-associated crescentic glomerulonephritis with immune deposits. Kidney Int 2003;64(1):140-148.

218.   Xiao H, Heeringa P, Hu P et al. Antineutrophil cytoplasmic autoantibodies specific for myeloperoxidase cause glomerulonephritis and vasculitis in mice. J Clin Invest 2002;110(7):955-963.

219.   Huugen D, Xiao H, van Esch A et al. Aggravation of anti-myeloperoxidase antibody-induced glomerulonephritis by bacterial lipopolysaccharide: Role of tumor necrosis factor-alpha. Am J Pathol 2005;167(1):47-58.

220.   Little MA, Smyth CL, Yadav R et al. Antineutrophil cytoplasm antibodies directed against myeloperoxidase augment leukocyte-microvascular interactions in vivo. Blood 2005;106(6):2050-2058.

221.   Heeringa P, Little MA. In vivo approaches to investigate anca-associated vasculitis: Lessons and limitations. Arthritis Res Ther 2011;13(1):204.

222.   van der Geld YM, Hellmark T, Selga D et al. Rats and mice immunised with chimeric human/mouse proteinase 3 produce autoantibodies to mouse pr3 and rat granulocytes. Ann Rheum Dis 2007;66(12):1679-1682.

223.   Little MA, Al-Ani B, Ren S et al. Anti-proteinase 3 anti-neutrophil cytoplasm autoantibodies recapitulate systemic vasculitis in mice with a humanized immune system. PLoS One 2012;7(1):e28626.

224.   Primo VC, Marusic S, Franklin CC et al. Anti-pr3 immune responses induce segmental and necrotizing glomerulonephritis. Clin Exp Immunol 2010;159(3):327-337.

225.   Charles LA, Caldas ML, Falk RJ et al. Antibodies against granule proteins activate neutrophils in vitro. J Leukoc Biol 1991;50(6):539-546.

226.   Brooks CJ, King WJ, Radford DJ et al. Il-1 beta production by human polymorphonuclear leucocytes stimulated by anti-neutrophil cytoplasmic autoantibodies: Relevance to systemic vasculitis. Clin Exp Immunol 1996;106(2):273-279.

227.   Ewert BH, Jennette JC, Falk RJ. Anti-myeloperoxidase antibodies stimulate neutrophils to damage human endothelial cells. Kidney Int 1992;41(2):375-383.

228.   Savage CO, Pottinger BE, Gaskin G et al. Autoantibodies developing to myeloperoxidase and proteinase 3 in systemic vasculitis stimulate neutrophil cytotoxicity toward cultured endothelial cells. Am J Pathol 1992;141(2):335-342.

229.   Root-Bernstein R, Couturier J. Antigenic complementarity in the origins of autoimmunity: A general theory illustrated with a case study of idiopathic thrombocytopenia purpura. Clin Dev Immunol 2006;13(1):49-65.

230.   Pendergraft WF, 3rd, Preston GA, Shah RR et al. Autoimmunity is triggered by cpr-3(105-201), a protein complementary to human autoantigen proteinase-3. Nat Med 2004;10(1):72-79.

231.   Stegeman CA, Tervaert JW, de Jong PE et al. Trimethoprim-sulfamethoxazole (co-trimoxazole) for the prevention of relapses of wegener's granulomatosis. Dutch co-trimoxazole wegener study group. N Engl J Med 1996;335(1):16-20.

232.   Tadema H, Kallenberg CG, Stegeman CA et al. Reactivity against complementary proteinase-3 is not increased in patients with pr3-anca-associated vasculitis. PLoS One 2011;6(3):e17972.

233.   Savige J, Nassis L, Cooper T et al. Antineutrophil cytoplasmic antibody (anca)-associated systemic vasculitis after immunisation with bacterial proteins. Clin Exp Rheumatol 2002;20(6):783-789.

234.   Saftig P, Klumperman J. Lysosome biogenesis and lysosomal membrane proteins: Trafficking meets function. Nat Rev Mol Cell Biol 2009;10(9):623-635.

235.   Kain R, Matsui K, Exner M et al. A novel class of autoantigens of anti-neutrophil cytoplasmic antibodies in necrotizing and crescentic glomerulonephritis: The lysosomal membrane glycoprotein h-lamp-2 in neutrophil granulocytes and a related membrane protein in glomerular endothelial cells. J Exp Med 1995;181(2):585-597.

236.   Kain R, Tadema H, McKinney EF et al. High prevalence of autoantibodies to hlamp-2 in anti-neutrophil cytoplasmic antibody-associated vasculitis. J Am Soc Nephrol 2012;23(3):556-566.

237.   Roth AJ, Brown MC, Smith RN et al. Anti-lamp-2 antibodies are not prevalent in patients with antineutrophil cytoplasmic autoantibody glomerulonephritis. J Am Soc Nephrol 2012;23(3):545-555.

238.   Brinkmann V, Reichard U, Goosmann C et al. Neutrophil extracellular traps kill bacteria. Science 2004;303(5663):1532-1535.

239.   Kessenbrock K, Krumbholz M, Schonermarck U et al. Netting neutrophils in autoimmune small-vessel vasculitis. Nat Med 2009;15(6):623-625.

240. Fuchs TA, Abed U, Goosmann C et al. Novel cell death program leads to neutrophil extracellular traps. J Cell Biol 2007;176(2):231-241.

241. Xiao H, Heeringa P, Liu Z et al. The role of neutrophils in the induction of glomerulonephritis by anti-myeloperoxidase antibodies. Am J Pathol 2005;167(1):39-45.

242. Krumbholz M, Specks U, Wick M et al. Baff is elevated in serum of patients with wegener's granulomatosis. J Autoimmun 2005;25(4):298-302.

243. Scapini P, Bazzoni F, Cassatella MA. Regulation of b-cell-activating factor (baff)/b lymphocyte stimulator (blys) expression in human neutrophils. Immunol Lett 2008;116(1):1-6.

