Identification of genetic biomarkers of response to conventional and biologic therapies in rheumatoid arthritis

Michael David Morgan

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Medicine

Institute of Rheumatic and Musculoskeletal Medicine

November 2013

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given with the thesis where reference has been made to the work of others.

Chapter 4: The meta-analysis of *MTHFR* SNPs has been submitted for publication in a jointly authored manuscript:

Morgan. MD., Al-Shaarawy. N., Martin S., Robinson JI., Twigg S., YEAR Consortium., Magdy AA., Omar AS., Ghattas MH., Emery P., Barrett JH., & Morgan AW., *MTHFR functional genetic variation and methotrexate treatment response in rheumatoid arthritis: A meta-analysis* (2014) , **Pharmacogenomics 15 (4), pp 467-475**

The candidate performed part of the SNP genotyping and DNA extraction with assistance from Drs Al-Sharawy and Robinson and Mr Martin. All statistical analyses were performed by the candidate with advice from Prof Barrett. The study was co-ordinated by Prof Morgan and Barrett, samples were collected and co-ordinated by Dr Twigg, Prof Emery and members of the YEAR Consortium. International student supervision was provided by Drs Magdy, Omar and Ghattas. Manuscript writing was undetaken by the candidate and overseen by Profs Morgan and Barrett.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgment.

© 2013 The University of Leeds and Michael David Morgan

The right of Michael David Morgan to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988

“There is grandeur in this view of life…”

Charles Robert Darwin (1809-1882)

Acknowledgements

Credit where credit is due. The work presented in this thesis was made possible with the aid of several notable individuals. My thanks go to Mr Stephen Martin, Mrs Lubna Haroon-Rashid and Dr Robinson for their assistance and support in the lab; to these and other members of the Morgan lab, past and present, my gratitude for putting up with me during my more morose moments. Thanks particularly go to Dr Sarah Twigg and Dr Elizabeth Hensor for the thankless task of collating together clinical data scattered across the four corners of the NHS.

Statistical advice and support was provided by Mr John Taylor; I am grateful for his infinite patience for my less intelligible and thought-through questions.

This doctoral thesis would not have been possible were it not for the patients that provided their consent and DNA samples for research. Likewise the army of nurses and physicians for collecting and co-ordinating the cohorts included in this doctoral work are recognised for their tireless work.

This work is the product of collaboration with many scientists and researchers from other academic institutions and pharmaceutical companies. Genotyping the RA susceptibility SNPs was performed as part of the UKRAG Consortium, and thanks particularly go to Mr Eddy Flynn, Dr Stephen Eyre and Prof Jane Worthington at the University of Manchester. Prof Gerry Wilson from the University of Sheffield, a member of the UKRAG Consortium, kindly provided the healthy control DNA samples used in Chapter 5. DNA on the BRAGGSS patients was provided by the members of the BRAGGSS consortium, with particular thanks to Dr Darren Plant for providing the clinical data used in Chapter 5. Many thanks go to Dr Jianmei Wang of Roche Products Ltd for collating and providing the clinical and genetic data on the Roche Clinical Trials patients. Genome-wide genotyping was performed in

collaboration with Dr Tim Bongartz at the Mayo Clinic, Rochester, USA as part of the PAMERA study; genotyping was performed at the RIKEN Centre for Genomic Medicine. I am thankful to Prof Annette van der Helm-van Mil of Leiden Medical Centre for generously providing baseline joint damage data on the EAC cohort. Many thanks to Dr Julian Knight at the Wellcome Trust Centre for Human Genetics, Oxford for kindly donating gene expression and genotyping data on the primary B cell and monocyte healthy controls.

Ultimately I am thankful for the endless support, training, advice and discussions provided by my supervisors Prof Ann Morgan and Prof Jenny Barrett. Thank you.

Abstract

Around 30-40% of rheumatoid arthritis (RA) patients fail to respond adequately to disease-modifying anti-rheumatic drugs (DMARDs). These individuals are at an elevated risk of irreversible joint damage resulting in functional impairment. Thus, there is an unmet need for markers that predict response to therapies so that treatment can be optimised before irreversible joint damage occurs.

Genetic variation provides a potential pool of biomarkers and has been the subject of investigation across drugs, including methotrexate (MTX) and TNF antagonists. This doctoral work sought to identify single nucleotide polymorphisms (SNPs) that were statistically associated with response to DMARD therapy based on their association with disease susceptibility. Nominal association was found between *HLA-DRB1* shared epitope alleles and worse response. These susceptibility SNPs were investigated in the context of early joint damage in RA; associations were observed with alleles at the *AFF3, C8orf13-BLK, IL2RB* and *KIF5A-PIP4K2C* loci. Publicly available expression data from healthy primary B cells and monocytes was used to investigate *cis* acting expression quantitative trait loci (*cis*-eQTL) over these regions. Multiple independent isoform specific *cis*-eQTLs were identified with *AFF3* transcripts.

A candidate gene association study investigated SNPs across the genomic regions encoding proteins involved in MTX metabolism and transport. A non-synonymous SNP, rs8923, in the *MTHFS* gene provides a putative causal candidate for further investigation.

TNF antagonist biologics containing Fc regions are likely to bind IgG receptors expressed on cells of the immune system. Several genes encoding the low-affinity IgG receptors (FcγR) are subject to copy number variation (CNV). An assay to measure CNV of *FCGR2C* was developed and applied in an association study of *FCGR* genetic variation with response to

TNF antagonist therapy. Association was observed between *FCGR2A* alleles and response to biologic therapies in a drug class dependent manner. These data highlight the complexity involved in identifying treatment response biomarkers.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

%ΔCRP	Percentile ranked improvement in natural log transformed CRP
%ΔDAS28	Percentile ranked improvement in DAS28-CRP
%ΔESR	Percentile ranked improvement in ESR
%ΔSJC	Percentile ranked improvement in SJC
ΔDAS28-CRP	Improvement in DAS28-CRP
ΔlnCRP	Improvement in natural log transformed CRP
ΔSJC28	Improvement in SJC28
5-ASA	5-aminosalicyclic acid
95% CI	Ninety five per cent confidence intervals
ABC	ATP-binding cassette
aCGH	Array comparative genome hybridisation
ACPA	Anti-citrullinated peptide antibodies
ACR	American College of Rheumatology
ACR20/50/70	ACR 20/50/70% improvement
ADCC	Antibody-dependent cell-mediated cytotoxicity
ADR	Adverse drug reaction
AICAR	5-aminoimidazole-4-carboxamide ribonucleotide
APC	Antigen presenting cell
ATIC	AICAR transformylase
ATP	Adenosine triphosphate
AUC	Area under the curve
BCR	B cell receptor
bFGF	Basic fibroblast growth factor
BLAST	Basic local alignment search tool
BLAT	BLAST-like alignment tool
BMI	Body mass index
bp	Base pair
BRAGGSS	Biologics in rheumatoid arthritis genetics and genomics syndicate study
BSRBR	British Society of Rheumatology biologics register
cAMP	Cyclic adenosine monophosphate
CCP	Cyclic citrullinated peptide
CD	Cluster of differentiation

CDAI	Clinical disease activity index
cDNA	Complementary DNA
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
CH2	Common heavy chain region 2
CH3	Common heavy chain region 3
CHD	Coronary heart disease
chIP	Chromatin immunoprecipitation
CN	Copy number
CNP	Copy number polymorphism
CNV	Copy number variation
CNVR	Copy number variable region
CRF	Case report form
CRP	C-reactive protein
CYP	Cytochrome P450
d.f.	Degrees of freedom
DAS	Disease activity score
DAS28	DAS with 28 joint count
ddNTP	dideoxy nucleotide triphosphate
DHF	Dihydrofolates
DMARD(s)	Disease-modifying anti-rheumatic drug(s)
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
DZ	Dizygotic
EAC	Early arthritis clinic
EDTA	Ethylenediaminetetraacetic acid
EGF	Endothelial growth factor
EIRA	Epidemiological Investigation of RA
EM	Expectation-maximisation
ENCODE	Encyclopaedia of DNA elements
eQTL	Expression quantitative trait loci
ERAN	Early RA Network
ERAS	Early RA Study
ERN	Erosion score
eSNP	eQTL SNP

ESR	Erythrocyte sedimentation rate
EtBr	Ethidium bromide
EULAR	European League Against Rheumatism
Fab	Fragment antigen binding
Fc	Fragment crystallisable
FCGR	Fc gamma receptor
FcγR	Fc γ receptor
FLS	Fibroblast-like synoviocyte
GBR	British in England and Scotland
GCTA	Genome-wide complex trait analysis
gDNA	Genomic DNA
GGH	γ-glutamyl hydrolase
GH-VAS	Global health on a 10cm visual analogue scale
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GOI	Gene of interest
GORA	Genetics of rheumatoid arthritis
GPI	Glycosyl phosphatidylinositol
GWAS	Genome-wide association study
HAQ-DI	Health assessment questionnaire - disability index
HCQ	Hydroxychloroquine
HCV	Hepatitis C virus
HLA	Human Leukocyte Antigen
hNA	Human neutrophil antigen
HWE	Hardy-Weinberg equilibrium
IBD	Identity by descent
IFN-γ	Interferon γ
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	Interleukin
IMGT	International Immunogenetics Project
IQR	Interquartile range
IRR	Incidence rate ratio
ITAM	Immunoreceptor tyrosine-based activatory motif
ITIM	Immunoreceptor tyrosine-based inhibitory motif

JSN	Joint space narrowing score
Kb	Kilobases
LCL	Lymphoblastoid cell line
LD	Linkage disequilibrium
LTα	Lymphotoxin α
mAb	Monoclonal antibody
MADGE	Microtitre array diagonal gel electrophoresis
MAF	Minor allele frequency
Mb	Megabases
MCP	Metacarpophalangeal
MHC	Major Histocompatibility Complex
MI	Myocardial infarction
MIP	Macrophage inflammatory protein
MLPA	Multiplex ligation dependent probe amplification
MMP	Matrix metalloproteinase
MREC	Multicentre research ethics committee
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
mTNF	Membrane TNF
MTP	Metatarsophalangeal
MTX	Methotrexate
MTXpg	Methotrexate polyglutamates
MuTHER	Multiple tissue human expression resource
MZ	Monozygotic
NARAC	North American RA Consortium
NAT2	*N*-acetyltransferase 2
NFκB	Nuclear factor κ-B
NGS	Next generation sequencing
NHGRI	National Human Genome Research Institute
NHS	National Health Service
NICE	NHS Institute for Health and Care Excellence
NK	Natural killer cell
NLRP3	Nod-like receptor family, Pyrin domain containing 3
NPH	Normalised peak height

NSAID(s)	Non-steroidal anti-inflammatory drug(s)
OR	Odds ratio
ORF	Open reading frame
PAMERA	Pharmacogenomics of methotrexate in rheumatoid arthritis
PBMC	Peripheral blood mononuclear cells
PCA	Principal components analysis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PharmGKB	Pharmacogenetics knowledge base
PIP	Proximal interphalangeal
PMN	Polymorphonuclear
PRT	Paralog ratio test
PSV	Paralogous sequence variant
QC	Quality control
QQ plot	Quantile-quantile plot
QSV	Quantitative sequence variant
RA	Rheumatoid arthritis
RANK	Receptor activator of nuclear factor kappa-B
RANKL	RANK ligand
RF	Rheumatoid factor
RFC	Reduced folate carrier
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RORγt	Retinoic acid-related orphan receptor γt
ROS	Reactive oxygen species
RPH	Relative peak height
RR	Risk ratio
RTX	Rituximab
s.d.	standard deviation
SA-HRP	Streptavidin horseradish peroxidase
SAM	S-adenosyl methionine
SD	Segmental duplication
SDAI	Simplified disease activity index
SE	Shared epitope

SHS	modified van der Heijde/Sharp score
SJC	Swollen joint count
SLC/OATP	Solute carrier organic anion transporter
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SP	Sulphapyridine
SSA	Sulphasalazine
sTNF	Soluble TNF
SVR	Sustained virological response
T1D	Type 1 diabetes
TBE	Tris borate EDTA
TCR	T-cell receptor
TF	Transcription factor
TFBS	Transcription factor binding site
TH1	T helper 1
THF	Tetrahydrofolates
TJC	Tender joint count
TLR	Toll-like receptor
Tm	Melting temperature
TNF	Tumour necrosis factor
TNFR1	TNF receptor type 1
TNFR2	TNF receptor type 2
Treg	Regulatory T-cell
TriPSV	Tri-allelic paralogous sequence variant
TSI	Toscani in Italia
TSS	Transcriptional start site
TYMS	Thymidylate synthase
UCSC	University California Santa Cruz
UKRAG	UK rheumatoid arthritis genetics
US	Ultrasound
US FDA	United States Food and Drug Administration
UTR	Untranslated region
VCAM1	Vascular endothelial adhesion molecule 1
VEGF-A	Vascular endothelial growth factor A

WGTP	Whole genome tiling path
WTCCC	Wellcome Trust Case Control Consortium
YEAR	Yorkshire Early Arthritis Register
ZINB	Zero inflated negative binomial

Chapter 1 Introduction – Rheumatoid arthritis, complex genetics and therapeutic treatment response

1.1 Rheumatoid arthritis

1.1.1 Rheumatoid arthritis epidemiology

Rheumatoid arthritis (RA) is a heterogeneous autoimmune disease characterised by progressive symmetrical synovitis leading to physical joint damage and functional impairment if inflammation is insufficiently managed. It is estimated to affect 0.8% of the UK population (1.16% female, 0.44% male)[1] and demonstrates a complex aetiology involving multiple genetic and environmental influences. Whilst the joints of the hands and feet, and to a lesser extent large joints of the knees, elbows and shoulders are affected, RA commonly displays extra-articular manifestations that belie its systemic nature [2]. Current first-line pharmacological treatment of RA in the UK rheumatology clinic involves the use of synthetic disease-modifying anti-rheumatic drugs (DMARDs), primarily methotrexate (MTX) and sulphasalazine (SSA), though other DMARDs have been, and continue to be, used such as hydroxychloroquine (HCQ) and leflunomide [3, 4]. Anti-citrullinated peptide antibodies (ACPA) are specific to RA, whilst immunoglobulin M (IgM) rheumatoid factor (RF) autoantibodies directed against IgG are common and frequently measured in clinical practice, but less specific to the disease itself [5].

A number of environmental exposures have been associated with the development of RA such as silica exposure and smoking (reviewed in [6]). Onset of disease generally occurs around, or after, the 5th decade though disease onset at an earlier age may also occur. This suggests that the development of RA, much like many other complex diseases, is the result of both static (i.e. genetic) and dynamic exposures, both spatially and temporally. For instance chronic exposure to cigarette smoke is a temporal influence, whilst ultraviolet radiation exposure is more often geographical.

The genetic influence on RA was speculated upon some two hundred years ago [7], though it remained contentious well into the 20th century. A systematic approach in the 1960s yielded consistent results of familial aggregation for both the occurrence and severity of RA from populations of Northern England and native American Indians [8], which suggested a polygenic inheritance. This genetic influence was also suggested by the higher prevalence amongst twins, both mono- and dizygotic [9-11].

1.1.2 Clinical and extra-articular manifestations

RA presents in the UK clinic as persistent musculoskeletal symptoms with evidence of synovitis (swelling, and tender joint upon palpation), early morning stiffness and motion impairment. The primary joints affected are the small joints of the hands, wrists and feet, but large joints such as the elbows, knees, shoulders and hips may also be affected later in disease, in addition to the cervical spine. Current imaging technologies, such as x-ray radiography, ultrasound (US – reviewed in [12]) and nuclear magnetic resonance imaging (MRI) are able to define the extent of sub-clinical synovitis, as well as the presence of bone osteitis (inflammation of the bone; MRI only) and joint damage [13, 14]. Synovitis in RA is driven by complex milieu of cytokine, inflammatory mediators and tissue remodelling molecules. This leads to the infiltration of inflammatory cells (macrophages, leucocytes, etc.) and development of pannus tissue by fibroblast-like synoviocytes, which may form germinal centres [15].

Extra-articular features of RA are relatively common (37% in a UK inception cohort over 5 years) [16]. Young & Koduri note that some extra-articular manifestations are classified as complications and *vice versa*, indicating the heterogeneity of disease trajectory as well as aetiology [2]. Table 1.1 lists these manifestations and complications, grouped into either systemic or organ system where appropriate. Rheumatoid nodules are a common (~30%) extra-articular manifestation which are often asymptomatic [2]. Many manifestations are

rare (Felty's syndrome 0.3-0.5%, secondary amyloidosis 0.6% of RA patients), particularly with more aggressive treatment in recent years, whilst others are more common and may reflect the systemic aspect of RA (localised and systemic vasculitis 4%, Sjögren's syndrome 7% [16, 17]). Cardiovascular disease is one of the greatest risks for patients diagnosed with RA, as well as other autoimmune rheumatic diseases [18]. A comparison of coronary heart disease (CHD) risk between RA and non-RA patients adjusting for known CHD risk factors found an increased risk of hospitalization for myocardial infarction (MI) or angina in the two years preceding diagnosis with RA [19].

Table 1.1. Clinical extra-articular manifestations, complications of RA and co-morbidities	
Organ affected	Clinical manifestation
Non-specific	Fever Weight loss Anorexia Malaise
Rheumatoid nodules	Subcutaneous Lung parenchymal
Cardiovascular	Vasculitis Pericardial inflammation and effusion Myocarditis Pericarditis Mitral valve disease Raynaud's phenomenon Accelerated atherosclerosis
Pulmonary	Pleural effusions Pulmonary nodules Interstitial fibrosis Pneumonitis Arteritis
Ocular	Keratoconjuctivitis sicca Episcleritis Scleritis Peripheral ulcerative keratitis
Neurological	Carpal tunnel syndrome Cervical myelopathy Mononeuritis multiplex
Dermal	Distal leg ulcers Palmar erythema Cutaneous vasculitis
Haematological	Anaemia Thrombocytosis Granulocytopenia Eosinophilia Neutropenia Hyper viscosity
Renal	Glomerulonephritis Vasculitis Secondary amyloidosis

Radiographic joint damage, detected by x-rays, US or MRI, can affect the joint through erosion of peri-articular cortical bone (erosions), demineralization of bone (peri-articular osteoporosis) or cartilage loss (joint space narrowing). The most recent European League Against Rheumatism (EULAR) recommendation for use in the 2010 ACR/EULAR RA classification criteria states that erosive disease is defined by the presence of cortical break (erosion) in at least 3 separate joints of either the wrist, proximal interphalangeal (PIP), metacarpophalangeal (MCP) or metatarsophalangeal (MTP) joints [20]. Joint damage

progression is driven through inflammatory processes in the rheumatoid joint (Figure 1.1), which may continue in the absence of clinically detectable inflammation possibly due to sub-clinical synovitis [21].

1.1.3 Autoimmunity and autoantibodies in rheumatoid arthritis

Autoimmunity in RA is characterised by the presence of two major autoantibody types; RF and ACPA. RFs are directed against the Fc region of IgG, and interact preferentially with hypoglycosylated IgG [22]; agalacotsyl IgG levels are increased in RA patients with active disease relative to age-matched healthy controls and patients with other autoimmune and inflammatory diseases [23, 24]. ACPA are highly specific to RA and recognise a number of citrullinated peptide epitopes against matrix proteins such as fibrinogen, vimentin and type II collagen and extra-cellular enzymes such as alpha enolase [25, 26]. Citrullination is a post-translational modification of arginine residues by protein-arginine deiminases that occurs by a number of biological processes, including inflammation, periodontitis and smoking, which may partially explain their association with RA, and more specifically the presence of ACPAs [27]. ACPAs are capable of activating the immune system through a number of mechanisms including both the classical and alternative complement pathways [28], and macrophage activation via Fc γ receptors (FcγR) [29]. Pre-clinical ACPA positive individuals initially present with lower antibody titres and limited reactivity to very few epitopes (epitope fine specificity) [30]. Detectable autoantibody reactivity to a broadening range of citrullinated epitopes in the same protein or even additional proteins (epitope spreading) is associated with higher titres and precedes clinical manifestation of RA [30].

1.1.4 Rheumatoid arthritis pathogenesis, cells of the immune system and the "rheumatoid" joint

Despite decades of research the aetiology of RA remains largely unknown, particularly in the early initiating stages of the disease, though key drivers of disease have been identified,

such as tumour necrosis factor (TNF). Large-scale genetic studies have highlighted a number of key cell types, and molecular and cellular pathways implicated in pathogenesis [31, 32], particularly with respect to B cell and T-cell signalling which may predispose to the development of autoimmunity *per se*, rather than RA specifically [33, 34]. The development of autoantibodies preceding disease onset by up to 10 years implies a loss of tolerance is necessary, but not sufficient, to initiate the development of RA [35]. Indeed the presence of T-cell clones with T cell receptors (TCR) reactive to self-antigens in healthy unaffected individuals forms part of the development of regulatory T-cells (T_{regs}) in the thymus [36]; increased numbers of T_{regs} are associated with disease predisposing alleles of T-cell signalling genes . This phenomenon is reinforced by the presence of detectable ACPAs in healthy controls and unaffected first degree relatives of RA patients [37].

The progression from healthy individual to autoreactivity to arthritis is complex and poorly understood. The current conjecture is that an environmental trigger, possibly an infection, in genetically and environmentally at risk individuals promotes autoreactivity and a loss of tolerance, resulting in the production of ACPA. The resulting inflammatory response would resolve in individuals not at risk of disease. However, a proportion of the population may harbour dysfunctions in the pathways involved in inflammation resolution, possibly at the genetic level. These individuals may also be at high risk from environmental influences, such as chronic smoke exposure through smoking and increasing age. At this stage of pre-clinical disease the unresolved inflammatory response may have progressed beyond antigen specificity into a more generalised response reflected by epitope spreading. Thus synovitis may represent a disease process that is already well established. Evidence suggests that citrullinated vimentin ACPAs can promote bone loss through osteoclastogenesis *in vitro*, and that later structural damage due to synovitis is secondary to this process [38]. Further evidence suggests the presence of cortical bone loss in ACPA positive individuals who are otherwise healthy, i.e. no signs of synovitis or musculoskeletal

complaint [39]. This implies bone loss precedes synovitis in ACPA positive patients. The presence of cortical fenestration in these individuals without clinical disease also implies the possibility that ACPA-producing plasma cells preside in the bone marrow. Osteitis, inflammation of the bone marrow, detected by MRI has been observed in ACPA positive arthralgia patients [40]; 12 of the 21 patients studied went on to fulfil the 2010 ACR/EULAR classification criteria within 12months.

The development of synovitis is due to the local activation and influx of immune cells (T-cells, B cells, macrophages, neutrophils) into the synovium (Figure 1.1). The transition to a clinical diagnosis of RA from undifferentiated arthritis varies between individuals and may be related to ACPA titre and epitope spreading [30, 41]. This heterogeneity in disease pathogenesis is reflected in the clinical course and progression of the disease, with patients responding differentially to pharmacological intervention, as well as in their manifestation of structural damage and extra-articular features.

RA pathogenesis and progression involves immune cells of both the adaptive and innate immune system infiltrating into the synovial joint, facilitated by angiogenesis within the synovium (Figure 1.1). Macrophages, activated via pattern recognition receptors such as the Toll-like receptors (TLRs), act as antigen-presenting cells to T-cells via MHC class II molecules. These MHC:peptide complexes form an immunological synapse with T-cells via complementary TCRs and accessory molecules CD4 or CD8. T-cell activation requires additional co-stimulatory signals, such as that provided by CD80/CD86 molecule binding to CD28. Control of TCR/CD28 T-cell activation is mediated via cytotoxic T-lymphocyte (CTLA) 4 binding to CD80/CD86 which initiates an inhibitory signal and CD80/CD86 internalisation [42, 43]. Activation of CD4+ T helper 1 (T_H1) cells leads to release of IL-2 and pro-inflammatory IFN-γ, which in turn activate T-cells and macrophages, respectively. Under non-pathological conditions IL-2 stimulated CD4+CD25+ T_{regs} [43] act as a check point to

down regulate T-cell activation by directly inhibiting T_H1 activation and acting as a sink for IL-2 (reviewed in [44, 45]). Thus T_{regs} are able to maintain a state of peripheral tolerance which is dysfunctional in RA [46]. T_H1 activated cytokine release also stimulates B cells, fibroblasts and osteoclasts, the latter of which are responsible for the breakdown of bone and the development of structural damage in RA via T-cell secreted and membrane bound RANK ligand (RANKL) [47] (Figure 1.1).

B cells in RA contribute to disease pathology via antibody production, antigen presentation and cytokine secretion. Antigen presentation by B cells can perpetuate a cycle of T-cell activation, which in turn acts upon B cells stimulating antibody production and secretion [48, 49]. Activated B cells differentiate into plasma cells which secrete the autoantibodies RF and ACPA in the rheumatoid joint [50, 51], which are in turn associated with disease severity and joint damage progression [52]. Immune complex formation between ACPA, RF and their antigen can activate macrophages through the Fc IgG receptor (FcγR) and complement [29] leading to TNF, IL-6 and IL-1 release [29, 53].

In addition to cells of the immune system, stromal cells contribute to disease pathology; fibroblast-like synoviocyte (FLS) release of pro-inflammatory cytokines such as TNF, matrix metalloproteinase (MMP) secretion and prostaglandin synthesis [54, 55] all add to the inflammatory milieu. MMP-3 levels are directly correlated with C-reactive protein (CRP) levels, and correlate strongly with a high rate of joint damage progression in early disease [56].

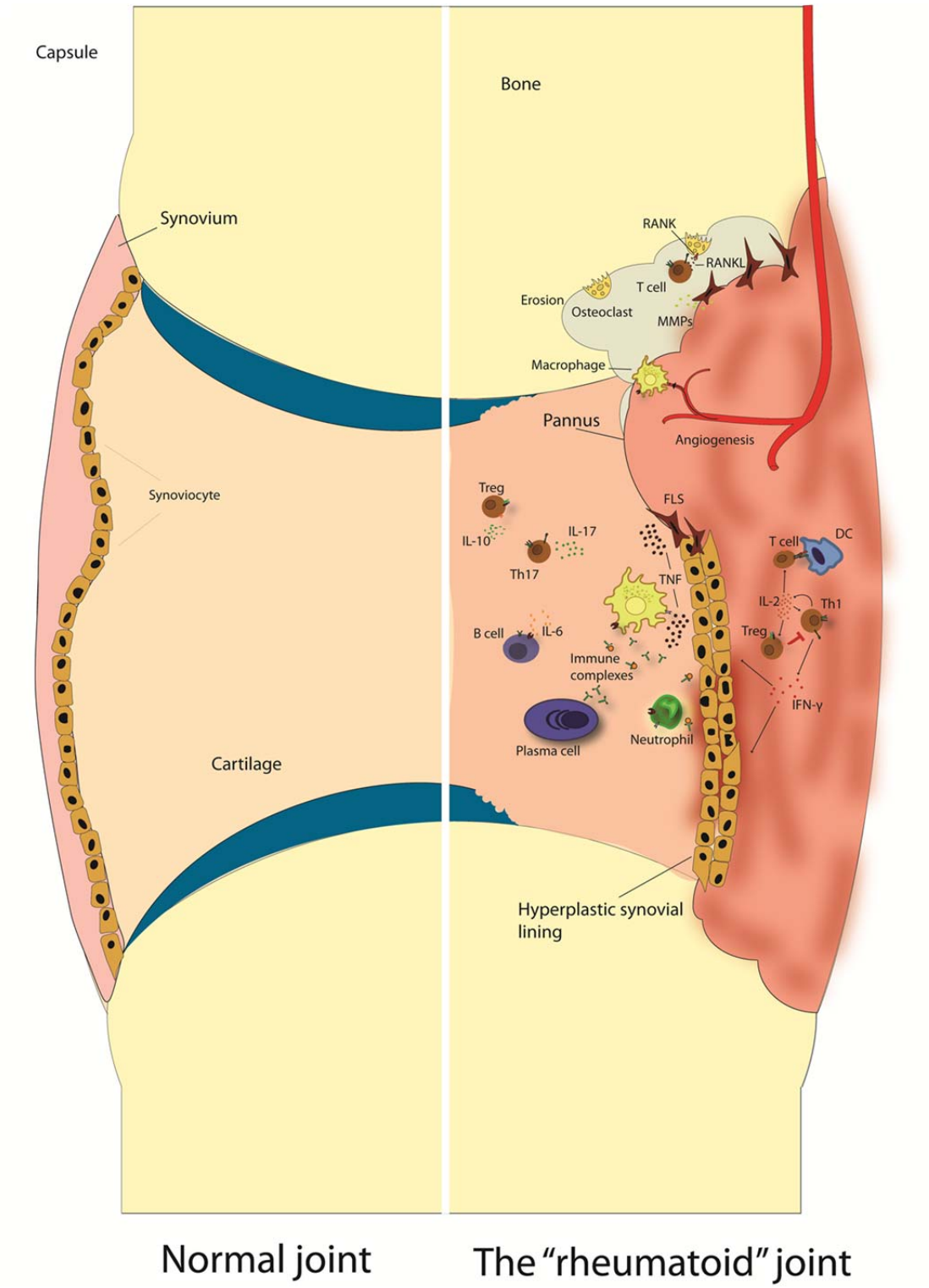


Figure 1.1. The rheumatoid joint. Synovitis results from the infiltration of T-cells, B cells, macrophages and neutrophils into the synovium by either local activation or influx enabled by angiogenesis. Activation of T-cells promotes IL-2 and IFN-γ release within the synovium. The synovial lining becomes hyperplastic, expanding the synovial membrane whilst the osteoclast enriched pannus promotes structural damage via activation of osteoclasts and secretion of degradative MMPs and other enzymes by FLS. B cells mature into plasma cells, and autoantibody production is promoted by T_H1 and BCR co-stimulation with IL-6. IL-17A released by T_H17 subsets also promotes inflammation and structural damage by activating neutrophils. Pro-inflammatory cytokine release, such as TNF and IL-1β, is propagated by immune complex deposition and activation of macrophages and neutrophils, as well as complement activation through both the classical and alternative pathways.

1.2 Rheumatoid arthritis genetics

1.2.1 Human Leukocyte Antigen and the shared epitope hypothesis

The major histocompatibility complex (MHC) locus, located on human chromosome 6p22.1-21.3, is recognised as the single strongest genetic risk factor for RA. The principle association at this locus derives from alleles of the *HLA-DRB1* gene that encodes the β chain of the HLA-DR molecule. The first recorded genetic link with HLA antigen and RA came from the HLA-Dw4 antigen detected in mixed lymphocyte culture reaction assays and HLA-DRw4 detected from separated B-cells [57, 58]. The association with these, and other, *HLA-DRB1* alleles was later hypothesised to result from a common amino acid motif that lies within the 3rd hyper-variable region of the DRβ chain at positions 70-74 (QKRAA, QRRAA, RRRAA); this motif was dubbed the "shared epitope" (SE) [59]. These SE alleles have been strongly linked with the presence of ACPA, the principal autoantibody present in RA. Evidence suggests the genetic associations with RA susceptibility may be due to a more general susceptibility to ACPA positivity as patients that are seropositive for this autoantibody display a consistently increased risk of disease [60], whilst ACPA negative disease display a different pattern of genetic risk at this locus [61]. The statistical association between RA and *HLA-DRB1* alleles is robust and has consistently been shown to be the strongest genetic risk factor for RA through both candidate gene and genome-wide association studies (GWAS). Whilst this is the case, it has been apparent from genome-wide and family-based analyses that the SE, and more generally allelic variation of *HLA-DRB1*, does not fully explain the genetic association between the MHC and RA susceptibility [62, 63]. This is partially evident from differences in risk attributed to *HLA-DRB1* alleles that contain the "SE" [64]. A number of classification systems for *HLA-DRB1* risk allele stratification have been proposed based on the amino acids at positions 67-74 [64-66]. In a recent large scale analysis of nearly 20,000 individuals (5018 ACPA+ RA patients and 14974

controls) Raychaudhuri *et al* refined the MHC association with RA to 3 principal regions; *HLA-DRB1*, *HLA-B* and *HLA-DPB1* [67]. Further to this, they demonstrated a stratification of risk based upon 4 key amino acid positions within the DRβ1 protein that corresponded to the base of the peptide binding groove of the HLA-DR molecule (Table 1.2). This included amino acid positions 11 and 13, not previously part of the shared epitope sequence. The non-*DRB1* susceptibility associations were refined to position 9 of the HLA-B protein (highest risk Asp vs. His and Tyr) which had previously been attributed to a long ancestral haplotype that also encompassed the *HLA-DRB1*03* genetic background [62, 68], and position 9 of the HLA-DPβ1 chain (highest risk for Phe vs. His and Tyr).

Table 1.2 – *HLA-DRB1* associations with RA attributed to amino acid positions 11, 13, 71 and 74 compared to SE risk category from Mackie *et al*

HLA-DRB1 allele	Classical SE allele	amino acid position				OR	"Risk" category by Mackie *et al* [64]
		11	13	71	74		
*0401	Yes	V	H	K	A	4.44	High
*0408, *0405, *0404, *1001	Yes	V	H/F	R	A	4.22	High (*0404) Medium (*1001)
*0102, *0101	Yes	L	F	R	A	2.17	Medium (*0101)
*1601	No	P	R	R	A	2.04	Non-SE D^{70} positive
*0403,*0407	No	V	H	R	Q	1.65	Neutral
*0901	No	N	F	R	Q	1.65	Non-SE D^{70} negative
*0402	No	V	H	Q	A	1.43	Non-SE D^{70} positive
*0801, *0804	No	S	G	R	L	0.71	Non-SE D^{70} positive
*0301	No	S	S	K	R	0.63	Non-SE D^{70} negative
*1102, *1103, *1301, *1302	No	S	S	Q	A	0.59	Non-SE D^{70} positive

Table 1.2. *HLA-DRB1* associations with RA attributed to amino acid positions 11, 13, 71 and 74 compared to SE risk category from Mackie *et al*. *HLA-DRB1* alleles are grouped by the amino acids at positions 11, 13, 71 and 74 with the published ORs from Raychaudhuri et al [67]. SE – shared epitope, OR – odds ratio.

1.2.2 Candidate gene and linkage studies in rheumatoid arthritis

Whilst the genetic link between the MHC and RA was consistent, establishing linkage between other loci and disease was less successful. Family-based linkage studies were often plagued by inconsistent findings, both within and between studies due to the genetic complexity, small effects of individual variants and the heterogeneity of RA [69-72]. Whilst HLA-disease associations tend to be allele specific and do not overlap between different

diseases, which may reflect differences in auto-antigen recognition and binding specificity, early evidence did show overlap between regions of linkage for RA and type 1 diabetes mellitus (T1D) outside of the MHC [73]. Candidate linkage studies based on known animal model genetic loci provided evidence of a replicable peak of linkage on chromosome 17q21-25 [74]; however, the genes and variants responsible were not identified.

Studies targeting candidate genes linked to T-cell function, specifically *CTLA4*, proved to be more successful. Initial studies lacked the statistical power to detect the modest effects of common variants on complex disease, but larger cohorts and meta-analysis increased power [75]. Plenge *et al* tested 17 candidate SNPs in a large combined cohort from the North American RA Consortium (NARAC) and the Swedish Epidemiological Investigation of RA (EIRA). They found evidence of modest associations with SNPs in *PTPN22*, *CTLA4* and *PADI4* [76], the latter of which was associated with RA in a Japanese cohort previously [77].

Whilst a number of other candidate association studies identified susceptibility loci, including SNPs across the locus containing *TRAF1* and *C5* genes, both potential susceptibility candidates [78], progress in understanding the genetic basis of RA required two major advances; the development of high-throughput genome-wide genotyping chips and large well powered cohorts provided by multi-centre collaborative consortia. Genome-wide association studies (GWAS) have enabled researchers to understand the common genetic architecture of common complex diseases that was not possible with family based linkage studies.

1.2.3 Genome-wide association studies

The Wellcome Trust Case Control Consortium (WTCCC) published it's seminal GWAS findings in 2007, covering 7 common complex diseases including RA [79]. The findings from the WTCCC confirmed the necessity of large well powered studies to probe the genetic basis of complex traits based on the modest effect sizes they observed. As a proof of

principle they replicated the strong associations between RA and *HLA-DRB1* and *PTPN22* alleles at their genome-wide significance threshold ($p=5.0x10^{-7}$), the two single largest genetic risk factors. They also reported an association with a SNP on 7q32 in a sex-differentiated analysis and 10p15 in a combined analysis with T1D patients. Moderate associations ($p>5x10^{-7}$) were reported at a further 9 loci including on 6q23, which was later confirmed by two independent studies as harbouring 2 independent intergenic risk variants between *OLIG3* and *TNFAIP3.* The latter of these two genes is a potential candidate gene that encodes the A20 protein, a nuclear factor-κB (NFκB) regulating ubiquitinase [80-82], and a variant within the coding region of *TNFAIP3* was subsequently confirmed as a third independent susceptibility variant [83, 84]. In the intervening years 10 GWAS have identified more than 30 loci associated with RA in both European and Asian populations [79, 80, 85-92]. The most recent high-density genotyping chip (Immunochip) study of RA increased this number to 60 variants across 54 loci associated with susceptibility to RA [32] (Table 1.3 – RA susceptibility loci). A number of loci were previously associated only in the Japanese population, but now show evidence of affecting disease susceptibility in patients of European ancestry [32, 88, 93]. These include the protein tyrosine phosphatase gene *PTPN2*, which shows a strong autoimmune phenotype in the T-cell conditional knock out mouse model of this gene, and is a negative regulator of TCR signalling [94]. The recent association with variants in the *PADI4* gene also demonstrates the genetic overlap in disease between European and Japanese populations [32, 88]. The enzyme encoded by this gene, peptidyl arginine deiminase type IV, citrullinates arginine residues and may be involved in the generation of citrullinated protein epitopes targeted by ACPAs.

There are regions of the genome that are not amenable to genotyping by high-throughput genotyping chips reliant on hybridisation chemistry. These regions of the genome are generally overlooked on genome-wide genotyping chips, or at the least the probes used for genotyping are not able to discriminate between highly homologous stretches of the

human genome. Some of these regions are of immunological relevance and may impact upon disease susceptibility. In particular one locus on chromosome 1q23 that encodes the low affinity IgG receptors (*FCGR* locus) has shown evidence of harbouring multiple susceptibility variants. Genomic complexity prevents accurate genotyping by genome-wide genotyping chips in this region [95]. This locus will be discussed in more depth later in Chapter 5.

Table 1.3. Reported Caucasian RA susceptibility loci identified and/or confirmed in GWAS			
Gene	**Cytoband**	**Study**	**Comment**
PTPN22	1p13	WTCCC [79], Plenge et al[76], Karlson et al [96], Stahl et al [91], Eyre et al [32], Raychaudhuri et al [90]	Strongest non-MHC risk locus for RA
CD2	1p13	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	
POU3F1	1p34	Eyre et al [32]	
PADI4	1p36	Suzuki et al [77] Eyre et al [32]	Originally identified as susceptibility locus in Japanese population
MMEL1-TNFRSF14	1p36	WTCCC [79], Barton et al [98], Raychaudhuri et al [90], Stahl et al [91], Eyre et al [32]	Possible causative SNP affecting splice site identified in *MMEL1*
IL6R	1q21	Stahl et al [91], Eyre et al [32]	Non-synonymous SNP associated with circulating *IL6R* concentrations
FCGR2A	1q23.3	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	Non-synonymous variants in strong LD with lead SNP within a region of high homology. There are multiple independent susceptibility variants at this locus not tagged in GWAS.
CD247	1q24.2	Stahl et al [91], Zhernakova et al [92]	
PTPRC	1q31.3	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	
SPRED2	2p14	Stahl et al [91], Eyre et al [32]	
REL	2p16	Gregersen et al [86], Eyre et al [99], Stahl et al [91], Eyre et al [32]	Multiple independent associations
AFF3	2q11	Barton et al [100], Stahl et al [91], Eyre et al [32]	
STAT4	2q32	Remmers et al [101], Barton et al [102], Raychaudhuri et al [90], Stahl et al [91], Zhernakova [92], Eyre et al [32]	Multiple independent associations
CD28	2q33	Stahl et al [91], Raychaudhuri et al [97], Eyre et al [32]	Multiple independent associations
CTLA4	2q33	Plenge et al [76], Raychaudhuri et al [90], Gregersen et al [86], Karlson et al [96], Stahl et al [91], Eyre et al [32]	
DNASEIL3-PXK	3p14	Stahl et al [91], Eyre et al [32]	Predicted non-synonymous causative variant in *DNASEIL3*
RBPJ	4p15	Stahl et al [91], Eyre et al [32]	
IL2/IL21	4q27	Zhernakova et al [103], Raychaudhuri [90], Karlson et al [96], Stahl et al [91], Eyre et al [32]	
ANKRD55	5q11	Stahl et al [91], Eyre et al [32]	
C5orf30	5q21	Stahl et al [91], Eyre et al [32]	
PRDM1	6q21	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	

Gene	Cytoband	Study	Comment
TNFAIP3	6q23	WTCCC [79], Thomson et al [81], Plenge et al [80], Barton et al [98], Raychaudhuri et al [90], Orozco et al [84], Shimane et al [104], Karlson et al [96], Stahl et al [91], Eyre et al [32]	Three independent effects, two intergenic and one within *TNFAIP3*
TAGAP	6q25	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	
CCR6	6q27	Kochi et al [87], Stahl et al [91], Eyre et al [32]	Functional regulatory dinucleotide polymorphism affecting *CCR6* expression and IL-17 secretion
CDK6	7q21	Raychaudhuri et al [90]	
IRF5	7q32	Stahl et al [91], Eyre et al [32]	
PODXL	7q32	WTCCC [79], Barton et al [98]	
BLK	8p23	Gregersen et al [86], Stahl et al [91], Eyre et al [32]	
CCL21	9p13	Raychaudhuri et al [90], Stahl et al [91], Eyre et al [32]	
TRAF1/C5	9q33	Kurreeman [78], Plenge et al [105], Barton et al [102], Chang et al [106], Karlson et al [96], Stahl et al [91], Zhernakova et al [92], Eyre et al [32]	
GATA3	10p14	Eyre et al [32]	Statistical interaction with *PRKCQ* locus SNPs
IL2RA	10p15	WTCCC [79], Barton et al [98], Stahl et al [91], Eyre et al [32]	
PRKCQ	10p15	Raychaudhuri et al [90], Barton et al [98], Stahl et al [91], Eyre et al [32]	Statistical interaction with *GATA3* locus SNPs
ARID5B	10q21	Okada et al [88], Eyre et al [32]	First identified in Japanese meta-analysis
TRAF6	11p12	Raychaudhuri et al [97], Stahl et al [91], Eyre et al [32]	
CD5	11q12	Eyre et al [32]	Predicted causative non-synonymous SNP in *CD5* coding region
DDX6	11q23.3	Zhernakova et al [92], Eyre et al [32]	
KIF5A-PIP4K2C	12q13	Raychaudhuri et al [90], Barton et al [98], Stahl et al [91], Eyre et al [32]	Susceptibility association signal refined to region near *OS9*
SH2B3	12q24	Stahl et al [91], Zhernakova et al [92]	
CRYL1	13q12	WTCCC [79], Barton et al[98]	
BATF	14q24	Stahl et al [91]	
RASGRP1	15q14	Eyre et al [32]	Multiple independent associations
TLE3	15q23	Eyre et al [32]	

Gene	Cytoband	Study	Comment
IRF8	16q24	Okada et al [88], Eyre et al [32]	Originally identified in a Japanese meta-analysis
IKZF3	17q12	Stahl et al [91], Eyre et al [32]	
PTPN2	18p11	Okada et al [88], Cobb et al [93]	Originally identified in a Japanese meta-analysis
TYK2	19p13	Eyre et al [32]	Multiple independent associations and possible causative non-synonymous variant *TYK2* coding region
CD40	20q13	Raychaudhuri et al [90], Karlson et al [96], Orozco et al [107], Stahl et al [91], Eyre et al [32]	
RCAN1	21q22	Eyre et al [32]	
RUNX1	21q22	Eyre et al [32]	
UBASH3A	21q22	Zhernakova et al [92], Stahl et al [91]	
UBE2L3	22q11.2	Zhernakova et al [92], Stahl et al [91]	
IL2RB	22q13	WTCCC[79], Barton et al [98], Stahl et al [91], Eyre et al [32]	
IRAK1	Xq28	Eyre et al [32]	First identified non-autosomal RA association

Table 1.3. RA susceptibility loci identified and/or confirmed in GWASs. Loci are listed by cytoband as associations often occur in intergenic regions or in genomic intervals with wide-spanning blocks of LD. Fine-mapping studies, such as the Immunochip study, have refined some of the association signals to smaller genomic intervals or single candidate causative SNPs indicated in the comments column.