244. Holden NJ, Williams JM, Morgan MD et al. Anca-stimulated neutrophils release blys and promote b cell survival: A clinically relevant cellular process. Ann Rheum Dis 2011;70(12):2229-2233.

245. Mellbye OJ, Mollnes TE, Steen LS. Igg subclass distribution and complement activation ability of autoantibodies to neutrophil cytoplasmic antigens (anca). Clin Immunol Immunopathol 1994;70(1):32-39.

246. Bolton WK, Innes DJ, Jr., Sturgill BC et al. T-cells and macrophages in rapidly progressive glomerulonephritis: Clinicopathologic correlations. Kidney Int 1987;32(6):869-876.

247. Mueller A, Holl-Ulrich K, Lamprecht P et al. Germinal centre-like structures in wegener's granuloma: The morphological basis for autoimmunity? Rheumatology (Oxford) 2008;47(8):1111-1113.

248. Tipping PG, Holdsworth SR. T cells in crescentic glomerulonephritis. J Am Soc Nephrol 2006;17(5):1253-1263.

249. Sanders JS, Huitma MG, Kallenberg CG et al. Plasma levels of soluble interleukin 2 receptor, soluble cd30, interleukin 10 and b cell activator of the tumour necrosis factor family during follow-up in vasculitis associated with proteinase 3-antineutrophil cytoplasmic antibodies: Associations with disease activity and relapse. Ann Rheum Dis 2006;65(11):1484-1489.

250. McKinney EF, Lee JC, Jayne DR et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. Nature 2015;523(7562):612-616.

251.   McKinney EF, Lyons PA, Carr EJ et al. A cd8+ t cell transcription signature predicts prognosis in autoimmune disease. Nat Med 2010;16(5):586-591, 581p following 591.

252.   Klapa S, Mueller A, Csernok E et al. Lower numbers of foxp3 and ccr4 co-expressing cells in an elevated subpopulation of cd4+cd25high regulatory t cells from wegener's granulomatosis. Clin Exp Rheumatol 2010;28(1 Suppl 57):72-80.

253.   Free ME, Bunch DO, McGregor JA et al. Patients with antineutrophil cytoplasmic antibody-associated vasculitis have defective treg cell function exacerbated by the presence of a suppression-resistant effector cell population. Arthritis Rheum 2013;65(7):1922-1933.

254.   Morgan MD, Day CJ, Piper KP et al. Patients with wegener's granulomatosis demonstrate a relative deficiency and functional impairment of t-regulatory cells. Immunology 2010;130(1):64-73.

255.   Chang J, Eggenhuizen P, O'Sullivan KM et al. Cd8+ t cells effect glomerular injury in experimental anti-myeloperoxidase gn. J Am Soc Nephrol 2017;28(1):47-55.

256.   Ruth AJ, Kitching AR, Kwan RY et al. Anti-neutrophil cytoplasmic antibodies and effector cd4+ cells play nonredundant roles in anti-myeloperoxidase crescentic glomerulonephritis. J Am Soc Nephrol 2006;17(7):1940-1949.

257.   Anolik JH, Looney RJ, Lund FE et al. Insights into the heterogeneity of human b cells: Diverse functions, roles in autoimmunity, and use as therapeutic targets. Immunol Res 2009;45(2-3):144-158.

258.   Mauri C. Regulation of immunity and autoimmunity by b cells. Curr Opin Immunol 2010;22(6):761-767.

259.   Flores-Borja F, Bosma A, Ng D et al. Cd19+cd24hicd38hi b cells maintain regulatory t cells while limiting th1 and th17 differentiation. Sci Transl Med 2013;5(173):173ra123.

260.   Bunch DO, McGregor JG, Khandoobhai NB et al. Decreased cd5(+) b cells in active anca vasculitis and relapse after rituximab. Clin J Am Soc Nephrol 2013;8(3):382-391.

261.   Todd SK, Pepper RJ, Draibe J et al. Regulatory b cells are numerically but not functionally deficient in anti-neutrophil cytoplasm antibody-associated vasculitis. Rheumatology (Oxford) 2014;53(9):1693-1703.

262.   Wilde B, Thewissen M, Damoiseaux J et al. Regulatory b cells in anca-associated vasculitis. Ann Rheum Dis 2013;72(8):1416-1419.

263. Xiao H, Schreiber A, Heeringa P et al. Alternative complement pathway in the pathogenesis of disease mediated by anti-neutrophil cytoplasmic autoantibodies. Am J Pathol 2007;170(1):52-64.

264. Huugen D, van Esch A, Xiao H et al. Inhibition of complement factor c5 protects against anti-myeloperoxidase antibody-mediated glomerulonephritis in mice. Kidney Int 2007;71(7):646-654.

265. Schreiber A, Xiao H, Jennette JC et al. C5a receptor mediates neutrophil activation and anca-induced glomerulonephritis. J Am Soc Nephrol 2009;20(2):289-298.

266. Xiao H, Dairaghi DJ, Powers JP et al. C5a receptor (cd88) blockade protects against mpo-anca gn. J Am Soc Nephrol 2014;25(2):225-231.

267. Jayne DR, Bruchfeld AN, Harper L et al. Randomized trial of c5a receptor inhibitor avacopan in anca-associated vasculitis. J Am Soc Nephrol 2017.

268. Franssen CF, Stegeman CA, Kallenberg CG et al. Antiproteinase 3- and antimyeloperoxidase-associated vasculitis. Kidney Int 2000;57(6):2195-2206.

269. Hilhorst M, van Paassen P, Tervaert JW et al. Proteinase 3-anca vasculitis versus myeloperoxidase-anca vasculitis. J Am Soc Nephrol 2015;26(10):2314-2327.

270. Cornec D, Cornec-Le Gall E, Fervenza FC et al. Anca-associated vasculitis - clinical utility of using anca specificity to classify patients. Nat Rev Rheumatol 2016;12(10):570-579.