1.2.4 Rheumatoid arthritis susceptibility loci and disease pathogenesis

GWAS have been phenomenally successful at identifying statistical associations between RA and common genetic variants across the human genome; the genes implicated by these studies point towards a disease initiated, and possibly driven by, T-cell function, development, and signalling. Further evidence of a pathogenic role for T-cells in RA has come from a number of studies that have investigated how gene transcriptional regulatory elements associated with susceptibility loci are enriched in particular cells of the immune system in mice [108] and humans [31]. Figure 1.2 shows the overlap between cellular pathways involved in RA pathogenesis and genes implicated in RA susceptibility from GWAS. The hypothesis that genetically susceptible individuals develop autoimmunity through a breakdown in central tolerance is supported by the genetic evidence. Multiple susceptibility variants lie within, or proximal to, genes involved in TCR signalling (*PTPN22*), interactions between T-cells and APCs (*HLA-DRB1, CTLA4, CD28*), including during thymic development, and T-cell activation cytokines, or their receptors (*IL2, IL2RA, IL2RB*). There is also evidence that cytokine imbalance either from effector or target cells is also affected in genetically at risk individuals (*IL6R, TYK2, and STAT4*). The importance of TNF as a central driver of RA is not to be overlooked, with genes that code for proteins involved in TNF response and signalling also associated with disease susceptibility (*TRAF6, REL, and TNFAIP3*). For most susceptibility loci, the causative variant(s) are yet to be identified, let alone characterised for their impact on human biology and thus how they affect disease. What is clear however is that these susceptibility loci implicate biological processes in disease pathology that may be, and in some instances are, targets of therapeutic agents (*CTLA4* – abatacept, *IL6R* - tocilizumab).

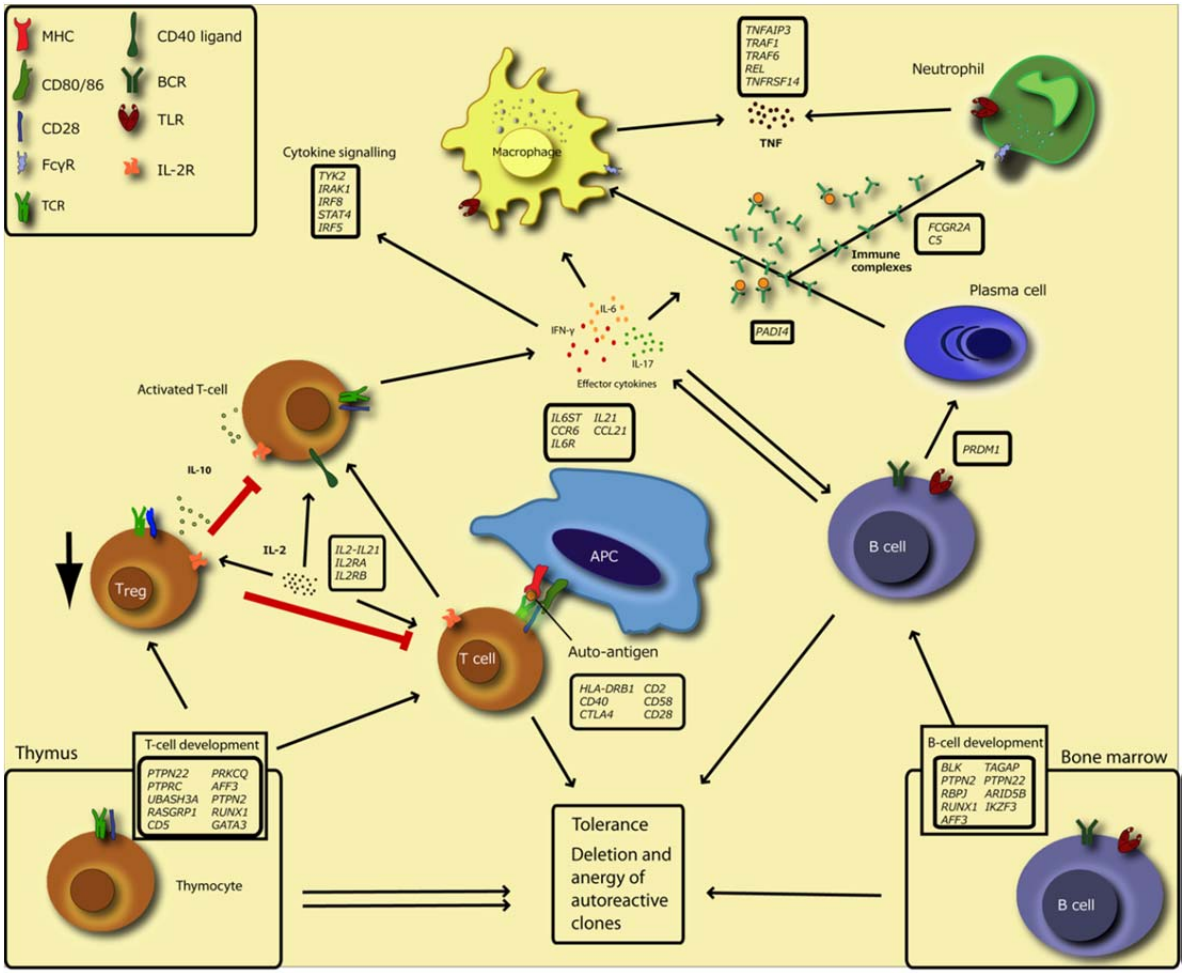


Figure 1.2. Genetic susceptibility loci overlap with RA pathogenesis associated cellular pathways and processes. Genetic susceptibility loci overlap with B-cell and T-cell development, and antigen receptor signalling pathways which may lead to the development of an autoreactive T-cell and B-cell repertoire in genetically susceptible individuals (e.g. *PTPN22, TAGAP, AFF3*). This process may be facilitated by interactions between naive T-cells and antigen presenting cells (APCs) during thymic selection or within lymph nodes where they then migrate to the periphery (e.g. *CTLA4, CD28, CD40*). Peripheral tolerance is maintained by a combination of T_{reg} activity and cytokine signalling (e.g. *IL2, IL2RA, IL2RB*). Effector cytokine signalling on target cells may also be affected by genetic susceptibility variants (e.g. *TYK2, IRF5, STAT4*). Innate immune cells and mechanisms are also important in disease pathogenesis and progression, particularly neutrophils and macrophage stimulation and TNF signalling pathways (*REL, TRAF6*). The central importance of autoantibodies and immune complexes are highlighted by susceptibility variants near the genes *PADI4* (citrullination of arginine residues) and *FCGR2A* (receptor for Fc region of IgG).

1.3 Rheumatoid arthritis treatment

1.3.1 Treating rheumatoid arthritis and the "window of opportunity"

The early treatment paradigm for RA was based around a pyramid of treatment starting with physiotherapy and escalating through non-steroidal anti-inflammatory drugs (NSAIDs) and analgesia prescription as disease severity warranted it, gauged by the treating physician. Administration of prednisolone was reserved for severe disease and DMARD therapy was often only given after several years of persistent disease. This treatment strategy was thrown out in favour of a "window of opportunity" hypothesis that aimed to test early aggressive treatment as a means to prevent structural joint damage and thus limit patient disability [109, 110]. The theoretical basis was that there existed a particular window in the disease progression that would be most amenable to DMARD therapy to limit its chronicity and severity; however, the definition of this "window" was not straightforward. The consensus was that the first 12 weeks of symptom onset constituted the optimal window, but this was confounded by differences in definition of symptom onset itself, be it by the patient due to painful joints or by the physician with indication of joint swelling and synovitis. In order to undertake a systematic review of symptom duration and outcome van Nies *et al* used a qualitative approach to determine whether the window of opportunity existed, and if possible when [111]. They looked at two outcomes - remission and severity of radiographic joint damage - and found a shorter symptom duration was associated with less severe joint damage progression. Additionally they found treatment initiation after 12 weeks (post-"window of opportunity") was associated with a lower probability of achieving DMARD-free remission. However, these are based on limited data and account for neither the possibility of publication bias nor the heterogeneity in symptom onset definition which precluded a full meta-analysis of all studies identified. Evidence of a synovial cytokine profile that differentiates between early

RA and other arthritides, including established RA may lend weight to the "window of opportunity". These individuals displayed a transient profile of T-cell, macrophage and stromal cytokines, including interleukin (IL)-2, IL-4, IL-13, IL-17, macrophage inflammatory protein (MIP)-1α, granulocyte-macrophage colony-stimulating factor (GM-CSF), endothelial growth factor (EGF) and basic fibroblast growth factor (bFGF) within 3 months of symptom onset [112].

With the change in treatment paradigm has come an increased use of synthetic DMARDs. The ultimate goal of drug-free remission in RA through reprogramming the immune system remains largely out of reach, though achieving states of low disease activity or clinical remission are more realistic. Rapid dose escalation with DMARDs and more intensive use has bought about this change in disease management.

1.3.2 Disease modifying anti-rheumatic drugs

1.3.2.1 Methotrexate

Introduced comprehensively into rheumatology clinics in Europe and the US in the 1980's, MTX, a folate antagonist, has now been adopted as the first line therapy for RA in the UK [4]. MTX is used to treat multiple haematological malignancies within the context of oncology as a cytotoxic agent, however, it is administered at considerably higher doses than that used to treat RA (5000mg/week vs. 15-25mg/week). Hence it is often referred to as low-dose MTX in the rheumatology literature. This is an important distinction as the primary mechanism of action may be altered depending on the magnitude of dose given.

At high doses MTX inhibits the intra-cellular enzyme dihydrofolate reductase (encoded by *DHFR*) which catabolises the reduction of dihydrofolates (DHFs) to tetrahydrofolates (THFs). These THFs then act as 1-carbon donors for purine and pyrimidine synthesis; disruption of this metabolic process interrupts DNA replication, repair and synthesis

(reviewed in [113]). Thus rapidly dividing cells are sensitive to the antagonistic effects of this drug, giving it its potent anti-proliferative potential.

Despite the known molecular targets of MTX, its exact anti-inflammatory mechanism of action is yet to be fully elucidated. A number of mechanisms have been put forward, with varying levels of evidence in support of each, including disruption of *de novo* purine and pyrimidine synthesis as described above, polyamine formation inhibition, and the accumulation and extra-cellular release of adenosine.

After absorption MTX competes with reduced folates for active transport across the cell membrane via the reduced folate carrier (RFC-1). Once transported into the cell MTX undergoes intracellular metabolism to its (reversible) polyglutamated form by γ-glutamyl hydrolase (encoded by *GGH*) [114], which act as more potent inhibitors of dihydrofolate reductase, 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase and thymidylate synthase than the parent drug [115]. MTX polyglutamation also inhibits efflux from the cell by ATP-binding cassette (ABC) transporters, maintaining intracellular concentrations [116].

An increase in the concentration of polyamines has been observed in the synovium and other bodily fluids, including peripheral blood mononuclear cells (PBMCs), of RA patients [117]. Polyamines, such as spermine and spermidine, are thought to be metabolised into lymphotoxic hydrogen peroxide and amines by monocytes [118]. THF acts as a methyl-group donor for the formation of methionine, which is converted to the methyl donor S-adenosyl methionine (SAM). SAM is required for transmethylation reactions, which include the formation of the polyamines spermine and spermidine. Whilst this evidence appears compelling there are no reports as to the clinical efficacy of a transmethylation reaction inhibitor, 3-deazaadenosine, despite initial promising *in vitro* evidence [119].

Given the low-dosage of MTX used to treat RA it is has been suggested that the anti-proliferative effects of dihydrofolate reductase inhibition are unlikely to account for the anti-inflammatory effects observed [120]. Rather AICAR transformylase inhibition leading to the accumulation of adenosine may be the principle mechanism of action. Purified AICAR transformylase was shown to be inhibited in a dose dependent manner by increasing MTX polyglutamates (MTX_{Glu1-5}; 0.057 – 143.9μM), with 2500-fold more potent inhibition by the pentaglutamate moiety of MTX compared to the parent compound. This indicates AICAR transformylase is strongly inhibited by MTX_{pg} at least *in vitro* [121].

Cronstein and colleagues noted marked differences in extracellular adenosine accumulation by human dermal fibroblasts and umbilical vein endothelial cells in response to both MTX and stimulated neutrophils, including a marked increase in adherence of stimulated neutrophils [122]. Following these observations they tested the effects of MTX induced adenosine release in a murine air-pouch model of inflammation. In a dose-dependent manner, MTX was able to diminish leukocyte accumulation, increase AICAR concentrations from splenocytes of treated mice and increase adenosine in the exudates of the same mice. The anti-inflammatory effects were diminished by the addition of adenosine deaminase, which irreversibly converts adenosine to its metabolically inert metabolite inosine. Further, they provided preliminary evidence that the anti-inflammatory effects observed were mediated through the adenosine receptor A_2, but not A_1 [123]. The principle metabolic pathways are shown in Figure 1.3, including the genes encoding the relevant enzymes and receptors implicated in MTX mechanism of action.

The effect of MTX as a folate antagonist has the drawback of drug induced toxicity involving multiple organ systems. Morgan *et al* hypothesised that the similarity between low-dose MTX treatment toxicities and folate deficiency could be alleviated by the addition of supplemental folic acid with low-dose MTX [124]. They found a marked reduction in drug

toxicity in the folate supplemented arm of their clinical trial (1mg/day) which did not appear to affect drug efficacy. This effect on reduced toxicity was replicated in a second larger clinical trial by the same group [125]. Folic acid prescribed as a supplement to low-dose MTX has now been widely adopted in the rheumatology clinic. The lack of a large influence on MTX efficacy with folate supplementation suggests that inhibition of dihydrofolate reductase is not central to the mechanism of action of low-dose MTX in RA.

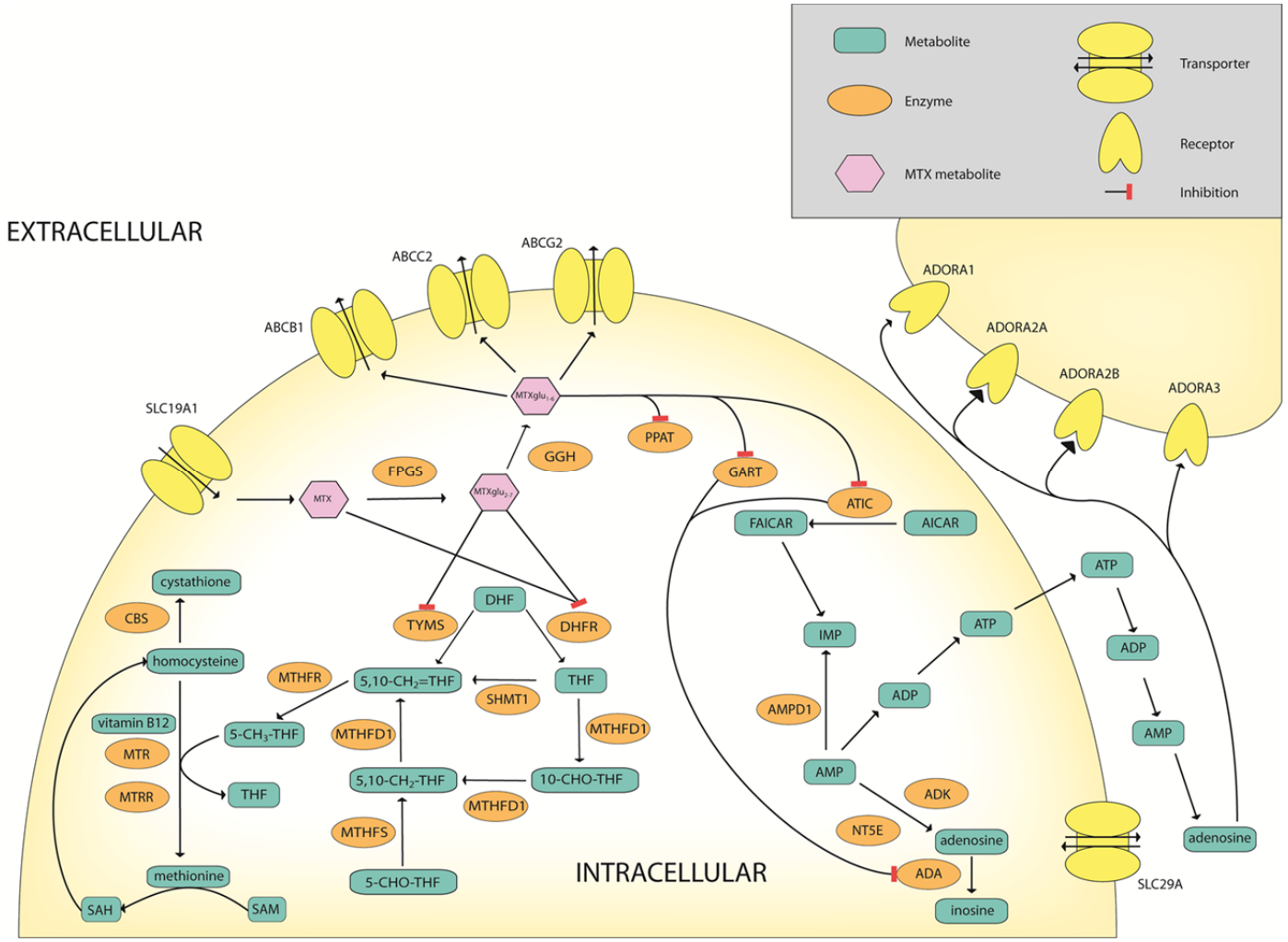


Figure 1.3. Metabolic pathways involved in MTX mechanism of action. MTX enters cells through the reduced folate carrier (SLC19A1) where it undergoes reversible polyglutamation by FPGS which maintains intracellular concentrations and inhibits export by ATP-binding cassette (ABC) transporters. MTXpgs inhibit AICAR transformylase (ATIC) which leads to an accumulation of AICAR and subsequently adenosine or adenosine nucleotides; the latter of which are metabolised to adenosine by ecto 5' nucleotidase (NT5E). Adenosine binds to its cognate extracellular receptors. The principle anti-inflammatory effect is mediated through adenosine receptor 2 (ADORA2A, ADORA2B) which is indicated by large arrows. MTXpg also disrupt *de novo* pyrimidine synthesis by direct inhibition of dihydrofolate reductase (DHFR) and thymidylate synthase (TYMS). Folate antagonism also interrupts polyamine synthesis and may lead to the accumulation of homocysteine, a substrate that may accumulate in the cerebrospinal fluid leading to neurological drug toxicity. Adapted from [308].

1.3.2.2 Sulphasalazine

Prior to dose-escalated MTX use in RA, SSA was a commonly used DMARD, and is often used in a contemporary setting in combination with MTX, or where MTX is not tolerated due to iatrogenic toxicity. Initial trials in long-standing RA patients showed promising results with SSA not only as an anti-inflammatory agent, but also as a DMARD in its capacity to reduce acute-phase reactants after 6 weeks treatment [126]. A small fraction of orally administered SSA is absorbed (30%), the majority undergoes colonic metabolism by bacterial azo-reductases where it is cleaved into its two active moieties 5-aminosalicyclic acid (5-ASA) and sulphapyridine (SP) [127]. Whilst 5-ASA is largely excreted, it is thought to be responsible for the therapeutic benefit seen in inflammatory bowel disease, whereas SP shows evidence of anti-inflammatory and disease-modifying activity in RA [128]. SP is also metabolised in the liver by *N*-acetyltransferanse 2 (encoded by *NAT2*), thus it may be affected by the genetically determined acetylator phenotypes of this enzyme [129].

The functional effects of SSA in RA are largely unknown, however, evidence points towards the most potent anti-inflammatory effects being attributable to the parent molecule despite the extensive colonic metabolism to 5-ASA and SP. SSA is a more potent inhibitor of AICAR transformylase *in vitro* than MTX which may contribute towards its anti-inflammatory effects in a similar way to MTX [130]. The SSA parental molecule may also affect cytokine production as patients show a decline in serum TNF and IL-1α associated with clinical improvement; however this may be secondary to the overall anti-inflammatory effects of SSA by other mechanisms as serum IL-1β and IL-6 were unaffected [131]. This anti-inflammatory effect appears to be mediated through an inhibition of the pro-inflammatory NF-κB transcription factor activation *in vitro* that was not displayed by either SP or 5-ASA [132].

1.3.3 Biologics in rheumatoid arthritis

Synthetic DMARDs remain the first-line treatment for RA in the UK clinic, though approximately 30% of patients may fail to achieve an adequate clinical response [133]. NHS Institute for Health and Care Excellence (NICE) guideline CG79 recommends anti-TNF therapy for patients refractory to at least two previous DMARDs (including MTX) and with active disease (DAS28>5.1). Biologic therapeutic agents that target specific cytokines and molecules involved in disease pathology are available and have helped to revolutionise treatment of RA and other autoimmune diseases. The development of these drugs has been driven by advances in the understanding of disease pathology, notably of TNF.

1.3.3.1 Anti-tumour necrosis factor biologics

The first anti-TNF therapy to be tested in RA was infliximab, which demonstrated both the clinical efficacy of this chimeric monoclonal antibody (mAb) and the central role of TNF as a driver of disease pathology in humans [134]. Currently 5 TNF blockade biologics are licensed for the treatment of RA in the UK (Table 1.4), including 3 mAbs, a recombinant TNF receptor 2 (TNFR2) fusion protein (etanercept) and a TNF-specific Fab fragment fused to polyethylene glycol (certolizumab pegol). All 5 biologics are able to bind both the transmembrane (mTNF) and soluble (sTNF) forms of TNF *in vitro* [135], though not on CD3/CD28 stimulated peripheral blood lymphocytes from healthy control volunteers [136]. Etanercept is also capable of binding to lymphotoxin α, as this also binds to TNFR2 [137]. Each mAb is bivalent which means it can bind two TNF monomers; TNF exists as a homotrimer which implies that it can bind three mAbs [138]; etanercept is only able to bind the homotrimer of TNF [138]. Patients with an inadequate response to etanercept may benefit from an alternative anti-TNF biologic, such as infliximab [139]. Etanercept rapidly dissociates and re-associates with sTNF, whilst infliximab associates with sTNF and does not dissociate in the presence of a number of competitive binder molecules [138]. This

difference in the stability of TNF inhibition by these two biologics may partially explain why switching drugs proves effective.

Table 1.4. Anti-TNF biologics licensed for treatment of RA				
Infliximab	Adalimumab	Golimumab	Etanercept	Certolizumab pegol
Human-murine Chimeric mAb IgG	Fully humanised mAb IgG	Fully humanised mAb IgG (humanised infliximab)	TNFR2 fusion to human IgG1 Fc hinge, CH2 and CH3 region	TNF-specific Fab bound to PEG
Bivalent	Bivalent	Bivalent	Monovalent	Monovalent
mTNF and sTNF	mTNF and sTNF	mTNF and sTNF	mTNF, sTNF, mLTα and sLTα	mTNF and sTNF
IgG1	IgG1	IgG1	IgG1	NA

1.3.3.2 Rituximab

The role of B cells in the pathology of RA is suggested by the central pathological role of autoantibodies in the disease, their presence in the synovium and secretion of pro-inflammatory cytokines. Currently rituximab (RTX), which targets the B cell marker CD20, is the only B cell depletion therapy approved for use in RA; this biologic is also approved for use in non-Hodgkin lymphoma. RTX targets the non-proliferating pool of CD20+ memory B cells and cell populations derived from them as progenitors, including autoreactive Ig-producing plasmablasts. Early recovery of memory B cells, specifically IgD+CD27+ cells, is indicative of a poor response after an initial cycle of RTX therapy [140]. RTX is often administered in two cycles; the second is given 2 weeks after the first. Complete peripheral blood depletion after the first infusion is associated with a persistently better EULAR response [141]. Whilst antibody titres decline upon treatment with RTX, these correlate poorly with clinical response [142, 143], suggesting alternative B cell intrinsic pathways in disease pathology. B cell depletion is associated with reduction in HLA-DR+ B cells in the peripheral blood and bone marrow which suggests a reduction in antigen presentation may be involved in the clinical efficacy of RTX [144]. IL-17A producing T-cells are thought to play a significant role in the inflammatory process in RA; RTX can inhibit the Th17 response *in*

vitro. Patients treated with RTX also show reduced synovial expression of the Th17 cytokines IL-17A and IL-22, as well as the transcription factor retinoic acid-related orphan receptor γt (RORγt) [145]. Together these imply the efficacy of RTX is complex. That RTX induces B cell depletion is indisputable; however, how it achieves this is a matter of some debate. Initial studies of RTX mechanism of action relied on cells from cancer patients and immortalised B cell lines [146], but were later confirmed in primary cells from RA patients and healthy controls; RTX induces apoptosis in target B cells [147].

RTX is able to induce apoptosis of B cells through antibody-dependent cell-mediated cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). RTX bound CD20+ B cells become targets of Fc γ receptor III (FcγRIII; CD16) mediated ADCC by NK cells, monocytes and macrophages [148]. C1q bound RTX is able to activate complement through the classical pathway which may contribute to the immediate effectiveness of this biologic [149, 150].

1.4 Pharmacogenetics

1.4.1 Pharmacogenetics – an introduction

Pharmacogenetics seeks to explain the genetic contribution towards individual variation in response to drugs and susceptibility to potential drug toxicities. In genetic epidemiology the initial approach to understanding the genetic aetiology of a complex trait often takes the form of family-based studies that seek to measure the general strength of the heritable component (e.g. through the familial relative risk, λ). The gold standard for identifying the level of genetic liability in a trait is to study monozygotic (MZ) and dizygotic (DZ) twins, comparisons between the two can also help to determine the level of shared environment that contributes to the trait in question. With pharmacogenetics twin studies and even other family-based studies are far less feasible. Where common diseases do run in families it is unusual that they will receive the same treatment regimes. As such, collecting cohorts large enough to study the heritable component of pharmacological response and toxicity (even more so where toxicity is generally a *rare* event) becomes prohibitive and unfeasible. Therefore what evidence is there of a genetic influence on pharmacological response and propensity to adverse drug reactions (ADRs)?

The basis for this is two-fold: historical and functional. In the 1950's, observations of drug response to isoniazid, primaquine and succinylcholine appeared in the literature, largely dealing with idiosyncratic drug reactions [151]. It emerged that variations in the metabolic profiles of patients was linked with a common adverse reaction when treated with isoniazid for tuberculosis [152], and that susceptibility to the ADR was higher in those with slow acetylation of the parent compound [153]. A more severe adverse reaction, apnoea, a prolonged paralysis during surgery, was noted in patients receiving the muscle relaxant succinylcholine [154]. A deficiency of a cholinesterase enzyme appeared to be inherited as an autosomal recessive trait, with a number of mutations now known to lead to

succinylcholine apnoea (reviewed in [155]). These documented Mendelian conditions suggest that variation in drug metabolism may sometimes have a strong genetic component. In addition, because drug metabolism and transport are catalysed by enzymes encoded by genes, genetic variation within these genes may lead to differences in pharmacological profile, response and susceptibility to ADRs. The heritability and phenotypic variance of a complex trait can also be estimated in samples of unrelated individuals from high-density genotyping data (e.g. genome-wide) by programs such as genome-wide complex trait analysis (GCTA) software [156]. This approach has been applied to pharmacogenetic measures of response in asthma clinical trials patients [157]. The heritability traits estimated using GCTA were in agreement with published estimates from pedigree data. Together this evidence suggests a, perhaps variable, role for genetic variation in response to a wide range of pharmacological interventions for many different conditions.

Most drugs undergo two major phases of metabolism: hydrolysis, oxidation and/or reduction (phase I), and conjugation (phase II). The cytochrome P450s are the primary mediators of phase I metabolism, and genetic variation within the genes, such as *CYP2D6*, that encode these proteins are known to determine metabolic phenotypes for a number of important pharmaceuticals, including opioid analgesics (*CYP3A4*) [158], warfarin (*CYP2C9*) [159] and tamoxifen (*CYP2D6*) [160]. Drug transport proteins, including the ATP-binding cassette (ABC) and solute carrier organic anion transporter (SLC/OATP) families have been causally linked to a number of differential drug responses and ADRs [161, 162], suggesting an important role for this class of protein in clinical response to a number of therapeutics. Genetic predictors of response are not confined to germline variants, but also extend to acquired mutations in the context of malignancies. Mutations clustering in codons 12, 13 and 61 in exons 2 and 3 of the *KRAS* gene provide a means to stratify patient treatment for the anti-epidermal growth factor (anti-EGF) biologic agents Cetuximab and Panitumumab

(reviewed in [163]). Further evidence of the clinical utility of pharmacogenetic biomarkers can be found in the treatment of sporadic breast cancer with tamoxifen, with a number of commercialised products approved by the US FDA (reviewed in [164]).

1.4.2 Methodological approaches to pharmacogenetic association

The highly variable nature of pharmacological response and susceptibility to ADRs between individuals suggest they are likely to be under the influence of a number of genetic variants (polygenic), in addition to other factors.

Two principal methodological approaches have been adopted in order to understand the genetic aetiology of complex traits: candidate gene analysis and GWAS. Candidate gene approaches adopt a focused method towards association testing by selecting specific genes and/or variants which are likely candidates based on previous published evidence or putative functional roles in the trait of interest. This allows the researcher to select genetic variants across the candidate region(s) in the most cost-effective manner. However, even if a region appears as a good candidate, genetic variation within the gene of interest may not be a factor in its aetiology. In addition this approach may miss variation that contributes to the trait of interest that lies in a region lacking intuitive or logical candidate genes.

A candidate gene approach was used to identify genetic variation across the gene encoding vitamin K epoxide reductase 1 (*VKORC1*) that affects warfarin response [165]. This gene was selected because it is the known drug target of coumarin compounds, the class of anti-coagulant drugs to which warfarin belongs. Likewise variants in the cytochrome P450 *CYP2C9* gene were investigated based on this gene's role in warfarin metabolism; non-synonymous variation was associated with lower dose requirement and higher rate of bleeding episodes [166].

The GWAS is a more recent method that has only become feasible due to the technological advent of genome-wide profiling gene chips that have the ability to genotype hundreds of thousands of SNPs across the entire human genome. GWAS allow a hypothesis-free approach to genetic association that overcomes the major pitfall of the candidate gene approach. GWAS suffer from their own caveats that are important when interpreting their results. Firstly genotyping thousands of markers across the genome introduces the likelihood of many false-positive associations occurring; an adjustment to account for this multiple testing burden needs to be applied. A common multiple testing adjustment is a Bonferroni correction ($p = \frac{\alpha}{n}$, where n = number of tests performed) which is considered conservative as it assumes each separate statistical test is independent. This correction reduces the probability of false-positive associations at the expense of statistical power [167]. The WTCCC applied a Bayesian approach of *a priori* assumption about the expected number of positive associations and the power of their study to detect a given effect size; a genome-wide significance level was estimated ($5x10^{-7}$) which has now been largely adopted for GWAS [79].

Many pharmacogenetic association studies have adopted the candidate gene approach based on the known pharmacological action of the drug. There have also been a number of GWAS of drug response, highlighting the advantages of this approach. A GWAS of hepatitis C virus treatment with pegylated interferon α-2b/2a and ribavirin identified a polymorphism near the *IL28B* gene encoding interferon-λ-3, associated with a sustained virological response (SVR) [168]. This SNP, rs12979860, is estimated to explain around half of the differences in response rates between different populations. The magnitude of association for this single SNP exceeds other common clinical factors associated with SVR (OR 7.3 [5.1-10.4]). The large effect size for this variant is unlikely to be representative of

pharmacogenetic variation, particularly for complex diseases where treatment response is a complex mix of genetic, clinical and environmental factors.

The identification of predictive and prognostic biomarkers of response is an unmet need in RA where around 30% of patients fail to respond to their first DMARD, leading to a delay in effective treatment and therefore the accrual of irreversible structural damage. Further to this, patients that show a moderate response to treatment may still be at risk of developing functional disability [169].

1.4.3 Methotrexate pharmacogenetics

Historically the major findings of MTX pharmacogenetics have been within the context of the haematological malignancies, with links to the MTX metabolising dihydrofolate reductase (*DHFR*) and folylpolyglutamate synthase (*FPGS*), and toxicity related methylene tetrahydrofolate reductase (*MTHFR*) genes [170]. Within the context of RA these are logical candidate genes, particularly with two well described functional variants within the *MTHFR* gene [171, 172] (rs1801133 and rs1801131). Knowledge of biochemical pathways and proposed mechanisms of action of MTX in RA have lead many groups to take a candidate gene approach and select those genetic variants most likely to have functional consequences, i.e. non-synonymous and regulatory SNPs. The two *MTHFR* functional variants, rs1801133 and rs1801131, have been tested for association with both toxicity and efficacy, with both positive [173-180] and null [174, 176, 178, 179, 181-186] findings for both. The positive associations with efficacy, however, are limited to two single studies [177, 180], results of which have failed to be replicated in other independent cohorts [176, 181-183, 185].

Many studies have looked for associations with variants within a number of other candidate genes. Focus has been on key enzymes within the folate metabolism, MTX transport and adenosine accumulation and release pathways. Table 1.5 gives an overview

of the genetic variants associated with MTX efficacy to date (2013). A limited number of genetic associations with MTX efficacy have been replicated in independent cohorts. SNPs in the gene encoding AICAR formyltransferase (*ATIC*) have been tested for association in 6 cohorts with a range of outcome measures used to define treatment response [180, 185, 187-190]. Three of these found associations with SNPs associated with the odds of successful treatment, gauged by the respective outcome measure used. Owens *et al* [189] tested SNPs across the *ATIC* gene and found several SNPs with nominal associations ($p<0.05$), the strongest of which lies within the 8th intron of *ATIC* (rs3821353); conditional logistic regression analysis in their cohort did not find multiple independent signals. A non-synonymous SNP, rs2372536 that was previously investigated for its role in MTX response was not in LD with the associated SNPs in their data, neither did they find an association with this SNP despite previous association with MTX response in three smaller cohorts that used the DAS44/DAS28 low-disease activity cut-off for response (Table 1.5). Two other studies failed to find any association between this non-synonymous variant and MTX response [188, 191]. Very few associations have been replicated in well powered cohorts, though the minor allele of a SNP in the ATP-binding cassette transporter gene *ABCB1* (rs1045642) has been associated with an increased response; the TT genotype was associated with both good EULAR and ACR20 response [192, 193]. This SNP was also found to be associated with a higher maintenance dose in a small Japanese cohort where treatment response was defined as a MTX maintenance dose at 1 year <6mg/week [194]. It is clear that a number of genetic variants have been associated with MTX response however, many of the associations have not been replicated in larger cohorts with sufficient power to detect modest effect sizes (Table 1.5).

Indeed, all of the studies published to date on MTX pharmacogenetics have taken a candidate gene approach, with the largest study containing 309 patients [195]. It is also worth noting that all of these studies are cross-sectional; one prospective study has been

published, but with very low numbers (n=48) [196]. Dervieux *et al* developed a toxicogenetic index as a predictor of the occurrence of ADRs, both gastrointestinal and neurological. A combination of *GGH* -401CC, *ATIC* rs2372536GG, *MTHFR* rs1801131 AC/CC, *MTR* 2756AA and *MTRR* 66GG genotypes was shown to be strongly associated with the percentage of period with ADRs in the 4-6 week periods assessed per patient. This use of combining associated genetic variants into pharmacogenetic and toxicogenetic indices has been tried in three other cohorts with mixed success in terms of their predictive ability [176, 190, 197]. Only one of these combinatorial approaches has incorporated non-genetic parameters, i.e. clinical characteristics [197]. Based upon the efficacy of this model (60% of responders and non-responders were classified correctly) it is highly likely that there are a number of other as yet undiscovered genetic and non-genetic factors that influence MTX response.

Owen *et al* attempted a replication of three of these combined indices [179, 190, 196], both pharmacogenetic and toxicogenetic, and failed to achieve the same results despite similar sample sizes [198]. In the original studies the partitioning into genotype combinations reduces the sample size of each group drastically, as they note in this short communication; this is cited as likely to contribute to the failure to replicate. It is also worth noting that the original publications may be examples of false positive associations as there are no other published studies that replicate these findings, to date. Small sample sizes are more likely to give rise to false positive results. Another factor may be publication bias, wherein only positive results are seen to be publishable. The fact that Owen and colleagues do not replicate the three indices that fail to take into account clinical covariates may also indicate a significant role for these factors in MTX response, and that any predictive index will require a combination of genetic markers and clinical variables.

It is important to note that the majority of studies published with regard to MTX pharmacogenetics have several important flaws which may have contributed to their lack of replication:

- Very few studies have reported *a priori* power and sample size calculations for the effect sizes they have reported. By reporting the statistical power the reader is better able to interpret any null findings.
- The largest study of MTX pharmacogenetics described here contains just 309 patients. With such small sample sizes there is low statistical power to detect modest effect sizes, e.g. OR<2.0.
- Heterogeneity between study outcomes and definitions of toxicities can hinder the ability to independently replicate previous findings.
- Differences between the make-up of the cohorts studied, such as their clinical demographics, i.e. age, disease duration, drug dose, can also contribute significantly to disparity in findings, reducing the likelihood of replication.
- Many studies are conducted in samples from a number of ethnicities. If a particular genetic marker is enriched within a population, but not accounted for, it can lead to false positive associations. Accounting for stratification is thus of great importance, especially when investigating the roles of small effect common variants.
- Taking into account multiple hypothesis testing has been poorly conducted. Whilst valid arguments can be made as to whether conservative corrections such as a Bonferroni correction are appropriate to genetic studies, the need for some form of multiple testing correction remains.

- All of the studies published to date have taken a candidate gene approach. Whilst this targeted approach is a logical starting point for pharmacogenetic studies, it raises the distinct possibility that many predictive markers are missed by this approach. GWAS have their own specific flaws and pitfalls but their strength lies in their hypothesis-free approach, potentially leading to novel understandings particularly in the context of disease-associations.

In a paper outlining issues of concern in pharmacogenetic studies of warfarin therapy, Jorgensen & Williamson [199] devised a set of criteria with which to judge the rigour and quality of pharmacogenetic studies. Whilst the paper in question deals specifically with Warfarin dosing pharmacogenetics, its content is applicable to other areas of pharmacogenetics research. Many of the points raised here feature in these issues of concern, and thus highlight a great need for more rigorous study design, application and analysis in the field of MTX pharmacogenetics.

Table 1.5. Pharmacogenetic associations with candidate genes and response to MTX in RA							
Gene	**Cytoband**	**Variant**	**Outcome measure**	**N**	**Effect Size (OR/RR [95% CI])**	**Association**	**Ref**
AMPD1	1p13.2	rs17602729 (C>T)	DAS28≤3.2 DAS44≤2.4	211 186	3.8 [1.3-10.6] (CT+TT) 2.1 [1.0-4.5] (T)	CT and TT genotype patients had higher odds of low-disease activity T allele associated with higher odds of good response	[200, 201]
MTHFR	1p36.22	rs1801131 (1298A>C)	MTX dose <6mg/week ΔDAS44>1.2	208 205	1.8 [1.1-3.0] (AA) 2.3 [1.2-4.4] (AA)	AA genotype patients more likely to require higher MTX dose after 1 year AA genotype associated with increased response at 6months post treatment	[177, 180]
		rs1801133 (677 C>T)	ΔDAS44>0.6	205	2.7 [1.0-7.3] (CC)	CC genotype associated with higher odds of greater DAS44 improvement at 6months post treatment	[180]
ATIC	2q35	rs4673993 (T>C)	DAS28≤3.2	120	3.9 [1.5-9.9] (C)	C allele associated with higher odds of low-disease activity	[187]
		rs2372536 (C>G)	DAS44≤2.4 SJC, VAS-GH	186 108	2.5 [1.3-4.7] (CC) -	CC genotype associated with higher odds of good response G allele associated with lower SJC and lower patient and physician VAS-GH	[201]
		rs3821353 (G>T)	Response vs. Inefficacy failure	248	0.5 [0.3-0.8] (T)	T allele associated with lower odds of response failure	[189]

Gene	Cytoband	Variant	Outcome measure	N	Effect Size (OR/RR [95% CI])	Association	Ref
ABCB1	7q21.12	rs1045642 (3435 C>T)	ACR20 EULAR response MTX dose <6mg/week	255 281 124	2.1 [1.3-3.4] (T) 0.3 [0.1-0.8] (CC) -	T allele associated with higher odds of ACR20 response CC genotype associated with lower risk of poor EULAR response TT genotype associated with higher MTX maintenance dose	[188, 194, 202]
GGH	8q12.3	rs12681874 (C>T)	Response vs. inefficacy failure	248	0.5 [0.3-1.0] (T)	T allele carriage associated with protection against inefficacy failure	[189]
FPGS	9q34.11	rs1544105 (G>A)	EULAR response	281	1.6 [1.0-2.4] (G)	G allele associated with poor response	[193]
MTHFD1	14q23.3	rs2236225 (G>A)	DAS28≤3.2	211	4.7 [1.3-17.3] (GG)	GG genotype associated with higher odds of low-disease activity	[200]
TYMS	18p11.32	rs2853539	EULAR response	281	0.3 [0.2-0.6] (Del)	Deletion allele associated with higher odds of poor EULAR response	[188]
ITPA	20p13	rs1127354 (C>A)	DAS44≤2.4 EULAR response	186 255	2.7 [1.1-8.1] (CC) 3.0 [1.4-6.4] (A)	CC genotype associated with higher odds of good response A allele associated with good EULAR response	[176, 201]
ADA	20q13.12	rs244076 (A>G)	EULAR response	281	1.7 [1.0-2.8] (G)	G allele associated with poor EULAR response	[188]
ADORA2A	22q11.23	rs5751876 (T>C)	EULAR response	281	1.6 [1.0-2.4] (T)	T allele associated with poor EULAR response	[188]

Gene	Cytoband	Variant	Outcome measure	N	Effect Size (OR/RR [95% CI])	Association	Ref
SLC19A1	21q22.3	rs1051266 (G>A)	ACR20	174	1.8 [1.1-2.81] (A)	A allele carriage associated with increased odds of achieving ACR20 response	[203]
		rs2274808 (C>T)	Response vs. inefficacy failure	248	1.8 [1.2-2.7] (T)	T allele associated with a higher risk of treatment inefficacy failure	[189]

Table 1.5. Pharmacogenetic associations with candidate genes and response to MTX in RA. Candidate gene association studies investigating genetic variation across a number of candidate genes involved in MTX metabolism and mechanism of action have identified SNPs associated with variable response to MTX monotherapy. Outcome measures are heterogeneous across studies but are generally categorical and based on either the DAS/DAS28 (EULAR response , DAS28≤3.2) or use components of the DAS (ACR20). DAS – disease activity score, ACR – American College of Rheumatology, EULAR – European League Against Rheumatism,

1.4.4 Sulphasalazine and combination disease-modifying anti-rheumatic drug therapy

Response and occurrence of toxicities in rheumatoid patients treated with SSA has been linked with the fast and slow acetylator phenotypes [204, 205]. The genetic basis for this difference in SSA metabolism is now established by the identification of the variants associated with the specific phenotypes in the gene *NAT2*. These genetic variants are also associated with SSA response and ADRs [129, 206-208]. Transport of drugs greatly affects their therapeutic potential, and this is no less the case with SSA treatment, where a member of the ATP-binding cassette transporters (*ABCG2*) is known to affect SSA response and toxicity [129, 209]. However, no genetic variation within these genes has been tested in the context of RA. James *et al* performed an association study on a small cohort of patients treated with a combination DMARD dose escalation regime for 1 year, and tested several candidate gene SNPs based on the shared mechanism of action of MTX and SSA [210]. SNPs in the genes *SLC19A1*, *MTR* and *TYMS* all showed evidenced of either allelic or genotype association with differences in response. They also showed a combination of alleles resulted in differences in the plateau level of the DAS28 over the 1 year studied. Whether these are due to effects on MTX, SSA or a combination of the two has not been investigated. Other genes involved in purine metabolism (*TPMT*) [211], mechanism of action (*MTHFR*)[212] and disease state (*TLR4*)[213] have also been implicated in treatment response. Laivoranta-Nyman *et al* found differences in the rate of remission between the two arms of the FIN-RACo study which compared SSA monotherapy, with or without prednisolone, and combination DMARD therapy according to the *TNF* b microsatellite, including a difference in DAS28 improvement in the combination therapy arm between patients with and without the b1 minor allele [214].