271. Cudworth AG, Woodrow JC. Letter: Hl-a antigens and diabetes mellitus. Lancet 1974;2(7889):1153.

272. Schlosstein L, Terasaki PI, Bluestone R et al. High association of an hl-a antigen, w27, with ankylosing spondylitis. N Engl J Med 1973;288(14):704-706.

273. Singal DP, Blajchman MA. Histocompatibility (hl-a) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes 1973;22(6):429-432.

274. Stastny P. Mixed lymphocyte cultures in rheumatoid arthritis. J Clin Invest 1976;57(5):1148-1157.

275. Hugot JP, Chamaillard M, Zouali H et al. Association of nod2 leucine-rich repeat variants with susceptibility to crohn's disease. Nature 2001;411(6837):599-603.

276. Ogura Y, Bonen DK, Inohara N et al. A frameshift mutation in nod2 associated with susceptibility to crohn's disease. Nature 2001;411(6837):603-606.

277.	Lowe CE, Cooper JD, Brusko T et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the il2ra region in type 1 diabetes. Nat Genet 2007;39(9):1074-1082.

278.	Vella A, Cooper JD, Lowe CE et al. Localization of a type 1 diabetes locus in the il2ra/cd25 region by use of tag single-nucleotide polymorphisms. Am J Hum Genet 2005;76(5):773-779.

279.	Gough SC, Walker LS, Sansom DM. Ctla4 gene polymorphism and autoimmunity. Immunol Rev 2005;204:102-115.

280.	Visscher PM, Wray NR, Zhang Q et al. 10 years of gwas discovery: Biology, function, and translation. Am J Hum Genet 2017;101(1):5-22.

281.	Howie B, Fuchsberger C, Stephens M et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44(8):955-959.

282.	Li Y, Willer CJ, Ding J et al. Mach: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34(8):816-834.

283.	Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet 2018;103(3):338-348.

284.	International HapMap C, Altshuler DM, Gibbs RA et al. Integrating common and rare genetic variation in diverse human populations. Nature 2010;467(7311):52-58.

285.	Genomes Project C, Auton A, Brooks LD et al. A global reference for human genetic variation. Nature 2015;526(7571):68-74.

286.	McCarthy S, Das S, Kretzschmar W et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48(10):1279-1283.

287.	Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet 2008;9(4):255-266.

288.	Pritchard JK, Cox NJ. The allelic architecture of human disease genes: Common disease-common variant...Or not? Hum Mol Genet 2002;11(20):2417-2423.

289.	Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 2010;11(6):415-425.

290.	Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69(1):124-137.

291. Hunt KA, Mistry V, Bockett NA et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 2013;498(7453):232-235.

292. Fisher RA. The genetical theory of natural selection. Oxford: Oxford University Press; 1930.

293. Yang J, Benyamin B, McEvoy BP et al. Common snps explain a large proportion of the heritability for human height. Nat Genet 2010;42(7):565-569.

294. Yang J, Lee SH, Goddard ME et al. Gcta: A tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88(1):76-82.

295. Yang J, Manolio TA, Pasquale LR et al. Genome partitioning of genetic variation for complex traits using common snps. Nat Genet 2011;43(6):519-525.

296. Davies G, Tenesa A, Payton A et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 2011;16(10):996-1005.

297. Deary IJ, Yang J, Davies G et al. Genetic contributions to stability and change in intelligence from childhood to old age. Nature 2012;482(7384):212-215.

298. Lee SH, DeCandia TR, Ripke S et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common snps. Nat Genet 2012;44(3):247-250.

299. Gibson G. Hints of hidden heritability in gwas. Nat Genet 2010;42(7):558-560.

300. Eichler EE, Flint J, Gibson G et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010;11(6):446-450.

301. Feldman MW, Lewontin RC. The heritability hang-up. Science 1975;190(4220):1163-1168.

302. Blanco-Gomez A, Castillo-Lluva S, Del Mar Saez-Freire M et al. Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. Bioessays 2016;38(7):664-673.

303. Solovieff N, Cotsapas C, Lee PH et al. Pleiotropy in complex traits: Challenges and strategies. Nat Rev Genet 2013;14(7):483-495.

304. Mokry LE, Ross S, Ahmad OS et al. Vitamin d and risk of multiple sclerosis: A mendelian randomization study. PLoS Med 2015;12(8):e1001866.

305. Rhead B, Baarnhielm M, Gianfrancesco M et al. Mendelian randomization shows a causal effect of low vitamin d on multiple sclerosis risk. Neurol Genet 2016;2(5):e97.

306.   Gibson G. Rare and common variants: Twenty arguments. Nat Rev Genet 2012;13(2):135-145.

307.   Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet 2015;24(R1):R111-119.

308.   Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. Nat Rev Genet 2009;10(10):681-690.

309.   Replication DIG, Meta-analysis C, Asian Genetic Epidemiology Network Type 2 Diabetes C et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46(3):234-244.

310.   Consortium EP. The encode (encyclopedia of DNA elements) project. Science 2004;306(5696):636-640.

311.   Adams D, Altucci L, Antonarakis SE et al. Blueprint to decode the epigenetic signature written in blood. Nat Biotechnol 2012;30(3):224-226.

312.   Bernstein BE, Stamatoyannopoulos JA, Costello JF et al. The nih roadmap epigenomics mapping consortium. Nat Biotechnol 2010;28(10):1045-1048.

313.   Andersson R, Gebhard C, Miguel-Escalada I et al. An atlas of active enhancers across human cell types and tissues. Nature 2014;507(7493):455-461.

314.   Maurano MT, Humbert R, Rynes E et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 2012;337(6099):1190-1195.

315.   Hnisz D, Abraham BJ, Lee TI et al. Super-enhancers in the control of cell identity and disease. Cell 2013;155(4):934-947.

316.   Vahedi G, Kanno Y, Furumoto Y et al. Super-enhancers delineate disease-associated regulatory nodes in t cells. Nature 2015;520(7548):558-562.