1.4.5 Tumour necrosis factor antagonist biologics

The advent of TNF-blockade therapy has allowed many patients with DMARD unresponsive disease to receive a more effective treatment. As with response to conventional DMARDs the rate of unresponsiveness to anti-TNF therapy is estimated at 30-40% [133], underpinning the need to identify predictive biomarkers so that patients are treated appropriately without unnecessary delay. The search for genetic biomarkers of anti-TNF response initially focussed on genetic variation in and around the *TNF* gene itself, including a functional promoter polymorphism (rs1800629; -308G>A) [215]. The association between anti-TNF treatment and this polymorphism has been contentious, with several cohorts observing an effect [216, 217] where other larger meta-analyses failed to find a consistent association [218, 219]. The class of anti-TNF may have influenced this inconsistency in association, as Maxwell *et al* found the association between the minor AA genotype and worse response was confined to the etanercept treated arm of their cohort [220]; this drug-specific effect was not observed in a meta-analysis by O'Reilly *et al* [221]. A more recent meta-analysis that included several larger anti-TNF treated cohorts did not find a consistent association between rs1800629 and treatment response [219].

Similar to MTX pharmacogenetics studies, the search for genetic predictors of anti-TNF therapy have largely focussed on candidate genes, including genes involved in TNF function [222-224], NLRP3 inflammasome [225], RA susceptibility [226-228] and TLR signalling [229, 230].

The first GWAS of anti-TNF response was performed on a small cohort (n=89) and found a number of associations at a significance threshold of $p<1x10^{-4}$ [231]. Many of their associations appeared to have large effect sizes (OR>4); however, the wide confidence intervals suggest these are not precise estimates. None of the loci associated with anti-TNF response have been replicated in subsequent GWASs. A recent GWAS of anti-TNF response

found a number of SNPs associated with response [232], none of which reach genome-wide significance ($p<5x10^{-7}$). The largest GWAS to date also failed to find a single SNP that reached genome-wide significance; however, based on an overlapping *cis*-eQTL, they identified a SNP associated with etanercept response in the *CD84* gene [233]. Currently there are no consistent genetic predictors of large effect of anti-TNF treatment response. A number of factors may be responsible for this:

- The influence of common genetic variation has very small effects on treatment response, studies to date have therefore been underpowered to detect these associations
- Combining all anti-TNF therapies together masks drug-specific effects; the recent association between etanercept and *CD84* is an example of a drug class-specific association
- Differences between patient cohorts in disease duration and concomitant DMARD/prednisolone use introduces unaccounted for heterogeneity into analyses
- The DAS28 is a composite score used to determine treatment response which includes subjective measures – a more objective measure of drug response may be under greater genetic control than the DAS/DAS28

1.4.6 Rituximab

The B cell depleting biologic RTX is usually given to RA patients who fail to respond to, or have a contraindication for, anti-TNF therapy. These patients therefore have long-standing RA that is refractory to therapy with both conventional DMARDs and one or more anti-TNF biologics. The ability to identify patients for whom RTX may be the most appropriate therapy early in the disease course would aid in the prevention of functional disability and progression of structural damage. A number of candidate genes have been investigated for their role in response to RTX in RA based on their role in B cell biology [234-236] and known

mechanism of action of RTX [237-239] (Table 1.6). All of these association studies have been performed in small cohorts, the largest with 212 patients [238]. The most consistent association published to date is between EULAR response and a non-synonymous SNP (rs396991) in the *FCGR3A* gene, this particular SNP has also been associated with susceptibility to RA [240, 241] and treatment response to RTX in Non-Hodgkin lymphoma [242].

Table 1.6. Summary of pharmacogenetic associations with RTX response in RA								
SNP	**Gene**	**Cytoband**	**Outcome measure**	**Genotype/allele**	**OR [95% CI]**	**p**	**n**	**ref**
rs396991	*FCGR3A*	1q23.3	EULAR response	VF vs. FF VV vs. VF	2.4 [1.0-5.4] 4.0 [1.2-13.9]	0.028 0.016	177	[243]
			EULAR response	V allele carriage	4.6 [1.5-13.6]	0.006	111	[237]
			EULAR response	V allele frequency	3.2 [1.2-11.1]	0.01	132	[239]
			EULAR response	VV vs. VF/FF	4.4 [1.4-13.5]	0.01	212	[238]
rs1800795	*IL6*	7p15.3	EULAR response	CC vs. CG/GG	4.0 [0.8-17.0]	0.049	144	[244]
			EULAR response	CC vs. CG/GG	2.8 [1.1-7.3]	0.031	158	[236]
rs9514828	*BAFF (TNFSF13B)*	13q33.3	EULAR response	C allele carriage	3.2 [1.1-9.2]	0.027	115	[235]
Haplotype*			EULAR response	TTTT	4.0 [1.0-15.1]	0.05[a]	152	[234]

Table 1.6. Summary of pharmacogenetic associations with RTX response in RA. Three loci have shown association with treatment response in at least two studies in RA. * A haplotype of SNPs in the *TNFSF13B* promoter – rs9514827 (T>C), rs3759467 (T>C), rs1041569 (T>A), rs9514828 (C>T). a – The strongest association was seen in RF seropositive patients who had previously failed anti-TNF therapy (p=0.04).

1.5 Rheumatoid arthritis treatment response and disease severity measures

1.5.1 Disease activity and treatment response definition

In order to study the pharmacogenetic influence of common genetic variation, treatment response must be defined in the most objective way possible. The success of warfarin and hepatitis C viral (HCV) infection pharmacogenetics is partly due to the objective measures of treatment monitoring and response (international normalised ratio (INR) with warfarin, and SVR for HCV treatment). In RA there is a concept of 'disease activity' that reflects the severity of the disease that is responsive to treatment. In the rheumatology clinic there does not exist one single measure of treatment response, rather several are defined based on clinical requirements, which also incorporate the concept of 'disease activity'. The ACR improvement criteria were introduced to formalise the comparison of therapeutic interventions in clinical trials [245], whilst the DAS was developed to formalise physician clinical judgement and decision making [246, 247]. Both of these measures incorporate both objective measures of inflammation and acute-phase reactants (CRP and erythrocyte sedimentation rate (ESR)), and subjective components (patient and physician global assessment questionnaires). During the development of both the DAS and the ACR improvement criteria it was noted that physician judgement was heavily weighted by the number of swollen and tender joints for a patient (swollen joint count, SJC; tender joint count, TJC). This suggests treatment response and potential changes in therapy are influenced by joint synovitis; joint swelling in particular correlates well with ultrasound detected joint effusion, but not soft tissue changes [248].

The DAS, and a modification using a restricted set of 28 joints in the SJC and TJC components (DAS28), are composite measures that are weighted based upon the original

findings of van der Heijde *et al* [246]. Two simplified non-weighted measures of disease activity have also been proposed for use in clinical practice; the clinical disease activity index (CDAI) which lacks a measure of acute phase reactants, and the simplified disease activity index (SDAI) [249].

Clinical trials in RA require validated measures of disease activity that are sensitive to changes brought about by therapeutic intervention. Many clinical trials define an end-point at which to assess the difference between treatment arms which often use a categorical outcome. The EULAR response classifications define three response categories, Good, Moderate and No/Poor, using the DAS or DAS28 that accounts for both the final score at the trial end point as well as the magnitude of improvement over the course of the trial. The magnitude of change was based on a difference twice that of the standard error of the DAS (1.2 units), whilst end-point cut-offs of high and low disease activity were defined by the lower and upper 25th percentiles [250]. These DAS and DAS28 definitions for the EULAR response criteria can be seen in Figure 1.4.

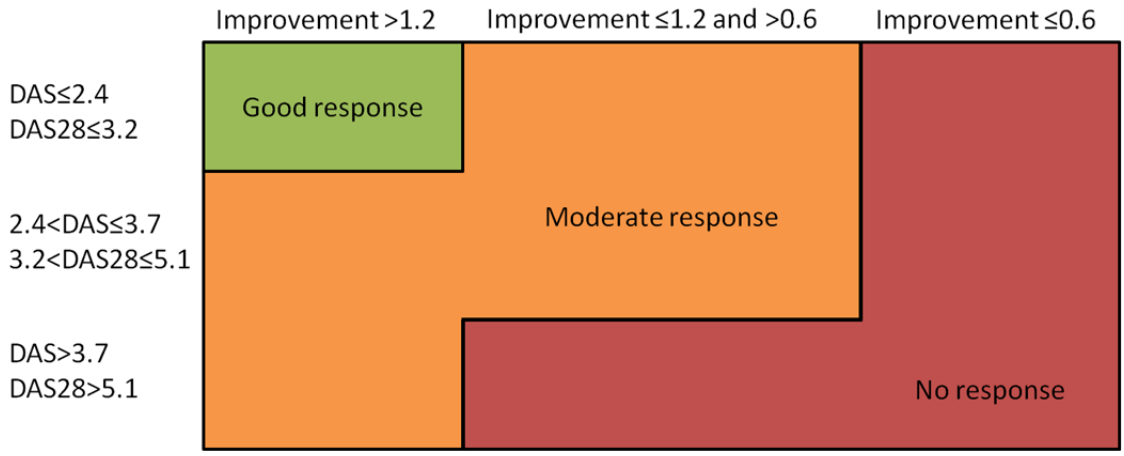


Figure 1.4. EULAR treatment response criteria for RA. The EULAR treatment response criteria are defined by the improvement in DAS/DAS28 (2x standard error) and the end-point disease activity. The DAS and DAS28 are not directly comparable so they have different cut-off points for the low, moderate and high disease activity categories which are defined by the lower and upper 25th percentiles of the distribution of the DAS or DAS28. DAS – disease activity score, DAS28 – disease activity score with 28 joints.

1.5.2 Structural joint damage and progression

The ultimate goal of RA treatment is the recapitulation of immune tolerance and correction of inappropriate inflammatory responses which lead to and drive structural joint damage progression. Therefore, the gold standard in treatment would be to prevent the onset and worsening of joint damage. Structural damage can be detected by a number of imaging modalities; historically and contemporarily x-rays are used to image both bone erosions and cartilage destruction, which is detected by narrowing of the joint space between bones. Several scoring systems for measuring joint damage severity have been developed that varyingly take into account bone erosions, joint space narrowing and peri-articular osteoporosis. The most commonly used methods for quantifying joint damage are the Larsen [251] and Sharpe methods [252] and modifications thereof [253-255]. The modified Larsen method grades the x-rays of each of 32 joints on a progressive scale of 0-5, where 0 represents no damage and 5 is mutilating damage where the original bony outlines have been destroyed; this score incorporates erosions and joint space narrowing together [251, 256]. Sharp *et al* originally developed a scoring method that assessed erosions and joint space narrowing separately for the joints of the hands and wrists [252]; the exact joints measured were defined later [253]. A total of 17 areas are scored in each hand and wrist for both the presence and severity of erosions based on the extent of surface area involvement on a scale of 0-5. Joint space narrowing is graded in 18 joint areas on a scale of 0-4, ranging from none (0) to ankylosis (4). van der Heijde *et al* proposed a modification of the Sharp method that included the feet and altered the joint scoring sites, removing sites with low inter-observer agreement [255]. In this modification erosions were scored from 0 (no erosion) to 5 (complete bone collapse), whilst the joint space narrowing also accounted for subluxation (joint malalignment). The total possible erosion score using this modified method is 280 (160 for hands and 120 for feet) and 168 for joint space narrowing (120 for hands and 48 for feet). A global measure of joint damage can be obtained by

summing the two scores across all joints to give a score that ranges from 0 to 448. This modified van der Heijde/Sharp score (SHS) does not measure soft tissue damage or peri-articular osteoporosis that was assessed in the original method of Larsen *et al* [256]. The SHS has been largely adopted in clinical trials based on its ability to detect subtle changes on an individual basis compared to the Larsen score [257].

Chapter 2 – Materials and Methods

This Materials and Methods chapter describes the techniques and methods employed herein in a general sense and are those methods which are ubiquitous throughout this doctoral work. They describe patient cohorts, statistical concepts and analytical approaches and broad laboratory protocols and techniques. Specific assays and analytical pipelines will be described in detail in the relevant chapters. As a rule of thumb, if a technique or analysis appears in more than one chapter then it is described in this chapter in a broad sense and the details are expanded upon in the relevant results chapters. All PCR primer sequences and summary information, and all buffers and reagents are detailed in Appendix I.

2.1 Patient cohorts and data

2.1.1 Yorkshire Early Arthritis Register

Yorkshire Early Arthritis Register (YEAR) patients were recruited from 14 rheumatology centres in Yorkshire, UK between 1997 and 2009 in three phases that represent changes in clinical best practise (YEAR A; 1997-2000, YEAR B; 2000-2002, YEAR C; 2002-2009); only patients recruited into YEAR B and YEAR C were included in this doctoral work [169, 258]. All patients were required to be 18 years or older at symptom onset with a maximum symptom duration at study entry of 24 months (initially 12 months), of Caucasian ancestry (self-reported) and had a consultant diagnosis of RA. Clinical and demographic data were collected at baseline and 3-monthly intervals for YEAR C and 6-monthly intervals for YEAR B intervals by either a physician or trained research nurse using a standard case report form (CRF) for a maximum of 2 years follow-up. These data included date of birth, gender, date of symptom onset, medication history, health assessment questionnaire disability index (HAQ-DI), 28 swollen (SJC28) and tender joint count (TJC28), CRP and/or ESR and patient assessment of global health on a 10cm visual analogue scale (GH-VAS). The DAS28-CRP was calculated from the TJC28, SJC28, GH-VAS and CRP according to the modified DAS28 [246,

259]. X-rays of the hands (including wrists) and feet were taken at the baseline visit, then at the 1 year and 2 year follow-up visits. X-rays were scored using the modified Sharp/van der Heijde system for erosions and joint space narrowing by a trained physician in collaboration with Professor Desireé van der Heijde [255]. Individual x-rays were scored in chronological order for each person without duplicate x-rays. All patients provided written informed consent and ethical approval was granted by the Northern and Yorkshire Multi-Centre Research Ethics Committee (MREC/99/3/48). IgM RF was measured by a standard nephelometric technique, whilst ACPA were measured in thawed serum samples using either the DIASTAT^{TM} anti-CCP ELISA platform (Axis-Shield Diagnostics Ltd, Cambs, UK) or the ELiA CCP kit on the ImmunoCAP 100 system (Phadia AB, Uppsala, Sweden); serum antibody typing was performed at either the Arthritis Research UK Epidemiology Unit at the University of Manchester or at Chapel Allerton Hospital by trained technical staff [81, 260]. Seropositivity for RF was defined as titre≥40 U/ml at any time point. ACPA positivity was defined as ≥5.5 U/ml using the DIASTAT^{TM} system, and ≥10 U/ml for the Phadia platform. CRP and ESR were measured at individual recruiting centres in the routine NHS laboratories.

2.1.2 Early Rheumatoid Arthritis Study

The Early Rheumatoid Arthritis Study (ERAS) is an inception cohort of early RA patients recruited from 9 centres around the United Kingdom between 1986 – 1998 [261]. Patients had a symptom duration of 2 years or less, a clinical diagnosis of RA and aged 18 years or more. Follow-up visits were at regular 3 month intervals up to 12months, then annually thereafter. Clinical data collected includes tender and swollen joint counts (TJC and SJC), HAQ-DI, BMI, treatment regime and duration of DMARD, steroid use, patient perception of pain on a 10cm VAS (Pain-VAS) and ESR. The DAS was calculated from the SJC, TJC, and ESR. Clinical data were kindly provided by Prof Adam Young (West Hertfordshire Hospital, Hertford, UK) on behalf of the Early RA Network (ERAN).

2.1.3 Roche clinical trials

Clinical data were provided from the methotrexate + placebo treated arms of 3 phase III clinical trials by Hoffman-La Roche; Ambition (tocilizumab) [262], IMAGE (rituximab)[263] and FILM (ocrelizumab). Patients recruited into the Ambition trial had diagnosed RA, moderate to severe disease for 3 or more months and age 18 years and above. Active disease was defined as SJC≥6, TJC≥8, and either CRP≥1 mg/dl or ESR≥28mm/h. Patients were excluded based on previous anti-TNF biologic treatment, and if treated with methotrexate in the 6 months prior to study entry or had failed to respond to methotrexate previously (including due to adverse side effects). Patients enrolled in the IMAGE trial were between 18-80 years old, had RA for at least 2 months, received outpatient treatment and were methotrexate naive. Patients were excluded if they had any other autoimmune comorbidity, surgery within 12 weeks of study entry, previous treatment with other biologic or cell-depleting therapy, and concurrent treatment with any other biologic or DMARD other than methotrexate. Patients enrolled into the FILM trial were aged 18 years or more, had diagnosed RA for 3months to 5 years and were methotrexate naive. Patients were excluded on the basis of autoimmune or inflammatory joint disease other than RA, received previous biologic treatment for RA and on concurrent DMARD therapy. The FILM trial was terminated due to safety concerns following ocrelizumab treatment. Clinical and demographic data collected on these clinical trial patients includes Body-mass index (BMI), CRP, ESR, HAQ-DI, SJC28, TJC28, ethnicity, NSAID use, smoking and GH-VAS at baseline study entry and 6 month follow-up. The DAS28-CRP was calculated from the SJC28, TJC28, GH-VAS and CRP.

2.2 Laboratory methods

2.2.1 Deoxyribonucleic acid extraction

Blood samples from patients recruited into the YEAR cohort were taken in ethylene diaminetetraacetic acid (EDTA) vacutainer tubes in the clinic or district hospital by a trained health professional. Samples were frozen at -20°C until DNA extraction. Extraction was initially performed in the regional tissue typing laboratory using a salt precipitation protocol, but was later changed to the Life Technologies Gene Catcher method (Life Technologies, Carlsbad, California, USA) in the Morgan laboratory. This extraction protocol utilised charged magnetic beads which bind the DNA and only release it upon a decrease in pH. The manufacturer's protocol was followed with several alterations. Falcon tubes containing 3ml of thawed whole blood was lysed in buffer containing guanidinium isothiocyanate, which acts as a chaotropic agent and aids in denaturing proteins, as well as a detergent agent that disrupts cell membranes. Positively charged magnetic beads are added with the lysis buffer so that DNA and other negatively charged particles are bound and retained. The non-bound supernatant was then removed using a Pasteur pipette and discarded into Virkon. The bound pellet was washed twice more with lysis buffer to try and maximise the DNA capture. The protocol only requires this step is performed once, however, it was performed twice as this generated higher 260/230nm ratios and 260/280nm ratios on the Nanodrop spectrophotometer, and thus higher quality DNA. Once the pellet had been washed, a protease digestion was carried out at 65°C for 10min using the protease and buffer solution provided in the kit. Some blood samples required longer incubation due to denser pellets, which did not disperse upon vortexing for 30s. Following the protein digest the samples were allowed to cool down to room temperature, which generally took 30min. The DNA was precipitated from solution by the addition of an equal volume of 100% alcohol. The total solution volume was first measured then an

equivalent volume of 100% isopropanol was added to each sample. At this stage a precipitate of DNA could be observed, and thus the success of the extraction could be judged. If the precipitate contained long strands this indicated DNA had been successfully extracted. However, if only broken clumps of pellet were observed the extraction was deemed to have been unsuccessful as DNA forms long strands upon precipitation with alcohol. The DNA was then washed with a standardised volume of 50% isopropanol which helps to bring any remaining protein into solution, which was then discarded. Several washing steps were carried out using a standard kit wash buffer. This removed remaining alcohol and salts prior to elution of the DNA from the magnetic beads by the addition of a lower pH solution. This protocol therefore allowed the isolation of DNA whilst removing cellular material and protein contaminants. The technique provided a high yield (>800µl) of high concentration (>50ng/µl) DNA suitable for many downstream PCR and hybridisation-based applications. It had been noted in-house that this method generated DNA of short fragments which makes it inappropriate for use as template material in long polymerase chain reactions (PCR) (>1kb). As an alternative technique that allowed the amplification of longer DNA fragments, the Qiagen QIAamp spin mini protocol (Qiagen, Düsseldorf, Germany) was used which provided an order of magnitude lower yield of DNA, but of a less fragmented nature. The Qiagen protocol uses lysed whole blood, which is incubated with a protease to digest all of the protein. Ethanol (100%) is added to precipitate the nucleic acids out of solution and applied to the QIAmp Mini spin column membrane where the DNA is adsorbed onto the column membrane in a salt concentration and pH dependent manner. Residual contaminants are removed by two wash and spin steps with sequentially lower concentrations of ethanol containing buffer. The purified nucleic acids are then removed from the membrane by incubation with a specific elution buffer and centrifuged through the column and into solution. All purified DNA from both the Life Technologies and Qiagen protocols were stored at 4°C.

2.2.2 Polymerase chain reaction and Sanger sequencing

2.2.2.1 Polymerase Chain Reaction

PCR was used to amplify target sequences for sequencing, genotyping and copy number variation assays. For each PCR-based assay primer pairs were designed to anneal within a GC% range 45-60% and a melting temperature (Tm) 50-72°C. Specific primer pairs are listed in (Appendix 1; Table 1).

PCR conditions were first optimised using an annealing temperature gradient that varied based on the predicted Tm of the primers. For instance, primers with a predicted Tm 52°C were optimised in an annealing temperature gradient 50-62°C. An exact temperature was selected based on the apparent specificity from an agarose electrophoresis gel. Further optimisation of reaction conditions was performed on an assay by assay basis and was specific to the assay in question.

All PCR reactions were run using 10X $MgCl_2$ reaction buffer (Promega, Southampton, UK) (15mM $MgCl_2$, 500mM KCl, 100mM Tris-HCl (pH 9.0)). Deoxy-nucleotide triphosphates (dNTPS; Bioline, London, UK) were included at a stock concentration of 10mM each and primers were used at a stock concentration of 10µM (Integrated DNA Technologies, Leuven, Belgium). Reactions were catalysed by a recombinant Klenow fragment thermostable Taq DNA polymerase at 1U per reaction (Promega, Southampton, UK). PCR reagent concentrations per reaction are shown in Table 2.1 for a 50µl volume.

Table 2.1. Polymerase chain reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	40.5µl
10X Mg+ buffer	1.5mM $MgCl_2$	5.0µl
dNTPs (10mM)	200µM each dNTP	1.0µl
Forward primer (10µM)	200nM	1.0µl
Reverse primer (10µM)	200nM	1.0µl
Taq polymerase	1U	0.5µl
gDNA	20ng	1.0µl
Total		50µl

Thermocycling conditions were run as follows for all PCR reactions: denaturing at 95°C 30s then 30 cycles of denaturing at 95°C 15s, annealing temperature (assay specific) 15s, elongation at 72°C 30-60s (depending on amplicon size), followed by a single round of elongation of 72°C for 5minutes. All PCR reactions were run on a Tetrad or Dyad Peltier thermocycler (BioRad, Hercules, California, USA). All PCR reactions were checked by agarose gel electrophoresis to confirm successful PCR amplification. A 1:1 mixture of PCR amplicon and 5X electrophoresis loading dye was loaded into each well of the agarose gel containing 0.0005% ethidium bromide and visualised using an ultraviolet GelDoc imager (BioRad, Hercules, California, USA). An appropriate sized DNA marker was run for each set of PCR reactions (New England Biolabs Inc., Hitchin, Hertfordshire, UK).

2.2.2.2 Charge Switch amplification product purification

Amplicons for subsequent dideoxy (Sanger) sequencing were purified using the Charge Switch PCR clean up kit (Life Technologies, Carlsbad, California, USA). This PCR amplicon purification chemistry works on the same basis as the GeneCatcher DNA extraction method (see above). The remaining PCR reaction volume (after an aliquot was removed for gel electrophoresis) was purified with an equal volume of purification buffer (N5) and 1.5µl magnetic beads. The purification buffer, magnetic beads and amplicon were pipette mixed and incubated at room temperature for 1 minute to allow the DNA and beads to bind, before being placed on a magnetic rack. The supernatant was removed with a clean

pipette tip and the pellet was washed with 150µl dH_2O and pipette mixed again. The tubes were then placed back on the magnetic rack to allow a pellet to form. This was repeated once more before elution of PCR products with 10µl elution buffer (E5) which allows the DNA to elute into the solution. This was then used as the template for further Sanger sequencing based assays.

2.2.2.3 Di-deoxy chain termination (Sanger) sequencing

Purified PCR amplicons were used as templates in Sanger sequencing reactions using the BigDye Terminator v3.1 cycle sequencing chemistry (Life Technologies, Carlsbad, California, USA). Sequencing was performed with either the forward or reverse primer from the PCR amplification, and was dictated by the presence of insertion/deletion polymorphisms and paralogous sequence variants (PSVs). The reaction components and conditions are shown in Table 2.2 in a total reaction volume of 10µl. Thermocycling was performed on either a Dyad or Tetrad Peltier thermocycler (BioRad, Hercules, California, USA). Sequencing reactions used the following thermocycling conditions: denature at 96°C for 1mins, then 25 cycles of denaturing at 96°C, primer annealing at 50°C for 5s and extension at 60°C for 2mins. All temperature changes were set to ramp at 1°C per second. Sanger sequencing reaction products were ethanol precipitated prior to capillary gel electrophoresis. EDTA at concentration of 125mM was added to each well to halt the sequencing reactions. EDTA chelates metal ions such as Mg^{2+} which is required as a co-factor by thermostable DNA polymerases. 100% ethanol was added to each well and pipette mixed before incubation at room temperature for 10mins. All samples were centrifuged at 3900rpm for 30mins to precipitate the DNA. Supernatants were discarded and the pellet was washed with 70% ethanol (v/v) then incubated for 5mins at room temperature. The supernatant was discarded and the pellet was air-dried at room temperature in the dark for a minimum of 30mins. Pellets were re-suspended in Hi-Di formamide (Life Technologies, Carlsbad, California, USA) to stabilise the single stranded DNA prior to loading on an Applied

Biosystems 3130*xl* Genetic Analyzer (Life Technologies, Carlsbad, California, USA). Base calling of sequencing products following capillary electrophoresis was performed in Sequencing Analysis 5.2 software (Life Technologies, Carlsbad, California, USA).

Table 2.2. Sanger sequencing reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	4.5µl
5X BigDye Sequencing buffer	1X	2.0µl
Sequencing primer (16nM)	1.6pmol	1.5µl
BigDye v3.1 Ready reaction mix	1X	1.0µl
Template DNA	1-50ng	1.0µl
Total		10µl

2.2.3 Genome-wide genotyping chips

Genome-wide genotyping was performed using several genotyping chips on a set of the YEAR MTX treated patient samples and the patients from the placebo arms of the Roche clinical trials. Genotyping on the YEAR MTX patients was performed by the RIKEN Centre for Genomic Medicine (Yokohama, Japan) in collaboration with the Mayo-PPII NIH-Pharmacogenetics Research Network and Pharmacogenetics of methotrexate in rheumatoid arthritis (PAMERA) consortium. These samples were genotyped on the HumanOmniExpressExome BeadChip (Illumina, San Diego, California, USA). Roche clinical trials patients were genotyped on the Illumina Human 1M-Duo (Ambition) and HumanOmni1-Quad BeadChip (IMAGE and FILM) by Hoffman La-Roche.

2.3 Statistical methods

2.3.1 Distributions of treatment outcome measures

The distribution of a given set of data can define the statistical tools used to analyse it. Broadly speaking the outcomes analysed in this thesis can be categorised as either continuous or binary. In all instances the hypothesis testing required the estimation of an effect size in order to gauge the magnitude of a genetic variant's effect on the outcome in question. To this end appropriate regression models were used based upon the distribution of the data; rather than non-parametric tests that may be largely agnostic to the distribution of the data but do not yield effect size estimates, beyond possibly correlation statistics (e.g. Spearman's ρ). Largely data were normally distributed, that is they did not deviate markedly from the expected normal distribution. The fit of a variable to the normal distribution was gauged using histograms and normal distribution probability plots (Figure 2.1 below).

Some dependent variables, such as the baseline CRP measurements shown in Figure 2.1, displayed a non-normal distribution. In these cases it is permissible to use either a non-parametric test that does not rely on the axiom that the data is normally distributed, or transform the data in such a way that it becomes normally distributed. For CRP measurements the curve of the data follow an apparently logarithmic shape, thus using a log transformation would result in data that more closely fits a normal/Gaussian distribution.

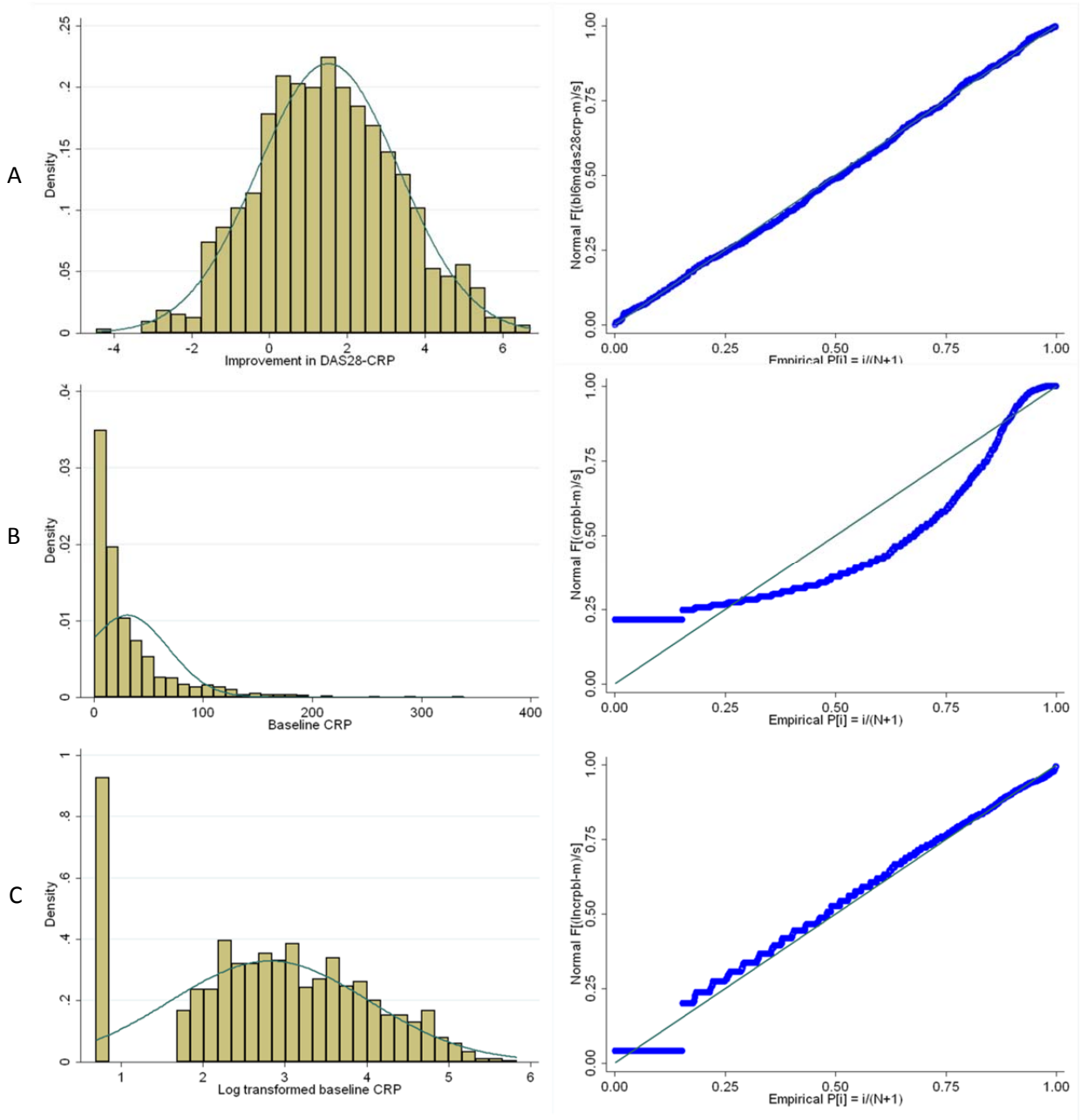


Figure 2.1 Histograms and normal distribution probability plots of clinical variables. The top panel (A) shows an example of a normally distributed variable, baseline DAS28-CRP. The green line shows the expected normal distribution. Panel B shows a non-normally distributed variable, baseline CRP, demonstrated by its deviation from the expected normal probability distribution. Logarithmic transformation of non-normally distributed data can be used to alleviate this as shown in panel C. The exception here is the presence of zero-inflation in the baseline CRP data that arises because of the way CRP is measured in NHS laboratories (i.e. values <5 are treated as 0).

Consideration must be given to how the distribution of values is generated. For instance a count variable may be treated as arising by a particular process such as a Poisson process. A Poisson process describes the probability distribution of the number of independent events occurring over a given time interval, i.e. a counting process. Thus it may be used to model data that can be thought of as count variables. A pertinent example would be the swollen joint count (SJC) used to calculate the DAS. The underlying theory behind this is that the number of swollen joints is a function of the severity and/or activity of the disease, so the more severe the disease the more joints are affected, though damage at one joint may be independent of damage at another. Alternatively, to bring in a time interval explanation, one would expect the joint count (if we consider each affected joint as a single event) to change as the disease develops or is managed by medical intervention, i.e. the patient is given treatment for the disease. The Poisson distribution is defined:

$$\Pr(Y_i = y) = \frac{e^{-\lambda_i}\lambda_i^y}{y!} \qquad \text{Equation 1.}$$

where;
$\lambda_i = \log(\mu)$
$\mu = \alpha + \beta x$
$\alpha =$ intercept
$\beta =$ parameter estimate of the explanatory variable
$x =$ explanatory variable
$Y =$ number of events
$y =$ any value ≥ 0 (dependent variable)
$i = i$th case

This distribution requires the conditional mean and variance to be equal; violation of this, where the variance is greater than the mean is called overdispersion. A generalization of this distribution can be used that allows for the incorporation of this overdispersion which can be used to model count data called the negative binomial distribution. In the example above, the SJC28 fits a negative binomial distribution better than a Poisson distribution because of the overdispersion present (Figure 2.2).

Thus the negative binomial distribution can be used to estimate the effects of a given predictor variable on the dependent variable without resorting to data transformation.

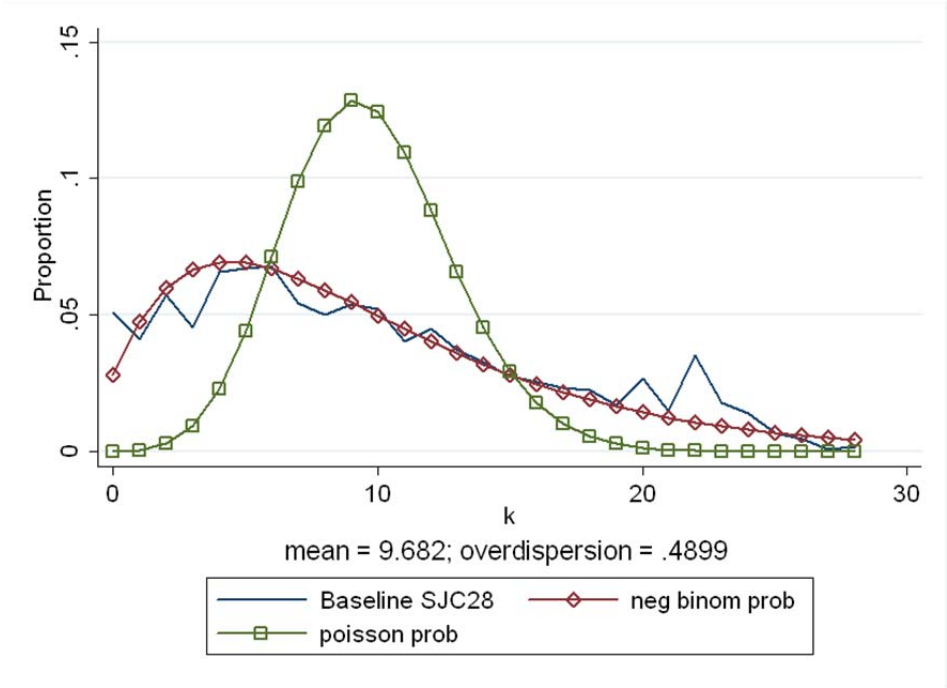


Figure 2.2 Comparison of the clinical variable SJC28 against the expected Poisson and negative binomial disitributions generated in Stata. The green connected line represents the expected Poisson distribution, which is clearly a poor fit for the observed data (blue connected line). The red connected line represents the negative binomial distributions which the SJC28 appears to approximate very closely.

2.3.2 Zero-inflated negative binomial regression

An interesting phenomenon described is the presence of excess numbers of 0's in a dataset called zero inflation. Statistical adaptations of regression models are available to account for zero-inflated data for both the Poisson and negative binomial regression models. These adaptations assume that the 0's in the data can arise by two processes; genuine 0's part of the expected distribution of the data and systematic zero's or data points that can only take a zero value for a particular reason. This phenomenon of zero-inflation and a non-normal distribution was observed in the radiographic damage variable the Sharp/van der Heijde score (SHS) (Figure 2.3).

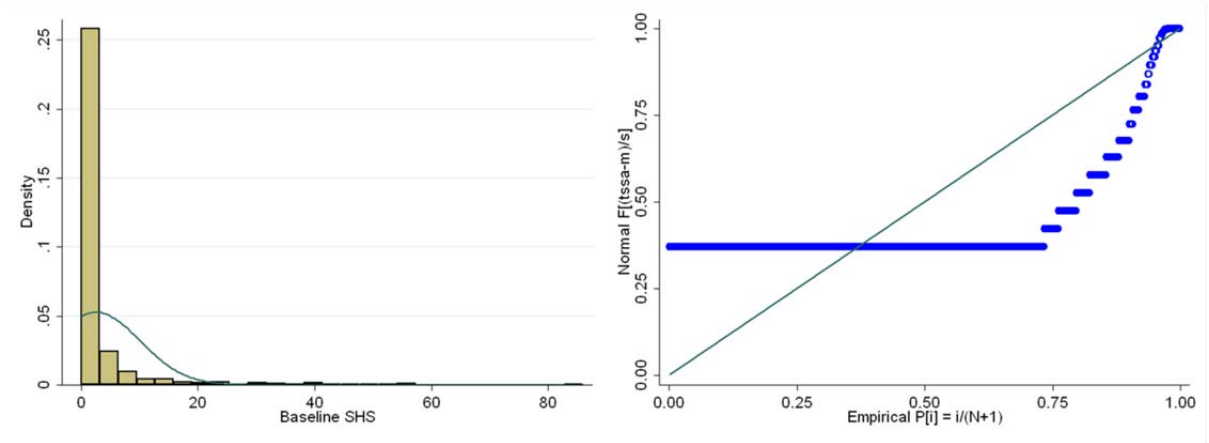


Figure 2.3 Zero-inflation is highly evident in the clinical variable Sharp score used to quantify joint damage in RA patients. The histogram in the left-hand panel shows the large proportion of 0's in the data, whilst the normal probability plot on the right shows both the non-normal distribution of the data and the effect of the zero-inflation within this variable.

The underlying theory behind the excess 0's in this data could be that certain patients cannot take a value other than 0 because of some clinical or biological factor. This will be explored in further detail in the results chapter on radiographic joint damage (Chapter 3.3.2). The joint damage variables SHS, erosions (ERN) and joint space narrowing (JSN) can be modelled using the negative binomial distribution (Figure 2.4), similar to the SJC, but with the incorporation of the zero-inflated data; this is the zero-inflated negative binomial (ZINB) regression model. A formal test of the fit of a zero-inflated model can be used to assess its appropriateness. In this instance a Vuong test of non-nested models of the negative binomial versus the zero-inflated negative binomial model gives a test statistic with a standard normal distribution, where large positive values indicate the zero-inflated model is a better fit [264].

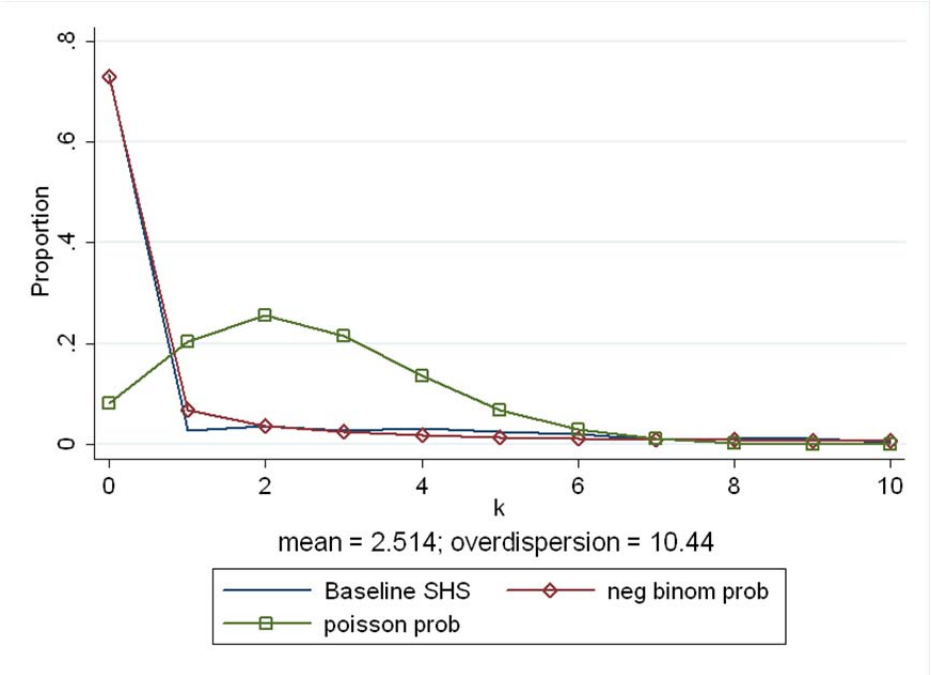


Figure 2.4 Comparison of the expected negative binomial and Poisson distributions against the SHS. The SHS very closely approximates the negative binomial distribution (red connected line), whilst the Poisson distribution (green connected line) would provide a poor fit, and thus an inappropriate regression model for predictors of joint damage.

The ZINB models the data in two parts; the count model with parameter estimates using the negative binomial distribution, and the zero-inflation model which performs a log-linked regression to estimate the log odds of the zero-inflation, where the dependent variable is the presence/absence of a zero value. Thus predictor variables can be separately entered into the count and zero-inflation equations to observe their effects.

2.3.3 Modified Poisson regression

The exponent of the estimate from the Poisson model (β) from the above equation (equation 1) is the incidence rate ratio (IRR), when applied to a binary variable this can be interpreted simply as the risk ratio (RR). In order to accurately estimate the RR for a binary outcome applying a Poisson regression it is necessary to apply a sandwich estimator of the standard error estimator, otherwise known as a robust error variance [265]. This allows the robust estimation of risk ratios in a cohort-based study, as opposed to the use of

logistic regression to estimate the log odds ratio which can overestimate the effect of a predictor where the outcome is common.

2.3.4 Linear regression and treatment response models

As discussed in Chapter 1.5.1 the definition of treatment response requires a strict prior definition based on the manner in which it is measured. The DAS28 is commonly used as a metric of response because it is familiar to most clinicians and has a simple algorithm for calculation [246]. The distribution of the DAS28 is approximately normal whether it is calculated using the CRP or ESR measurement, and whether it includes 4 variables or 3 (Figure 2.5). The magnitude of response can be calculated from individual time points, for instance in the first six months of treatment the response could be measured between the baseline visit and this time point by subtracting the value at 6months from the baseline (ΔDAS28-CRP). This gives a directional measure of response where positive values indicate an improvement in the patients DAS28, and a negative value indicates a worsening (Figure 2.6). The ΔDAS28-CRP has an approximately normal distribution which makes it amenable as a dependent variable in a linear regression analysis. The other requirements of a linear regression are that the residuals from the regression model are linear and have a constant variance, and are normally distributed. These assumptions can be checked by visualising the regression model output graphically. For instance a normal probability plot can detect deviation from expected values based on the normal distribution, whilst the linearity and constant variance can be graphed on a scatter plot of the residuals against the fitted values (Figure 2.7).

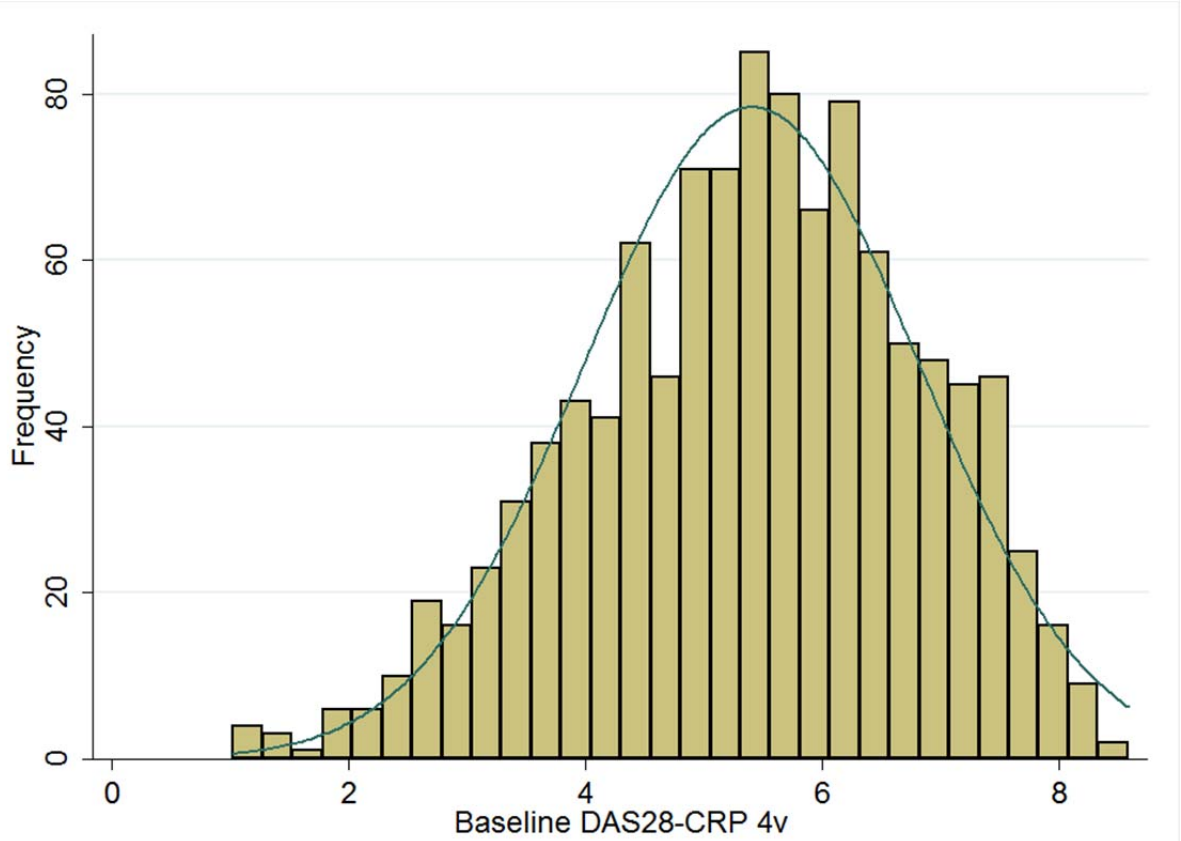


Figure 2.5 – Histogram of baseline DAS28-CRP calculated using 4 variables (SJC28, TJC28, CRP and VAS-Global Health). The DAS28-CRP is distributed in an approximately normal manner shown by the turquoise curve. Data are from the YEAR cohort with symptom duration of 24months or less.

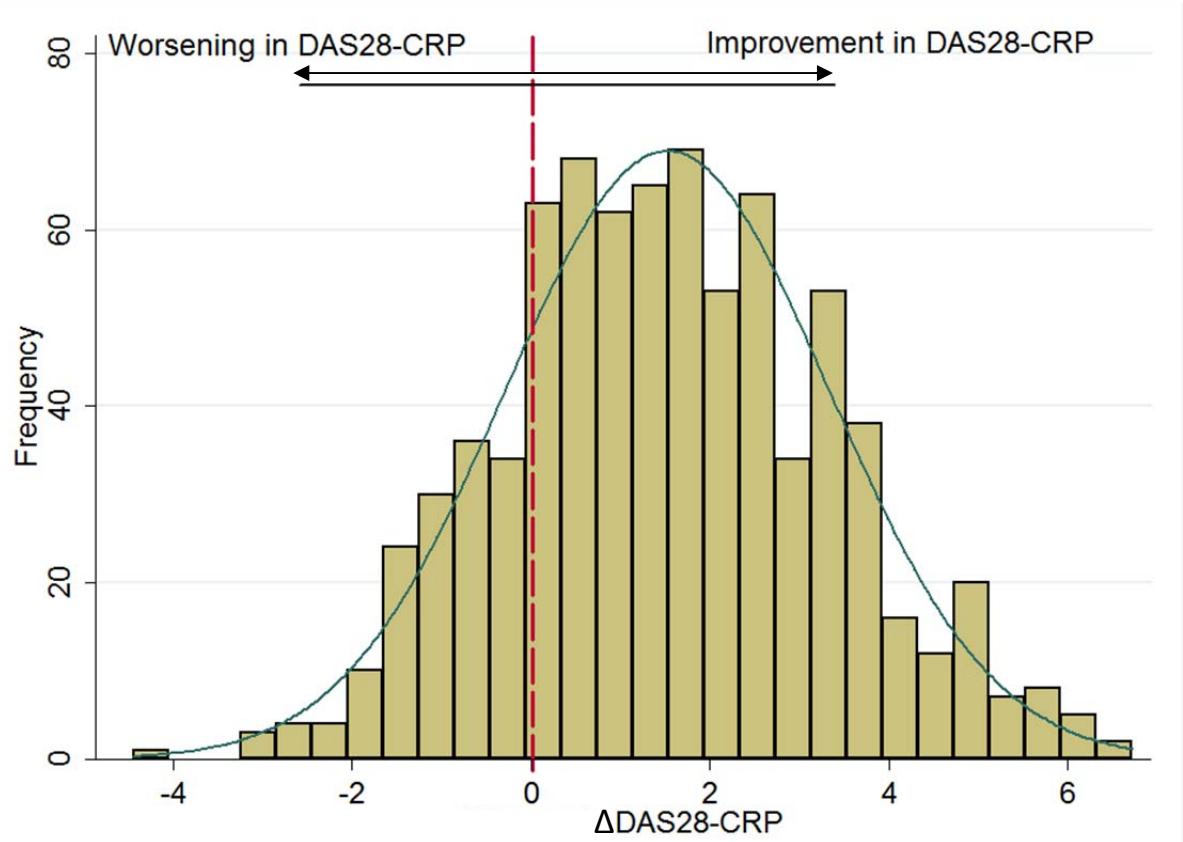


Figure 2.6 – Histogram of ΔDAS28-CRP from baseline to 6months in the YEAR cohort. The direction of the ΔDAS28-CRP indicates an improvement or worsening in disease activity for any given patient. The majority of patients (~80%) show an improvement in their DAS28 over this time course. The ΔDAS28-CRP is approximately normally distributed as shown by the turquoise curve. The red dashed reference line represents the 0 value of no improvement or worsening.

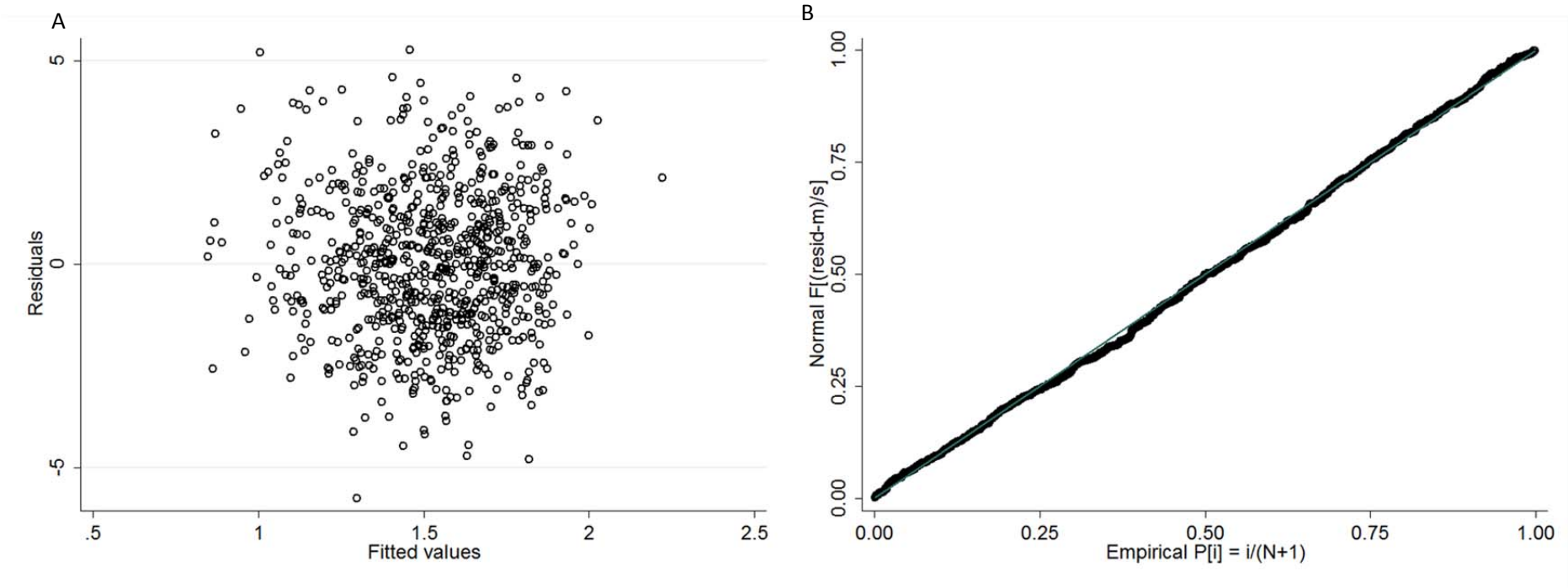


Figure 2.7 – Graphical visualisation of linear regression residuals to check for deviation from model assumptions. A – A scatter plot of the residuals and fitted values from an example linear regression model with the ΔDAS28-CRP as the dependent variable and patient baseline age as the predictor variable. No obvious patterns are observed in the data indicative of non-constant variance. B – A normal probability plot of the linear regression model residuals. There is very little deviation from the expected distribution for this model implying it is appropriate for this context.

Measuring the treatment response of a patient using such a measure at a single time point can fall foul of several confounders; 1) the patient may be experiencing a flare in their symptoms on a particular day, which may underestimate the magnitude of their response, 2) non-disease related factors can affect their response on the VAS-Global health questionnaire of the DAS28 thereby introducing subjective error into the measure, 3) weighting of individual components can affect the total DAS28 score differently. A patient may have few large inflamed joints which can lead to a highly elevated CRP/ESR whilst a patient with a high number of smaller affected joints may have a lower CRP/ESR. In order to deal with these possible confounders several approaches can be taken. Data can be measured on serial time points to draw a curve of disease activity. The area under the curve (AUC) would then reflect a smoothed average of the disease activity over time which would account for 1). The DAS28 can be calculated with 3 variables that does not include the VAS-Global Health, removing some confounding subjective error, this would partially account for 2). The weighting of the individual components is based upon their relative contribution to the original factor analysis in which the DAS28 was calculated, thus this confounder cannot be easily accounted for. One possible solution is to use multiple response measures which individually reflect different aspects of disease activity in RA, such as the separate components of the DAS28.

2.3.5 Genetic association models

The effects of biallelic SNPs were tested in regression models using an additive genetic model. This assumes an allele-dose effect of the minor allele on the dependent variable on a linear scale. This is denoted as the log additive model on a log scale. Alternative genetic models, such as recessive or dominant models, may fit the data better in some circumstances, and specifying an inappropriate genetic model results in a loss of statistical power. However, testing each SNP under each genetic model increases the multiple testing burden and thus the probability of type I errors. An additive model was selected as there is

little impact on statistical power if the true model is dominant or co-dominant. Though it lacks power to detect recessive effects with MAF<0.5 [167]. Genetic association testing was performed using Stata IC v11 (Stata Corp, College Station, Texas, USA) and PLINK [266].

2.3.6 Genetic analysis programs – Haploview and LocusZoom

GWAS visually display results using a Manhattan plot; a plot of $-\log_{10}$ p values (y axis) against SNP position, ordered by chromosome. A standard Manhattan plot can therefore display data from millions of markers across the human genome in an easy to interpret manner. However, it does not necessarily allow for the fine-scale interrogation of association signals on its own. It is possible to generate high-resolution Manhattan plots in a number of statistical packages, but it does not immediately allow for the integration of other data of interest, such as locus-wide recombination hotspots and linkage disequilibrium between SNP markers. A commonly used server to generate plots directly from genetic analysis software is LocusZoom. This online application utilizes the R statistical package to generate and integrate data from several public sources, including the HapMap and 1000 Genomes project [267]. LocusZoom was used to visualise data from all statistical association analyses that utilised large scale genotyping data, i.e. those derived from whole-genome genotyping chips.

Genotype data for association studies can also be used to generate LD plots and perform statistical association studies. Haploview 4.2 is capable of generating LD plots for a number of LD measures including r^2 and D' [268] and was used in this thesis for this purpose.

2.3.7 Genotype imputation

Genotype imputation was performed on pre-phased haplotypes using a set of reference haplotypes (1000 Genomes March 2012) in IMPUTE2 [269]. Imputation was performed on the Roche clinical trials, YEAR and healthy control primary cell gene expression data donated by Dr Julian Knight. All pre-phasing and imputation was carried out by Mr John

Taylor (Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds).

Chapter 3 – Rheumatoid arthritis susceptibility loci, treatment response and disease severity

3.1 Introduction

The principal aim of this doctoral work was to identify genetic biomarkers of response to drugs used to treat RA. Treatment response is multi-faceted as it can be affected by both the action of the drug and disposition within the body (pharmacokinetics and pharmacodynamics) and by the disease pathology (e.g. severity). GWAS and fine-mapping studies have identified around 60 variants across 54 loci that are associated with RA susceptibility as of 2013. Prior to the commencement of this doctoral work there were around 40 non-MHC SNPs associated with RA susceptibility. However, the effect of genetic variants at these loci on disease pathology and treatment response had not been reported, with limited reports of associations with disease severity in the form of radiographic damage progression [270]. This chapter describes the methodology and results for analyses of these pre-2013 RA susceptibility loci and *HLA-DRB1* SE alleles in relation to drug response, radiographic joint damage and measures of disease activity and severity. In order to aid the interpretation of these results in the context of joint damage and RA susceptibility loci a focussed *cis*-eQTL analysis was carried out utilising publicly available, and previously published, gene expression data from lymphoblastoid cell lines and primary B cells and monocytes. ENCODE annotations of SNPs were used to identify putative functional SNPs responsible for the disease and joint damage associations at these loci.

3.2 Methods

3.2.1 Rheumatoid arthritis susceptibility single nucleotide polymorphism genotyping

The RA susceptibility SNPs in this thesis were genotyped as part of a series of United Kingdom Rheumatoid Arthritis Genetics (UKRAG) Consortium studies that aimed to replicate putative RA susceptibility variants identified from a series of GWAS [81, 84, 98-

100, 102, 107, 271]. The genotyping was carried out as part of collaboration between Prof Ann Morgan and Prof Jane Worthington at the University of Manchester using a Sequenom MASSArray iPlex platform (Sequenom, Hamburg, Germany). This technology is able to multiplex more than 30 SNP PCR assays in a single reaction and relies on the principles of mass spectrometry to derive genotypes from PCR products. Genotyping was performed on the YEAR and ERAS cohort patient samples described in Chapters 2.1.1 and 2.1.2.

3.2.2 Rheumatoid arthritis susceptibility loci single nucleotide polymorphisms

RA susceptibility loci SNPs were selected for genotyping in the UKRAG consortium based initially on the top two arbitrary tiers of associated SNPs (defined by their p-values) from the WTCCC GWAS study [272], and an international meta-analysis of RA susceptibility GWAS in 2010 [91]. The RA susceptibility SNPs included in this doctoral work are listed in Table 3.1; details of the relevant association studies are introduced in Chapter 1.2.3. Several loci have multiple SNPs genotyped. This is due to the original UKRAG Consortium SNP selection criteria which were based on the smallest reported p-value or top hits from a study. These SNPs were often originally reported across different genotyping platforms and geographical populations. Where multiple SNPs were in strong LD ($r^2 \geq 0.8$) the SNP with the highest genotype call rate was selected. If they were equal then one was arbitrarily used in all subsequent analyses.

Genotyping quality control (QC) was carried out on these data on the YEAR cohort using a sample-wise cut-off of 90% and a SNP-wise call rate of 90% (a standard practise in UKRAGG studies). Samples or SNPs not meeting these thresholds were excluded from all further analyses. SNPs were also checked for their concordance with expected genotype proportions based on Hardy-Weinberg equilibrium (HWE) assumptions [273]. This was formally tested by a χ^2 test with 2 degrees of freedom.

Table 3.1. Rheumatoid arthritis susceptibility loci genotyped SNPs				
SNP	**Genetic locus**	**Cytoband**	**Minor allele**	**MAF**
rs11586238	*CD2/CD58/IGSF2*	1p13.1	C	0.246
rs4272626	*CASQ2*	1p13.1	C	0.355
rs2476601	*PTPN22*	1p13.2	G	0.141
rs10910099	*MMEL1/TNFRSF14*	1p36.2	A	0.309
rs12746613	*FCGR2A*	1q23.2	C	0.121
rs10919563	*PTPRC*	1q31.3	G	0.135
rs1160542	*AFF3*	2q11.2	A	0.481
rs10181656	*STAT4*	2q32.3	C	0.246
rs3087243	*CTLA4*	2q33.2	G	0.437
rs231775			A	0.397
rs1980422	*CD28*	2q33.2	T	0.236
rs333	*CCR5*	3p21.31	Del	0.121
rs4535211	*PLCL2*	3p24.3	G	0.489
rs231707	*FAM193A*	4p16.3	G	0.193
rs6822844	*IL2/IL21*	4q27	G	0.167
rs6897932	*IL7R*	5p13.2	C	0.278
rs1043730	*TRAF3IP2*	6q21	A	0.240
rs548234	*PRDM1*	6q21	T	0.345
rs13207033	*6q23*	6q23.3	G	0.258
rs6920220	*6q23*	6q23.3	G	0.243
rs5029937	*TNFAIP3*	6q23.3	G	0.056
rs394581	*TAGAP*	6q25.3	T	0.294
rs42041	*CDK6*	7q21.2	C	0.257
rs13277113	*BLK*	8p23.1	G	0.254
rs2812378	*CCL21*	9p13.3	T	0.362
rs10760130	*TRAF1/C5*	9q33.1	A	0.464
rs2900180			C	0.372
rs2104286	*IL2RA*	10p15.1	A	0.275
rs41295061			C	0.09
rs11594656			T	0.208
rs4750316	*DKFZp667F0711*	10p15.1	G	0.197
rs540386	*TRAF6*	11p12	C	0.137
rs1678542	*KIF5A*	12q13.3	C	0.399
rs775241	*DCTN2*	12q13.3	A	0.122
rs7234029	*PTPN2*	18p11.21	A	0.178
rs763361	*CD226*	18q22.2	C	0.493
rs4239702	*CD40*	20q13.12	C	0.258
rs3788013	*UBASH3A*	21q22.3	C	0.447
rs743777	*IL2RB*	22q12.3	A	0.327
rs3218258			C	0.455
rs229541	*C1QTNF6*		G	0.275

Table 3.1. RA susceptibility loci genotyped SNPs. SNPs were selected as described in the text. The gene symbol represents the closest candidate gene to the SNP, or where no obvious candidate exists, the nearest UCSC gene, based on the human genome reference build Hg19/GRCh37. Loci where multiple SNPs are listed are either due to moderate to low LD (r^2<0.8) or where multiple independent associations with RA susceptibility are reported (e.g. 6q23). MAF – minor allele frequency

3.2.3 Human Leukocyte Antigen genotyping

HLA-DRB1 genotyping was carried out by Mr Stephen Martin and Mrs Lubna Haroon-Rashid. YEAR patients were genotyped by a combination of PCR sequence-specific oligonucleotide probe linear array technology typing and high resolution typing by PCR and Sanger sequencing. A multiplex PCR reaction containing biotinylated primers of unknown sequence (provided by Roche Molecular Systems, Pleasanton, CA, USA) was performed to amplify regions in *HLA-DRB1*. PCR products were hybridised in buffer (4X NaP solution with $NaCl_2$ and EDTA (SSPE), 0.5% SDS) to linear strips and washed with specific wash buffer (1X SSPE, 0.5% SDS). Streptavidin horseradish peroxidase (SA-HRP) was conjugated to the biotinylated PCR products, washed (1X SSPE, 0.5% SDS), incubated with citrate for 5mins and allowed to develop. The strips were then washed twice in dH_20 before being scanned using a flatbed scanner and analysed with StripScan software 2.8. The banding pattern on each strip was indicative of *HLA-DRB1* allele which was used to derive 2-digit resolution typing.

Where 2-digit *HLA-DRB1* typing was available from the Roche strips a single initial PCR and sequencing reaction was performed using a common reverse primer (DRB95008b) and a forward primer specific to the 2-digit allele (e.g. DRB95001 for DRB1*01, etc.). Where no 2-digit *HLA-DRB1* data were available a PCR reaction was run with each of the allele-specific forward primers and the common reverse primer (8 reactions in total). Allele typing is confounded by sequence variation at codon 86 which is not part of the *HLA-DRB1* allele definition. Where heterozygosity was evident at two neighbouring nucleotides of this codon two further PCR reactions were performed using the relevant allele-specific forward primer and two reverse primers specific for the two codon 86 variants (primers DRB96001GT and DRB96002TG). All *HLA-DRB1* PCR were run with the reaction mix in Table 3.2 and the following thermocycling conditions: denaturing at 94°C for 2mins followed by 35 cycles of denaturing at 94°C for 30secs, primer annealing at 62°C for 30secs, and

extension at 72°C for 45secs; a final extension at 72°C was run for 5mins. Sanger sequencing reactions were performed using the allele-specific forward primer using the reaction conditions in Chapter 2.2.2.3. Allele typing from sequence electropherograms was performed by alignment to allele-specific reference sequences downloaded from the international Immunogenetics (IMGT) project website (http://www.ebi.ac.uk/imgt/hla/).

Table 3.2. *HLA-DRB1* PCR conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	15.6µl
10X Mg+ buffer	1.5mM $MgCl_2$	2.0µl
dNTPs (10mM)	200µM each dNTP	0.4µl
Allele specific forward primer (10µM)	200nM	0.4µl
Common reverse primer (10µM)	200nM	0.4µl
Taq polymerase	1U	0.2µl
gDNA	50ng	1.0µl
Total		20µl

3.2.4 Statistical analysis

3.2.4.1 Treatment response analysis

The definition of treatment response was laid out in Chapters 1 and 2. The analysis of the RA susceptibility loci SNPs used the ΔDAS28-CRP as the primary outcome measure with sub-analyses for the improvement in the natural log (ln) transformed CRP and the improvement in the SJC28, adjusted for their respective baseline measure. All patients recruited into YEAR B and C, regardless of treatment, were included in this analysis to test the hypothesis that RA susceptibility SNPs affect treatment response through disease-related mechanisms, as opposed to drug-specific ones. In order to test and replicate any drug-specific effects of these SNPs, patients were split into whether they were treated with MTX (n=551, genotyped n=298) or SSA (n=534, genotyped n=188) monotherapy (without glucocorticoids data) from baseline; some patients did not receive either DMARD and so

were not included in further analyses (n=92). These included patients receiving hydroxychloroquine, lefluonomide or combination therapy from baseline. Each SNP was tested for association with the 6month ΔDAS28-CRP, ΔlnCRP and ΔSJC28 in the two DMARD groups. This analysis was then repeated in the Roche clinical trials data set for the same SNPs and outcome measures as a second cohort for MTX-specific effects. A comparable analysis was carried out using patients recruited into ERAS and treated with SSA from baseline. The ERAS cohort does not have SJC28 and CRP data collected rather the swollen joint count of 44 joints (SJC44) and erythrocyte sedimentation rate (ESR) were recorded instead. The results from the YEAR and ERAS SSA-treated patients were not directly comparable due to the differences in outcome measure collected. Ranked percentiles of each variable were therefore generated and used in the association analysis. The effect estimates and standard errors from the linear regression models of the SSA and MTX analyses were used in a random effects meta-analysis with inverse variance weighting (Figure 3.1). Heterogeneity within each meta-analysis was quantified using the I^2 statistic [274].

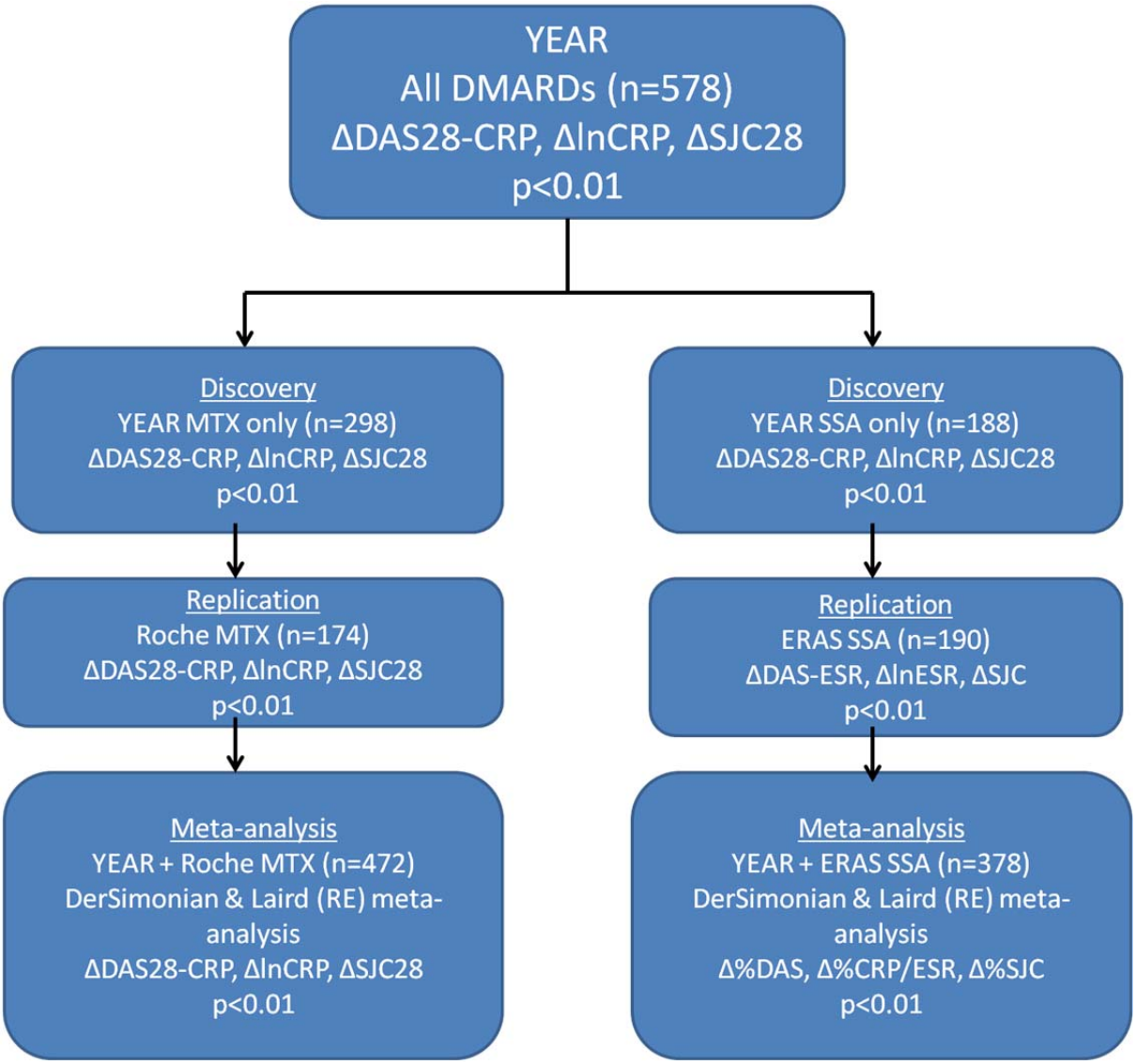


Figure 3.1. Work pipeline for RA susceptibility SNP association analysis. The initial association analysis across all DMARD treated patients in YEAR was carried out in 578 individuals using the three outcome measures ΔDAS28-CRP, ΔlnCRP and ΔSJC28 as described in the text. A sub-analysis of the specific DMARDs MTX and SSA was then carried out with the YEAR data acting as a discovery cohort and the Roche (MTX) and ERAS (SSA) acting as the replication cohorts. The results from the two analyses were then combined into two separate meta-analyses to give a set of SNP associations with DMARD response in RA. Δ%DAS, Δ%CRP/ESR and Δ%SJC represent the percentile improvement in outcome measure. p-values represent statistical significance thresholds.

3.2.4.2 Joint damage statistical analysis

The extent of joint damage of each patient recruited into the YEAR cohort was quantified using the SHS from x-rays taken of the feet and hands and wrists. Scoring was carried out

by Prof Desiree van der Heijde (Leiden Medical Centre, Netherlands) for patient x-rays at baseline, 1 year and 2 years where they were available. Scoring was carried out for ERN and JSN separately for the hands/wrists and feet. The maximum score for erosions from both the hands and feet is 280, whilst for joint space narrowing it is 168. The SHS combines the joint space narrowing and erosion scores for each joint and x-ray to give a total that ranges from 0-448. As noted in Chapter 2.3.2 the baseline SHS (and subsequently the ERN and JSN) was not normally distributed and more closely fits a negative binomial distribution, with considerable zero-inflation (SHS = 71.3%, ERN = 84.9%, JSN = 77.2%). Therefore a zero-inflated negative binomial (ZINB) regression model was used to model the effect of RA susceptibility SNPs on baseline joint damage in the YEAR cohort. The ZINB model consists of two component equations, for the count and the zero inflation (see Chapter 2.3.2 for details). In order to test the overall effect of each SNP on the outcome (i.e. SHS, ERN or JSN) a *post hoc* Wald-type test was performed. Clinical and demographic variables were entered into both components of the model and tested for an effect (2-sided Wald-type test). Clinical and demographic variables considered to be statistically significantly associated with joint damage (univariate $p<0.05$) were included in the model prior to the genetic association analysis. Each SNP was therefore tested in a multivariate model adjusted for the associated clinical and/or demographic variables found to be associated with the SHS. The sub-analyses for the ERN and JSN were carried out using the same model but with the appropriate dependent variable in each case, i.e. the other predictor variables were those associated with the baseline SHS. This was to maintain comparability and for ease of interpretation of results. For instance if a SNP is associated with the SHS, adjusted for age and symptom duration, a separate model that tests the effect of the SNP on the ERN with an additional adjustment for gender would not be directly comparable. Therefore any effect could not be wholly attributed to the SNP as the addition of different covariates may alter the final effect estimate and model statistics.

The effect of RA susceptibility SNPs on joint damage progression was tested using a binary variable for the presence of increased SHS, ERN and/or JSN after the baseline measurement. A modified Poisson regression model with a robust variance estimator was used to estimate the risk ratio of progression for each SNP using a log-additive genetic model [265]. The two year follow-up measurement of joint damage was used in this analysis as it provided the maximum possible time for new damage to occur, and previous damage to worsen, given the data collection for the YEAR cohort.

An additional SNP was included in the joint damage analysis at the *C5orf30* locus (rs26232) in a subset of individuals as part of collaboration with Prof Gerry Wilson and Dr Dawn Teare (Sheffield University) [275].

3.2.4.3 *HLA-DRB1* shared epitope analysis of drug response and joint damage

HLA-DRB1 alleles were assigned to either SE positive or negative groups and each YEAR patient was assigned a score based on the number of shared epitope alleles they carried (i.e. 0, 1 or 2). This additive coding for SE genotypes was entered into the treatment response and joint damage analysis regression models described above and in Chapter 2.3.

The *HLA-DRB1* alleles have been classified according to a stratification of disease susceptibility risk [64]. Patients were assigned to carriage of the number of alleles for each risk category: high risk, medium risk, neutral or protective. For instance an individual with two high risk alleles was coded '2' for this category and '0' for all others, whilst an individual with one high risk and one medium risk allele was coded '1' for each high and medium risk variables and '0' for all others. The *HLA-DRB1* alleles and risk category are shown in Table 3.3. The effect of these stratified risk groups was tested in regression models for DMARD response and both baseline joint damage and risk of joint damage progression, relative to the neutral risk category by including all risk category variables in

the same model except for the neutral risk group; this was to minimise the multiple testing burden. Each risk group was coded assuming an additive effect.

Table 3.3. *HLA-DRB1* stratified risk categories according to Mackie *et al* [64]		
Risk group	*HLA-DRB1* alleles	Frequency
High risk	*0401, *0404	0.30
Medium risk	*0101, *1001	0.16
Neutral	*0103, *0402, *07, *08, *11, *12, *13, *16	0.30
Protective	*0112, *03, *0448, *09, *14, *15	0.24

3.2.4.4 Correlation analysis and genetic association of disease severity measures

An exploratory analysis using pair-wise correlation between disease severity measures at baseline and 1 year was performed using Spearman's correlation coefficient (ρ). A threshold for statistically significant correlation was set at $p<0.05$ ($|\rho|>0.1$). Genetic association analysis of correlated measures was carried out for each outcome using a regression model appropriate to the distribution of that measure (e.g. negative binomial regression for baseline SJC28). Each regression model included the predictor SNP, assuming a (log) additive effect, adjusted for other correlated severity measures. Only the SNPs nominally associated with baseline joint damage were included in the regression models of baseline severity measures.

The SNPs included in the baseline analysis were also included in the analysis of 1 year severity measures, alongside SNPs nominally associated ($p<0.05$) with improvement in DAS28, SJC28 and/or CRP, and risk of joint damage progression. Effect estimates of each regression model were reported with 95% CI's and p-values. In the case of the ZINB models the p-values were calculated using a *post hoc* Wald-type test.

3.2.5 Expression quantitative trait loci association

Existing publicly available eQTL analysis results were accessed using the GeneVar viewer from the Wellcome Trust Sanger Institute website (http://www.sanger.ac.uk/resources/software/genevar/). *cis*-eQTLs were identified using the MuTHER data set (Grundberg 2012 release) [276]. Loci were selected on the basis of

nominally associated SNPs with radiographic joint damage or response to DMARD therapy (either combined, MTX or SSA monotherapy). Each trait-associated SNP was entered as a search term for association with expression levels of any genes within 0.5Mb ($p \leq 1x10^{-5}$). These gene symbols were then entered into GeneVar to refine the position of the *cis*-eQTL itself. Regression analysis results were downloaded directly from the GeneVar viewer and visualised with high-resolution Manhattan plots for each probe and tissue type. Putative *cis*-eQTLs (test $p \leq 1x10^{-5}$) were further investigated in a gene expression dataset derived from isolated primary monocyte and B cells which were kindly provided by Dr Julian Knight (Wellcome Trust Centre for Human Genetics, Oxford, UK). Full experimental and methodological details can be found in ref [277]. Processed expression data were downloaded from http://www.ebi.ac.uk/arrayexpress/, data set E-MTAB-945 (accessed April 2012). Genotype data were provided by file transfer protocol (FTP) via the European Genome-Phenome archive; https://www.ebi.ac.uk/ega/ (EGAS00000000109). Selected genomic regions corresponding to candidate genes of interest (GOI) were imputed within 0.5Mb of the transcriptional start site (TSS) of the GOI using IMPUTE2 (1000 Genomes March 2012 release) in the primary monocyte and B cell data by Mr John Taylor. Expression probes were mapped to the human genome using the BLAST-like alignment tool (BLAT) on the UCSC Genome Browser (Hg19). Probes predicted to align to more than one position were excluded. The Genome Browser was visually scrutinised for overlapping common SNPs with each expression probe (dbSNP 135 MAF>1%). Single *cis* SNP-probe association was carried out using linear regression for each probe. Conditional linear regression was used to investigate independent effects on gene expression using each lead associated SNP as a covariate in the regression model. Where evidence of multiple non-independent signals was present, assessed by linkage disequilibrium between SNP markers, conditional haplotype analysis was performed using the *--chap* command with specific SNPs. The default test for this command is an *H-1* degrees of freedom F-test of association

under the null (no haplotypes have different effects) and alternate (a difference in effect between haplotypes) hypothesis, where H is the number of phased haplotypes, one of which is chosen arbitrarily as the reference haplotype. The effect of individual haplotypes on probe expression at particular loci relative to the others formed by the specified SNPs was tested by using the *–specific-haplotype* sub-command. This sets the non-specified haplotypes as the reference. All eQTL analyses were carried out using PLINK [266]. High-resolution Manhattan plots used to visualize associations were generated using LocusZoom [267].

3.2.6 RegulomeDB and Encyclopaedia Of DNA Elements data

The Encyclopaedia Of DNA Elements (ENCODE) project was launched in 2003 to identify and characterise all of the functional DNA elements in the human genome following the completion of the Human Genome Project (HGP). The pilot phase was completed in 2007 and the first major data release in the second half of 2012, resulting in the publication of over 30 articles on an unprecedented scale [278, 279]. All references to ENCODE data therefore refer to the 2012 data freeze. The ENCODE data has been integrated into the UCSC Genome browser since the public release of the pilot phase data, and this allows the visualisation and data mining of ENCODE data to varying levels of detail across all areas of the human genome covered by the project [280]. The ENCODE project utilised a number of experimental and laboratory techniques to functionally characterise DNA elements, primarily techniques aimed at identifying and annotating regulatory elements, such as chromatin immunoprecipitation (chIP), chIP with high throughput next generation sequencing (chIP-seq), histone and CpG methylation marks, and DNase hypersensitivity assays. These techniques can be used to identify transcription factor bindings sites (TFBS) and regions of open chromatin that allow transcription factors and other components of the transcription machinery to directly access DNA. The ENCODE project also facilitated the development of computational and bioinformatics tools to interpret these data,

particularly within the context of how common genetic variation can lead to differences in transcriptional regulation. These data are likely to aid in the quest to interpret signals from genome-wide association studies which have overwhelmingly been located outside of protein coding regions of the genome [281].

RegulomeDB was developed as part of the ENCODE project to annotate SNPs that overlap with regions likely to affect transcription by integrating data from multiple experiments [282]. This server requires dbSNP rs IDs, Hg19 coordinates or files from next generation sequencing (NGS) variant calling as an input and displays a list of retrieved SNPs with a score from 1 to 6 based on the weight of evidence, where 1 = more evidence that the SNP affects transcription and 6 = minimal evidence [282]. Schaub *et al* published a slight amendment to this scale which collapsed down the scores onto a scale of 2 to 6 (all 1's were counted as 2's) in order to integrate additional sources of evidence from eQTL analyses [281]. The scores are sub-classified based on the experimental evidence for each (Table 3.4). The scores themselves are heavily weighted towards those with published eQTL associations, thus the Schaub *et al* amendment was used in this doctoral work to avoid overweighting SNPs with eQTLs in non-disease relevant tissue (i.e. neuronal specific eQTLs).

Table 3.4. RegulomeDB score and evidence			
Score	Evidence	Schaub *et al* score	Evidence
1a	eQTL, TF binding, matched TF motif, matched DNase footprint and DNase peak	2a	TF binding, matched TF motif, matched DNase footprint and DNase peak
1b	eQTL, TF binding, any motif, DNase footprint and DNase peak	2b	TF binding, any motif, DNase footprint and DNase peak
1c	eQTL, TF binding, matched motif, DNase peak	2c	TF binding, matched TF motif and DNase peak
1d	eQTL, TF binding, any motif and DNase peak	3a	TF binding, any motif and DNase peak
1e	eQTL, TF binding and matched TF motif	3b	TF binding and matched TF motif
1f	eQTL, TF binding or DNase peak	4	TF binding and DNase peak
2a	TF binding, matched TF motif, matched DNase footprint and DNase peak	5	TF binding or DNase peak
2b	TF binding, any motif, DNase footprint and DNase peak	6	other
2c	TF binding, matched TF motif and DNase peak		
3a	TF binding, any motif and DNase peak		
3b	TF binding and matched TF motif		
4	TF binding and DNase peak		
5	TF binding or DNase peak		
6	other		

Table 3.4. RegulomeDB score and evidence based on integrated ENCODE, eQTL and transcription factor binding site motif prediction. Each SNP is assigned a score based on the weight of evidence defined by overlap in mapping position between the SNP and ENCODE regulatory element data, transcription factor binding motif predictions from positional weight matrices and publicly available expression data from cerebellum, brain cortex, fibroblasts, frontal-cortex, liver, lymphoblastoid cell lines, monocytes, brain pons, T-cells and brain temporal cortex. These eQTL sources are essentially ignored in the Schaub *et al* amended scoring.

3.3 Results

3.3.1 Rheumatoid arthritis susceptibility single nucleotide polymorphisms and drug response

3.3.1.1 Rheumatoid arthritis susceptibility single nucleotide polymorphism association with response in Yorkshire Early Arthritis Register disease-modifying anti-rheumatic drug treated patients

SNP genotyping and treatment response data were available on 577 DMARD treated patients with diagnosed RA in the YEAR cohort. Mean patient age at baseline was 59.1 years, 68.7% were female and 71.6% were positive for RF and/or ACPA (Table 3.5). Mean (standard deviation) baseline DAS28-CRP was 5.3 (1.44) and mean symptom duration was 7.3 months; all patients had a symptom duration of 24 months or less. A total of 41 SNPs were selected for association following genotyping; a single SNP failed QC with a low call-rate (70.8%) at the *CCR5* locus (rs333). The summary statistics for the outcome measures are shown in Table 3.5; mean (SD) ΔDAS28-CRP was 1.44 (1.79), mean (s.d.) ΔSJC28 was 5.31 (7.32) and mean (s.d.) ΔlnCRP was 0.65 (1.19), all of which are moderate to strongly correlated (Spearman's ρ 0.29-0.69).