317.   Consortium GT. The genotype-tissue expression (gtex) project. Nat Genet 2013;45(6):580-585.

318.   Võsa U, Claringbould A, Westra H et al. Unraveling the polygenic architecture of complex traits using blood eqtl meta-analysis. bioRxiv 2018.

319.   Sun BB, Maranville JC, Peters JE et al. Genomic atlas of the human plasma proteome. Nature 2018;558(7708):73-79.

320.   Dekker J, Rippe K, Dekker M et al. Capturing chromosome conformation. Science 2002;295(5558):1306-1311.

321. Javierre BM, Burren OS, Wilder SP et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 2016;167(5):1369-1384 e1319.

322. Ray G, Longworth MS. Epigenetics, DNA organization, and inflammatory bowel disease. Inflamm Bowel Dis 2019;25(2):235-247.

323. Day N, Oakes S, Luben R et al. Epic-norfolk: Study design and characteristics of the cohort. European prospective investigation of cancer. Br J Cancer 1999;80 Suppl 1:95-103.

324. Tennessen JA, Bigham AW, O'Connor TD et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 2012;337(6090):64-69.

325. Minikel EV, Vallabh SM, Lek M et al. Quantifying prion disease penetrance using large population control cohorts. Sci Transl Med 2016;8(322):322ra329.

326. Consortium UK, Walter K, Min JL et al. The uk10k project identifies rare variants in health and disease. Nature 2015;526(7571):82-90.

327. Stenson PD, Mort M, Ball EV et al. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 2014;133(1):1-9.

328. Hoffmann TJ, Zhan Y, Kvale MN et al. Design and coverage of high throughput genotyping arrays optimized for individuals of east asian, african american, and latino race/ethnicity using imputation and a novel hybrid snp selection algorithm. Genomics 2011;98(6):422-430.

329. Chang CC, Chow CC, Tellier LC et al. Second-generation plink: Rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.

330. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55(4):997-1004.

331. Yang J, Zaitlen NA, Goddard ME et al. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet 2014;46(2):100-106.

332. Loh PR, Kichaev G, Gazal S et al. Mixed-model association for biobank-scale datasets. Nat Genet 2018;50(7):906-908.

333. Loh PR, Tucker G, Bulik-Sullivan BK et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015;47(3):284-290.

334.   Das S, Forer L, Schonherr S et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48(10):1284-1287.

335.   Loh PR, Danecek P, Palamara PF et al. Reference-based phasing using the haplotype reference consortium panel. Nat Genet 2016;48(11):1443-1448.

336.   Pistis G, Porcu E, Vrieze SI et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: Implications for cost-effective study designs. Eur J Hum Genet 2015;23(7):975-983.

337.   Motyer AM, Vukcevic D, Dilthey A et al. Practical use of methods for imputation of hla alleles from snp genotype data.

338.   Berden A, Goceroglu A, Jayne D et al. Diagnosis and management of anca associated vasculitis. BMJ 2012;344:e26.

339.   Lee SH, Yang J, Goddard ME et al. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 2012;28(19):2540-2542.

340.   Loh PR, Bhatia G, Gusev A et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 2015;47(12):1385-1392.

341.   Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet 2014;94(4):559-573.

342.   Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics 2016;32(2):283-285.

343.   Wellcome Trust Case Control C, Maller JB, McVean G et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 2012;44(12):1294-1301.

344.   Wakefield J. Bayes factors for genome-wide association studies: Comparison with p-values. Genet Epidemiol 2009;33(1):79-86.

345.   Giambartolomei C, Vukcevic D, Schadt EE et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10(5):e1004383.

346.   Astle WJ, Elding H, Jiang T et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell 2016;167(5):1415-1429 e1419.

347.   Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23(R1):R89-98.

348.   Mifsud B, Tavares-Cadete F, Young AN et al. Mapping long-range promoter contacts in human cells with high-resolution capture hi-c. Nat Genet 2015;47(6):598-606.

349.   Schofield EC, Carver T, Achuthan P et al. Chicp: A web-based tool for the integrative and interactive visualization of promoter capture hi-c datasets. Bioinformatics 2016;32(16):2511-2513.

350.   Staley JR, Blackshaw J, Kamat MA et al. Phenoscanner: A database of human genotype-phenotype associations. Bioinformatics 2016;32(20):3207-3209.

351.   Watanabe K, Taskesen E, van Bochoven A et al. Functional mapping and annotation of genetic associations with fuma. Nat Commun 2017;8(1):1826.

352.   Lyons PA, Koukoulaki M, Hatton A et al. Microarray analysis of human leucocyte subsets: The advantages of positive selection and rapid purification. BMC Genomics 2007;8:64.

353.   Haslett C, Guthrie LA, Kopaniak MM et al. Modulation of multiple neutrophil functions by preparative methods or trace concentrations of bacterial lipopolysaccharide. Am J Pathol 1985;119(1):101-110.

354.   Berro AI, Perry GA, Agrawal DK. Increased expression and activation of cd30 induce apoptosis in human blood eosinophils. J Immunol 2004;173(3):2174-2183.

355.   Kitchen E, Rossi AG, Condliffe AM et al. Demonstration of reversible priming of human neutrophils using platelet-activating factor. Blood 1996;88(11):4330-4337.

356.   Farahi N, Singh NR, Heard S et al. Use of 111-indium-labeled autologous eosinophils to establish the in vivo kinetics of human eosinophils in healthy subjects. Blood 2012;120(19):4068-4071.

357.   Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative pcr and the 2(-delta delta c(t)) method. Methods 2001;25(4):402-408.

358.   Rainbow DB, Esposito L, Howlett SK et al. Commonality in the genetic control of type 1 diabetes in humans and nod mice: Variants of genes in the il-2 pathway are associated with autoimmune diabetes in both species. Biochem Soc Trans 2008;36(Pt 3):312-315.