The 40 SNPs that passed QC were tested for association with the three outcomes measures ΔDAS28-CRP, ΔSJC28 and ΔlnCRP. The minor allele of a SNP at the *TRAF6* locus, rs540386, was associated with a lower ΔSJC28 (β -1.14 95% CI [-1.99--0.29], p=0.009; Table 3.6), but not with ΔDAS28 or ΔlnCRP. No non-MHC SNP associations were found consistently across all 3 outcome measures.

A nominal association was found between *HLA-DRB1* SE alleles and lower ΔDAS28-CRP (β -0.19 95% CI [-0.38--0.01], p=0.036). This borderline effect appeared principally to be due to an effect on lower ΔlnCRP (Table 3.6; β -0.19 95% CI [-0.31--0.07], p=0.002); there was no

statistically significant association with ΔSJC28, though the effect direction did suggest a lower improvement in SJC28. Association testing using the risk categories observed and defined by Mackie *et al* did not find a stratification of effect on DMARD treatment response measured by ΔDAS28-CRP in the YEAR cohort (Table 3.7). A stronger effect was observed in the high risk group in the analysis of ΔlnCRP compared to the effect in the medium risk group (Table 3.7).

Table 3.5. Clinical variables linear regression by outcome variable at 6months													
Variable	Summary stat.	ΔDAS28-CRP				ΔlnCRP				ΔSJC28			
		β	95% CI	p-value	n	β	95% CI	p-value	n	β	95% CI	p-value	n
Baseline age (years), mean (s.d.)	59.1 (13.2)	0.01	0.00-0.02	0.036	578	0.00	0.00-0.01	0.27	575	0.03	-0.01-0.06	0.11	577
Gender, % female	68.7%	-0.20	-0.48-0.08	0.16	578	-0.08	-0.26-0.10	0.37	575	-0.78	-1.63-0.11	0.09	577
Baseline HAQ-DI, median (IQR)	1.2 (0.6)	-0.52	-0.78--0.26	$7.78x10^{-5}$	568	-0.03	-0.18-0.12	0.66	565	-0.80	-1.51--0.09	0.027	567
Autoantibodies, % positive	71.6%	-0.28	-0.57-0.01	0.06	561	-0.31	-0.49--0.12	0.001	558	-0.79	-1.70-0.13	0.09	560
RF, % positive	64.7%	-0.29	-0.57--0.02	0.037	553	-0.31	-0.49--0.14	$4.9x10^{-4}$	552	-0.85	-1.72-0.03	0.06	553
ACPA, % positive	61.7%	-0.43	-0.73--0.13	0.006	432	-0.31	-0.50--0.12	0.002	429	-1.13	-2.09--0.18	0.02	431
Symptom duration (months), mean (s.d.)	7.3 (4.1)	0.00	-0.03-0.03	0.98	578	-0.01	-0.03-0.01	0.17	575	0.09	-0.01-0.19	0.08	577
Cohort (B vs. C)	64.9% YEAR C	0.01	-0.26-0.29	0.92	578	0.18	0.01-0.36	0.04	575	1.16	0.30-2.00	0.008	577
Outcome	Summary stat.	ρ		p	n	ρ		p	n	ρ		p	n
Baseline DAS28-CRP, mean (SD)	5.26 (1.44)	0.46		$5.6x10^{-31}$	577	0.33		$3.9x10^{-16}$	574	0.50		$3.4x10^{-38}$	576
Baseline SJC28, median (range)	8 (0-28)	0.42		$1.8x10^{-25}$	576	0.20		$8.9x10^{-7}$	574	0.76		$5.4x10^{-109}$	576
Baseline CRP, median (range)	15 (0.12-178)	0.25		$5.5x10^{-10}$	576	0.50		$4.2x10^{-38}$	574	0.22		$1.4x10^{-7}$	576
ΔDAS28-CRP, mean (SD)	1.44 (1.79)	-		-	-	0.53		$3.7x10^{-43}$	574	0.69		$6.2x10^{-83}$	576
ΔlnCRP, mean (SD)	0.65 (1.19)	0.53		$3.7x10^{-43}$	574	-		-	-	0.29		$1.2x10^{-12}$	574
ΔSJC28, mean (SD)	5.31 (7.32)	0.69		$6.2x10^{-83}$	576	0.29		$1.2x10^{-12}$	574	-		-	-

Table 3.5. YEAR cohort demographics and clinical measure summary. Summary statistics for each demographic and clinical variable are shown with linear regression estimates and p-values for each clinical variable on each treatment outcome measure. The spearman correlation between treatment outcome measures and their baseline time point measurement are also shown to demonstrate the relationship between these three clinical variables. SD – standard deviation, IQR – inter-quartile range, RF – rheumatoid factor, ACPA – anti-citrullinated peptide antibodies, DAS28-CRP – disease activity score with 28 joint count and CRP, CRP – C-reactive protein, SJC28 – swollen joint count of 28 joints, β – effect estimate from linear regression model, ρ – Spearman correlation coefficient, 95% CI – 95% confidence intervals.

Table 3.6. MHC and SNP association with ΔDAS28-CRP4v, ΔCRP and ΔSJC28 from baseline to 6months adjusted for baseline measurement

SNP	Cytoband	Locus	MAF	Allele	ΔDAS28-CRP				ΔlnCRP				ΔSJC28			
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N
SE	**6p21**	***HLA-DRB1***	**0.46**	***0401,*0404, *0101, *1001**	**-0.19**	**-0.38--0.01**	**0.036**	**605**	**-0.19**	**-0.31--0.07**	**0.002**	**597**	**-0.42**	**-0.98-0.15**	**0.149**	**600**
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.10	-0.30-0.11	0.34	578	-0.13	-0.26-0.01	0.06	575	-0.27	-0.90-0.37	0.41	577
rs4272626	**1p13.1**	***CASQ2***	**0.349**	**T**	**0.18**	**-0.01-0.37**	**0.07**	**578**	**0.10**	**-0.02-0.23**	**0.11**	**575**	**0.67**	**0.07-1.28**	**0.029**	**577**
rs2476601	1p13.2	*PTPN22*	0.138	A	-0.06	-0.33-0.21	0.67	576	-0.11	-0.28-0.07	0.23	573	-0.05	-0.90-0.79	0.90	575
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	0.04	-0.16-0.24	0.68	576	-0.04	-0.17-0.09	0.56	573	0.12	-0.50-0.75	0.70	575
rs12746613	1q23.2	*FCGR2A*	0.119	T	-0.20	-0.49-0.08	0.16	578	-0.05	-0.23-0.14	0.62	575	-0.53	-1.42-0.35	0.24	577
rs10919563	1q31.3	*PTPRC*	0.129	A	0.15	-0.12-0.43	0.27	577	0.04	-0.14-0.22	0.67	574	0.54	-0.33-1.40	0.23	576
rs1160542	2q11.2	*AFF3*	0.493	A	-0.10	-0.28-0.08	0.28	577	-0.06	-0.18-0.05	0.27	574	0.18	-0.37-0.74	0.52	576
rs1980422	2q33.2	*CD28*	0.232	C	-0.13	-0.36-0.09	0.25	578	-0.09	-0.24-0.05	0.21	575	0.02	-0.68-0.72	0.96	577
rs10181656	2q32.2	*STAT4*	0.251	G	-010	-0.30-0.11	0.36	577	0.02	-0.11-0.16	0.73	574	-0.22	-0.86-0.42	0.50	576
rs3087243	2q33.2	*CTLA4*	0.430	A	-0.16	-0.34-0.01	0.07	577	-0.02	-0.13-0.10	0.77	574	-0.42	-0.97-0.13	0.13	576
rs231775	2q33.2	*CTLA4*	0.409	G	0.12	-0.06-0.30	0.21	576	0.03	-0.09-0.15	0.62	573	0.18	-0.39-0.75	0.53	575
rs4535211	3p24.3	*PLCL2*	0.464	A	-0.01	-0.19-0.17	0.90	578	0.03	-0.09-0.15	0.62	575	0.06	-0.51-0.62	0.84	577
rs231707	4p16.3	*FAM193A*	0.191	A	-0.07	-0.30-0.17	0.59	578	0.04	-0.12-0.19	0.63	575	0.08	-0.66-0.82	0.83	577
rs2069778	4q27	*IL2/IL21*	0.155	T	-0.04	-0.28-0.21	0.76	575	-0.11	-0.27-0.04	0.16	572	-0.22	-0.98-0.54	0.57	574
rs6897932	5p13.2	*IL7R*	0.282	C	-0.04	-0.23-0.16	0.72	578	-0.01	-0.13-0.12	0.90	575	-0.17	-0.77-0.44	0.59	577
rs1043730	6q21	*TRAF3IP2*	0.245	C	0.07	-0.14-0.28	0.53	571	-0.02	-0.16-0.11	0.75	568	0.23	-0.43-0.90	0.49	570
rs548234	6q21	*PRDM1*	0.347	C	0.02	-0.18-0.21	0.56	578	-0.04	-0.17-0.08	0.51	575	0.26	-0.35-0.87	0.40	577
rs13207033	6q23.3	*6q23*	0.262	A	0.00	-0.21-0.20	0.98	578	-0.03	-0.11-0.16	0.71	575	-0.25	-0.89-0.39	0.44	577
rs6920220	6q23.3	*6q23*	0.240	A	0.02	-0.19-0.24	0.83	578	-0.04	-0.18-0.10	0.55	575	0.06	-0.61-0.73	0.86	577
rs5029937	6q23.3	*TNFAIP3*	0.055	T	-0.02	-0.42-0.39	0.94	576	0.01	-0.25-0.27	0.94	573	-0.20	-1.47-1.07	0.75	575
rs394581	6q25.3	*TAGAP*	0.294	C	-0.13	-0.34-0.08	0.22	578	-0.04	-0.18-0.09	0.52	575	-0.22	-0.88-0.43	0.51	577
rs42041	7q21.2	*CDK6*	0.259	G	0.03	-0.19-0.24	0.80	577	0.00	-0.14-0.14	1.00	574	-0.20	-0.87-0.46	0.55	576
rs13277113	8p23.1	*BLK*	0.256	A	0.13	-0.08-0.35	0.23	571	-0.01	-0.15-0.13	0.85	568	-0.08	-0.76-0.60	0.81	570
rs2812378	9p13.3	*CCL21*	0.362	C	-0.03	-0.30-0.17	0.75	575	0.00	-0.12-0.13	0.95	572	-0.11	-0.73-0.51	0.73	574

SNP	Cytoband	Locus	MAF	Allele	ΔDAS28-CRP				ΔlnCRP				ΔSJC28			
					β	95% CI	P	N	β	95% CI	p	N	β	95% CI	p	N
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.20	0.01-0.39	0.04	568	0.05	-0.07-0.18	0.39	565	0.38	-0.22-0.97	0.21	567
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.12	-0.08-0.31	0.25	577	0.08	-0.05-0.21	0.22	574	0.21	-0.40-0.82	0.51	576
rs2104286	10p15.1	*IL2RA*	0.279	G	-0.08	-0.28-0.13	0.46	574	-0.01	-0.15-0.12	0.83	571	-0.15	-0.78-0.49	0.65	573
rs11594656	10p15.1	*IL2RA*	0.205	A	0.11	-0.11-0.32	0.33	569	0.00	-0.14-0.14	1.00	566	0.02	-0.66-0.69	0.96	568
rs41295061	10p15.1	*IL2RA*	0.093	A	-0.08	-0.40-0.24	0.61	576	-0.02	-0.22-0.19	0.88	573	-0.37	-1.38-0.63	0.47	575
rs4750316	10p15.1	*10p15*	0.195	C	0.13	-0.08-0.35	0.23	577	0.01	-0.13-0.16	0.86	574	0.25	-0.44-0.94	0.47	576
rs540386	**11p12**	***TRAF6***	**0.134**	**T**	**-0.20**	**-0.48-0.07**	**0.15**	**578**	**0.03**	**-0.14-0.21**	**0.70**	**575**	**-1.14**	**-1.99--0.29**	**0.009**	**577**
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	0.00	-0.18-0.18	0.99	578	-0.08	-0.20-0.04	0.19	575	0.21	-0.35-0.78	0.46	577
rs775241	12q13.3	*DCTN2*	0.119	G	0.06	-0.21-0.34	0.65	577	0.05	-0.12-0.23	0.55	574	0.43	-0.43-1.29	0.33	576
rs7234029	18p11.21	*PTPN2*	0.175	G	0.05	-0.20-0.30	0.69	575	-0.07	-0.23-0.09	0.37	572	0.16	-0.62-0.93	0.69	574
rs763361	18q22.2	*CD226*	0.487	C	0.01	-0.16-0.19	0.87	574	0.05	-0.06-0.17	0.34	571	0.23	-0.31-0.78	0.40	573
rs4239702	20q13.12	*CD40*	0.261	T	0.01	-0.22-0.23	0.96	569	0.06	-0.08-0.21	0.38	566	0.02	-0.68-0.72	0.96	568
rs3788013	21q22.3	*UBASH3A*	0.443	A	-0.16	-0.35-0.03	0.10	574	-0.05	-0.18-0.07	0.40	571	0.20	-0.40-0.79	0.52	573
rs743777	22q12.3	*IL2RB*	0.327	G	-0.06	-0.25-0.14	0.56	576	0.02	-0.10-0.15	0.73	573	0.11	-0.49-0.72	0.71	575
rs3218258	22q12.3	*IL2RB*	0.270	A	-0.14	-0.34-0.06	0.16	576	0.01	-0.12-0.14	0.90	573	-0.30	-0.92-0.32	0.35	575
rs229541	22q12.3	*C1QTNF6*	0.461	T	0.02	-0.17-0.20	0.86	571	0.00	-0.12-0.12	0.97	568	0.26	-0.32-0.83	0.38	570

Table 3.6. Linear regression association analysis of RA susceptibility SNPs and *HLA-DRB1* shared epitope alleles on treatment response to first line DMARDs. Linear regression summary of 40 RA susceptibility SNPs tested for association with ΔDAS28-CRP, ΔlnCRP and ΔSJC28. SNPs with a p<0.01 in any of the three outcome measures are highlighted in bold.

Table 3.7. *HLA-DRB1* shared epitope risk category association with YEAR DMARD treatment response														
Category	MAF	Alleles	ΔDAS28-CRP				ΔlnCRP				ΔSJC28			
			β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N
High risk	0.30	*0401, *0404	-0.29	-0.52--0.06	0.012		-0.21	-0.35--0.06	0.005		-0.58	-1.29-0.13	0.110	
Medium risk	0.16	*0101, *1001	-0.32	-0.60--0.02	0.035	605	-0.17	-0.35-0.02	0.077	597	-0.46	-1.36-0.44	0.316	600
Protective	0.24	*0112, *03, *0448, *09, *14, *15	-0.26	-0.50--0.01	0.042		-0.02	-0.18-0.14	0.814		-0.31	-1.07-0.45	0.421	

Table 3.7. *HLA-DRB1* shared epitope risk category association with DMARD treatment response. Risk categories are defined by the stratification of risk of RA observed and reported in [64].

3.3.1.2 Rheumatoid arthritis susceptibility single nucleotide polymorphisms associated with response to methotrexate monotherapy

Patients with symptom duration ≤ 24 months who initiated treatment on MTX as their first DMARD were identified in the YEAR cohort (n=551). Complete genotyping and response outcome measure data were available on up to 298 patients. The baseline demographics of the MTX-monotherapy patients were very similar to the whole YEAR cohort (Table 3.5 vs. Table 3.8).

Table 3.8. Clinical and demographic summary of YEAR MTX treated patients, n=297	
Variable	Summary statistic
Baseline age, mean (s.d.)	59.3 (12.0)
Gender (% female)	71.1%
Baseline HAQ-DI, median (IQR)	1.3 (2.7)
Autoantibodies	73.3%
RF	66.0%
ACPA	67.8%
Symptom duration, mean (s.d.)	7.3 (4.1)
Baseline DAS28-CRP, mean (s.d.)	5.3 (1.4)
Baseline SJC28, median (range)	7 (0-24)
Baseline CRP, median (range)	15.5 (0.12-178.3)
ΔDAS28-CRP, mean (s.d.)	1.51 (1.73)
ΔSJC28, mean (s.d.)	5.6 (7.3)
ΔlnCRP, mean (s.d.)	0.8 (1.23)

Table 3.8. Clinical and demographic variable summary of MTX-monotherapy YEAR patients. S.d. – standard deviation, IQR – inter-quartile range, HAQ-DI – health assessment questionnaire disability index, RF – rheumatoid factor, ACPA – anti-citrullinated peptide antibodies, DAS28-CRP – disease activity score with 28 joint count and CRP, CRP – C-reactive protein, SJC28 – swollen joint count of 28 joints.

Each of the 40 RA susceptibility SNPs that passed QC were tested for association in this sub-cohort of YEAR with each outcome measure as described in the previous section. No SNPs were associated with ΔDAS28-CRP, ΔSJC28 or ΔlnCRP in these patients with $p<0.01$ (Table 3.9 and Table 3.10).

A second cohort of patients recruited into the placebo arms of 3 clinical trials treated with MTX monotherapy, sponsored by Hoffman La-Roche were used (n=174). The RA susceptibility SNPs were tested for association using the same method as for the YEAR MTX

cohort. The minor allele of rs4750316 (10p15) was associated with an increase in ΔDAS28-CRP (β 0.54 95% CI [0.13-0.94], p=0.009), but not ΔCRP or ΔSJC28, though they all showed a consistent direction of effect.

Following association testing in these two cohorts the data were combined in a random effects meta-analysis; several SNPs, highlighted in red in Tables 3.10, 3.11 and 3.12, demonstrated high heterogeneity (I^2>50%). No SNP was associated in any of the meta-analyses with p<0.01. There was very high heterogeneity (76-87%) with the SNPs at the *IL2RB* and 10p15 loci; these were the SNPs most strongly associated with MTX treatment response in the Roche clinical trials cohort (p<0.05).

The strongest association in the combined DMARD analysis with ΔSJC28 was in the same direction of effect in the YEAR MTX monotherapy analysis, but did not reach statistical significance (*TRAF6*; rs540386 β-0.99 95% CI [-2.02-], p=0.06). No effect was observed in the Roche clinical trials MTX cohort or the combined meta-analysis.

Table 3.9. MTX-YEAR, Roche and meta-analysis SNP association with ΔDAS28-CRP from baseline to 6months, adjusted for baseline DAS28-CRP																	
SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.16	-0.43-0.11	0.24	298	-0.15	-0.56-0.27	0.48	174	-0.16	-0.38-0.07	0.17	472	0
rs4272626	1p13.1	*CASQ2*	0.349	T	0.12	-0.13-0.38	0.35	298	-0.06	-0.38-0.26	0.71	174	0.05	-0.15-0.25	0.62	472	0
rs2476601	1p13.2	*PTPN22*	0.138	A	-0.21	-0.57-0.15	0.25	296	0.17	-0.24-0.58	0.42	174	-0.03	-0.40-0.34	0.86	470	46.4
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	0.03	-0.24-0.30	0.84	297	0.06	-0.29-0.40	0.75	174	0.04	-0.17-0.25	0.72	471	0
rs12746613	**1q23.2**	***FCGR2A***	**0.119**	**T**	**-0.35**	**-0.72-0.03**	**0.07**	**298**	**-0.33**	**-0.83-0.18**	**0.20**	**174**	**-0.34**	**-0.64--0.04**	**0.027**	**472**	**0**
rs10919563	1q31.3	*PTPRC*	0.129	A	0.11	-0.27-0.48	0.58	298	0.41	-0.13-0.95	0.14	174	0.21	-0.10-0.51	0.19	472	0
rs1160542	2q11.2	*AFF3*	0.493	A	-0.05	-0.29-0.19	0.68	297	-0.20	-0.53-0.14	0.25	174	-0.10	-0.29-0.09	0.31	471	0
rs1980422	2q33.2	*CD28*	0.232	C	-0.11	-0.39-0.17	0.45	298	-0.15	-0.50-0.20	0.40	174	-0.12	-0.34-0.09	0.26	472	0
rs10181656	2q32.2	*STAT4*	0.251	G	-0.21	-0.49-0.08	0.16	298	0.18	-0.14-0.51	0.27	174	-0.02	-0.40-0.36	0.92	472	67.7
rs3087243	2q33.2	*CTLA4*	0.430	A	-0.16	-0.40-0.08	0.20	298	-0.22	-0.55-0.11	0.19	174	-0.18	-0.38-0.01	0.07	472	0
rs231775	2q33.2	*CTLA4*	0.409	G	0.18	-0.06-0.42	0.15	297	0.17	-0.16-0.50	0.31	174	0.18	-0.02-0.37	0.08	471	0
rs4535211	3p24.3	*PLCL2*	0.464	A	0.03	-0.22-0.27	0.83	298	0.12	-0.18-0.43	0.42	174	0.06	-0.12-0.25	0.50	472	0
rs231707	4p16.3	*FAM193A*	0.191	A	-0.10	0.42-0.22	0.53	298	0.05	-0.34-0.45	0.79	174	-0.04	-0.29-0.21	0.75	472	0
rs2069778	4q27	*IL2/IL21*	0.155	T	0.18	-0.15-0.51	0.28	297	-0.18	-0.65-0.29	0.45	174	0.04	-0.31-0.39	0.82	471	35.8
rs6897932	5p13.2	*IL7R*	0.282	C	0.12	-0.14-0.38	0.36	298	-0.01	-0.35-0.33	0.96	174	0.07	-0.13-0.28	0.48	472	0
rs1043730	6q21	*TRAF3IP2*	0.245	C	0.14	-0.15-0.42	0.36	296	-0.22	-0.58-0.14	0.24	174	-0.02	-0.37-0.32	0.89	470	55.8
rs548234	6q21	*PRDM1*	0.347	C	-0.17	-0.44-0.10	0.21	298	-0.03	-0.34-0.28	0.85	174	-0.11	-0.31-0.09	0.28	472	0
rs13207033	6q23.3	*6q23*	0.262	A	-0.26	-0.53-0.02	0.07	298	-0.01	-0.35-0.32	0.94	174	-0.15	-0.39-0.09	0.21	472	21.2
rs6920220	6q23.3	*6q23*	0.240	A	0.08	-0.20-0.36	0.59	298	0.19	-0.20-0.57	0.34	174	0.12	-0.11-0.34	0.32	472	0
rs5029937	6q23.3	*TNFAIP3*	0.055	T	0.01	-0.59-0.61	0.98	297	0.28	-0.62-1.18	0.54	174	0.09	-0.40-0.59	0.72	471	0
rs394581	6q25.3	*TAGAP*	0.294	C	-0.18	-0.47-0.12	0.24	298	-0.04	-0.41-0.33	0.84	174	-0.12	-0.35-0.10	0.29	472	0
rs42041	7q21.2	*CDK6*	0.259	G	-0.05	-0.33-0.22	0.70	297	0.24	-0.14-0.62	0.22	174	0.06	-0.22-0.34	0.66	471	33.4
rs13277113	8p23.1	*BLK*	0.256	A	0.06	-0.22-0.34	0.66	293	0.15	-0.22-0.51	0.42	174	0.09	-0.13-0.32	0.40	467	0
rs2812378	9p13.3	*CCL21*	0.362	C	0.04	-0.21-0.29	0.75	296	0.02	-0.31-0.35	0.89	174	0.03	-0.17-0.23	0.74	470	0
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.13	-0.13-0.38	0.33	295	-0.12	-0.46-0.21	0.46	174	0.02	-0.22-0.27	0.85	469	29.3
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.11	-0.15-0.38	0.39	297	0.04	-0.28-0.36	0.81	174	0.08	-0.12-0.29	0.42	471	0
rs2104286	10p15.1	*IL2RA*	0.279	G	-0.13	-0.40-0.14	0.36	295	0.32	-0.06-0.7	0.09	174	0.08	-0.36-0.52	0.73	469	72.5
rs11594656	10p15.1	*IL2RA*	0.205	A	0.17	-0.12-0.45	0.25	293	-0.10	-0.50-0.30	0.62	174	0.07	-0.18-0.32	0.58	467	12.9
rs41295061	10p15.1	*IL2RA*	0.093	A	-0.28	-0.69-0.13	0.18	297	-0.01	-0.52-0.55	0.97	174	-0.19	-0.51-0.14	0.278	471	0

SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs4750316	**10p15.1**	***10p15***	**0.195**	**C**	**0.01**	**-0.29-0.31**	**0.95**	**298**	**0.54**	**0.13-0.94**	**0.009**	**174**	**0.26**	**-0.26-0.77**	**0.33**	**472**	**76.3**
rs540386	11p12	*TRAF6*	0.134	T	-0.18	-0.53-0.16	0.30	298	-0.01	-0.62-0.60	0.97	174	-0.14	-0.44-0.16	0.36	472	0
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	-0.11	-0.35-0.12	0.35	298	0.00	-0.33-0.33	0.99	174	-0.08	-0.27-0.12	0.44	472	0
rs775241	12q13.3	*DCTN2*	0.119	G	-0.04	-0.41-0.34	0.86	298	-0.10	-0.57-0.37	0.68	174	-0.06	-0.36-0.23	0.69	472	0
rs7234029	18p11.21	*PTPN2*	0.175	G	-0.15	-0.50-0.19	0.38	295	-0.22	-0.61-0.17	0.27	174	-0.18	-0.44-0.07	0.16	469	0
rs763361	18q22.2	*CD226*	0.487	C	-0.03	-0.27-0.19	0.75	294	0.23	-0.04-0.56	0.18	174	0.07	-0.18-0.32	0.59	468	38.7
rs4239702	20q13.12	*CD40*	0.261	T	0.00	-0.30-0.31	0.99	295	0.08	-0.28-0.43	0.68	174	0.03	-0.20-0.26	0.78	469	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	-0.20	-0.47-0.06	0.13	296	-0.08	-0.39-0.23	0.61	174	-0.15	-0.35-0.05	0.14	470	0
rs743777	**22q12.3**	***IL2RB***	**0.327**	**G**	**-0.18**	**-0.44-0.08**	**0.18**	**297**	**0.43**	**0.06-0.80**	**0.022**	**174**	**0.11**	**-0.48-0.71**	**0.72**	**471**	**85.8**
rs3218258	**22q12.3**	***IL2RB***	**0.270**	**A**	**-0.22**	**-0.49-0.05**	**0.11**	**296**	**0.42**	**0.05-0.78**	**0.026**	**174**	**0.09**	**-0.54-0.71**	**0.79**	**470**	**87**
rs229541	22q12.3	*C1QTNF6*	0.461	T	0.06	-0.18-0.31	0.60	296	0.01	-0.31-0.33	0.95	174	0.05	-0.15-0.24	0.65	470	0

Table 3.9. RA susceptibility SNP association summary of ΔDAS28-CRP in YEAR and Roche MTX monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with ΔDAS28-CRP. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either of the cohorts or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

Table 3.10. MTX-YEAR, Roche and meta-analysis SNP association with ΔlnCRP from baseline to 6months, adjusted for baseline lnCRP																
SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.13	-0.31-0.06	0.18	295	0.04	-0.20-0.28	0.73	178	-0.06	-0.22-0.10	0.46	473	16.5
rs4272626	1p13.1	*CASQ2*	0.349	T	0.07	-0.10-0.25	0.41	295	0.12	-0.07-0.30	0.21	178	0.09	-0.03-0.22	0.14	473	0
rs2476601	1p13.2	*PTPN22*	0.138	A	-0.16	-0.40-0.08	0.20	293	-0.16	-0.39-0.01	0.16	178	-0.16	-0.33-0.01	0.06	471	0
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	-0.04	-0.23-0.14	0.66	294	-0.06	-0.26-0.14	0.56	178	-0.05	-0.19-0.09	0.48	472	0
rs12746613	1q23.2	*FCGR2A*	0.119	T	-0.09	-0.34-0.16	0.49	295	-0.25	-0.55-0.04	0.09	178	-0.16	-0.35-0.03	0.10	473	0
rs10919563	1q31.3	*PTPRC*	0.129	A	0.14	-0.12-0.40	0.30	295	-0.03	-0.34-0.28	0.86	178	0.07	-0.13-0.27	0.50	473	0
rs1160542	2q11.2	*AFF3*	0.493	A	-0.03	-0.19-0.13	0.70	294	-0.05	-0.24-0.15	0.63	178	-0.04	-0.16-0.09	0.54	472	0
rs1980422	2q33.2	*CD28*	0.232	C	0.09	-0.10-0.28	0.34	295	0.05	-0.15-0.25	0.63	178	0.07	-0.07-0.21	0.30	473	0
rs10181656	2q32.2	*STAT4*	0.251	G	-0.11	-0.30-0.09	0.28	295	0.05	-0.14-0.24	0.58	178	-0.03	-0.18-0.13	0.76	473	26.4
rs3087243	2q33.2	*CTLA4*	0.430	A	-0.09	-0.25-0.08	0.29	295	-0.14	-0.33-0.05	0.16	178	-0.11	-0.23-0.01	0.08	473	0
rs231775	2q33.2	*CTLA4*	0.409	G	0.15	-0.02-0.31	0.08	294	0.08	-0.11-0.27	0.40	178	0.12	-0.01-0.24	0.06	472	0
rs4535211	3p24.3	*PLCL2*	0.464	A	0.09	-0.07-0.25	0.28	295	0.10	-0.08-0.27	0.28	178	0.09	-0.03-0.21	0.13	473	0
rs231707	4p16.3	*FAM193A*	0.191	A	0.00	-0.22-0.21	0.99	295	-0.02	-0.24-0.21	0.89	178	-0.01	-0.16-0.15	0.92	473	0
rs2069778	4q27	*IL2/IL21*	0.155	T	-0.05	-0.27-0.18	0.69	294	-0.06	-0.33-0.21	0.66	178	-0.05	-0.22-0.12	0.55	472	0
rs6897932	5p13.2	*IL7R*	0.282	C	0.09	-0.08-0.27	0.29	295	0.00	-0.20-0.20	1.00	178	0.05	-0.08-0.18	0.43	473	0
rs1043730	6q21	*TRAF3IP2*	0.245	C	0.02	-0.17-0.22	0.82	293	0.03	-0.18-0.24	0.79	178	0.03	-0.12-0.17	0.73	471	0
rs548234	6q21	*PRDM1*	0.347	C	-0.16	-0.34-0.02	0.08	295	0.01	-0.17-0.19	0.93	178	-0.08	-0.24-0.09	0.36	473	39.4
rs13207033	6q23.3	*6q23*	0.262	A	-0.10	-0.28-0.09	0.31	295	-0.08	-0.28-0.11	0.40	178	-0.09	-0.22-0.04	0.19	473	0
rs6920220	6q23.3	*6q23*	0.240	A	-0.04	-0.23-0.14	0.64	295	-0.13	-0.35-0.09	0.23	178	-0.08	-0.23-0.06	0.26	473	0
rs5029937	6q23.3	*TNFAIP3*	0.055	T	0.00	-0.40-0.41	0.98	294	-0.45	-0.98-0.07	0.09	178	-0.19	-0.64-0.25	0.39	472	46
rs394581	6q25.3	*TAGAP*	0.294	C	-0.07	-0.27-0.13	0.49	295	0.10	-0.12-0.31	0.39	178	0.01	-0.15-0.17	0.94	473	19.3
rs42041	7q21.2	*CDK6*	0.259	G	-0.08	-0.27-0.11	0.39	294	0.20	-0.02-0.42	0.08	178	0.05	-0.22-0.33	0.71	472	72.6
rs13277113	8p23.1	*BLK*	0.256	A	-0.04	-0.23-0.15	0.68	290	-0.01	-0.23-0.20	0.90	178	-0.03	-0.17-0.11	0.69	468	0
rs2812378	9p13.3	*CCL21*	0.362	C	0.04	-0.13-0.21	0.66	293	-0.04	-0.23-0.15	0.68	178	0.00	-0.12-0.13	0.96	471	0
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.10	-0.08-0.27	0.27	292	0.08	-0.11-0.27	0.42	178	0.09	-0.04-0.22	0.17	470	0
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.11	-0.07-0.28	0.24	294	0.08	-0.10-0.27	0.38	178	0.10	-0.03-0.22	0.14	472	0
rs41295061	10p15.1	*IL2RA*	0.093	A	-0.17	-0.45-0.11	0.23	294	0.02	-0.31-0.35	0.90	178	-0.09	-0.30-0.12	0.40	472	0
rs4750316	10p15.1	*10p15*	0.195	C	0.11	-0.10-0.32	0.32	295	0.18	-0.06-0.42	0.14	178	0.14	-0.02-0.30	0.08	473	0
rs540386	11p12	*TRAF6*	0.134	T	-0.06	-0.30-0.17	0.61	295	-0.06	-0.41-0.29	0.74	178	-0.06	-0.26-0.13	0.54	473	0

SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	-0.13	-0.29-0.03	0.12	295	0.13	-0.06-0.32	0.18	178	0.00	-0.26-0.25	0.97	473	75.9
rs775241	12q13.3	*DCTN2*	0.119	G	0.10	-0.16-0.36	0.44	295	-0.05	-0.33-0.22	0.71	178	0.03	-0.16-0.22	0.76	473	0
rs7234029	18p11.21	*PTPN2*	0.175	G	-0.13	-0.36-0.10	0.26	292	-0.18	-0.41-0.05	0.12	178	-0.16	-0.32-0.01	0.06	470	0
rs763361	18q22.2	*CD226*	0.487	C	0.05	-0.11-0.20	0.55	291	0.17	-0.02-0.36	0.09	178	0.10	-0.03-0.22	0.12	469	0
rs4239702	20q13.12	*CD40*	0.261	T	0.07	-0.14-0.27	0.53	292	0.04	-0.17-0.24	0.71	178	0.05	-0.09-0.20	0.48	470	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	-0.15	-0.33-0.03	0.11	293	0.06	-0.12-0.24	0.53	178	-0.05	-0.25-0.16	0.66	471	60.7
rs743777	22q12.3	*IL2RB*	0.327	G	0.00	-0.17-0.18	0.97	294	0.07	-0.14-0.29	0.51	178	0.03	-0.11-0.17	0.65	472	0
rs3218258	22q12.3	*IL2RB*	0.270	A	-0.04	-0.22-0.14	0.66	293	0.03	-0.19-0.25	0.77	178	-0.01	-0.15-0.13	0.88	471	0
rs229541	**22q12.3**	***C1QTNF6***	**0.461**	**T**	**0.06**	**-0.10-0.23**	**0.46**	**293**	**0.20**	**0.01-0.38**	**0.038**	**178**	**0.12**	**-0.01-0.25**	**0.07**	**471**	**14.1**
rs2104286	10p15.1	*IL2RA*	0.279	G	-0.02	-0.21-0.16	0.79	292	-0.02	-0.24-0.20	0.85	178	-0.02	-0.17-0.12	0.75	470	0
rs11594656	10p15.1	*IL2RA*	0.205	A	0.07	-0.12-0.27	0.46	290	0.05	-0.19-0.28	0.70	178	0.06	-0.09-0.21	0.41	468	0

Table 3.10. RA susceptibility SNP association summary of ΔlnCRP in YEAR and Roche MTX monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with ΔlnCRP. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either the YEAR or Roche cohort or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

Table 3.11. MTX-YEAR, Roche and meta-analysis SNP association with ΔSJC28 from baseline to 6months, adjusted for baseline SJC28																
SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.08	-0.89-0.74	0.85	297	-0.81	-2.26-0.64	0.27	176	-0.25	-0.96-0.45	0.48	473	0
rs4272626	**1p13.1**	***CASQ2***	**0.349**	**T**	**0.80**	**0.02-1.57**	**0.04**	**297**	**0.27**	**-0.86-1.39**	**0.64**	**176**	**0.63**	**-0.01-1.26**	**0.05**	**473**	**0**
rs2476601	1p13.2	*PTPN22*	0.138	A	-0.22	-1.30-0.86	0.69	295	0.72	-0.70-2.14	0.32	176	0.14	-0.76-1.03	0.77	471	7.5
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	0.18	-0.64-0.99	0.67	296	0.55	-0.66-1.76	0.37	176	0.29	-0.38-0.97	0.39	472	0
rs12746613	**1q23.2**	***FCGR2A***	**0.119**	**T**	**-1.45**	**-0.26--0.33**	**0.011**	**297**	**-0.04**	**-1.82-1.75**	**0.97**	**176**	**-0.92**	**-2.26-0.42**	**0.18**	**473**	**42.9**
rs10919563	1q31.3	*PTPRC*	0.129	A	0.66	-0.46-1.79	0.25	297	1.76	-0.11-3.63	0.07	176	0.96	0.00-1.92	0.05	473	0
rs1160542	2q11.2	*AFF3*	0.493	A	0.17	-0.55-0.89	0.64	296	0.20	-0.98-1.38	0.73	176	0.18	-0.43-0.79	0.57	472	0
rs1980422	2q33.2	*CD28*	0.232	C	-0.30	-1.13-0.54	0.48	297	-0.63	-1.86-0.60	0.31	176	-0.40	-1.09-0.28	0.25	473	0
rs10181656	2q32.2	*STAT4*	0.251	G	-0.42	-1.27-0.44	0.34	297	0.55	-5.93-1.69	0.34	176	-0.01	-0.94-0.92	0.99	473	43.6
rs3087243	2q33.2	*CTLA4*	0.430	A	-0.25	-0.99-0.48	0.50	297	-0.23	-1.38-0.92	0.70	176	-0.25	-0.86-0.37	0.44	473	0
rs231775	2q33.2	*CTLA4*	0.409	G	0.32	-0.40-1.05	0.38	296	0.11	-1.05-1.27	0.85	176	0.26	-0.35-0.88	0.40	472	0
rs4535211	3p24.3	*PLCL2*	0.464	A	0.03	-0.70-0.75	0.95	297	0.14	-0.92-1.21	0.79	176	0.06	-0.53-0.66	0.84	473	0
rs231707	4p16.3	*FAM193A*	0.191	A	0.11	-0.85-1.06	0.83	297	0.78	-0.60-2.15	0.27	176	0.33	-0.46-1.11	0.41	473	0
rs2069778	4q27	*IL2/IL21*	0.155	T	0.10	-0.88-1.08	0.84	296	-0.77	-2.40-0.87	0.36	176	-0.13	-0.97-0.71	0.76	472	0
rs6897932	5p13.2	*IL7R*	0.282	C	0.05	-0.72-0.82	0.90	297	-0.15	-1.34-1.04	0.80	176	-0.01	-0.65-0.64	0.98	473	0
rs1043730	6q21	*TRAF3IP2*	0.245	C	0.26	-0.60-1.12	0.55	295	-0.50	-1.75-0.75	0.45	176	0.02	-0.69-0.72	0.96	471	0
rs548234	6q21	*PRDM1*	0.347	C	-0.34	-0.83-0.76	0.93	297	-0.18	-1.28-0.93	0.76	176	-0.28	-0.92-0.36	0.39	473	0
rs13207033	6q23.3	*6q23*	0.262	A	-0.11	-0.94-0.72	0.80	297	1.04	-1.22-2.21	0.08	176	0.39	-0.73-1.51	0.49	473	60.4
rs6920220	6q23.3	*6q23*	0.240	A	-0.02	-0.86-0.82	0.96	297	1.29	-0.02-2.59	0.05	176	0.53	-0.73-1.80	0.41	473	63.7
rs5029937	6q23.3	*TNFAIP3*	0.055	T	-0.11	-1.90-1.68	0.91	296	0.09	-3.09-3.27	0.96	176	-0.06	-1.62-1.49	0.94	472	0
rs394581	6q25.3	*TAGAP*	0.294	C	-0.62	-1.48-0.24	0.16	297	-0.02	-1.33-1.29	0.97	176	-0.44	-1.15-0.28	0.23	473	0
rs42041	7q21.2	*CDK6*	0.259	G	0.04	-0.79-0.87	0.92	296	0.03	-1.33-1.38	0.97	176	0.04	-0.67-0.74	0.92	472	0
rs13277113	8p23.1	*BLK*	0.256	A	0.11	-0.73-0.95	0.80	292	0.31	-0.98-1.60	0.63	176	0.17	-0.53-0.87	0.64	468	0
rs2812378	9p13.3	*CCL21*	0.362	C	0.13	-0.62-0.89	0.73	295	-0.24	-1.40-0.93	0.69	176	0.02	-0.61-0.65	0.95	471	0
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	-0.23	-0.99-0.53	0.56	294	-0.39	-1.54-0.75	0.50	176	-0.28	-0.91-0.35	0.39	470	0
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	-0.14	-0.93-0.65	0.72	296	0.21	-0.91-1.33	0.71	176	-0.03	-0.67-0.62	0.94	472	0
rs2104286	10p15.1	*IL2RA*	0.279	G	0.36	-0.45-1.18	0.38	294	1.10	-0.21-2.43	0.11	176	0.57	-0.13-1.26	0.11	470	0
rs11594656	10p15.1	*IL2RA*	0.205	A	-0.02	-0.87-0.84	0.97	292	-0.22	-1.64-1.20	0.76	176	-0.07	-0.80-0.66	0.84	468	0
rs41295061	10p15.1	*IL2RA*	0.093	A	-0.55	-0.18-0.68	0.38	296	-0.07	-2.05-1.90	0.94	176	-0.42	-1.46-0.63	0.43	472	0

SNP	Cytoband	Locus	MAF	Allele	YEAR				Roche				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs4750316	10p15.1	*10p15*	0.195	C	-0.20	-1.11-0.71	0.66	297	1.26	-0.19-2.70	0.09	176	0.42	-1.00-1.83	0.56	473	64.8
rs540386	11p12	*TRAF6*	0.134	T	-0.99	-2.02-0.05	0.06	297	-0.02	-2.18-2.14	0.98	176	-0.81	-1.73-0.12	0.09	473	0
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	0.15	-0.55-0.86	0.66	297	0.06	-1.11-1.23	0.92	176	0.13	-0.47-0.73	0.67	473	0
rs775241	12q13.3	*DCTN2*	0.119	G	0.11	-1.02-1.25	0.84	297	0.74	-0.91-2.40	0.38	176	0.32	-0.62-1.25	0.51	473	0
rs7234029	18p11.21	*PTPN2*	0.175	G	0.02	-1.00-1.04	0.96	294	-0.77	-2.13-0.60	0.27	176	-0.26	-1.07-0.55	0.53	470	0
rs763361	18q22.2	*CD226*	0.487	C	-0.28	-0.98-0.41	0.43	293	0.38	-0.80-1.55	0.53	176	-0.11	-0.70-0.49	0.72	469	0
rs4239702	20q13.12	*CD40*	0.261	T	-0.10	-1.01-0.81	0.82	294	0.39	-0.85-1.62	0.54	176	0.07	-0.66-0.80	0.85	470	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	0.08	-0.72-0.87	0.85	295	-0.78	-1.87-0.32	0.16	176	-0.27	-1.09-0.55	0.53	471	35.5
rs743777	**22q12.3**	***IL2RB***	**0.327**	**G**	**-0.25**	**-1.30-0.30**	**0.54**	**296**	**1.40**	**0.11-2.68**	**0.033**	**176**	**0.50**	**-1.11-2.10**	**0.55**	**472**	**78.5**
rs3218258	**22q12.3**	***IL2RB***	**0.270**	**A**	**-0.50**	**-1.03-0.54**	**0.22**	**295**	**1.58**	**0.29-2.87**	**0.016**	**176**	**0.48**	**-1.56-2.52**	**0.65**	**471**	**86.3**
rs229541	22q12.3	*C1QTNF6*	0.461	T	0.42	-0.30-1.15	0.25	295	-0.57	-1.69-0.56	0.32	176	0.03	-0.92-0.98	0.96	471	52.5

Table 3.11. RA susceptibility SNP association summary of ΔSJC28 in YEAR and Roche MTX monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with ΔSJC28. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either cohort or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

3.3.1.3 Rheumatoid arthritis single nucleotide polymorphism association with sulphasalazine monotherapy treatment response

During the recruitment into the YEAR cohort clinical best practise moved towards escalation MTX-monotherapy treatment; previously the first-line treatment was SSA. Patients treated with SSA monotherapy were identified based on their starting DMARD (n=534), of which 188 had matched clinical and genotyping data. Ranked percentiles of ΔDAS28-CRP (%ΔDAS), ΔSJC28 (%ΔSJC) and ΔlnCRP (%ΔCRP) were calculated to facilitate the meta-analysis of data from an additional data set of SSA treated patients derived from the ERAS cohort (see methods). The SSA treated YEAR patients were broadly similar to the MTX-treated patients, with the exception of fewer ACPA positive patients in the SSA cohort (67.8% vs. 57.8%) and a higher median baseline CRP (15.5 vs. 20.0).