359.   de Joode AA, Sanders JS, Stegeman CA. Renal survival in proteinase 3 and myeloperoxidase anca-associated systemic vasculitis. Clin J Am Soc Nephrol 2013;8(10):1709-1717.

360.   Hilhorst M, Wilde B, van Paassen P et al. Improved outcome in anti-neutrophil cytoplasmic antibody (anca)-associated glomerulonephritis: A 30-year follow-up study. Nephrol Dial Transplant 2013;28(2):373-379.

361.   Mandel H, Szargel R, Labay V et al. The deoxyguanosine kinase gene is mutated in individuals with depleted hepatocerebral mitochondrial DNA. Nat Genet 2001;29(3):337-341.

362.   Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 2011;13(1):7-13.

363.   Yang R, Qu C, Zhou Y et al. Hydrogen sulfide promotes tet1- and tet2-mediated foxp3 demethylation to drive regulatory t cell differentiation and maintain immune homeostasis. Immunity 2015;43(2):251-263.

364.   Zhang Q, Zhao K, Shen Q et al. Tet2 is required to resolve inflammation by recruiting hdac2 to specifically repress il-6. Nature 2015;525(7569):389-393.

365.   Langemeijer SM, Aslanyan MG, Jansen JH. Tet proteins in malignant hematopoiesis. Cell Cycle 2009;8(24):4044-4048.

366.   Xue S, Liu C, Sun X et al. Tet3 inhibits type i ifn production independent of DNA demethylation. Cell Rep 2016;16(4):1096-1105.

367.   Patel DR, Richardson BC. Epigenetic mechanisms in lupus. Curr Opin Rheumatol 2010;22(5):478-482.

368.   Bednash JS, Weathington N, Londino J et al. Targeting the deubiquitinase stambp inhibits nalp7 inflammasome activity. Nat Commun 2017;8:15203.

369.   Yang W, Tang H, Zhang Y et al. Meta-analysis followed by replication identifies loci in or near cdkn1b, tet3, cd80, dram1, and arid5b as associated with systemic lupus erythematosus in asians. Am J Hum Genet 2013;92(1):41-51.

370.   Kichaev G, Bhatia G, Loh PR et al. Leveraging polygenic functional enrichment to improve gwas power. Am J Hum Genet 2019;104(1):65-75.

371.   Murea M, Lu L, Ma L et al. Genome-wide association scan for survival on dialysis in african-americans with type 2 diabetes. Am J Nephrol 2011;33(6):502-509.

372.   Gaunt TR, Shihab HA, Hemani G et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol 2016;17:61.

373.   Ji SG, Juran BD, Mucha S et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet 2017;49(2):269-273.

374.   Jin Y, Andersen G, Yorgov D et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet 2016;48(11):1418-1424.

375.   Berndt SI, Camp NJ, Skibola CF et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. Nat Commun 2016;7:10933.

376.   Hirota T, Takahashi A, Kubo M et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the japanese population. Nat Genet 2012;44(11):1222-1226.

377.   Baurecht H, Hotze M, Brand S et al. Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. Am J Hum Genet 2015;96(1):104-120.

378.   Sudlow C, Gallacher J, Allen N et al. Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.

379.   Bouillet P, Metcalf D, Huang DC et al. Proapoptotic bcl-2 relative bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity. Science 1999;286(5445):1735-1738.

380.   Kotzin JJ, Spencer SP, McCright SJ et al. The long non-coding rna morrbid regulates bim and short-lived myeloid cell lifespan. Nature 2016;537(7619):239-243.

381.   Liberg D, Sigvardsson M, Akerblad P. The ebf/olf/collier family of transcription factors: Regulators of differentiation in cells originating from all three embryonal germ layers. Mol Cell Biol 2002;22(24):8389-8397.

382.   Liao D. Emerging roles of the ebf family of transcription factors in tumor suppression. Mol Cancer Res 2009;7(12):1893-1901.

383.   Wu C, Luo Z, Pang B et al. Associations of pulmonary fibrosis with peripheral blood th1/th2 cell imbalance and ebf3 gene methylation in uygur pigeon breeder's lung patients. Cell Physiol Biochem 2018;47(3):1141-1151.

384.   Park SH, Kim SK, Choe JY et al. Hypermethylation of ebf3 and irx1 genes in synovial fibroblasts of patients with rheumatoid arthritis. Mol Cells 2013;35(4):298-304.

385.   Zabriskie JB. Essential clinical immunology. Cambridge University Press; 2009.

386.   Watson CT, Glanville J, Marasco WA. The individual and population genetics of antibody immunity. Trends Immunol 2017;38(7):459-470.

387.   Avnir Y, Watson CT, Glanville J et al. Ighv1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with ighv utilization shifts and varies by ethnicity. Sci Rep 2016;6:20842.

388.   Glanville J, Kuo TC, von Budingen HC et al. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. Proc Natl Acad Sci U S A 2011;108(50):20066-20071.

389.   Wang C, Liu Y, Cavanagh MM et al. B-cell repertoire responses to varicella-zoster vaccination in human identical twins. Proc Natl Acad Sci U S A 2015;112(2):500-505.

390.   Wheatley AK, Whittle JR, Lingwood D et al. H5n1 vaccine-elicited memory b cells are genetically constrained by the ighv locus in the recognition of a neutralizing epitope in the hemagglutinin stem. J Immunol 2015;195(2):602-610.

391.   Ekiert DC, Bhabha G, Elsliger MA et al. Antibody recognition of a highly conserved influenza virus epitope. Science 2009;324(5924):246-251.

392.   Wrammert J, Koutsonanos D, Li GM et al. Broadly cross-reactive antibodies dominate the human b cell response against 2009 pandemic h1n1 influenza virus infection. J Exp Med 2011;208(1):181-193.

393.   Pappas L, Foglierini M, Piccoli L et al. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. Nature 2014;516(7531):418-422.