Table 3.12. Clinical and demographic summary of YEAR SSA treated patients, n=188	
Variable	Summary statistic
Baseline age, mean (s.d.)	60.0 (13.3)
Gender	63.8%
Baseline HAQ-DI, median (IQR)	1.3 (2.7)
Autoantibodies	71.0%
RF	64.9%
CCP	57.8%
Symptom duration, mean (s.d.)	7.3 (4.1)
Baseline DAS28-CRP, mean (s.d.)	5.6 (1.4)
Baseline SJC28, median (range)	10 (1-28)
Baseline CRP, median (range)	20 (0.16-144)
ΔDAS28-CRP, mean (s.d.)	1.61 (1.85)
ΔSJC28, mean (s.d.)	5.8 (7.6)
ΔlnCRP, mean (s.d.)	0.61 (1.08)

Table 3.12. Clinical and demographic variable summary of SSA-monotherapy YEAR patients. S.d. – standard deviation, IQR – inter-quartile range, HAQ-DI – health assessment questionnaire disability index, RF – rheumatoid factor, ACPA – anti-citrullinated peptide antibodies, DAS28-CRP – disease activity score with 28 joint count and CRP, CRP – C-reactive protein, SJC28 – swollen joint count of 28 joints.

RA susceptibility SNPs were tested for association with SSA response measured using %ΔDAS, %ΔSJC and %ΔCRP (Tables 3.13, 3.14 and 3.15). No single SNP was associated with any of the three outcome measures ($p<0.01$) in the YEAR data.

The ERAS cohort recruitment began in 1989 and patients continue to be followed up to date. As such, many of the patients in this cohort received SSA monotherapy as the first-line DMARD during that recruitment period. SSA monotherapy patients were identified from the ERAS cohort with matching RA susceptibility SNP genotype and clinical data (n=190). This dataset served as a second cohort for the YEAR SSA cohort, which were then combined in a random effects meta-analysis. The minor allele of a SNP was associated with increased %ΔDAS in the ERAS data (Table 3.13) at the *TRAF3IP2* locus (rs1043730 β 10.5 95% CI [4.20-16.9], p=0.001). The minor allele of rs6920220 at 6q23 was associated with a lower %ΔESR (β -9.73 95% CI [-16.7--2.77], p=0.006), but not %ΔSJC or %ΔDAS.

A random-effects meta-analysis of the YEAR and ERAS data found no SNPs associated with %ΔDAS (Table 3.13), %ΔSJC (Table 3.14) or %ΔCRP/ESR (Table 3.15). High heterogeneity was observed for several SNPs, including rs1043730 at the *TRAF3IP2* locus which was only associated in the ERAS dataset (I^2 ≈75).

The *TRAF6* SNP rs540386 minor allele associated in the combined DMARD analysis did not show any evidence of association in the SSA monotherapy YEAR, ERAS or meta-analysis results.

Table 3.13. SSA-YEAR, ERAS and meta-analysis SNP association with %ΔDAS from baseline to 6months, adjusted for baseline DAS28/DAS

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	0.61	-5.83-7.06	0.85	188	5.97	-3.78-12.3	0.07	190	3.32	-1.93-8.57	0.22	378	26.8
rs4272626	**1p13.1**	***CASQ2***	**0.349**	**T**	**2.95**	-3.16-9.05	**0.34**	**188**	**6.12**	0.71-11.5	**0.027**	**190**	**4.73**	**0.70-8.75**	**0.021**	**378**	**0**
rs2476601	1p13.2	*PTPN22*	0.138	A	4.97	-3.09-13.0	0.23	188	2.71	-4.46-9.88	0.46	189	3.71	-1.61-9.03	0.17	377	0
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	-0.53	-6.78-5.72	0.87	187	2.07	-3.79-7.92	0.49	190	0.85	-3.39-5.10	0.69	377	0
rs12746613	1q23.2	*FCGR2A*	0.119	T	-1.72	-9.90-6.47	0.68	188	1.95	-6.59-10.5	0.65	184	0.04	-5.83-5.91	0.99	372	0
rs10919563	1q31.3	*PTPRC*	0.129	A	4.47	-3.80-12.7	0.29	187	1.93	-6.55-10.4	0.65	184	3.23	-2.65-0.91	0.28	371	0
rs1160542	2q11.2	*AFF3*	0.493	A	-1.80	-73.1-3.70	0.52	188	-0.69	-6.19-4.82	0.81	187	-1.24	-5.11-0.26	0.53	375	0
rs1980422	2q33.2	*CD28*	0.232	C	-5.51	-11.8-0.80	0.09	188	-0.32	-6.48-5.84	0.92	189	-2.87	-7.96-2.21	0.27	377	25.8
rs10181656	2q32.2	*STAT4*	0.251	G	-0.38	-7.53-6.76	0.92	188	-1.43	-7.94-5.08	0.67	190	-0.10	-5.74-3.82	0.70	378	0
rs3087243	2q33.2	*CTLA4*	0.430	A	-3.52	-8.56-1.51	0.17	187	1.94	-3.50-7.39	0.48	189	-0.89	-0.62-4.46	0.74	376	52.7
rs231775	2q33.2	*CTLA4*	0.409	G	3.38	-2.28-9.04	0.24	187	-1.50	-6.69-3.69	0.57	188	0.81	-3.96-5.58	0.74	375	36.2
rs4535211	3p24.3	*PLCL2*	0.464	A	-1.83	-7.42-3.76	0.52	188	-0.52	-5.19-4.88	0.85	190	-1.15	-5.01-2.71	0.56	378	0
rs231707	4p16.3	*FAM193A*	0.191	A	-2.07	-9.09-4.96	0.56	188	-3.91	-11.4-3.60	0.31	190	-2.93	-8.02-2.17	0.26	378	0
rs2069778	4q27	*IL2/IL21*	0.155	T	-2.94	-13.3-1.38	0.11	186	-0.42	-6.91-6.07	0.90	190	-1.53	-6.36-3.30	0.53	376	0
rs6897932	5p13.2	*IL7R*	0.282	C	-0.45	-6.71-5.82	0.89	188	2.33	-4.15-8.82	0.49	189	0.89	-0.36-5.37	0.70	377	0
rs1043730	**6q21**	***TRAF3IP2***	**0.245**	**C**	**-0.12**	**-6.60-6.35**	**0.97**	**184**	**10.53**	**4.20-16.9**	**0.001**	**190**	**5.23**	**-5.22-15.7**	**0.33**	**374**	**81.4**
rs548234	6q21	*PRDM1*	0.347	C	1.68	-4.04-7.40	0.56	188	-1.07	-6.71-4.58	0.71	189	0.29	-3.70-4.28	0.89	377	0
rs13207033	6q23.3	*6q23*	0.262	A	3.08	-2.95-9.12	0.32	188	0.45	-5.68-6.59	0.88	190	1.79	-2.48-6.06	0.41	378	0
rs6920220	6q23.3	*6q23*	0.240	A	-0.88	-7.60-5.84	0.80	188	-6.45	-13.1-0.17	0.06	190	-3.70	-9.15-1.76	0.18	378	26.2
rs5029937	6q23.3	*TNFAIP3*	0.055	T	3.16	-8.88-15.2	0.61	188	1.84	-8.68-12.4	0.73	189	2.41	-5.46-10.3	0.55	377	0
rs394581	6q25.3	*TAGAP*	0.294	C	1.26	-5.18-7.70	0.70	188	-0.38	-6.08-5.31	0.89	189	0.34	-3.90-4.57	0.88	377	0
rs42041	7q21.2	*CDK6*	0.259	G	4.71	-2.08-11.5	0.17	188	3.42	-2.36-9.20	0.24	188	3.97	-0.41-8.34	0.08	376	0
rs13277113	8p23.1	*BLK*	0.256	A	4.33	-2.39-11.1	0.21	186	3.89	-1.91-9.69	0.19	190	4.08	-0.28-8.44	0.07	376	0
rs2812378	9p13.3	*CCL21*	0.362	C	-0.68	-7.16-5.80	0.84	187	-1.73	-7.74-4.28	0.57	189	-1.24	-5.62-3.13	0.58	376	0
rs10760130	**9q33.1**	***TRAF1/C5***	**0.466**	**G**	**6.42**	**0.56-12.3**	**0.03**	**183**	**2.74**	**-2.67-8.14**	**0.32**	**190**	**4.43**	**0.48-8.38**	**0.028**	**373**	**0**
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	3.15	-2.69-8.98	0.29	188	3.96	-1.66-9.58	0.17	190	3.57	-0.45-7.59	0.08	378	0
rs2104286	10p15.1	*IL2RA*	0.279	G	-0.97	-7.26-5.32	0.76	187	-0.48	-6.79-5.83	0.88	185	-0.73	-5.15-3.70	0.75	372	0
rs11594656	10p15.1	*IL2RA*	0.205	A	1.56	-5.11-8.23	0.65	185	5.57	-0.46-11.6	0.07	188	3.77	-0.68-8.21	0.10	373	0
rs41295061	10p15.1	*IL2RA*	0.093	A	0.69	-8.93-10.3	0.89	188	-0.08	-8.99-8.82	0.99	190	0.27	-0.62-6.77	0.94	378	0

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs4750316	10p15.1	*10p15*	0.195	C	4.19	-2.56-10.9	0.22	187	-1.49	-8.66-5.68	0.68	190	1.48	-4.08-7.04	0.60	377	22.8
rs540386	11p12	*TRAF6*	0.134	T	-6.94	-16.0-2.07	0.13	188	3.59	-4.94-12.1	0.41	190	-1.57	-11.9-8.75	0.77	378	64.3
rs1678542	**12q13.3**	***KIF5A/PIP4K2C***	**0.412**	**C**	**1.53**	**-4.17-7.23**	**0.60**	**188**	**7.41**	**1.53-13.3**	**0.014**	**190**	**4.42**	**-1.34-10.2**	**0.13**	**378**	**50.2**
rs775241	12q13.3	*DCTN2*	0.119	G	4.31	-3.86-12.5	0.30	187	3.44	-5.45-12.3	0.45	189	3.91	-2.07-9.89	0.20	376	0
rs7234029	18p11.21	*PTPN2*	0.175	G	5.32	-2.15-12.8	0.16	188	-2.99	-10.2-4.23	0.42	190	1.11	-7.03-9.25	0.79	378	59.8
rs763361	18q22.2	*CD226*	0.487	C	-0.26	-5.49-4.98	0.92	188	3.36	-2.06-8.78	0.22	190	1.49	-2.25-5.23	0.44	378	0
rs4239702	20q13.12	*CD40*	0.261	T	-2.89	-9.64-3.86	0.40	183	-5.01	-11.3-1.24	0.12	190	-4.03	-8.59-0.53	0.08	373	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	-0.19	-5.77-5.39	0.95	186	0.52	-4.80-2.85	0.85	190	0.18	-3.64-4.01	0.93	376	0
rs743777	22q12.3	*IL2RB*	0.327	G	-1.41	-7.21-4.39	0.63	188	5.32	-0.07-11.2	0.06	189	2.02	-4.57-8.61	0.55	377	63.7
rs3218258	22q12.3	*IL2RB*	0.270	A	-0.09	-5.74-5.57	0.98	187	5.55	-0.19-10.8	0.05	190	2.74	-2.78-8.26	0.33	377	48.6
rs229541	22q12.3	*C1QTNF6*	0.461	T	-0.29	-6.11-5.53	0.92	183	-2.20	-7.27-2.87	0.94	190	-1.38	-5.18-2.42	0.48	373	0

Table 3.13. RA susceptibility SNP association summary of %ΔDAS in YEAR and ERAS SSA monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with %ΔDAS. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either the ERAS or YEAR cohorts or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

Table 3.14. SSA-YEAR, ERAS and meta-analysis SNP association with %ΔSJC28 from baseline to 6months, adjusted for baseline SJC28/SJC

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.94	-5.83-3.94	0.70	188	5.11	-0.72-10.9	0.09	190	1.87	-4.05-7.79	0.54	378	59.5
rs4272626	1p13.1	*CASQ2*	0.349	T	0.66	-3.99-5.30	0.78	188	4.42	-0.56-9.39	0.08	190	2.43	-1.25-6.11	0.20	378	16
rs2476601	1p13.2	*PTPN22*	0.138	A	1.51	-4.61-7.64	0.63	188	2.35	-4.20-8.89	0.48	189	1.90	-2.54-6.35	0.40	377	0
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	-0.41	-5.12-4.30	0.86	187	3.58	-1.80-8.95	0.19	190	1.37	-2.52-5.26	0.49	377	17.7
rs12746613	1q23.2	*FCGR2A*	0.119	T	1.57	-4.62-7.76	0.62	188	2.01	-5.84-9.86	0.61	184	1.74	-3.09-6.57	0.48	372	0
rs10919563	1q31.3	*PTPRC*	0.129	A	2.02	-4.19-8.24	0.52	187	3.14	-4.69-11.0	0.43	184	2.46	-2.38-7.29	0.32	371	0
rs1160542	2q11.2	*AFF3*	0.493	A	0.64	-3.55-4.82	0.76	188	-0.86	-5.89-4.16	0.74	187	0.02	-3.18-3.22	0.99	375	0
rs1980422	2q33.2	*CD28*	0.232	C	-1.75	-6.55-3.05	0.47	188	-3.90	-9.86-2.06	0.20	190	-2.65	-6.28-0.99	0.15	378	0
rs10181656	2q32.2	*STAT4*	0.251	G	-0.86	-6.28-4.56	0.75	188	-1.24	-6.90-4.14	0.67	189	-1.03	-5.02-2.95	0.61	377	0
rs3087243	2q33.2	*CTLA4*	0.430	A	-2.76	-6.56-1.03	0.15	187	0.95	-4.06-5.96	0.71	189	-1.28	-4.84-2.29	0.48	376	26.3
rs231775	2q33.2	*CTLA4*	0.409	G	2.69	-1.57-6.94	0.21	187	0.35	-4.45-5.14	0.89	188	1.66	-1.5-04.82	0.30	375	0
rs4535211	3p24.3	*PLCL2*	0.464	A	-1.68	-5.92-2.57	0.44	188	0.66	-4.28-5.61	0.79	190	-0.69	-3.89-2.51	0.67	378	0
rs231707	4p16.3	*FAM193A*	0.191	A	0.67	-4.63-5.96	0.80	188	-6.06	-12.9-0.78	0.08	190	-2.34	-8.89-4.22	0.48	378	57.5
rs2069778	4q27	*IL2/IL21*	0.155	T	-3.85	-9.39-1.70	0.17	186	-2.20	-8.14-3.75	0.47	190	-3.08	-7.11-0.95	0.13	376	0
rs6897932	5p13.2	*IL7R*	0.282	C	-1.79	-6.52-2.94	0.46	188	3.29	-2.66-9.24	0.28	189	0.42	-4.51-5.36	0.87	377	42.4
rs1043730	**6q21**	***TRAF3IP2***	**0.245**	**C**	**2.18**	**-2.74-7.10**	**0.38**	**184**	**7.52**	**1.65-13.4**	**0.012**	**190**	**4.60**	**-0.61-9.81**	**0.08**	**374**	**47.2**
rs548234	6q21	*PRDM1*	0.347	C	-1.36	-5.70-2.97	0.54	188	-0.85	-6.03-4.33	0.75	189	-1.15	-4.45-2.15	0.49	377	0
rs13207033	6q23.3	*6q23*	0.262	A	-0.69	-5.29-3.91	0.77	188	-0.76	-6.40-4.88	0.79	190	-0.72	-4.26-2.82	0.69	378	0
rs6920220	6q23.3	*6q23*	0.240	A	-2.70	-7.76-2.37	0.30	188	-3.49	-9.59-2.61	0.26	190	-3.02	-6.89-0.85	0.13	378	0
rs5029937	6q23.3	*TNFAIP3*	0.055	T	1.96	-7.16-11.1	0.67	188	2.46	-7.17-12.1	0.62	189	2.20	-4.38-8.77	0.51	377	0
rs394581	6q25.3	*TAGAP*	0.294	C	0.55	-4.32-5.43	0.82	188	0.68	-4.53-5.90	0.80	189	0.61	-2.93-4.15	0.73	377	0
rs42041	7q21.2	*CDK6*	0.259	G	0.83	-4.35-6.00	0.75	188	2.28	-2.98-7.55	0.39	188	1.54	-2.13-5.21	0.41	376	0
rs13277113	8p23.1	*BLK*	0.256	A	-2.93	-8.04-2.17	0.26	186	4.60	-0.70-9.91	0.09	190	0.80	-6.58-8.18	0.83	376	75.5
rs2812378	9p13.3	*CCL21*	0.362	C	2.34	-2.52-7.21	0.34	187	-0.05	-5.58-5.47	0.99	189	1.30	-2.33-4.93	0.48	376	0
rs10760130	**9q33.1**	***TRAF1/C5***	**0.466**	**G**	**4.85**	**0.43-9.27**	**0.032**	**183**	**1.53**	**-3.43-6.49**	**0.55**	**190**	**3.38**	**0.10-6.66**	**0.04**	**373**	**0**
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	3.05	-1.35-7.45	0.17	188	2.82	-2.34-7.99	0.28	190	2.95	-0.37-6.28	0.08	378	0
rs2104286	10p15.1	*IL2RA*	0.279	G	-2.87	-7.60-1.87	0.23	187	1.87	-3.85-7.58	0.52	185	-0.78	-5.39-3.83	0.74	372	36.9
rs11594656	10p15.1	*IL2RA*	0.205	A	-1.42	-6.46-3.62	0.58	185	4.60	-0.91-10.1	0.10	188	1.48	-4.42-7.38	0.62	373	60.5
rs41295061	10p15.1	*IL2RA*	0.093	A	-2.41	-9.67-4.85	0.51	188	2.63	-5.55-10.8	0.53	190	-0.19	-5.58-5.21	0.95	378	0

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs4750316	10p15.1	*10p15*	0.195	C	3.06	-2.09-8.22	0.24	187	-0.94	-7.52-5.63	0.78	190	1.54	-2.49-5.57	0.45	377	0
rs540386	11p12	*TRAF6*	0.134	T	-3.74	-10.6-3.10	0.28	188	0.57	-7.28-8.43	0.89	190	-1.88	-7.01-3.25	0.47	378	0
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	0.28	-4.03-4.59	0.90	188	2.39	-3.06-7.84	0.39	190	1.09	-2.27-4.45	0.52	378	0
rs775241	12q13.3	*DCTN2*	0.119	G	5.86	-0.31-12.0	0.06	187	2.22	-5.90-10.4	0.59	189	4.53	-0.35-9.41	0.07	376	0
rs7234029	18p11.21	*PTPN2*	0.175	G	3.48	-2.18-9.13	0.23	188	-3.64	-10.3-2.98	0.28	190	0.13	-6.83-7.09	0.97	378	61.5
rs763361	18q22.2	*CD226*	0.487	C	0.24	-3.73-4.21	0.91	188	2.10	-2.89-7.09	0.41	190	0.96	-2.13-4.05	0.54	378	0
rs4239702	20q13.12	*CD40*	0.261	T	-2.08	-7.24-3.09	0.43	183	-4.03	-9.78-1.71	0.17	190	-2.95	-6.77-0.87	0.13	373	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	0.71	-3.47-4.89	0.74	186	0.90	-3.99-5.79	0.72	190	0.79	-2.37-3.95	0.62	376	0
rs743777	22q12.3	*IL2RB*	0.327	G	2.77	-1.49-7.03	0.20	187	2.98	-2.10-8.07	0.25	189	2.86	-0.42-6.13	0.09	376	0
rs3218258	22q12.3	*IL2RB*	0.270	A	1.30	-3.07--5.68	0.56	188	3.11	-2.07-8.30	0.24	190	2.07	-1.22-5.37	0.22	378	0
rs229541	22q12.3	*C1QTNF6*	0.461	T	0.98	-3.44-5.39	0.66	183	-2.83	-7.48-1.81	0.23	190	-0.86	-4.58-2.87	0.65	373	27.2

Table 3.14. RA susceptibility SNP association summary of %ΔSJC in YEAR and ERAS SSA monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with %ΔSJC. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either cohort or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

Table 3.15. SSA-YEAR, ERAS and meta-analysis SNP association with %ΔCRP/ESR from baseline to 6months, adjusted for baseline lnCRP/lnESR

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	-0.76	-6.81-5.29	0.81	188	3.03	-3.80-9.87	0.38	190	0.91	-3.59-5.41	0.69	378	0
rs4272626	**1p13.1**	***CASQ2***	**0.349**	**T**	**1.98**	**-3.81-7.78**	**0.50**	**188**	**6.63**	**0.90-12.4**	**0.024**	**190**	**4.33**	**-0.23-8.88**	**0.06**	**378**	**20.8**
rs2476601	1p13.2	*PTPN22*	0.138	A	-2.46	-10.0-5.12	0.52	188	1.41	-6.30-9.13	0.72	189	0.38	-3.50-4.26	0.86	377	0
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	-2.24	-8.07-3.60	0.45	187	-0.22	-6.48-6.04	0.95	190	-1.30	-5.54-2.94	0.55	377	0
rs12746613	1q23.2	*FCGR2A*	0.119	T	2.87	-4.78-10.5	0.46	188	-3.68	-12.9-5.54	0.43	184	0.11	-6.22-6.45	0.97	372	14
rs10919563	1q31.3	*PTPRC*	0.129	A	2.91	-4.47-10.6	0.45	187	-3.56	-13.0-5.84	0.46	184	0.27	-5.96-6.50	0.93	371	9.7
rs1160542	2q11.2	*AFF3*	0.493	A	-0.29	-5.45-4.87	0.91	188	1.05	-4.73-6.83	0.72	187	0.31	-3.52-4.13	0.88	375	0
rs1980422	2q33.2	*CD28*	0.232	C	-3.15	-9.08-2.79	0.30	188	0.28	-6.67-7.23	0.94	190	-1.6	-5.97-2.77	0.47	378	0
rs10181656	2q32.2	*STAT4*	0.251	G	-3.42	-10.1-3.26	0.31	188	-3.73	-10.3-2.81	0.26	189	-3.57	-8.35-1.22	0.14	377	0
rs3087243	2q33.2	*CTLA4*	0.430	A	1.07	-3.69-5.83	0.66	187	2.48	-3.33-8.29	0.40	189	1.64	-2.02-5.30	0.38	376	0
rs231775	2q33.2	*CTLA4*	0.409	G	-0.14	-7.45-3.17	0.43	187	-2.21	-7.74-3.32	0.43	188	-1.14	-4.94-2.67	0.56	375	0
rs4535211	3p24.3	*PLCL2*	0.464	A	-1.69	-6.92-3.54	0.52	188	0.88	-4.86-6.62	0.76	190	-0.53	-4.37-3.31	0.79	378	0
rs231707	4p16.3	*FAM193A*	0.191	A	1.05	-5.52-7.63	0.75	188	3.73	-4.23-11.7	0.36	190	2.14	-2.90-7.18	0.41	378	0
rs2069778	4q27	*IL2/IL21*	0.155	T	-5.36	-12.2-1.51	0.13	186	3.37	-3.47-10.2	0.33	190	-0.99	-9.54-7.57	0.82	376	68.3
rs6897932	5p13.2	*IL7R*	0.282	C	-1.19	-7.14-4.60	0.69	188	1.05	-2.89-8.00	0.77	189	-0.24	-4.73-4.25	0.92	377	0
rs1043730	6q21	*TRAF3IP2*	0.245	C	-1.35	-7.43-4.73	0.66	184	8.11	1.27-15.0	0.02	190	3.25	-6.02-12.51	0.49	374	75.9
rs548234	6q21	*PRDM1*	0.347	C	0.93	-4.43-6.29	0.73	188	-2.86	-8.90-3.18	0.35	189	-0.74	-4.72-3.24	0.72	377	0
rs13207033	6q23.3	*6q23*	0.262	A	4.61	-0.95-10.2	0.10	188	0.23	-6.31-6.77	0.95	190	2.77	-1.47-7.01	0.20	378	1.3
rs6920220	**6q23.3**	***6q23***	**0.240**	**A**	**-1.92**	**-8.20-4.36**	**0.55**	**188**	**-9.73**	**-16.7--2.77**	**0.006**	**190**	**-5.68**	**-13.33-1.97**	**0.15**	**378**	**63**
rs5029937	6q23.3	*TNFAIP3*	0.055	T	7.09	-4.14-18.3	0.22	188	-0.46	-11.7-10.8	0.94	189	5.44	-0.69-11.56	0.08	377	42.7
rs394581	6q25.3	*TAGAP*	0.294	C	-2.49	-8.45-3.48	0.41	188	2.72	-3.30-8.74	0.37	189	0.10	-5.00-5.20	0.97	377	32
rs42041	7q21.2	*CDK6*	0.259	G	5.53	-0.82-11.9	0.09	188	-0.48	-6.65-5.68	0.88	188	2.48	-3.41-8.37	0.41	376	44.3
rs13277113	8p23.1	*BLK*	0.256	A	-0.07	-6.42-6.28	0.98	186	1.50	-4.71-7.71	0.63	190	0.73	-3.68-5.14	0.75	376	0
rs2812378	9p13.3	*CCL21*	0.362	C	-1.66	-7.80-4.47	0.59	187	-0.62	-7.05-5.81	0.85	189	-1.17	-5.58-3.24	0.60	376	0
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	1.55	-3.97-7.06	0.58	183	3.81	-1.92-9.55	0.19	190	2.63	-1.31-6.58	0.19	373	0
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	1.27	-4.20-6.74	0.65	188	3.02	-2.99-9.02	0.32	190	2.07	-1.95-6.08	0.31	378	0
rs2104286	10p15.1	*IL2RA*	0.279	G	-1.33	7.21-4.54	0.66	187	1.21	-5.48-7.90	0.72	185	-0.23	-4.61-4.16	0.92	372	0
rs11594656	10p15.1	*IL2RA*	0.205	A	-0.96	-7.19-5.26	0.76	185	2.43	-3.96-8.81	0.45	188	0.69	-3.74-5.12	0.70	373	0
rs41295061	10p15.1	*IL2RA*	0.093	A	7.66	-1.25-16.6	0.09	188	0.71	-8.79-10.2	0.88	190	4.38	-2.42-11.2	0.211	378	9.8

SNP	Cytoband	Locus	MAF	Allele	YEAR				ERAS				RE Meta-analysis				
					β	95% CI	p	N	β	95% CI	p	N	β	95% CI	p	N	I^2
rs4750316	10p15.1	*10p15*	0.195	C	-1.44	-7.8-4.91	0.65	187	-2.79	-10.4-4.81	0.47	190	-2.00	-6.85-2.85	0.42	377	0
rs540386	11p12	*TRAF6*	0.134	T	2.28	-6.21-10.8	0.60	188	2.49	-6.62-11.6	0.59	190	2.38	-3.79-8.55	0.45	378	0
rs1678542	12q13.3	*KIF5A/PIP4K2C*	0.412	C	-2.06	-7.39-3.28	0.45	188	4.89	-1.50-11.3	0.13	190	1.19	-5.60-7.98	0.73	378	63.2
rs775241	12q13.3	*DCTN2*	0.119	G	-0.34	-8.04-7.35	0.93	187	1.44	-8.04-10.9	0.77	189	0.37	-5.57-6.30	0.90	376	0
rs7234029	18p11.21	*PTPN2*	0.175	G	0.02	-6.99-7.04	1.00	188	-1.96	-9.67-5.74	0.62	190	-0.88	-6.03-4.28	0.74	378	0
rs763361	18q22.2	*CD226*	0.487	C	2.78	-2.12-7.68	0.27	188	4.15	-1.60-9.89	0.16	190	3.36	-0.35-7.06	0.08	378	0
rs4239702	20q13.12	*CD40*	0.261	T	-0.49	-6.87-5.89	0.88	183	-2.64	-9.40-4.07	0.44	190	-1.53	-6.11-3.08	0.52	373	0
rs3788013	21q22.3	*UBASH3A*	0.443	A	1.14	-4.07-6.35	0.67	186	1.95	-3.71-7.61	0.50	190	1.51	-2.30-5.32	0.44	376	0
rs743777	22q12.3	*IL2RB*	0.327	G	2.46	-2.10-8.75	0.36	187	1.32	-4.60-7.25	0.66	189	1.96	-1.97-5.88	0.33	376	0
rs3218258	22q12.3	*IL2RB*	0.270	A	3.33	-2.83-7.76	0.23	188	2.06	-3.98-8.09	0.50	190	2.76	-1.25-6.77	0.18	378	0
rs229541	**22q12.3**	***C1QTNF6***	**0.461**	**T**	**-5.35**	**-10.8-0.08**	**0.05**	**183**	**-3.80**	**-9.21-1.61**	**0.17**	**190**	**-4.57**	**-8.38--0.77**	**0.019**	**373**	**0**

Table 3.15. RA susceptibility SNP association summary of %ΔCRP/ESR in YEAR and ERAS SSA monotherapy patients. Linear regression estimates and summary presented for 40 RA susceptibility SNPs and their association with %ΔCRP/ESR. The results of the two analyses are presented with pooled effect estimates from a DerSimonian and Laird random effect meta-analysis. Heterogeneity is quantified with the I^2 statistic. SNPs highlighted in bold are those with a p<0.05 in either the results from the YEAR or ERAS cohort or the meta-analysis, whilst those highlighted in red show high heterogeneity (I^2>50%). MAF – minor allele frequency, RE – random effects.

3.3.2 Rheumatoid arthritis susceptibility single nucleotide polymorphisms and radiographic joint damage

3.3.2.1 Association of clinical and demographic variables with baseline radiographic joint damage

A total of 589 eligible YEAR patients had their baseline x-rays scored by the SHS (symptom duration ≤24 months); genotyping data were available on 425 of these patients after QC (Section 3.2.2). Clinical and demographic data were available on up to 425 of these patients (Table 3.16). The mean age of patients was 59 years (s.d. 12.8) and 300 were female (70.6%). RF positivity was recorded in 258/404 (63.9%) and 199/313 were anti-citrullinated peptide antibody (ACPA positive (63.6%). Each clinical variable listed in Table3.16 was tested for association with SHS in the ZINB model. Baseline patient age was the clinical variable with the strongest association, higher age being associated with both increased SHS and lower odds of zero score (Table 3.16). Symptom duration was found to be statistically significantly associated with a decreased probability of zero SHS (OR 0.91 95% CI [0.86-0.98]) so was also included in all ZINB models. Seropositivity for ACPA and/or RF was borderline associated with an increased SHS (IRR 1.64 95% CI [0.99-2.72]) and has been reported as a strong predictor of radiographic joint damage previously [283, 284]. The IRR estimates a 64% increase in joint damage in patients positive for either RF and/or ACPA. Therefore it was included in all further ZINB models of radiographic joint damage. Due to the zero-inflation in the SHS (and subsequently ERN and JSN), standard summary statistics (i.e. median, mean, etc.) are not a useful representation of the data. Summary statistics are therefore presented for those patients with SHS>0 along with the proportion of patients with recorded damage of each type (Table 3.17).

A total of 40 RA susceptibility SNPs were tested for association with SHS and in sub-analyses of ERN and JSN.

Table 3.16. YEAR clinical and demographic variables and ZINB association with SHS							
Variable		Count model		Zero-inflation model		N	p
		IRR	95% CI	OR	95%CI		
Age (years), mean (s.d.)	59.0 (12.8)	1.02	1.00-1.03	0.94	0.92-0.96	425	$3.9x10^{-9}$
Gender (% female)	70.6%	1.30	0.78-2.17	0.60	0.34-1.07	425	0.09
Seropositivity (% positive)	70.7%	1.64	0.99-2.72	0.85	0.47-1.54	413	0.09
RF (% positive)	63.9%	1.40	0.88-2.24	0.90	0.52-1.58	404	0.27
ACPA (% positive)	63.8%	1.46	0.89-2.41	0.84	0.45-1.52	313	0.16
Symptom duration (months), mean (s.d.)	7.4 (4.2)	1.01	0.97-1.05	0.91	0.86-0.98	425	0.017

Table 3.16. Summary statistics of the YEAR cohort analysed for RA susceptibility SNP association with radiographic joint damage. Data are presented as summary statistics for each variable with the model statistics for the count and zero-inflated components. p-values are calculated using a *post hoc* Wald-type test. s.d. – standard deviation, RF – rheumatoid factor, ACPA – anti-citrullinated peptide antibodies, IRR – incidence rate ratio, OR – odds ratio.

Table 3.17. YEAR radiographic joint damage summary			
Joint damage score	Proportion score>0 (n)	Median (range) score>0	n
SHS	28.7% (122)	5 (1-56)	425
ERN	15.1% (64)	3 (1-29)	425
JSN	22.8% (97)	5 (1-42)	425

Table 3.17. Joint damage summary statistics of the YEAR cohort analysed for RA susceptibility SNP association with radiographic joint damage. Summary statistics are presented in a number of ways. The proportion with damage are those with a score>0. The median joint damage for those individuals with score>0 are also shown with the range of values. This illustrates the long-tailed distribution of these data that are not normally distributed (see Chapter 2.3.1).

3.3.2.2 Association of rheumatoid arthritis susceptibility single nucleotide polymorphisms with baseline radiographic joint damage

In total 40 SNPs were tested for association with baseline radiographic joint damage using the ZINB model with the covariates outlined above. The results of this association can be seen in Table 3.18 for the SHS, Table 3.19 for ERN and Table 3.20 for JSN. At a significance threshold for association of $p<0.01$ 4 SNPs were associated with baseline SHS. Of these the minor allele of rs13277113 was also associated with increased ERN, and the minor allele of rs3218258 was associated with both lower ERN and lower JSN. In the ERN analysis the minor allele of an additional SNP at the *C5orf30* locus (rs26232) was associated with a

decreased ERN (IRR 0.43 95% CI [0.22-0.83], p=0.003) in a subset of patients for which this SNP was genotyped (it was not included in the original selection of SNPs for genotyping in this study). This SNP has been associated with joint damage in multiple cohorts, both longitudinal and cross-sectional, and these data are published [275]. Table 3.21 shows a summary of the 3 most strongly associated SNPs with baseline joint damage. There was no association between *HLA-DRB1* SE alleles and baseline joint damage.

Table 3.18. Zero-inflated negative binomial regression of SHS, adjusted for age and symptom duration at baseline and autoantibodies										
SNP	Cytoband	Locus	MAF	Allele	Count model		Zero-inflated model		N	P-value
					IRR	95% CI	OR	95%CI		
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101, *1001	1.17	0.83-1.65	0.89	0.61-1.29	441	0.46
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.246	G	1.04	0.75-1.44	0.98	-0.42-0.38	413	0.96
rs4272626	1p13.1	*CASQ2*	0.355	T	1.00	0.76-1.33	0.99	-0.38-0.37	413	1.00
rs2476601	1p13.2	*PTPN22*	0.141	A	0.94	0.58-1.53	1.17	-0.41-0.74	411	0.81
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.309	A	0.86	0.58-1.28	0.87	-0.54-0.30	411	0.71
rs12746613	1q23.2	*FCGR2A*	0.121	T	0.77	0.48-1.26	0.76	-0.85-0.30	413	0.45
rs10919563	1q31.3	*PTPRC*	0.135	A	0.92	0.56-1.49	0.86	-0.72-0.42	413	0.85
rs1160542	**2q11.2**	***AFF3***	**0.481**	**A**	**1.62**	**1.22-2.17**	**1.42**	**-0.02-0.72**	**413**	**0.0027**
rs10181656	2q32.2	*STAT4*	0.246	G	0.96	0.68-1.37	0.91	-0.52-0.35	413	0.92
rs1980422	2q33.2	*CD28*	0.236	C	0.67	0.45-1.01	0.84	-0.64-0.30	413	0.16
rs3087243	2q33.2	*CTLA4*	0.437	A	1.03	0.77-1.37	1.15	-0.21-0.49	412	0.74
rs231775	2q33.2	*CTLA4*	0.397	G	1.10	0.80-1.52	0.88	-0.48-0.22	413	0.57
rs4535211	3p24.3	*PLCL2*	0.489	A	0.97	0.68-1.39	0.69	-0.74-0.01	412	0.15
rs231707	4p16.3	*FAM193A*	0.193	A	0.80	0.52-1.23	1.02	-0.47-0.51	413	0.58
rs2069778	4q27	*IL2/IL21*	0.165	A	0.95	0.64-1.40	1.05	-0.45-0.56	412	0.93
rs6897932	5p13.2	*IL7R*	0.278	C	0.99	0.71-1.38	1.05	-0.35-0.45	413	0.96
rs1043730	6q21	*TRAF3IP2*	0.240	C	0.97	0.64-1.47	1.13	-0.33-0.56	412	0.85
rs548234	6q21	*PRDM1*	0.345	C	1.35	0.99-1.85	0.96	-0.43-0.34	413	0.14
rs13207033	6q23.3	*6q23*	0.258	A	0.86	0.59-1.25	0.93	-0.51-0.36	413	0.72
rs6920220	6q23.3	*6q23*	0.243	A	0.69	0.45-1.07	0.77	-0.72-0.20	413	0.20
rs5029937	6q23.3	*TNFAIP3*	0.056	T	0.68	0.30-1.53	1.27	-061-1.09	412	0.46
rs394581	6q25.3	*TAGAP*	0.294	C	1.13	0.80-1.59	0.96	-0.46-0.39	413	0.76
rs42041	7q21.2	*CDK6*	0.257	G	1.44	1.02-2.03	1.15	-0.25-0.54	413	0.09
rs13277113	**8p23.1**	***C8orf13-BLK***	**0.254**	**A**	**1.60**	**1.17-2.19**	**0.79**	**-0.65-0.17**	**410**	**0.0031**
rs2812378	9p13.3	*CCL21*	0.362	C	1.25	0.91-1.71	1.31	-0.13-0.67	410	0.20
rs10760130	9q33.1	*TRAF1/C5*	0.464	G	1.37	0.91-2.05	1.06	-0.33-0.45	408	0.32
rs2900180	9q33.1	*TRAF1/C5*	0.372	T	1.47	1.02-2.12	1.02	-0.37-0.42	412	0.12

SNP	Cytoband	Locus	MAF	Allele	Count model		Zero-inflated model		N	P-value
					IRR	95% CI	OR	95%CI		
rs2104286	10p15.1	*IL2RA*	0.275	G	1.25	0.89-1.75	1.07	-0.33-0.46	411	0.43
rs11594656	10p15.1	*IL2RA*	0.208	A	0.95	0.68-1.34	0.66	-0.83-0.02	411	0.17
rs41295061	10p15.1	*IL2RA*	0.09	A	0.82	0.49-1.36	1.04	-0.59-0.66	413	0.72
rs4750316	10p15.1	*10p15*	0.197	C	1.10	0.72-1.67	1.21	-0.27-0.65	413	0.70
rs540386	11p12	*TRAF6*	0.137	T	1.12	0.68-1.85	1.52	-0.14-0.98	413	0.34
rs1678542	**12q13.3**	***KIF5A-PIP4K2C***	**0.399**	**C**	**1.18**	**0.79-1.75**	**0.66**	**-0.80--0.01**	**412**	**0.047**
rs775241	12q13.3	*DCTN2*	0.122	G	0.89	0.56-1.41	0.80	-0.77-0.33	412	0.69
rs7234029	18p11.21	*PTPN2*	0.178	G	1.45	1.01-2.10	0.94	-0.53-0.41	411	0.11
rs763361	18q22.2	*CD226*	0.493	C	0.96	0.73-1.26	1.03	-0.32-0.38	410	0.93
rs4239702	20q13.12	*CD40*	0.258	T	0.84	0.58-1.23	0.90	-0.56-0.35	412	0.65
rs3788013	21q22.3	*UBASH3A*	0.447	A	0.83	0.59-1.18	0.85	-0.54-0.21	412	0.46
rs743777	**22q12.3**	***IL2RB***	**0.327**	**G**	**0.61**	**0.45-0.83**	**0.80**	**-0.61-0.18**	**413**	**0.0065**
rs229541	22q12.3	*IL2RB*	0.455	T	0.85	0.61-1.19	1.05	-0.31-0.42	410	0.57
rs3218258	**22q12.3**	***IL2RB***	**0.275**	**A**	**0.56**	**0.42-0.74**	**0.69**	**-0.77-0.04**	**412**	**0.0002**
rs26232	**5q21.1**	***C5orf30***	**0.314**	**T**	**0.62**	**0.39-0.97**	**1.22**	**-0.27-0.68**	**377**	**0.031**

Table 3.18. Summary of RA susceptibility SNP and *HLA-DRB1* shared epitope allele association with baseline Sharp/van der Heijde Score (SHS) in the YEAR cohort. Effect estimates from the ZINB regression model are presented from both the count and zero-inflation model. The exponential of the count model effect estimates are taken to give the incidence rate ratio (IRR) and the exponent of the zero-inflation model effect estimate is the odds ratio (OR) of zero-inflation. p-values are presented from a Wald-type test for the effect of the SNP on the total model. SNPs associated with baseline SHS p<0.01 are highlighted in bold. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.19. Zero-inflated negative binomial regression of baseline ERN, adjusted for age and symptom duration at baseline and autoantibodies

SNP	Cytoband	Locus	MAF	Allele	Count model		zero-inflated model		N	P-value
					IRR	95% CI	OR	95%CI		
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101,*1001	1.32	0.77-2.25	0.99	0.62-1.57	441	0.55
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.246	G	0.96	0.57-1.62	-0.07	-0.57-0.42	413	0.96
rs4272626	1p13.1	*CASQ2*	0.355	T	1.20	0.80-1.78	0.05	-0.41-0.51	413	0.68
rs2476601	1p13.2	*PTPN22*	0.141	A	1.27	0.60-2.67	0.33	-0.39-1.06	411	0.61
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.309	A	0.81	0.44-1.49	-0.19	-0.74-0.36	411	0.72
rs12746613	1q23.2	*FCGR2A*	0.121	T	1.30	0.59-2.84	0.27	-0.44-0.98	413	0.68
rs10919563	1q31.3	*PTPRC*	0.135	A	0.82	0.41-1.64	-0.43	-1.12-0.26	413	0.47
rs1160542	2q11.2	*AFF3*	0.481	A	1.64	1.04-2.59	0.08	-0.38-0.54	413	0.10
rs10181656	2q32.2	*STAT4*	0.246	G	0.81	0.46-1.44	0.10	0.47-0.67	413	0.63
rs1980422	2q33.2	*CD28*	0.236	C	0.67	0.37-1.20	-0.18	-0.75-0.39	413	0.39
rs3087243	2q33.2	*CTLA4*	0.437	A	1.16	0.76-1.78	0.23	-0.19-0.66	412	0.51
rs231775	2q33.2	*CTLA4*	0.397	G	1.00	0.62-1.61	-0.14	-0.58-0.30	413	0.80
rs4535211	3p24.3	*PLCL2*	0.489	A	1.46	0.78-2.74	0.04	-0.46-0.53	412	0.46
rs231707	4p16.3	*FAM193A*	0.193	A	0.52	0.29-0.93	-0.33	-0.97-0.30	413	0.09
rs2069778	4q27	*IL2*	0.165	A	0.95	0.56-1.60	-0.25	-0.84-0.34	412	0.71
rs6897932	5p13.2	*IL7R*	0.278	C	0.66	0.39-1.12	-0.12	-0.64-0.39	413	0.31
rs1043730	6q21	*TRAF3IP2*	0.240	C	1.19	0.64-2.22	0.35	-0.21-0.90	412	0.47
rs548234	6q21	*PRDM1*	0.345	C	1.49	0.97-2.31	-0.06	-0.53-0.41	413	0.15
rs13207033	6q23.3	*6q23*	0.258	A	0.80	0.50-1.31	-0.16	-0.68-0.37	413	0.64
rs6920220	6q23.3	*6q23*	0.243	A	0.67	0.38-1.17	-0.28	-0.81-0.25	413	0.31
rs5029937	6q23.3	*TNFAIP3*	0.056	T	0.61	0.18-2.15	0.25	-0.84-1.33	412	0.53
rs394581	6q25.3	*TAGAP*	0.294	C	1.03	0.64-1.65	-0.29	-0.81-0.22	413	0.49
rs42041	7q21.2	*CDK6*	0.257	G	1.20	0.74-1.97	-0.07	-0.55-0.41	413	0.67
rs13277113	**8p23.1**	***C8orf13-BLK***	**0.254**	**A**	**2.42**	**1.56-3.76**	**-0.32**	**-0.82-0.17**	**410**	**2.6×10^{-5}**
rs2812378	9p13.3	*CCL21*	0.362	C	1.13	0.69-1.83	0.13	-0.37-0.63	410	0.83
rs10760130	9q33.1	*TRAF1/C5*	0.464	G	1.23	0.68-2.22	-0.25	-0.73-0.23	408	0.30

SNP	Cytoband	Locus	MAF	Allele	Count model		zero-inflated model		N	P-value
					IRR	95% CI	OR	95%CI		
rs2900180	9q33.1	*TRAF1/C5*	0.372	T	1.43	0.85-2.43	-0.24	-0.72-0.24	412	0.14
rs2104286	10p15.1	*IL2RA*	0.275	G	1.19	0.70-2.01	-0.18	-0.66-0.30	411	0.51
rs11594656	10p15.1	*IL2RA*	0.208	A	0.82	0.51-1.32	-0.39	-0.92-0.15	411	0.34
rs41295061	10p15.1	*IL2RA*	0.09	A	0.65	0.27-1.58	0.07	-0.77-0.90	413	0.54
rs4750316	10p15.1	*10p15*	0.197	C	1.04	0.51-2.13	0.14	-0.46-0.74	413	0.90
rs540386	11p12	*TRAF6*	0.137	T	1.42	0.67-3.01	0.43	-0.29-1.14	413	0.41
rs1678542	**12q13.3**	***KIF5A-PIP4K2C***	**0.399**	**C**	**2.29**	**0.96-5.44**	**-0.24**	**-0.80-0.33**	**412**	**0.012**
rs775241	12q13.3	*DCTN2*	0.122	G	1.15	0.56-2.37	0.08	-0.60-0.76	412	0.92
rs7234029	**18p11.21**	***PTPN2***	**0.178**	**G**	**2.05**	**1.19-3.55**	**-0.23**	**-0.79-0.33**	**411**	**0.012**
rs763361	18q22.2	*CD226*	0.493	C	0.99	0.66-1.50	0.08	-0.35-0.51	410	0.92
rs4239702	20q13.12	*CD40*	0.258	T	0.87	0.51-1.46	-0.27	-0.83-0.28	412	0.61
rs3788013	21q22.3	*UBASH3A*	0.447	A	0.86	0.52-1.40	0.09	-0.37-0.54	412	0.70
rs743777	**22q12.3**	***IL2RB***	**0.327**	**G**	**0.49**	**0.30-0.79**	**-0.18**	**-0.70-0.34**	**413**	**0.013**
rs3218258	**22q12.3**	***IL2RB***	**0.275**	**A**	**0.45**	**0.29-0.70**	**-0.25**	**-0.78-0.28**	**412**	**0.0019**
rs229541	22q12.3	*C1QTNF6*	0.455	T	1.02	0.63-1.66	-0.05	-0.51-0.41	410	0.96
rs26232	**5q21.1**	***C5orf30***	**0.314**	**T**	**0.43**	**0.22-0.83**	**0.34**	**-0.33-1.01**	**377**	**0.0027**

Table 3.19. Summary of RA susceptibility SNP and *HLA-DRB1* shared epitope allele association with baseline erosion score (ERN) in the YEAR cohort. Effect estimates from the ZINB regression model are presented from both the count and zero-inflation model. The exponential of the count model effect estimates are taken to give the incidence rate ratio (IRR) and the exponent of the zero-inflation model effect estimate is the odds ratio (OR) of zero-inflation. p-values are presented from a Wald-type test for the effect of the SNP on the total model. SNPs associated with baseline SHS $p<0.01$ are highlighted in bold. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.20. Zero-inflated negative binomial regression of baseline JSN, adjusted for age and symptom duration at baseline and autoantibodies

SNP	Cytoband	Locus	MAF	Allele	Count model		Zero-inflated model		N	P-value
					IRR	95% CI	OR	95% CI		
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101,*1001	1.15	0.83-1.59	0.85	0.57-1.25	443	0.42
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.246	G	1.01	0.75-1.35	-0.06	-0.48-0.35	415	0.95
rs4272626	1p13.1	*CASQ2*	0.355	T	0.94	0.72-1.22	-0.10	-0.48-0.29	415	0.82
rs2476601	1p13.2	*PTPN22*	0.141	A	0.76	0.49-1.19	0.09	-0.52-0.69	413	0.44
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.309	A	1.03	0.70-1.52	0.09	-0.33-0.52	413	0.91
rs12746613	1q23.2	*FCGR2A*	0.121	T	0.75	0.48-1.18	-0.37	-0.95-0.22	415	0.27
rs10919563	1q31.3	*PTPRC*	0.135	A	0.82	0.51-1.32	-0.17	-0.77-0.42	415	0.65
rs1160542	**2q11.2**	***AFF3***	**0.481**	**A**	**1.43**	**1.09-1.88**	**0.25**	**-0.13-0.63**	**415**	**0.025**
rs10181656	2q32.2	*STAT4*	0.246	G	0.96	0.68-1.34	-0.05	-0.51-0.41	415	0.95
rs1980422	2q33.2	*CD28*	0.236	C	0.74	0.50-1.10	-0.03	-0.52-0.46	415	0.32
rs3087243	2q33.2	*CTLA4*	0.437	A	0.93	0.71-1.21	-0.04	-0.40-0.32	414	0.86
rs231775	2q33.2	*CTLA4*	0.397	G	1.15	0.86-1.52	-0.03	-0.39-0.33	415	0.61
rs4535211	3p24.3	*PLCL2*	0.489	A	0.92	0.67-1.27	-0.33	-0.71-0.05	414	0.24
rs231707	4p16.3	*FAM193A*	0.193	A	0.91	0.60-1.36	0.02	-0.49-0.52	415	0.89
rs2069778	4q27	*IL2/IL21*	0.165	A	0.92	0.62-1.35	0.24	-0.30-0.77	414	0.58
rs6897932	5p13.2	*IL7R*	0.278	C	0.97	0.73-1.30	-0.15	-0.56-0.26	415	0.77
rs1043730	6q21	*TRAF3IP2*	0.240	C	0.97	0.65-1.45	0.09	-0.37-0.55	414	0.90
rs548234	6q21	*PRDM1*	0.345	C	1.07	0.79-1.44	-0.24	-0.65-0.17	415	0.43
rs13207033	6q23.3	*6q23*	0.258	A	0.87	0.60-1.26	-0.01	-0.45-0.43	415	0.76
rs6920220	6q23.3	*6q23*	0.243	A	0.84	0.55-1.28	-0.03	-0.48-0.42	415	0.73
rs5029937	6q23.3	*TNFAIP3*	0.056	T	0.71	0.33-1.53	0.21	-0.70-1.12	414	0.54
rs394581	6q25.3	*TAGAP*	0.294	C	1.15	0.82-1.60	0.08	-0.37-0.53	415	0.69
rs42041	**7q21.2**	***CDK6***	**0.257**	**G**	**1.58**	**1.17-2.14**	**0.18**	**-0.23-0.60**	**415**	**0.011**
rs13277113	8p23.1	*C8orf13-BLK*	0.254	A	1.32	1.00-1.76	-0.15	-0.58-0.27	412	0.09
rs2812378	9p13.3	*CCL21*	0.362	C	1.26	0.95-1.69	0.09	-0.32-0.50	412	0.28
rs10760130	9q33.1	*TRAF1/C5*	0.464	G	1.38	0.93-2.03	0.13	-0.27-0.54	410	0.26

SNP	Cytoband	Locus	MAF	Allele	Count model		Zero-inflated model		N	P-value
					IRR	95% CI	OR	95% CI		
rs2900180	9q33.1	*TRAF1/C5*	0.372	T	1.20	0.85-1.71	-0.10	-0.52-0.31	414	0.47
rs2104286	10p15.1	*IL2RA*	0.275	G	1.41	1.04-1.91	0.24	-0.17-0.65	413	0.06
rs11594656	10p15.1	*IL2RA*	0.208	A	1.04	0.75-1.44	-0.17	-0.60-0.25	413	0.69
rs41295061	10p15.1	*IL2RA*	0.09	A	1.03	0.65-1.64	0.12	-0.50-0.75	415	0.92
rs4750316	10p15.1	*10p15*	0.197	C	1.16	0.79-1.70	0.18	-0.29-0.66	415	0.60
rs540386	11p12	*TRAF6*	0.137	T	1.12	0.71-1.79	0.35	-0.23-0.93	415	0.47
rs1678542	12q13.3	*KIF5A-PIP4K2C*	0.399	C	0.95	0.69-1.33	-0.32	-0.71-0.08	414	0.29
rs775241	12q13.3	*DCTN2*	0.122	G	0.80	0.53-1.21	-0.25	-0.81-0.31	414	0.45
rs7234029	**18p11.21**	***PTPN2***	**0.178**	**G**	**1.54**	**1.10-2.15**	**0.20**	**-0.29-0.70**	**413**	**0.038**
rs763361	18q22.2	*CD226*	0.493	C	0.94	0.73-1.21	-0.07	-0.43-0.30	412	0.85
rs4239702	20q13.12	*CD40*	0.258	T	0.85	0.59-1.25	0.08	-0.39-0.55	414	0.64
rs3788013	21q22.3	*UBASH3A*	0.447	A	0.92	0.65-1.29	-0.19	-0.57-0.20	414	0.60
rs743777	**22q12.3**	***IL2RB***	**0.327**	**G**	**0.69**	**0.52-0.91**	**-0.20**	**-0.60-0.20**	**415**	**0.027**
rs3218258	**22q12.3**	***IL2RB***	**0.275**	**A**	**0.64**	**0.49-0.84**	**-0.24**	**-0.63-0.16**	**414**	**0.004**
rs229541	22q12.3	*C1QTNF6*	0.455	T	0.74	0.54-1.00	0.04	-0.34-0.41	412	0.13
rs26232	5q21.1	*C5orf30*	0.314	T	0.83	0.55-1.27	0.24	-0.24-0.72	379	0.35

Table 3.20. Summary of RA susceptibility SNP and *HLA-DRB1* shared epitope allele association with joint space narrowing score (JSN) in the YEAR cohort. Effect estimates from the ZINB regression model are presented from both the count and zero-inflation model. The exponential of the count model effect estimates are taken to give the incidence rate ratio (IRR) and the exponent of the zero-inflation model effect estimate is the odds ratio (OR) of zero-inflation. p-values are presented from a Wald-type test for the effect of the SNP on the total model. SNPs associated with baseline SHS $p<0.01$ are highlighted in bold. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.21. Summary of zero-inflated negative binomial regression results of baseline joint damage scores										
SNP	Cytoband	Locus	Allele	Outcome	Count model		Zero-inflation model		N	p-value
					IRR	95% CI	OR	95% CI		
rs1160542	2q11.2	*AFF3*	A	SHS	1.62	1.22-2.17	1.42	0.98-2.05	413	$2.7x10^{-3}$
				ERN	1.64	1.04-2.59	1.08	0.68-1.72	413	0.110
				JSN	1.43	1.09-1.88	1.28	0.88-1.88	415	0.026
rs13277113	8p23.1	*C8orf13-BLK*	A	SHS	1.60	1.17-2.19	0.79	0.52-1.19	410	$3.0x10^{-3}$
				ERN	2.42	1.56-3.76	0.73	0.44-1.19	410	$1.0x10^{-4}$
				JSN	1.32	1.00-1.76	0.86	0.56-1.31	412	0.089
rs3218258	22q12.3	*IL2RB*	A	SHS	0.56	0.42-0.74	0.69	0.46-1.04	412	$2.0x10^{-4}$
				ERN	0.45	0.29-0.70	0.78	0.46-1.32	412	$4.5x10^{-3}$
				JSN	0.64	0.49-0.84	0.79	0.53-1.17	414	$6.1x10^{-3}$

Table 3.21. Summary of RA susceptibility SNP associations with baseline radiographic joint damage in the YEAR cohort. Effect estimates from the ZINB regression model are presented from both the count and zero-inflation model for each of the three joint damage measures. The exponential of the count model effect estimates are taken to give the incidence rate ratio (IRR) and the exponent of the zero-inflation model effect estimate is the odds ratio (OR) of zero-inflation. p-values are presented from a Wald-type test for the effect of the SNP on the total model.

3.3.2.3 Replication analysis of baseline joint damage single nucleotide polymorphism associations

In order to further test the association of some of these SNPs with joint damage a replication data set (Leiden EAC, n=521) of baseline joint damage was kindly provided by Prof Annette van der Helm-van Mil (Leiden Medical Centre, Netherlands). The distribution of these data appeared a close fit to a negative binomial distribution, with little evidence of zero-inflation (SHS - 11.6% zeroes). A negative binomial regression model was used, adjusted for symptom duration and autoantibodies to test the effect of 3 SNPs in Table 3.21 and a fourth nominally associated SNP with baseline ERN at the *KIF5A-PIP4K2C* locus (rs1678542). No evidence of association was found at 3 of the loci however, the minor allele of rs1678542 displayed a nominal association with increased baseline SHS in this data set (Table 3.22). This association was seen at its strongest in the JSN score analysis (SHS p=0.017 vs. JSN p=0.003). This is in contrast to the effect observed in the YEAR data where the effect was strongest in the ERN score analysis (Table 3.19), though the direction of effect is consistent between the data sets.

Table 3.22. Negative binomial regression of log additive SNP model on joint damage, adjusted for symptom duration and autoantibodies														
SNP	Locus	Cytoband	SHS				ERN				JSN			
			IRR	95% CI	p-value	n	IRR	95% CI	p-value	n	IRR	95% CI	p-value	n
rs1160542	*AFF3*	2q11.2	1.01	0.88-1.15	0.903	521	1.07	0.92-1.24	0.367	521	0.90	0.72-1.12	0.350	521
rs13277113	*C8orf13-BLK*	8p23.1	1.08	0.93-1.27	0.314	521	1.12	0.95-1.33	0.174	521	1.01	0.78-1.30	0.949	521
rs1678542	***KIF5A-PIP4K2C***	**12q13.3**	**1.22**	**1.05-1.41**	**0.008**	**521**	**1.12**	**0.96-1.32**	**0.157**	**521**	**1.44**	**1.13-1.82**	**0.003**	**521**
rs3218258	*IL2RB*	22q12.3	0.98	0.83-1.15	0.776	521	0.91	0.77-1.09	0.317	521	1.12	0.85-1.49	0.413	521

Table 3.22. Replication of baseline joint damage associated SNPs in an independent cohort from Leiden Medical Centre. SNPs tested for replication in the Leiden EAC cohort are presented from the negative binomial regression model. Effect estimates from the regression model are presented; the exponential of the effect estimates are taken to give the incidence rate ratio (IRR). A single SNP was also associated with SHS and JSN in this cohort with $p<0.01$. SHS – Sharp/van der Heijde score, ERN – erosion score, JSN – joint space narrowing score, 95% CI – 95% confidence intervals.

3.3.2.4 Association of rheumatoid arthritis susceptibility loci with joint damage progression

Joint damage progression was measured as a binary variable at two years follow-up, defined as having a SHS, ERN or JSN score higher at two years than at baseline. Therefore this included patients that accrued new damage, as well as those whose existing damage worsened. Of the clinical and demographic variables tested the strongest predictors were autoantibody positivity and symptom duration (Table 3.23). These effects were constant across both the ERN and JSN progression, though the increased risk of erosion progression with increasing symptom duration was borderline significant compared to the association with JSN progression.

The effect of the RA susceptibility SNPs on joint damage progression was tested using a modified Poisson regression model with robust error variance to estimate the risk ratio. Data are presented adjusted for presence of damage at baseline, and as a multivariate regression model adjusted for autoantibodies and symptom duration. No SNPs were associated with 2 year SHS progression at a p-value threshold <0.01 in the unadjusted analysis (Table 3.24). In the ERN progression multivariate analysis 2 SNPs were nominally associated (p=0.01, Table 3.25). The minor alleles of rs231707 (FAM193A) and rs2900180 (*TRAF1/C5*) were both nominally associated with a lower risk of erosion progression in both the adjusted and unadjusted analyses. No SNPs were associated with either an increased or decreased risk of JSN progression (Table 3.26). The effects of these SNPs across SHS, ERN and JSN progression are summarised in Table 3.27.

HLA-DRB1 SE alleles were associated with an increased risk of SHS progression in the unadjusted analysis which was not observed in the multivariate model adjusted for autoantibodies and symptom duration (Table 3.24). The multivariate model was adjusted for autoantibodies which includes ACPA; SE susceptibility to RA has been described as specific to ACPA positive disease, which suggests the effect of *HLA-DRB1* SE on risk of joint

damage progression is probably due to the presence of ACPA rather than an effect from the SE alleles *per se*. This association in the unadjusted model was also observed with risk of JSN progression (Table 3.26; p=0.004), but was nominal in the ERN progression analysis (Table 3.25; p=0.03).

Table 3.23. YEAR clinical and demographic variable association with 2 year joint damage progression												
Variable	SHS				ERN				JSN			
	RR	95% CI	P	N	RR	95% CI	P	N	RR	95% CI	P	N
Age	1.00	0.99-1.02	0.786	269	1.00	0.98-1.02	0.984	274	1.00	0.98-1.02	0.893	276
Gender	1.03	0.67-1.60	0.884	276	1.24	0.68-2.27	0.480	274	1.31	0.71-2.39	0.387	276
Seropositivity	2.51	1.37-4.60	0.003	269	3.55	1.31-9.62	0.013	270	3.21	1.51-6.82	0.002	272
RF	1.51	0.97-2.33	0.066	262	1.43	0.78-2.63	0.251	263	1.90	1.09-3.30	0.024	265
ACPA	3.05	1.61-5.80	0.001	216	4.92	1.56-15.5	0.006	217	3.43	1.63-7.20	0.001	219
Symptom duration	1.04	1.00-1.08	0.043	269	1.05	1.00-1.10	0.050	274	1.07	1.02-1.13	0.003	276

Table 3.23. Association of clinical and demographic variables with joint damage progression in the YEAR cohort. Modified Poisson regression model estimates are presented as risk ratios (RR) for several clinical and demographic variables for the presence of radiographic progression at 2years defined as a SHS/ERN/JSN score at 2 years higher than the baseline measurement.

Table 3.24. modified Poisson regression of SHS progression presence/absence, 0-24months											
SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101,*1001	1.23	0.96-1.58	0.11	289	1.38	1.08-1.76	0.009
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	1.01	0.78-1.32	0.93	269	1.04	0.79-1.38	0.78
rs4272626	1p13.1	*CASQ2*	0.349	T	1.13	0.90-1.42	0.31	269	1.22	0.97-1.55	0.09
rs2476601	1p13.2	*PTPN22*	0.138	A	1.05	0.70-1.57	0.83	269	1.11	0.74-1.67	0.62
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	0.93	0.69-1.24	0.61	268	0.89	0.67-1.19	0.44
rs12746613	1q23.2	*FCGR2A*	0.119	T	0.66	0.43-1.01	0.06	269	0.73	0.48-1.11	0.15
rs10919563	1q31.3	*PTPRC*	0.129	A	0.65	0.43-0.99	0.045	269	0.75	0.48-1.16	0.19
rs1160542	2q11.2	*AFF3*	0.493	A	1.15	0.86-1.53	0.35	269	1.09	0.81-1.46	0.56
rs10181656	2q32.2	*STAT4*	0.251	G	1.11	0.84-1.47	0.47	269	1.11	0.83-1.49	0.46
rs1980422	2q33.2	*CD28*	0.232	C	0.91	0.65-1.26	0.55	269	1.00	0.70-1.41	0.99
rs3087243	2q33.2	*CTLA4*	0.430	A	0.89	0.70-1.13	0.34	269	0.92	0.71-1.18	0.49
rs231775	2q33.2	*CTLA4*	0.409	G	1.11	0.84-1.45	0.46	269	1.04	0.79-1.37	0.76
rs4535211	3p24.3	*PLCL2*	0.464	A	0.94	0.73-1.21	0.63	269	0.96	0.75-1.25	0.78
rs231707	4p16.3	*FAM193A*	0.191	A	0.83	0.57-1.21	0.33	269	0.84	0.57-1.25	0.40
rs2069778	4q27	*IL2/IL21*	0.155	T	1.33	0.92-1.93	0.13	268	1.29	0.88-1.89	0.20
rs6897932	5p13.2	*IL7R*	0.282	C	0.87	0.63-1.20	0.41	269	0.82	0.59-1.13	0.23
rs1043730	6q21	*TRAF3IP2*	0.245	C	1.28	0.91-1.80	0.16	268	1.38	0.98-1.95	0.07
rs548234	6q21	*PRDM1*	0.347	C	0.85	0.64-1.13	0.25	269	0.81	0.61-1.08	0.15
rs13207033	6q23.3	*6q23*	0.262	A	0.96	0.71-1.31	0.81	269	0.98	0.73-1.32	0.90
rs6920220	6q23.3	*6q23*	0.240	A	0.87	0.63-1.21	0.40	269	0.88	0.63-1.23	0.45
rs5029937	6q23.3	*TNFAIP3*	0.055	T	0.60	0.29-1.27	0.18	268	0.62	0.30-1.27	0.19
rs394581	6q25.3	*TAGAP*	0.294	C	0.92	0.68-1.25	0.60	269	0.86	0.63-1.18	0.36
rs42041	7q21.2	*CDK6*	0.259	G	1.09	0.85-1.41	0.50	269	1.11	0.85-1.46	0.45
rs13277113	8p23.1	*C8orf13-BLK*	0.256	A	0.79	0.58-1.06	0.11	267	0.81	0.60-1.09	0.16
rs2812378	9p13.3	*CCL21*	0.362	C	1.11	0.86-1.43	0.42	267	1.08	0.84-1.40	0.54
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.81	0.62-1.06	0.13	265	0.82	0.62-1.09	0.17
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.79	0.61-1.03	0.08	268	0.80	0.61-1.07	0.13
rs2104286	10p15.1	*IL2RA*	0.279	G	0.85	0.64-1.12	0.25	268	0.80	0.60-1.08	0.15

SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
rs11594656	10p15.1	*IL2RA*	0.205	A	0.98	0.74-1.31	0.92	268	0.99	0.73-1.33	0.94
rs41295061	10p15.1	*IL2RA*	0.093	A	0.86	0.51-1.45	0.57	269	0.81	0.47-1.39	0.44
rs4750316	10p15.1	*10p15*	0.195	C	1.15	0.85-1.57	0.37	269	1.26	0.93-1.70	0.13
rs540386	11p12	*TRAF6*	0.134	T	0.95	0.64-1.41	0.80	269	0.92	0.62-1.37	0.69
rs1678542	12q13.3	*KIF5A-PIP4K2C*	0.412	C	1.02	0.76-1.38	0.88	268	1.01	0.76-1.36	0.93
rs775241	12q13.3	*DCTN2*	0.119	G	0.87	0.58-1.30	0.50	269	0.94	0.63-1.41	0.77
rs7234029	18p11.21	*PTPN2*	0.175	G	0.94	0.66-1.35	0.75	268	0.93	0.64-1.35	0.69
rs763361	18q22.2	*CD226*	0.487	C	1.08	0.85-1.38	0.51	266	1.07	0.84-1.37	0.60
rs4239702	20q13.12	*CD40*	0.261	T	1.10	0.81-1.50	0.53	268	1.03	0.76-1.41	0.83
rs3788013	21q22.3	*UBASH3A*	0.443	A	0.98	0.74-1.30	0.90	268	0.97	0.73-1.30	0.86
rs743777	22q12.3	*IL2RB*	0.327	G	1.09	0.79-1.49	0.60	269	1.09	0.80-1.47	0.58
rs3218258	22q12.3	*IL2RB*	0.270	A	0.94	0.68-1.31	0.71	268	0.90	0.66-1.24	0.53
rs229541	22q12.3	*C1QTNF6*	0.461	T	1.10	0.85-1.44	0.46	266	1.11	0.86-1.44	0.41
rs26232	5q21.1	*C5orf30*	0.314	T	1.18	0.81-1.71	0.38	256	1.17	0.81-1.68	0.40

Table 3.24. Association of RA susceptibility SNPs and *HLA-DRB1* shared epitope alleles with progression of Sharp/van der Heijde score in the YEAR cohort. Modified Poisson regression model estimates are presented as risk ratios (RR) for RA susceptibility SNPs for the presence of radiographic progression at 2years defined as a SHS score at 2 years higher than the baseline measurement. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.25. modified Poisson regression of ERN progression presence/absence, 0-24months											
SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101,*1001	1.25	0.88-1.79	0.214	282	1.45	1.04-2.03	0.03
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	1.14	0.81-1.62	0.44	270	1.18	0.82-1.69	0.38
rs4272626	1p13.1	*CASQ2*	0.349	T	1.25	0.93-1.67	0.13	270	1.34	1.00-1.79	0.049
rs2476601	1p13.2	*PTPN22*	0.138	A	0.13	0.70-1.80	0.62	270	1.18	0.74-1.89	0.49
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	0.81	0.54-1.21	0.31	269	0.80	0.54-1.17	0.25
rs12746613	1q23.2	*FCGR2A*	0.119	T	0.49	0.27-0.89	0.019	270	0.56	0.30-1.02	0.059
rs10919563	1q31.3	*PTPRC*	0.129	A	0.72	0.45-1.15	0.17	270	0.80	0.48-1.33	0.39
rs1160542	2q11.2	*AFF3*	0.493	A	1.27	0.90-1.80	0.19	270	1.15	0.81-1.64	0.43
rs10181656	2q32.2	*STAT4*	0.251	G	1.30	0.92-1.84	0.14	270	1.30	0.92-1.84	0.14
rs1980422	2q33.2	*CD28*	0.232	C	0.90	0.60-1.36	0.62	270	1.00	0.64-1.54	0.99
rs3087243	2q33.2	*CTLA4*	0.430	A	0.94	0.69-1.27	0.68	270	0.96	0.69-1.32	0.79
rs231775	2q33.2	*CTLA4*	0.409	G	1.12	0.80-1.57	0.50	270	1.09	0.77-1.54	0.63
rs4535211	3p24.3	*PLCL2*	0.464	A	1.08	0.76-1.53	0.66	270	1.13	0.80-1.59	0.48
rs231707	4p16.3	*FAM193A*	0.191	A	0.44	0.23-0.82	0.010	270	0.43	0.23-082	0.011
rs2069778	4q27	*IL2/IL21*	0.155	T	0.98	0.60-1.62	0.95	269	0.91	0.55-1.49	0.70
rs6897932	5p13.2	*IL7R*	0.282	C	0.94	0.61-1.47	0.80	270	0.90	0.58-1.41	0.66
rs1043730	6q21	*TRAF3IP2*	0.245	C	0.93	0.58-1.49	0.77	269	1.07	0.67-1.73	0.77
rs548234	6q21	*PRDM1*	0.347	C	1.00	0.71-1.42	0.99	270	0.96	0.68-1.37	0.84
rs13207033	6q23.3	*6q23*	0.262	A	1.08	0.74-1.56	0.70	270	1.11	0.77-1.62	0.57
rs6920220	6q23.3	*6q23*	0.240	A	0.83	0.56-1.22	0.34	270	0.83	0.56-1.25	0.39
rs5029937	6q23.3	*TNFAIP3*	0.055	T	0.55	0.21-1.43	0.22	269	0.56	0.21-1.51	0.25
rs394581	6q25.3	*TAGAP*	0.294	C	0.78	0.52-1.18	0.25	270	0.72	0.48-1.08	0.11
rs42041	7q21.2	*CDK6*	0.259	G	1.18	0.86-1.62	0.30	270	1.26	0.90-1.76	0.19
rs13277113	8p23.1	*C8orf13-BLK*	0.256	A	0.87	0.60-1.26	0.46	268	0.84	0.57-1.23	0.36
rs2812378	9p13.3	*CCL21*	0.362	C	1.09	0.76-1.58	0.64	268	1.09	0.75-1.56	0.66
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.79	0.55-1.15	0.22	266	0.78	0.52-1.16	0.21
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.61	0.42-0.89	0.010	269	0.60	0.40-0.89	0.012
rs2104286	10p15.1	*IL2RA*	0.279	G	0.77	0.54-1.11	0.17	269	0.72	0.49-1.04	0.08

SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
rs11594656	10p15.1	*IL2RA*	0.205	A	0.81	0.53-1.24	0.33	269	0.81	0.52-1.26	0.35
rs41295061	10p15.1	*IL2RA*	0.093	A	0.96	0.48-1.91	0.90	270	0.88	0.44-1.77	0.72
rs4750316	10p15.1	*10p15*	0.195	C	1.03	0.67-1.59	0.89	270	1.19	0.77-1.84	0.43
rs540386	11p12	*TRAF6*	0.134	T	1.13	0.70-1.83	0.61	270	1.08	0.66-1.76	0.76
rs1678542	12q13.3	*KIF5A-PIP4K2C*	0.412	C	1.01	0.71-1.44	0.95	269	0.99	0.70-1.40	0.95
rs775241	12q13.3	*DCTN2*	0.119	G	0.87	0.48-1.56	0.63	270	0.97	0.54-0.73	0.91
rs7234029	18p11.21	*PTPN2*	0.175	G	0.97	0.60-1.58	0.91	269	0.91	0.57-1.46	0.70
rs763361	18q22.2	*CD226*	0.487	C	0.97	0.70-1.33	0.83	267	0.95	0.68-1.33	0.77
rs4239702	20q13.12	*CD40*	0.261	T	1.10	0.75-1.60	0.63	269	0.99	0.66-1.50	0.98
rs3788013	21q22.3	*UBASH3A*	0.443	A	0.76	0.53-1.10	0.14	269	0.77	0.52-1.14	0.19
rs743777	22q12.3	*IL2RB*	0.327	G	0.90	0.60-1.35	0.60	270	0.92	0.62-1.36	0.68
rs3218258	22q12.3	*IL2RB*	0.270	A	0.83	0.54-1.26	0.38	269	0.83	0.55-1.23	0.35
rs229541	22q12.3	*C1QTNF6*	0.461	T	0.83	0.59-1.17	0.28	267	0.84	0.60-1.17	0.31
rs26232	5q21.1	*C5orf30*	0.314	T	0.89	0.55-1.46	0.66	257	0.92	0.56-1.53	0.75

Table 3.25. Association of RA susceptibility SNPs and *HLA-DRB1* shared epitope alleles with progression of erosion score in the YEAR cohort. Modified Poisson regression model estimates are presented as risk ratios (RR) for RA susceptibility SNPs for the presence of radiographic progression at 2years defined as an ERN score at 2 years higher than the baseline measurement. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.26. modified Poisson regression of JSN progression presence/absence, 0-24months											
SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
SE	6p21.32	*HLA-DRB1*	0.46	*0401,*0404, *0101,*1001	1.30	0.97-1.75	0.079	285	1.54	1.15-2.06	0.004
rs11586238	1p13.1	*CD2/CD58/IGSF2*	0.241	G	0.95	0.70-1.29	0.73	272	0.99	0.71-1.39	0.97
rs4272626	1p13.1	*CASQ2*	0.349	T	1.00	0.76-1.32	0.99	272	1.12	0.84-1.48	0.44
rs2476601	1p13.2	*PTPN22*	0.138	A	1.03	0.50-2.10	0.94	272	0.99	0.50-1.98	0.98
rs10910099	1p36.2	*MMEL1/TNFRSF14*	0.319	A	1.02	0.73-1.42	0.91	271	1.01	0.73-1.40	0.95
rs12746613	1q23.2	*FCGR2A*	0.119	T	0.80	0.50-1.28	0.35	272	0.93	0.59-1.46	0.76
rs10919563	1q31.3	*PTPRC*	0.129	A	0.59	0.35-0.98	0.04	272	0.74	0.43-1.26	0.26
rs1160542	2q11.2	*AFF3*	0.493	A	0.97	0.68-1.39	0.88	272	0.97	0.68-1.37	0.85
rs10181656	2q32.2	*STAT4*	0.251	G	1.08	0.75-1.56	0.69	272	1.06	0.72-1.55	0.77
rs1980422	2q33.2	*CD28*	0.232	C	0.96	0.64-1.44	0.85	272	1.07	0.69-1.65	0.78
rs3087243	2q33.2	*CTLA4*	0.430	A	0.84	0.63-1.13	0.25	272	0.85	0.63-1.16	0.31
rs231775	2q33.2	*CTLA4*	0.409	G	1.06	0.77-1.46	0.72	272	1.02	0.75-1.39	0.92
rs4535211	3p24.3	*PLCL2*	0.464	A	0.94	0.68-1.30	0.72	272	1.00	0.73-1.36	0.98
rs231707	4p16.3	*FAM193A*	0.191	A	1.00	0.68-1.48	0.99	272	1.05	0.69-1.60	0.83
rs2069778	4q27	*IL2/IL21*	0.155	T	1.64	1.05-2.55	0.029	271	1.50	0.96-2.34	0.074
rs6897932	5p13.2	*IL7R*	0.282	C	0.74	0.51-1.07	0.11	272	0.70	0.48-1.03	0.07
rs1043730	6q21	*TRAF3IP2*	0.245	C	1.32	0.86-2.05	0.21	271	1.50	0.99-2.27	0.05
rs548234	6q21	*PRDM1*	0.347	C	0.84	0.60-1.19	0.33	272	0.81	0.58-1.15	0.24
rs13207033	6q23.3	*6q23*	0.262	A	0.99	0.66-1.49	0.96	272	0.99	0.67-1.47	0.97
rs6920220	6q23.3	*6q23*	0.240	A	0.88	0.56-1.37	0.57	272	0.84	0.53-1.35	0.48
rs5029937	6q23.3	*TNFAIP3*	0.055	T	0.61	0.27-1.36	0.23	271	0.63	0.29-1.39	0.26
rs394581	6q25.3	*TAGAP*	0.294	C	1.13	0.78-1.63	0.52	272	1.08	0.75-1.55	0.67
rs42041	7q21.2	*CDK6*	0.259	G	1.00	0.70-1.41	0.98	272	1.05	0.73-1.52	0.79
rs13277113	8p23.1	*C8orf13-BLK*	0.256	A	0.95	0.67-1.35	0.78	270	1.02	0.73-1.44	0.90
rs2812378	9p13.3	*CCL21*	0.362	C	1.08	0.81-1.45	0.60	270	1.06	0.77-1.46	0.73
rs10760130	9q33.1	*TRAF1/C5*	0.466	G	0.90	0.66-1.21	0.47	268	0.91	0.65-1.27	0.57
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	0.91	0.69-1.22	0.54	271	0.96	0.69-1.34	0.81
rs2104286	10p15.1	*IL2RA*	0.279	G	1.07	0.77-1.49	0.67	271	0.99	0.71-1.39	0.98

SNP	Cytoband	Locus	MAF	Allele	Adjusted			N	Unadjusted		
					RR	95% CI	p		RR	95% CI	p
rs11594656	10p15.1	*IL2RA*	0.205	A	0.95	0.65-1.38	0.79	271	0.96	0.65-1.42	0.83
rs41295061	10p15.1	*IL2RA*	0.093	A	1.02	0.57-1.84	0.94	272	0.92	0.50-1.68	0.78
rs4750316	10p15.1	*10p15*	0.195	C	1.16	0.81-1.67	0.42	272	1.27	0.88-1.83	0.20
rs540386	11p12	*TRAF6*	0.134	T	0.90	0.56-1.43	0.65	272	0.93	0.58-1.50	0.78
rs1678542	12q13.3	*KIF5A-PIP4K2C*	0.412	C	1.09	0.74-1.60	0.66	271	1.10	0.76-1.59	0.62
rs775241	12q13.3	*DCTN2*	0.119	G	0.99	0.64-1.53	0.95	272	1.09	0.70-1.70	0.71
rs7234029	18p11.21	*PTPN2*	0.175	G	1.00	0.63-1.60	0.99	271	0.97	0.60-1.59	0.92
rs763361	18q22.2	*CD226*	0.487	C	0.93	0.70-1.23	0.61	269	0.91	0.68-1.21	0.53
rs4239702	20q13.12	*CD40*	0.261	T	1.08	0.75-1.56	0.66	271	1.00	0.68-1.46	0.99
rs3788013	21q22.3	*UBASH3A*	0.443	A	1.07	0.76-1.51	0.69	271	1.09	0.77-1.55	0.64
rs743777	22q12.3	*IL2RB*	0.327	G	1.11	0.77-1.61	0.58	272	1.09	0.76-1.55	0.64
rs3218258	22q12.3	*IL2RB*	0.270	A	1.08	0.73-1.60	0.68	271	0.97	0.67-1.41	0.89
rs229541	22q12.3	*C1QTNF6*	0.461	T	1.21	0.88-1.66	0.23	269	1.22	0.91-1.65	0.19
rs26232	5q21.1	*C5orf30*	0.314	T	1.44	0.93-2.22	0.10	259	1.32	0.87-2.00	0.19

Table 3.26. Association of RA susceptibility SNPs and *HLA-DRB1* shared epitope alleles with progression of joint space narrowing score in the YEAR cohort. Modified Poisson regression model estimates are presented as risk ratios (RR) for RA susceptibility SNPs for the presence of radiographic progression at 2years defined as a JSN score at 2 years higher than the baseline measurement. MAF – minor allele frequency, 95% CI – 95% confidence intervals

Table 3.27. Summary of modified Poisson regression adjusted analysis of joint damage progression 0-24months									
SNP	Cytoband	Locus	MAF	Allele	Outcome	RR	95% CI	P-value	N
rs12746613	1q23.2	*FCGR2A*	0.119	T	SHS	0.66	0.43-1.01	0.057	269
					ERN	0.49	0.27-0.89	0.019	270
					JSN	0.80	0.50-1.28	0.352	272
rs10919563	1q31.3	*PTPRC*	0.129	A	SHS	0.65	0.43-0.99	0.045	269
					ERN	0.72	0.45-1.15	0.169	270
					JSN	0.59	0.35-0.98	0.040	272
rs2069778	4q27	*IL2/IL21*	0.155	T	SHS	1.33	0.92-1.93	0.133	268
					ERN	0.98	0.60-1.62	0.951	269
					JSN	1.64	1.05-2.55	0.029	271
rs231707	4p16.3	*FAM193A*	0.191	A	SHS	0.83	0.57-1.21	0.327	269
					ERN	0.44	0.23-0.82	0.010	270
					JSN	1.00	0.68-1.48	0.993	272
rs2900180	9q33.1	*TRAF1/C5*	0.374	T	SHS	0.79	0.61-1.03	0.081	268
					ERN	0.61	0.42-0.89	0.010	269
					JSN	0.91	0.69-1.22	0.543	271

Table 3.27. Association summary of RA susceptibility SNPs with joint damage progression in the YEAR cohort. Modified Poisson regression model estimates are presented as risk ratios (RR) for RA susceptibility SNPs associated with SHS, ERN or JSN progression at 2years defined as a SHS/ERN/JSN score at 2 years higher than the baseline measurement. Results are highlighted in bold to illustrate the specificity of associations for a particular type of joint damage. MAF – minor allele frequency, 95% CI – 95% confidence intervals

3.3.3 Joint damage, disease modifying anti-rheumatic drug treatment response and rheumatoid arthritis susceptibility loci

3.3.3.1 Correlation between disease activity and other severity measures

The relationship between disease activity, joint damage and cumulative inflammatory burden is not straightforward [285, 286]. This relationship is complicated by treatment regime, particularly whether treatment is aggressive or if there is delay before initial DMARD therapy [287]. An exploratory analysis of the relationship between clinical, disease activity and joint damage measures was undertaken in the YEAR data set. Pair-wise Spearman rank correlation coefficients were calculated for each pair-wise comparison (Table 3.28); correlations were calculated separately for baseline measures and 1 year follow-up. Increased age is correlated with lower baseline symptom duration, but a higher baseline CRP and more joint space narrowing. Baseline CRP is negatively correlated with both symptom duration and female gender, but positively correlated with SJC28. The moderate correlation between SJC28 and CRP is expected as they are both components of the DAS28-CRP. CRP is also positively associated with JSN, but not ERN, suggesting the development of this type of joint damage before first DMARD treatment is not directly driven by the cumulative inflammatory burden up to that point. Baseline ERN is weakly correlated with symptom duration and RF and ACPA ($\rho \sim 0.10$). Figure 3.2A shows the relationship between these baseline variables. Table 3.29 shows the same correlations with 1 year measures in order to test whether the relationships in Figure 3.2A are maintained over time. The relationship between 1 year follow-up measures is shown in Figure 3.2B. The AUC of the SJC28, CRP and HAQ-DI were used for this correlation analysis as they reflect changes over time in response to treatment. The SJC28 is correlated with both ERN and JSN at 1 year follow-up; however, whilst it is also correlated with the AUC of CRP, CRP itself is not correlated with either joint damage measure. ACPA is likewise

strongly correlated with erosion score at 1 year and the AUC of both CRP and SJC28, despite the lack of correlation between these inflammatory and disease activity markers and ERN. This lack of correlation between CRP, SJC28 and ERN indicate dissociation between disease activity and inflammation and accrued joint damage after the first year of DMARD treatment. Correlations that existed in the baseline data (Figure 3.2A) are altered following the first year of treatment. For instance the baseline ERN score was correlated with autoantibodies, JSN and symptom duration, however the ERN score after 1 year has formed a correlation with SJC28 (Figure 3.2B). Likewise symptom duration is no longer correlated with SJC28, CRP and ERN after 1 year.