394.   Yeung YA, Foletti D, Deng X et al. Germline-encoded neutralization of a staphylococcus aureus virulence factor by the human antibody repertoire. Nat Commun 2016;7:13376.

395.   Lupianez DG, Spielmann M, Mundlos S. Breaking tads: How alterations of chromatin domains result in disease. Trends Genet 2016;32(4):225-237.

396. So HC, Gui AH, Cherny SS et al. Evaluating the heritability explained by known susceptibility variants: A survey of ten complex diseases. Genet Epidemiol 2011;35(5):310-317.

397. van Timmeren MM, Heeringa P, Kallenberg CG. Infectious triggers for vasculitis. Curr Opin Rheumatol 2014;26(4):416-423.

398. Laudien M, Gadola SD, Podschun R et al. Nasal carriage of staphylococcus aureus and endonasal activity in wegener s granulomatosis as compared to rheumatoid arthritis and chronic rhinosinusitis with nasal polyps. Clin Exp Rheumatol 2010;28(1 Suppl 57):51-55.

399. Popa ER, Stegeman CA, Abdulahad WH et al. Staphylococcal toxic-shock-syndrome-toxin-1 as a risk factor for disease relapse in wegener's granulomatosis. Rheumatology (Oxford) 2007;46(6):1029-1033.

400. Zycinska K, Wardyn KA, Zielonka TM et al. Chronic crusting, nasal carriage of staphylococcus aureus and relapse rate in pulmonary wegener's granulomatosis. J Physiol Pharmacol 2008;59 Suppl 6:825-831.

401. Rhee RL, Sreih AG, Najem CE et al. Characterisation of the nasal microbiota in granulomatosis with polyangiitis. Ann Rheum Dis 2018;77(10):1448-1453.

402. Kronbichler A, Wagner J, Harrison E et al. 176.   The composition of the human nasal microbiome in granulomatosis with polyangiitis: A pilot study. Rheumatology 2019;58(Supplement_2).

403. Mazmanian SK, Skaar EP, Gaspar AH et al. Passage of heme-iron across the envelope of staphylococcus aureus. Science 2003;299(5608):906-909.

404. Vu NT, Moriwaki Y, Caaveiro JM et al. Selective binding of antimicrobial porphyrins to the heme-receptor isdh-neat3 of staphylococcus aureus. Protein Sci 2013;22(7):942-953.

405. Spirig T, Malmirchegini GR, Zhang J et al. Staphylococcus aureus uses a novel multidomain receptor to break apart human hemoglobin and steal its heme. J Biol Chem 2013;288(2):1065-1078.

406. Honsa ES, Maresso AW, Highlander SK. Molecular and evolutionary analysis of near-iron transporter (neat) domains. PLoS One 2014;9(8):e104794.

407. Pishchany G, Sheldon JR, Dickson CF et al. Isdb-dependent hemoglobin binding is required for acquisition of heme by staphylococcus aureus. J Infect Dis 2014;209(11):1764-1772.

408.   Torres VJ, Pishchany G, Humayun M et al. Staphylococcus aureus isdb is a hemoglobin receptor required for heme iron utilization. J Bacteriol 2006;188(24):8421-8429.

409.   Ebert T, Smith S, Pancari G et al. A fully human monoclonal antibody to staphylococcus aureus iron regulated surface determinant b (isdb) with functional activity in vitro and in vivo. Hum Antibodies 2010;19(4):113-128.

410.   Nishitani K, Beck CA, Rosenberg AF et al. A diagnostic serum antibody test for patients with staphylococcus aureus osteomyelitis. Clin Orthop Relat Res 2015;473(9):2735-2749.

411.   Lingwood D, McTamney PM, Yassine HM et al. Structural and genetic basis for development of broadly neutralizing influenza antibodies. Nature 2012;489(7417):566-570.

412.   Throsby M, van den Brink E, Jongeneelen M et al. Heterosubtypic neutralizing monoclonal antibodies cross-protective against h5n1 and h1n1 recovered from human igm+ memory b cells. PLoS One 2008;3(12):e3942.

413.   Duggal T, Segal P, Shah M et al. Antineutrophil cytoplasmic antibody vasculitis associated with influenza vaccination. Am J Nephrol 2013;38(2):174-178.

414.   Watanabe T. Vasculitis following influenza vaccination: A review of the literature. Curr Rheumatol Rev 2017;13(3):188-196.

415.   Chang DY, Li ZY, Chen M et al. Myeloperoxidase-anca-positive granulomatosis with polyangiitis is a distinct subset of anca-associated vasculitis: A retrospective analysis of 455 patients from a single center in china. Semin Arthritis Rheum 2019;48(4):701-706.

416.   Chen M, Yu F, Zhang Y et al. Antineutrophil cytoplasmic autoantibody-associated vasculitis in older patients. Medicine (Baltimore) 2008;87(4):203-209.

417.   Kim K, Bang SY, Lee HS et al. High-density genotyping of immune loci in koreans and europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis 2015;74(3):e13.

418.   Betz RC, Petukhova L, Ripke S et al. Genome-wide meta-analysis in alopecia areata resolves hla associations and reveals two new susceptibility loci. Nat Commun 2015;6:5966.

419.   Tanaka S, Jiang Y, Martinez GJ et al. Trim33 mediates the proinflammatory function of th17 cells. J Exp Med 2018;215(7):1853-1868.

420.   Wilde B, Thewissen M, Damoiseaux J et al. Th17 expansion in granulomatosis with polyangiitis (wegener's): The role of disease activity, immune regulation and therapy. Arthritis Res Ther 2012;14(5):R227.

421.   Nogueira E, Hamour S, Sawant D et al. Serum il-17 and il-23 levels and autoantigen-specific th17 cells are elevated in patients with anca-associated vasculitis. Nephrol Dial Transplant 2010;25(7):2209-2217.

422.   Gan PY, Steinmetz OM, Tan DS et al. Th17 cells promote autoimmune anti-myeloperoxidase glomerulonephritis. J Am Soc Nephrol 2010;21(6):925-931.

423.   Li X, Zhang R, Luo D et al. Tumor necrosis factor alpha-induced desumoylation and cytoplasmic translocation of homeodomain-interacting protein kinase 1 are critical for apoptosis signal-regulating kinase 1-jnk/p38 activation. J Biol Chem 2005;280(15):15061-15070.

424.   Guerra FM, Gommerman JL, Corfe SA et al. Homeodomain-interacting protein kinase (hipk)-1 is required for splenic b cell homeostasis and optimal t-independent type 2 humoral response. PLoS One 2012;7(4):e35533.

425.   Oliveira JB, Bidere N, Niemela JE et al. Nras mutation causes a human autoimmune lymphoproliferative syndrome. Proc Natl Acad Sci U S A 2007;104(21):8953-8958.

426.   Cooper JD, Smyth DJ, Smiles AM et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet 2008;40(12):1399-1401.

427.   Franke A, McGovern DP, Barrett JC et al. Genome-wide meta-analysis increases to 71 the number of confirmed crohn's disease susceptibility loci. Nat Genet 2010;42(12):1118-1125.

428.   International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011;476(7359):214-219.

429.   Roederer M, Quaye L, Mangino M et al. The genetic architecture of the human immune system: A bioresource for autoimmunity and disease pathogenesis. Cell 2015;161(2):387-403.

430.   Weinhold N, Johnson DC, Chubb D et al. The ccnd1 c.870g>a polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. Nat Genet 2013;45(5):522-525.

431.   Qiu C, Huang S, Park J et al. Renal compartment-specific genetic variation analyses identify new pathways in chronic kidney disease. Nat Med 2018;24(11):1721-1731.

432.   Shinnakasu R, Inoue T, Kometani K et al. Regulated selection of germinal-center cells into the memory b cell compartment. Nat Immunol 2016;17(7):861-869.

433. Muto A, Ochiai K, Kimura Y et al. Bach2 represses plasma cell gene regulatory network in b cells to promote antibody class switch. EMBO J 2010;29(23):4048-4061.

434. Muto A, Tashiro S, Nakajima O et al. The transcriptional programme of antibody class switching involves the repressor bach2. Nature 2004;429(6991):566-571.

435. Ochiai K, Katoh Y, Ikura T et al. Plasmacytic transcription factor blimp-1 is repressed by bach2 in b cells. J Biol Chem 2006;281(50):38226-38234.

436. Roychoudhuri R, Hirahara K, Mousavi K et al. Bach2 represses effector programs to stabilize t(reg)-mediated immune homeostasis. Nature 2013;498(7455):506-510.

437. Afzali B, Gronholm J, Vandrovcova J et al. Bach2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. Nat Immunol 2017;18(7):813-823.

438. Zhang A, Yeung PL, Li CW et al. Identification of a novel family of ankyrin repeats containing cofactors for p160 nuclear receptor coactivators. J Biol Chem 2004;279(32):33799-33805.

439. Koppen M, Metodiev MD, Casari G et al. Variable and tissue-specific subunit composition of mitochondrial m-aaa protease complexes linked to hereditary spastic paraplegia. Mol Cell Biol 2007;27(2):758-767.

440. Peters JE, Lyons PA, Lee JC et al. Insight into genotype-phenotype associations through eqtl mapping in multiple cell types in health and immune-mediated disease. PLoS Genet 2016;12(3):e1005908.

441. Nguyen T, Liu XK, Zhang Y et al. Btnl2, a butyrophilin-like molecule that functions to inhibit t cell activation. J Immunol 2006;176(12):7354-7360.

442. Chen GB, Lee SH, Brion MJ et al. Estimation and partitioning of (co)heritability of inflammatory bowel disease from gwas and immunochip data. Hum Mol Genet 2014;23(17):4710-4720.

443. Li YR, Zhao SD, Li J et al. Genetic sharing and heritability of paediatric age of onset autoimmune diseases. Nat Commun 2015;6:8442.

444. Duncan LE, Shen H, Ballon JS et al. Genetic correlation profile of schizophrenia mirrors epidemiological results and suggests link between polygenic and rare variant (22q11.2) cases of schizophrenia. Schizophr Bull 2018;44(6):1350-1361.

445.   Tan DS, Gan PY, O'Sullivan KM et al. Thymic deletion and regulatory t cells prevent antimyeloperoxidase gn. J Am Soc Nephrol 2013;24(4):573-585.

446.   Paust HJ, Riedel JH, Krebs CF et al. Cxcr3+ regulatory t cells control th1 responses in crescentic gn. J Am Soc Nephrol 2016;27(7):1933-1942.

447.   Rimbert M, Hamidou M, Braudeau C et al. Decreased numbers of blood dendritic cells and defective function of regulatory t cells in antineutrophil cytoplasmic antibody-associated vasculitis. PLoS One 2011;6(4):e18734.

448.   Anjos S, Nguyen A, Ounissi-Benkalha H et al. A common autoimmunity predisposing signal peptide variant of the cytotoxic t-lymphocyte antigen 4 results in inefficient glycosylation of the susceptibility allele. J Biol Chem 2002;277(48):46478-46486.

449.   Pruul K, Kisand K, Alnek K et al. Expression of b7 and cd28 family genes in newly diagnosed type 1 diabetes. Hum Immunol 2013;74(10):1251-1257.

450.   Ueda H, Howson JM, Esposito L et al. Association of the t-cell regulatory gene ctla4 with susceptibility to autoimmune disease. Nature 2003;423(6939):506-511.

451.   Kasela S, Kisand K, Tserel L et al. Pathogenic implications for autoimmune mechanisms derived by comparative eqtl analysis of cd4+ versus cd8+ t cells. PLoS Genet 2017;13(3):e1006643.

452.   Kerr JF, Wyllie AH, Currie AR. Apoptosis: A basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 1972;26(4):239-257.

453.   Abdgawad M, Pettersson A, Gunnarsson L et al. Decreased neutrophil apoptosis in quiescent anca-associated systemic vasculitis. PLoS One 2012;7(3):e32439.

454.   Harper L, Cockwell P, Adu D et al. Neutrophil priming and apoptosis in anti-neutrophil cytoplasmic autoantibody-associated vasculitis. Kidney Int 2001;59(5):1729-1738.

455.   O'Connor L, Strasser A, O'Reilly LA et al. Bim: A novel member of the bcl-2 family that promotes apoptosis. EMBO J 1998;17(2):384-395.

456.   Chow LT, Gelinas RE, Broker TR et al. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger rna. Cell 1977;12(1):1-8.

457.   Serena G, Yan S, Camhi S et al. Proinflammatory cytokine interferon-gamma and microbiome-derived metabolites dictate epigenetic switch between forkhead box protein 3 isoforms in coeliac disease. Clin Exp Immunol 2017;187(3):490-506.

458.   Masi AT, Hunder GG, Lie JT et al. The american college of rheumatology 1990 criteria for the classification of churg-strauss syndrome (allergic granulomatosis and angiitis). Arthritis Rheum 1990;33(8):1094-1100.

459.   Khoury P, Grayson PC, Klion AD. Eosinophils in vasculitis: Characteristics and roles in pathogenesis. Nat Rev Rheumatol 2014;10(8):474-483.

460.   Yousem SA, Lombard CM. The eosinophilic variant of wegener's granulomatosis. Hum Pathol 1988;19(6):682-688.

461.   Al-Hakami H, Al-Arfaj AS, Al-Sohaibani M et al. An eosinophilic variant granulomatosis with polyangiitis involving the dura, bilateral orbits, and mastoids. Saudi Med J 2016;37(6):690-693.

462.   Chan AS, Yu DL, Rao NA. Eosinophilic variant of wegener granulomatosis in the orbit. Arch Ophthalmol 2011;129(9):1238-1240.

463.   Henochowicz S, Eggensperger D, Pierce L et al. Necrotizing systemic vasculitis with features of both wegener's granulomatosis and churg-strauss vasculitis. Arthritis Rheum 1986;29(4):565-569.

464.   Kamali S, Kasapoglu E, Akturk F et al. Eosinophilia and hyperimmunoglobulinemia e as the presenting manifestations of wegener's granulomatosis. Clin Rheumatol 2003;22(4-5):333-335.

465.   Krupsky M, Landau Z, Lifschitz-Mercer B et al. Wegener's granulomatosis with peripheral eosinophilia. Atypical variant of a classic disease. Chest 1993;104(4):1290-1292.

466.   Potter MB, Fincher RK, Finger DR. Eosinophilia in wegener's granulomatosis. Chest 1999;116(5):1480-1483.

467.   Uematsu H, Takata S, Sueishi K et al. Polyangiitis overlap syndrome of granulomatosis with polyangiitis (wegener's granulomatosis) and eosinophilic granulomatosis with polyangiitis (churg-strauss syndrome). BMJ Case Rep 2014;2014.

468.   Balding CE, Howie AJ, Drake-Lee AB et al. Th2 dominance in nasal mucosa in patients with wegener's granulomatosis. Clin Exp Immunol 2001;125(2):332-339.

469.   Schnabel A, Csernok E, Braun J et al. Inflammatory cells and cellular activation in the lower respiratory tract in churg-strauss syndrome. Thorax 1999;54(9):771-778.

470.   Kay AB. The early history of the eosinophil. Clin Exp Allergy 2015;45(3):575-582.

471. Lee JJ, Jacobsen EA, McGarry MP et al. Eosinophils in health and disease: The liar hypothesis. Clin Exp Allergy 2010;40(4):563-575.

472. Liu LY, Bates ME, Jarjour NN et al. Generation of th1 and th2 chemokines by human eosinophils: Evidence for a critical role of tnf-alpha. J Immunol 2007;179(7):4840-4848.

473. Chu VT, Frohlich A, Steinhauser G et al. Eosinophils are required for the maintenance of plasma cells in the bone marrow. Nat Immunol 2011;12(2):151-159.

474. Hellmark T, Ohlsson S, Pettersson A et al. Eosinophils in anti-neutrophil cytoplasmic antibody associated vasculitis. BMC Rheumatol 2019;3:9.

475. Dandona P, Mohanty P, Hamouda W et al. Effect of dexamethasone on reactive oxygen species generation by leukocytes and plasma interleukin-10 concentrations: A pharmacodynamic study. Clin Pharmacol Ther 1999;66(1):58-65.

476. Schleimer RP, Bochner BS. The effects of glucocorticoids on human eosinophils. J Allergy Clin Immunol 1994;94(6 Pt 2):1202-1213.

477. Kim K, Hwang SM, Kim SM et al. Terminally differentiating eosinophils express neutrophil primary granule proteins as well as eosinophil-specific granule proteins in a temporal manner. Immune Netw 2017;17(6):410-423.

478. Melnikov A, Murugan A, Zhang X et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 2012;30(3):271-277.

479. Wiedenheft B, Sternberg SH, Doudna JA. Rna-guided genetic silencing systems in bacteria and archaea. Nature 2012;482(7385):331-338.

480. Mali P, Aach J, Stranges PB et al. Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 2013;31(9):833-838.

481. Gilbert LA, Larson MH, Morsut L et al. Crispr-mediated modular rna-guided regulation of transcription in eukaryotes. Cell 2013;154(2):442-451.

482. Jacobsen EA, Helmers RA, Lee JJ et al. The expanding role(s) of eosinophils in health and disease. Blood 2012;120(19):3882-3890.

483. Ochkur SI, Kim JD, Protheroe CA et al. A sensitive high throughput elisa for human eosinophil peroxidase: A specific assay to quantify eosinophil degranulation from patient-derived sources. J Immunol Methods 2012;384(1-2):10-20.