Table 3.28. Correlation of baseline clinical, inflammatory and joint damage measures, n=525											
	Age	Symptom duration	Gender	SJC28	CRP	RF	ACPA*	HAQ-DI	SHS	ERN	JSN
Age	0										
Symptom duration	-0.24	0									
Gender	-0.03	0.05	0								
SJC28	0.08	-0.19	-0.12	0							
CRP	0.24	-0.16	-0.11	0.33	0						
RF	-0.08	-0.02	0.06	-0.05	0.03	0					
ACPA*	-0.10	0.04	0.03	-0.02	0.02	0.59	0				
HAQ-DI	0.05	-0.12	0.12	0.30	0.23	-0.01	-0.04	0			
SHS	0.31	0.00	0.05	0.07	0.12	-0.01	0.02	0.07	0		
ERN	0.06	0.12	0.04	0.02	0.06	0.10	0.11	-0.01	0.69	0	
JSN	0.38	-0.04	0.06	0.07	0.11	-0.05	-0.04	0.09	0.87	0.40	0

Table 3.28. Correlation of clinical, inflammatory and joint damage measures at baseline. Spearman correlation coefficients are shown with $|\rho|>0.10$ highlighted in green as significant correlations ($p<0.05$).*ACPA n=332

Table 3.29. Correlation of 1 year follow-up clinical, inflammatory and joint damage measures, n=300											
	Age	Symptom duration	Gender	SJC28	CRP	RF	ACPA*	HAQ-DI	SHS	ERN	JSN
Age	0										
Symptom duration	-0.26	0									
Gender	0.01	0.08	0								
SJC28	0.02	0.05	0.16	0							
CRP	0.02	-0.07	-0.05	0.22	0						
RF	-0.06	-0.03	0.03	0.08	0.23	0					
ACPA*	0.00	-0.06	0.04	0.20	0.23	0.59	0				
HAQ-DI	0.05	-0.03	0.17	0.44	0.24	-0.05	0.00	0			
SHS	0.30	-0.05	0.08	0.26	0.05	0.00	0.14	0.15	0		
ERN	0.04	0.07	0.04	0.20	0.05	0.13	0.19	0.07	0.66	0	
JSN	0.35	-0.06	0.08	0.26	0.05	-0.02	0.10	0.15	0.92	0.42	0

Table 3.29. Correlation of clinical, inflammatory and joint damage measures after 1 year follow-up. Spearman correlation coefficients are shown with $|\rho|>0.10$ highlighted in green as significant correlations ($p<0.05$).*ACPA n=198

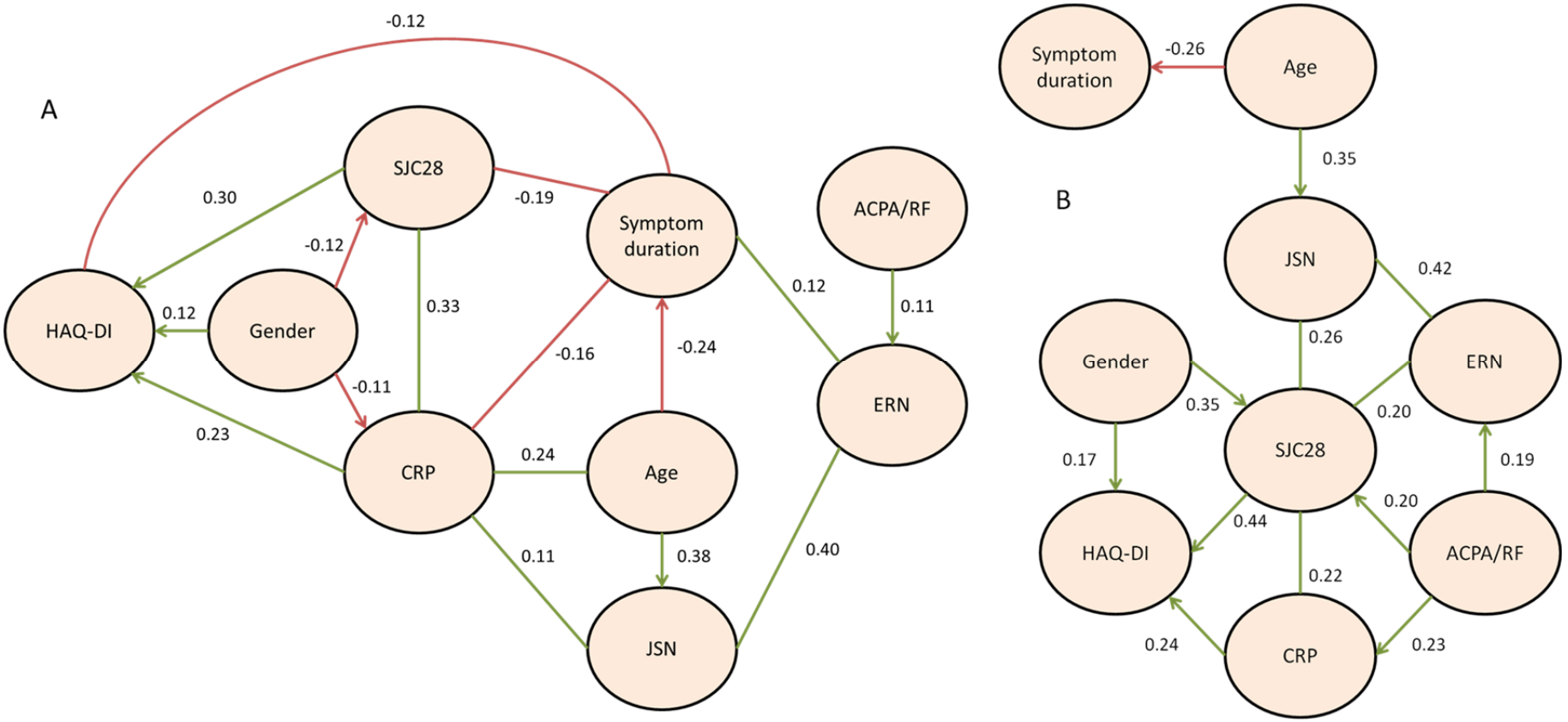


Figure 3.2 Correlation of clinical disease activity, inflammatory and joint damage measures. A – Pair-wise correlation of baseline clinical variables indicates inflammation is tied to JSN, but not ERN, and they are influenced by age and symptom duration. Erosions are moderately correlated with autoantibodies and symptom duration, and are distinct from markers of clinical severity and inflammation. B – The inflammatory burden and disease activity influence joint damage progression after 1 year as ERN are correlated with SJC28. Lines between clincial variables represent $|\rho|>0.1$ from Tables 3.28 and 3.29, directionality is implied from biological causality, i.e. patient age is not directly influence by their JSN score, however, age may affect the extent of joint space narrowing. Straight lines with no arrow head are correlations without known biological causality.

3.3.3.2 Rheumatoid arthritis susceptibility loci influence on correlated disease severity measures

RA susceptibility SNPs nominally associated ($p<0.05$) with baseline joint damage were selected for multivariate modelling. Each SNP was entered into a regression model based on the distribution of the dependent variable, adjusted for the clinical variables correlated with the dependent variable (Table 3.28 and Figure 3.2A). The same analysis was performed on the 1 year time point for these same clinical variables, adjusted for their baseline level; this analysis also included SNPs from section 3.3.1 associated with DMARD response outcome measures ($p<0.05$). No SNP associated with baseline joint damage displayed association with any of the other additional baseline measures in the multivariate analysis (Table 3.30), including baseline CRP which would reflect the pre-treatment inflammatory burden.

Three SNPs associated with baseline joint damage in section 3.3.2 showed a nominal association with an additional clinical measure at 1 year (Table 3.31). The minor allele of rs3218258 (*IL2RB*) was associated with lower AUC CRP ($p=0.02$), whilst the minor allele of rs13277113 (*C8orf13-BLK*) was associated with a lower AUC HAQ-DI ($p=0.023$) and rs42041 (*CDK6*) minor allele was associated with lower ERN ($p=0.007$).

SNPs that showed nominal association with treatment response were also nominally associated with 1 year disease severity. The minor allele of rs540386 (*TRAF6*), associated with a lower 6month ΔSJC28, was associated with increased SJC28 AUC ($p=0.006$); this suggests the effect of this minor allele continues past the first 6 months of treatment. The minor allele of rs10760130 (*TRAF1/C5*) was nominally associated with higher JSN at 1 year ($p=0.013$), and with a borderline increased 6 month ΔDAS28 ($p=0.04$).

A single SNP associated with a lower risk of ERN progression was also nominally associated with multiple outcomes at 1 year (Table 3.31). The minor allele of rs231707 (*FAM193A*) was associated with a lower AUC SJC28 (p=0.018), but an increased JSN (p=0.024)

Table 3.30. Summary of SNPs associated with baseline disease severity outcomes

SNP	Cytoband	Locus	Allele	Variable	Regression model	Coefficient	95% CI	p-value	n
rs1160542	2q11.2	*AFF3*	A	Baseline lnCRP	Linear	-0.03	-0.19-0.12	0.68	372
				Baseline SJC28	Negative binomial	1.01	0.93-1.08	0.89	762
				Baseline HAQ-DI	Linear	-0.03	-0.08-0.24	0.27	807
				Baseline ERN	ZINB	1.64	1.03-2.61	0.039	413
				Baseline JSN	ZINB	1.27	0.97-1.67	0.08	404
rs13277113	8p23.1	*C8orf13-BLK*	A	Baseline lnCRP	Linear	-0.09	-0.27-0.09	0.34	369
				Baseline SJC28	Negative binomial	1.00	0.91-1.09	0.92	753
				Baseline HAQ-DI	Linear	0.03	-0.04-0.09	0.42	798
				Baseline ERN	ZINB	2.70	1.66-4.39	$6.6x10^{-5}$	410
				Baseline JSN	ZINB	1.43	1.09-1.88	0.011	401
rs1678542	12q13.3	*KIF5A-PIP4K2C*	C	Baseline lnCRP	Linear	-0.07	-0.23-0.09	0.37	371
				Baseline SJC28	Negative binomial	1.00	0.93-1.07	0.99	762
				Baseline HAQ-DI	Linear	0.01	-0.04-0.06	0.74	807
				Baseline ERN	ZINB	2.87	1.45-5.67	0.002	412
				Baseline JSN	ZINB	0.89	0.64-1.20	0.42	403
rs3218258	22q12.3	*IL2RB*	A	Baseline lnCRP	Linear	-0.16	-0.32-0.00	0.05	371
				Baseline SJC28	Negative binomial	0.99	0.91-1.07	0.73	761
				Baseline HAQ-DI	Linear	-0.02	-0.08-0.04	0.54	805
				Baseline ERN	ZINB	0.53	0.35-0.80	0.003	412
				Baseline JSN	ZINB	0.70	0.54-0.91	0.008	403
rs26232	5q21.1	*C5orf30*	T	Baseline lnCRP	Linear	-0.03	-0.21-0.16	0.78	338
				Baseline SJC28	Negative binomial	1.01	0.93-1.10	0.74	709
				Baseline HAQ-DI	Linear	-0.03	-0.09-0.04	0.41	755
				Baseline ERN	ZINB	0.36	0.19-0.68	0.001	374
				Baseline JSN	ZINB	0.77	0.53-1.14	0.20	363

SNP	Cytoband	Locus	Allele	Variable	Regression model	Coefficient	95% CI	p-value	n
rs42041	7q21.2	*CDK6*	G	Baseline lnCRP	Linear	-0.04	-0.21-0.12	0.60	372
				Baseline SJC28	Negative binomial	1.01	0.93-1.10	0.78	762
				Baseline HAQ-DI	Linear	0.00	-0.06-0.06	0.97	807
				Baseline ERN	ZINB	1.19	0.73-1.95	0.48	413
				Baseline JSN	ZINB	1.35	1.02-1.80	0.036	404
rs7234029	18p11.21	*PTPN2*	G	Baseline lnCRP	Linear	0.11	-0.09-0.32	0.27	370
				Baseline SJC28	Negative binomial	1.03	0.93-1.14	0.53	760
				Baseline HAQ-DI	Linear	-0.01	-0.08-0.07	0.87	805
				Baseline ERN	ZINB	2.15	1.21-3.82	0.009	411
				Baseline JSN	ZINB	1.56	1.13-2.14	0.006	402

Table 3.30. Exploratory genetic association of baseline joint damage associated RA susceptibility SNPs with baseline disease severity measures. Each SNP was entered into an appropriate regression model based on the distribution of the dependent variable. Each model was adjusted for the other clinical variables and outcome measures highlighted in Table 3.28. All SNPs highlighted in bold are associated with baseline joint damage only. Coefficients are effect estimates from each regression model, linear regression = β estimate, negative binomial regression = IRR, ZINB regression = Count model IRR. lnCRP – natural log transformed C-reactive protein, SJC28 – swollen joint count with 28 joints, HAQ-DI – health assessment questionnaire disability index, ERN – erosion score, JSN – joint space narrowing score, ZINB – zero-inflated negative binomial regression.

Table 3.31. Summary of SNPs associated with multiple disease severity outcomes at 1 year follow-up, adjusted for baseline

SNP	Cytoband	Locus	Allele	Variable	Regression model	Coefficient	95% CI	p-value	n
rs1160542	2q11.2	*AFF3*	A	lnCRP AUC 1 year	Linear	0.37	-0.63-1.36	0.47	427
				SJC28 AUC 1 year	Linear	1.36	-4.06-6.78	0.62	201
				HAQ-DI AUC 1 year	Linear	1.43	-3.01-5.88	0.53	207
				ERN 1 year	ZINB	0.93	0.66-1.30	0.67	236
				JSN 1 year	ZINB	1.10	0.90-1.34	0.34	230
rs4272626	1p13.1	*CASQ2*	T	lnCRP AUC 1 year	Linear	-0.03	-1.08-1.02	0.96	428
				SJC28 AUC 1 year	Linear	-0.74	-6.12-4.65	0.79	201
				HAQ-DI AUC 1 year	Linear	-0.43	-4.88-4.02	0.85	207
				ERN 1 year	ZINB	0.92	0.68-1.24	0.58	236
				JSN 1 year	ZINB	0.90	0.76-1.08	0.26	230
rs12746613	1q23.2	*FCGR2A*	T	lnCRP AUC 1 year	Linear	-0.38	-1.93-1.17	0.63	428
				SJC28 AUC 1 year	Linear	6.07	-2.55-14.7	0.17	201
				HAQ-DI AUC 1 year	Linear	-1.20	-8.42-6.01	0.74	207
				ERN 1 year	ZINB	0.98	0.62-1.56	0.94	236
				JSN 1 year	ZINB	1.11	0.85-1.44	0.44	230
rs10919563	1q31.3	*PTPRC*	A	lnCRP AUC 1 year	Linear	-0.28	-1.80-1.24	0.72	428
				SJC28 AUC 1 year	Linear	1.10	-7.29-9.48	0.80	201
				HAQ-DI AUC 1 year	Linear	-0.37	-7.37-6.62	0.92	207
				ERN 1 year	ZINB	0.64	0.38-1.09	0.10	236
				JSN 1 year	ZINB	0.72	0.51-1.03	0.07	230
rs13277113	8p23.1	*C8orf13-BLK*	A	lnCRP AUC 1 year	Linear	0.15	-1.00-1.30	0.80	423
				SJC28 AUC 1 year	Linear	3.03	-3.06-9.12	0.33	199
				HAQ-DI AUC 1 year	Linear	-5.79	-10.8--0.82	0.023	205
				ERN 1 year	ZINB	1.19	0.84-1.69	0.33	233
				JSN 1 year	ZINB	0.89	0.70-1.13	0.34	227

SNP	Cytoband	Locus	Allele	Variable	Regression model	Coefficient	95% CI	p-value	n
rs1678542	12q13.3	*KIF5A-PIP4K2C*	C	lnCRP AUC 1 year	Linear	0.46	-00.53-1.46	0.36	427
				SJC28 AUC 1 year	Linear	-3.02	-8.26-2.22	0.26	200
				HAQ-DI AUC 1 year	Linear	-0.39	-4.81-4.03	0.86	206
				ERN 1 year	ZINB	0.92	0.63-1.33	0.65	235
				JSN 1 year	ZINB	1.05	0.85-1.31	0.63	229
rs3218258	22q12.3	*IL2RB*	A	lnCRP AUC 1 year	Linear	-1.28	-2.36--0.21	0.020	426
				SJC28 AUC 1 year	Linear	4.09	-1.43-9.61	0.15	200
				HAQ-DI AUC 1 year	Linear	-0.29	-4.94-4.37	0.90	206
				ERN 1 year	ZINB	1.06	0.77-1.46	0.70	235
				JSN 1 year	ZINB	0.95	0.78-1.17	0.65	229
rs26232	5q21.1	*C5orf30*	T	lnCRP AUC 1 year	Linear	-0.20	-1.34-0.94	0.74	402
				SJC28 AUC 1 year	Linear	6.12	0.26-11.98	0.041	190
				HAQ-DI AUC 1 year	Linear	-4.10	-8.96-0.76	0.10	194
				ERN 1 year	ZINB	1.09	0.77-1.55	0.61	220
				JSN 1 year	ZINB	1.28	0.96-1.70	0.09	210
rs42041	7q21.2	*CDK6*	G	lnCRP AUC 1 year	Linear	-1.07	-2.21-0.08	0.07	427
				SJC28 AUC 1 year	Linear	-0.55	-6.47-2.37	0.85	201
				HAQ-DI AUC 1 year	Linear	-1.87	-6.78-3.05	0.46	207
				ERN 1 year	ZINB	0.68	0.51-0.90	0.007	236
				JSN 1 year	ZINB	0.88	0.71-1.09	0.25	230
rs2900180	9q33.1	*TRAF1/C5*	T	lnCRP AUC 1 year	Linear	-0.56	-1.64-0.51	0.30	427
				SJC28 AUC 1 year	Linear	0.11	-5.57-5.80	0.97	200
				HAQ-DI AUC 1 year	Linear	-0.74	-5.46-3.97	0.76	206
				ERN 1 year	ZINB	0.78	0.55-1.12	0.19	235
				JSN 1 year	ZINB	1.18	0.94-1.49	0.14	229

SNP	Cytoband	Locus	Allele	Variable	Regression model	Coefficient	95% CI	p-value	n
rs10760130	9q33.1	*TRAF1/C5*	G	lnCRP AUC 1 year	Linear	-0.17	-1.21-0.87	0.75	424
				SJC28 AUC 1 year	Linear	-1.47	-6.92-3.98	0.60	199
				HAQ-DI AUC 1 year	Linear	-0.49	-5.01-4.04	0.83	205
				ERN 1 year	ZINB	0.77	0.52-1.12	0.17	233
				JSN 1 year	ZINB	1.34	1.06-1.68	0.013	227
rs7234029	18p11.21	*PTPN2*	G	lnCRP AUC 1 year	Linear	0.17	-1.22-1.56	0.81	426
				SJC28 AUC 1 year	Linear	-1.98	-9.68-5.72	0.61	199
				HAQ-DI AUC 1 year	Linear	4.59	-10.9-1.74	0.15	205
				ERN 1 year	ZINB	0.93	0.61-1.42	0.74	235
				JSN 1 year	ZINB	0.83	0.63-1.09	0.17	229
rs231707	4p16.3	*FAM193A*	A	lnCRP AUC 1 year	Linear	0.55	-0.75-1.85	0.40	428
				SJC28 AUC 1 year	Linear	-8.17	-14.9--1.41	0.018	201
				HAQ-DI AUC 1 year	Linear	4.88	-0.80-10.56	0.09	207
				ERN 1 year	ZINB	0.78	0.49-1.26	0.31	236
				JSN 1 year	ZINB	1.36	1.04-1.79	0.024	230
rs2069778	4q27	*IL2/IL21*	T	lnCRP AUC 1 year	Linear	0.35	-0.98-1.68	0.61	426
				SJC28 AUC 1 year	Linear	4.45	-2.51-11.4	0.21	200
				HAQ-DI AUC 1 year	Linear	2.72	-3.21-8.65	0.37	206
				ERN 1 year	ZINB	0.86	0.58-1.28	0.46	235
				JSN 1 year	ZINB	1.09	0.76-1.56	0.65	229
rs540386	11p12	*TRAF6*	T	lnCRP AUC 1 year	Linear	-1.01	-2.51-0.50	0.19	428
				SJC28 AUC 1 year	Linear	10.7	3.09-18.3	0.006	201
				HAQ-DI AUC 1 year	Linear	-5.80	-12.3-0.67	0.08	207
				ERN 1 year	ZINB	0.85	0.55-1.32	0.48	236
				JSN 1 year	ZINB	0.88	0.66-1.18	0.39	230

Table 3.31. Exploratory genetic association of 1 year joint damage and treatment response associated RA susceptibility SNPs with 1 year disease severity measures. Each SNP was entered into an appropriate regression model based on the distribution of the dependent variable. Each model was adjusted for the other clinical variables, their respective baseline measurement and outcome measures highlighted in Table 3.29. All SNPs highlighted in bold are associated with $p<0.05$. Coefficients are effect estimates from each regression model, linear regression = β estimate, negative binomial regression = IRR, ZINB regression = Count model IRR. lnCRP – natural log transformed C-reactive protein, SJC28 – swollen joint count with 28 joints, HAQ-DI – health assessment questionnaire disability index, ERN – erosion score, JSN – joint space narrowing score, ZINB – zero-inflated negative binomial regression.

3.3.4 *cis*-expression quantitative trait loci and disease severity associated loci

3.3.4.1 *cis*-expression quantitative trait loci analysis

Robust genetic associations are abstract statistical phenomena that likely reflect underlying biological phenomena, which could be simple or complex. Therefore the interpretation of genetic association signals requires a functional and/or biological context. One avenue is to look for biologically plausible functional changes such as non-synonymous variants, or overlapping signals of association with biological processes, such as gene expression. The genetic control of gene expression by common genetic variation is an established biological phenomenon, and is thought to contribute to a substantial proportion of disease association signals arising from GWAS. Approximately 40% of NHGRI GWAS catalogue SNPs show evidence of overlap with annotated regulatory elements [281].

SNPs nominally associated with baseline joint damage, DMARD response (including MTX and SSA monotherapy) and joint damage progression were input into the GeneVar viewer (Table 3.32). Statistically significant ($p \leq 1x10^{-5}$) *cis*-eQTLs were evident for 6 of 14 SNPs. Some of the genes were covered by multiple expression probes which showed evidence of differential eQTLs, e.g. *AFF3*. Several of the genes queried have been previously reported as *cis*-eQTLs in lymphoblastoid cell lines (LCLs) [288]. LD between gene expression-associated SNPs (eSNPs) and trait-associated SNPs was used to evaluate the relationship between the trait of interest and the *cis*-eQTLs identified (Table 3.33). Genotype data from the combined CEU and TSI populations were downloaded from the HapMap project website directly through Haploview 4.2. There was no data available on a number of eSNPs and for one of the trait-associated SNPs in the HapMap reference panel, so LD could not be assessed for these. Whilst moderate LD exists between several SNPs and eSNPs, the strongest effect observed is with rs13277113. This SNP in particular is strongly associated with the differential expression of both *C8orf13* and *BLK* in LCLs and primary B cells, the

details of which have been published previously [277, 288]. The association in the context of LCLs for *BLK* at least, was attributed to a haplotype spanning the 5' UTR of the *BLK* mRNA.

In order to further investigate the link between these eQTLs gene expression data from isolated primary cells from healthy control individuals was used. The details and results of these data have been published previously [277]. Several of the *cis*-eQTLs in Table 3.32 are also present in this data, including those with probes mapping to *BLK*, *C8orf13*, *C5* and *FAM119B* (*METTL21B*). No statistically significant *cis*-eQTL associations were found for any probe at the *IL2RB* locus in the primary B cells or monocytes data.

Table 3.32. MuTHER resource *cis*-eQTL query summary

Trait associated SNP	Gene	Trait	Cytoband	Probe IDs	Adipose tissue			Lymphocyte cell line			Skin biopsy			*cis*-eQTL
					Lead SNP	β	p	Lead SNP	β	p	Lead SNP	β	p	
rs4272626	*CASQ2*	SSA response %ΔDAS	1p13.1	-	-	-	-	-	-	-	-	-	-	x
rs12746113	*FCGR2A*	MTX response ΔDAS28	1q23.2	-	-	-	-	-	-	-	-	-	-	x
rs10919563	*PTPRC*	JSN progression	1q31.3	-	-	-	-	-	-	-	-	-	-	x
rs1160542	*AFF3*	Baseline SHS	2q11.2	ILMN_1677951 ILMN_1775235	- rs11691869	- -0.11	- $1.9x10^{-7}$	- rs6542920	- -0.12	- $6.2x10^{-20}$	- -	- -	- -	✓
rs2069778	*IL2/IL21*	JSN progression	4q27	-	-	-	-	-	-	-	-	-	-	x
rs26232	*PAM*	Baseline ERN	5q21.1	ILMN_2313901	-	-	-	rs17154942	-0.17	$2.3x10^{-16}$	-	-	-	✓
rs42041	*CDK6*	Baseline JSN	7q21.2	-	-	-	-	-	-	-	-	-	-	x
rs13277113	*C8orf13*	Baseline ERN	8p23.1	ILMN_1687213	-	-	-	rs13277113	-0.42	$6.2x10^{-44}$	rs2409773	0.15	$4.8x10^{-19}$	✓

Trait associated SNP	Gene	Trait	Cytoband	Probe IDs	Adipose tissue			Lymphocyte cell line			Skin biopsy			*cis*-eQTL
					Lead SNP	β	p	Lead SNP	β	p	Lead SNP	β	p	
rs13277113	*BLK*	Baseline ERN	8p23.1	ILMN_1668277	-	-	-	rs4840568	0.44	$3.6x10^{-27}$	-	-	-	✓
rs2900180	*C5*	ERN progression	9q33.1	ILMN_1746819	rs7040033	-0.18	$1.0x10^{-19}$	rs10760142	-0.12	$4.8x10^{-22}$	-	-	-	✓
rs2900180	*GSN*	ERN progression	9q33.1	ILMN_1787518 ILMN_1801043	- -	- -	- -	- rs2308681	- -0.09	- $5.7x10^{-9}$	- -	- -	- -	✓
rs2900180	*MEGF9*	ERN progression	9q33.1	ILMN_1658798 ILMN_2290118	rs10616 rs735110	0.09 0.13	$4.9x10^{-30}$ $4.9x10^{-35}$	- -	- -	- -	rs10985008 -	-0.04 -	$4.6x10^{-6}$ -	✓
rs540386	*TRAF6*	ΔSJC28	11p12	-	-	-	-	-	-	-	-	-	-	x
rs1678542	*METTL21B*	Baseline SHS	12q13.3	ILMN_1680231 ILMN_1723846	- rs11172343	- -0.19	- $1.2x10^{-44}$	- rs4646536	- 0.26	- $2.2x10^{-46}$	- rs1021469	- -0.13	- $8.6x10^{-22}$	✓
rs7234029	*PTPN2*	Baseline SHS	18p11.21	-	-	-	-	-	-	-	-	-	-	x
rs3218258	*IL2RB*	Baseline SHS	22q12.3	ILMN_1684349	-	-	-	rs3218253	-0.13	$2.1x10^{-15}$	rs228975	-0.27	$5.9x10^{-35}$	✓

Table 3.32. *cis*-eQTLs within 0.5Mb of joint damage and DMARD response associated SNPs. SNP IDs were input into the GeneVar viewer using the *cis*-eQTL search. *cis*-eQTLs within 0.5Mb and with association $p \leq 1x10^{-5}$ are shown. *cis*-eQTL results were downloaded from the GeneVar viewer based on the analysis of the MuTHER data set across adipose tissue, LCLs and skin biopsy tissue. The lead SNPs listed for each eQTL are those with the smallest association p-value for each tissue type. LCL – lymphoblastoid cell line, eQTL – expression quantitative trait loci.

Table 3.33. LD relationship between trait-associated SNPs and eSNPs					
Trait associated SNP	MAF	eSNP	MAF	Gene	LD (r^2)
rs1160542	0.493	rs11691869	NA	*AFF3*	NA
		rs6542920	0.315	*AFF3*	0.45
rs26232[a]	0.314	rs17154942	NA	*PAM*	NA
rs13277113	0.256	rs2409773	NA	*C8orf13*	NA
		rs13277113[b]	0.256	*C8orf13*	NA
		rs4840568	0.246	*BLK*	0.85
rs2900180	0.374	rs7040033	NA	*C5*	NA
		rs10760142	0.434	*C5*	0.17
		rs2308681	NA	*GSN*	NA
		rs10616	0.322	*MEGF9*	0.17
		rs735110	NA	*MEGF9*	NA
		rs10985008	0.322	*MEGF9*	0.21
rs1678542	0.412	rs4646536	0.305	*METTL21B*	0.21
		rs11172333	0.305	*METTL21B*	0.21
		rs1021469	0.288	*METTL21B*	0.16
rs3218258	0.270	rs3218253	NA	*IL2RB*	NA
		rs228975	NA	*IL2RB*	NA

Table 3.33. Pair wise LD between eSNPs and RA trait associated SNPs. LD between SNPs was calculated in Haploview 4.2 using the r^2 statistic. NA – no information available either for the trait-associated SNP or the eSNP in the HapMap reference populations CEU and TSI. Each gene represents the genomic interval to which the expression probes are designed to anneal to from Table 3.32. *METTL21B* is listed as *FAM119B* in the GeneVar record, as the gene symbol has been changed since the expression array was designed. a – genotype data for rs26232 is not available in these reference populations, b – the lead SNP for *C8orf13* in the LCL expression data is rs13277113. MAF – minor allele frequency, eSNP – *cis*-eQTL associated SNP, LD – linkage disequilibrium.

Gene expression data from both primary B cells and monocytes was extracted for the specific probes listed in Table 3.32 and matched to genotype data within 0.5Mb of each gene, including those genes with previously published eQTLs. Three expression probes mapping to the *AFF3* gene, ILMN_1775235, ILMN_1677951 and ILMN_1697308, were identified. Following QC of the expression probe data two SNPs were found to overlap with probe ILMN_1697308. Therefore association testing was performed with and without individuals carrying the minor allele of these two SNPs (rs17023562 and rs17023563). This probe was analysed in neither the MuTHER dataset nor the published primary immune cell data and corresponds to isoform 1 of the *AFF3* gene; two recorded mRNA isoforms exist for this gene (Isoform 1:NM_002285.2, Isoform 2:NM_001025108.1). From the 9 expression probes with eQTLs in the MuTHER data (Table 3.32) 7 showed evidence of SNP-probe

associations. The probe at the *AFF3* locus, ILMN_1697308, not tested in the MuTHER or published primary cell data, also displayed evidence of a *cis*-eQTL (Table 3.34).

Table 3.34. Primary B cell and monocyte eQTLs overlapping with MuTHER eQTLs										
Trait associated SNP	Gene	Probe	Cell type	Lead SNP	p-value	Gene expression change	Trait associated SNPLD (r^2)	Trait-SNP conditioned p-value[a]	Lead eSNP conditioned p-value	Conditional lead eSNP[b]
rs1160542	*AFF3*	ILMN_1775235	B cell	rs13421952	$2.5x10^{-5}$	Down	0.44	$1.7x10^{-3}$	$3.8x10^{-3}$	rs115325974
			Monocyte	rs10167279	$1.4x10^{-3}$	-	-	-	-	
		ILMN_1677951	B cell	rs11689199	$1.9x10^{-3}$	-	-	-	-	-
			Monocyte	rs57412166	$1.9x10^{-4}$	-	-	-	-	-
		ILMN_1697308	B cell	rs2309750	$8.6x10^{-34}$	Up	0.21	$8.9x10^{-35}$	$6.6x10^{-8}$	rs11692215
			Monocyte	rs78266048	$4.1x10^{-5}$	-	-	-	-	-
rs13277113	*BLK*	ILMN_1668277	B cell	rs1478901	$1.9x10^{-14}$	Down	0.99	$5.4x10^{-4}$	$4.8x10^{-4}$	rs118092403
			Monocyte	rs804273	$4.1x10^{-4}$	-	-	-	-	-
	C8orf13	ILMN_1687213	B cell	rs2736332	$1.4x10^{-87}$	Up	0.84	$3.2x10^{-16}$	$1.9x10^{-6}$	rs922483
			Monocyte	rs2736328	$5.4x10^{-14}$	Up	0.03	$2.9x10^{-16}$	$8.5x10^{-17}$	rs12680762
rs2900180	*MEGF9*	ILMN_1658798	B cell	rs914592	$5.9x10^{-3}$	-	-	-	-	-
			Monocyte	rs2416815	$4.8x10^{-4}$	-	-	-	-	-
		ILMN_2290118	B cell	rs7044106	3.010^{-10}	Down	0.18	$3.2x10^{-8}$	$5.1x10^{-3}$	rs4836827
			Monocyte	rs10984990	$1.1x10^{-11}$	Down	0.15	$8.2x10^{-9}$	$8.9x10^{-4}$	rs191047
	C5	ILMN_1746819	B cell	rs10818502	$2.5x10^{-12}$	Up	0.14	$2.3x10^{-9}$	$2.7x10^{-4}$	rs2416814
			Monocyte	rs10733650	$5.0x10^{-26}$	Up	0.23	$8.3x10^{-22}$	$2.9x10^{-4}$	rs2416814
	GSN	ILMN_1787518	B cell	rs12551209	$1.5x10^{-3}$	-	-	-	-	-
			Monocyte	rs10760159	$7.9x10^{-3}$	-	-	-	-	-
		ILMN_1801043	B cell	rs7869290	$1.5x10^{-2}$	-	-	-	-	-
			Monocyte	rs4836848	$6.3x10^{-32}$	Up	0.10[a]	$8.1x10^{-30}$	$3.0x10^{-3}$	rs4836825
rs1678542	*METTL21B*	ILMN_1680231	B cell	rs4760176	0.297	-	-	-	-	-
			Monocyte	rs7136981	$6.7x10^{-3}$	-	-	-	-	-
		ILMN_1723846	B cell	rs10877013	$1.0x10^{-10}$	Up	0.27[b]	$4.4x10^{-7}$	$2.9x10^{-3}$	rs34181394
			Monocyte	rs6581155	$2.4x10^{-57}$	Up	0.27[b]	$5.0x10^{-43}$	$2.2x10^{-2}$	rs35612896
rs3218258	*IL2RB*	ILMN_1684349	B cell	rs9610753	$3.0x10^{-3}$	-	-	-	-	-
			Monocyte	rs9610587	$3.2x10^{-5}$	-	-	-	-	-

Table 3.34. Primary B cell and monocyte eQTLs at RA trait-associated loci. Each eQTL from Table 3.31 was tested in a dataset of resting primary B cells and monocytes from a panel of 286 healthy control volunteers . SNPs within 0.5Mb of each gene were tested for association with probe intensity for each probe mapping to the relevant gene for each cell type separately. a – p-value of lead eSNP from linear regression of probe intensity conditioned on the trait associated SNP at that locus, b – The eSNP with the smallest p-value following linear regression of expression probe intensity conditioned on the original eSNP.

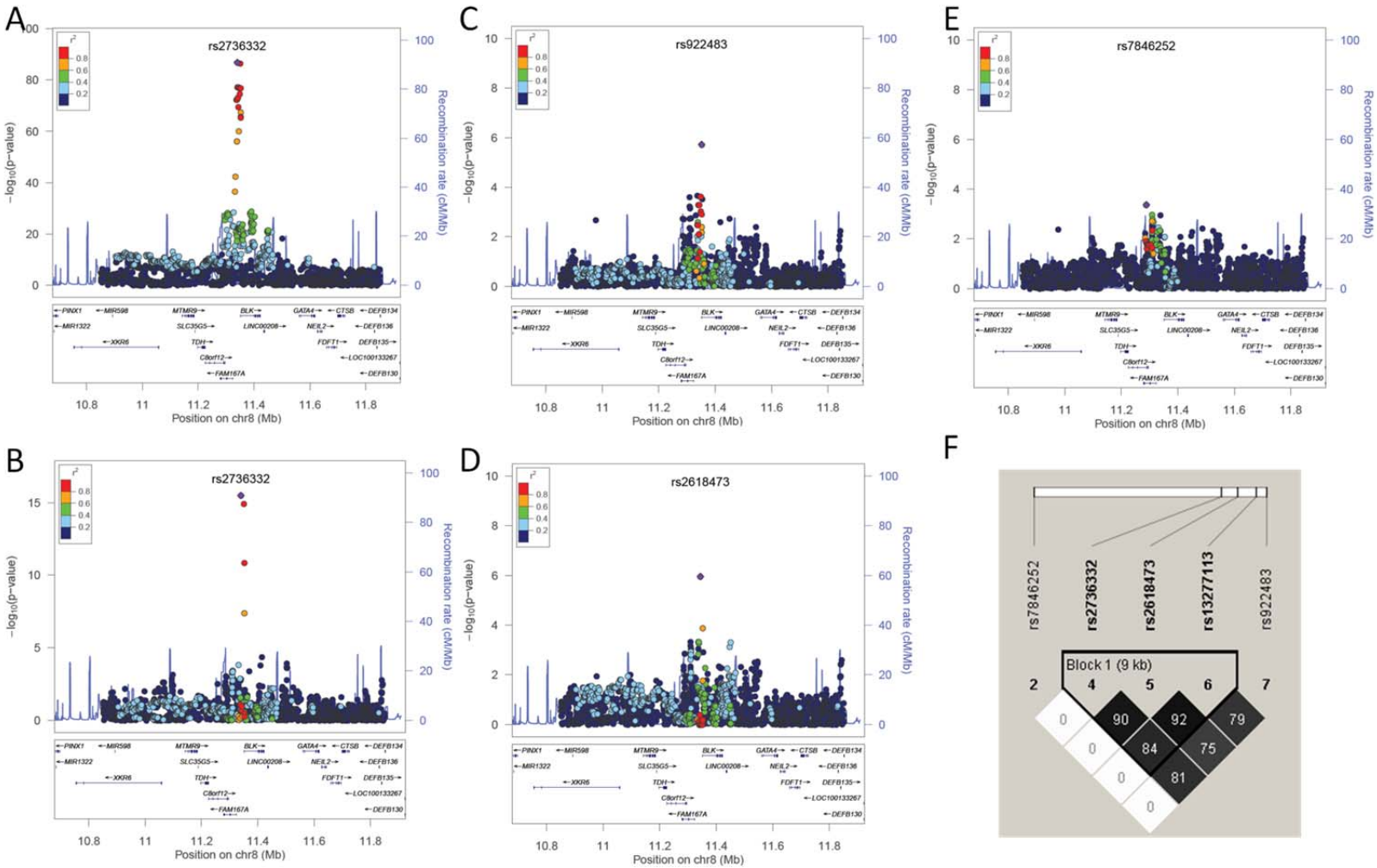


Figure 3.3. High resolution Manhattan plots of conditional linear regression analysis of B cell *C8orf13 cis*-eQTL association analysis. SNP-probe association p-values are plotted by physical position along the chromosome from the linear regression model. The lead SNP from each analysis is shown as a purple diamond and annotated with the relevant rs# ID. Colouring represents LD between the lead SNP and other SNPs included in the analysis, the key is shown in the top left corner of each plot. A – Primary analysis of *C8orf13* expression. A strong eQTL is present which was reported in the original publication [275]. B – The original eQTL is evident following conditioning on the baseline joint damage associated SNP rs13277113, however some of the signal has been accounted for by adjusting for this SNP. C – Conditional analysis on the lead SNP in panel A. rs922483 is in strong LD with rs13277113 and rs2736332 which suggests there may be a haplotype contributing to the observed eQTL at this locus. D – Further conditional analysis on both rs13277113 and rs2736322 reveals a third *cis*-eQTL signal with rs2618473. E – No further *cis*-eQTL signals are seen at the prescribed statistical significance level ($p \leq 1x10^{-5}$), F – LD matrix of B cell *C8orf13* eSNPs. Colour is by r^2 where increasing depth of shade indicates stronger LD; values are shown for each pair wise comparison. 4 SNPs are in strong LD, possibly pointing to a haplotype effect rather than independent SNP effects. *C8orf13* is also annotated *FAM167A*.

Each *cis*-eQTL analysis was repeated conditioned on the trait-associated SNP in order to test the contribution of the SNP towards the *cis*-eQTL. The *cis*-eQTL association signal in B cells with the *BLK* probe was markedly reduced, with no single SNP reaching the threshold for statistical significance in the conditional analysis ($p \leq 1x10^{-5}$). The *cis*-eQTL observed with the *C8orf13* probe in the B cell data, adjusted for the trait-associated SNP rs13277113, appeared to show evidence of an independent association signal (Figure 3.3B). In the linear regression model adjusted for the original eSNP an association signal was observed with a SNP in strong LD with rs13277113 (Table 3.34 and Figure 3.3C). A linear regression model conditioned on both rs13277113 and rs2736332 found a third signal (Figure 3.3D, rs2618473 $p=1.1x10^{-6}$). A further conditional linear regression adjusted for these 3 SNPs found only circumstantial evidence of a *cis*-eQTL signal (Figure 3.3E, rs7846252 $p=4.3x10^{-4}$). The strong LD between all 4 of these SNPs (including rs13277113) suggests they are not independent, but represent a haplotype effect modulating the expression of this gene in resting healthy B cells. Conditional haplotype analysis showed the presence of two haplotypes which explain over 67% of the variance in *C8orf13* expression levels in B cells ($p=4.6x10^{-70}$). Including the SNP rs7846252 into the haplotype analysis parses the two haplotypes into 3; however, there is no difference in the direction of effect for the minor allele of this SNP which suggests it has no additional effect on *C8orf13* B cell expression (Table 3.35).

The *C8orf13* eQTL in monocytes showed a large, locus-wide, enrichment for small p-values (Figure 3.4A). Conditioning on the lead SNP from this analysis showed a second strong association signal ~300kb centromeric to the *C8orf13* coding region with rs4840550 which appeared to be independent of the primary association (Figure 3.4C). A linear regression model conditioned on rs13277113 in these data did not have an appreciable effect on the overall eQTL signal at this locus (Figure 3.4B). Controlling for both rs2736328 and rs4840550 found an additional association downstream of *C8orf13* with rs4532565 (Figure

3.4D). Further conditioning on this SNP found no additional association signals. Weak LD exists between these 3 eSNPs and rs13277113 (Figure 3.4F), as well as between the eSNPs themselves. There are no common haplotypes (frequency ≥5%) formed by the 3 eSNPs which implies these are all independent *cis*-eQTL associations.

Table 3.35. B cell *C8orf13* eQTL conditional haplotype analysis									
Haplotype	Frequency	β	R^2	p-value	Haplotype	Frequency	β	R^2	p-value*
CTAT	0.244	1.80	0.71	$6.4x10^{-76}$	CCTAT	0.22	1.90	0.67	$3.4x10^{-69}$
GCGC	0.697	-1.66	0.68	$4.6x10^{-70}$	CGCGC	0.59	-1.13	0.33	$3.3x10^{-26}$
					TGCGC	0.11	-0.95	0.11	$2.11x10^{-8}$

Table 3.35. Conditional haplotype analysis of B cell *C8orf13* eSNPs. Two haplotypes of SNPs rs2736332, rs2618473, rs13277113 and rs922483 explain ~70% of the expression of *C8orf13* in resting healthy B cells, and show opposite directions of effect. Including the SNP rs7846252 parses the GCGC haplotype into two more but does not appear to have an effect on expression. Only haplotypes with a frequency ≥5% were tested for association.

High resolution Manhattan plots of the eQTL association results for the *AFF3* isoform 1 probe can be seen in Figure 3.5 from healthy control B cells. A strong eQTL was observed with the expression probe that maps to *AFF3* isoform 1 which was unaltered by the omission of individuals carrying the minor alleles of the two SNPs mapping to the same position as the probe (Figure 3.5A). Sequential conditional linear regression analyses showed evidence of 3 *cis*-eQTL signals with this isoform probe (Figure 3.5A-C). Further conditional analysis controlling for these three lead SNPs did not show any further evidence of independent effects with this isoform expression (Figure 3.5D). Adjusting the primary linear regression for rs1160542 did not appreciably alter the *cis*-eQTL signal (Figure 3.5E). Moderate LD exists across the locus, including between rs1160542 and the two eSNPs rs11690905 and rs11692215 (r^2=0.2-0.4). Indeed, as seen in Figure 3.5F moderate LD exists between these two eSNPs which may imply these two SNPs tag a haplotype that modulate *AFF3* isoform 1 expression.

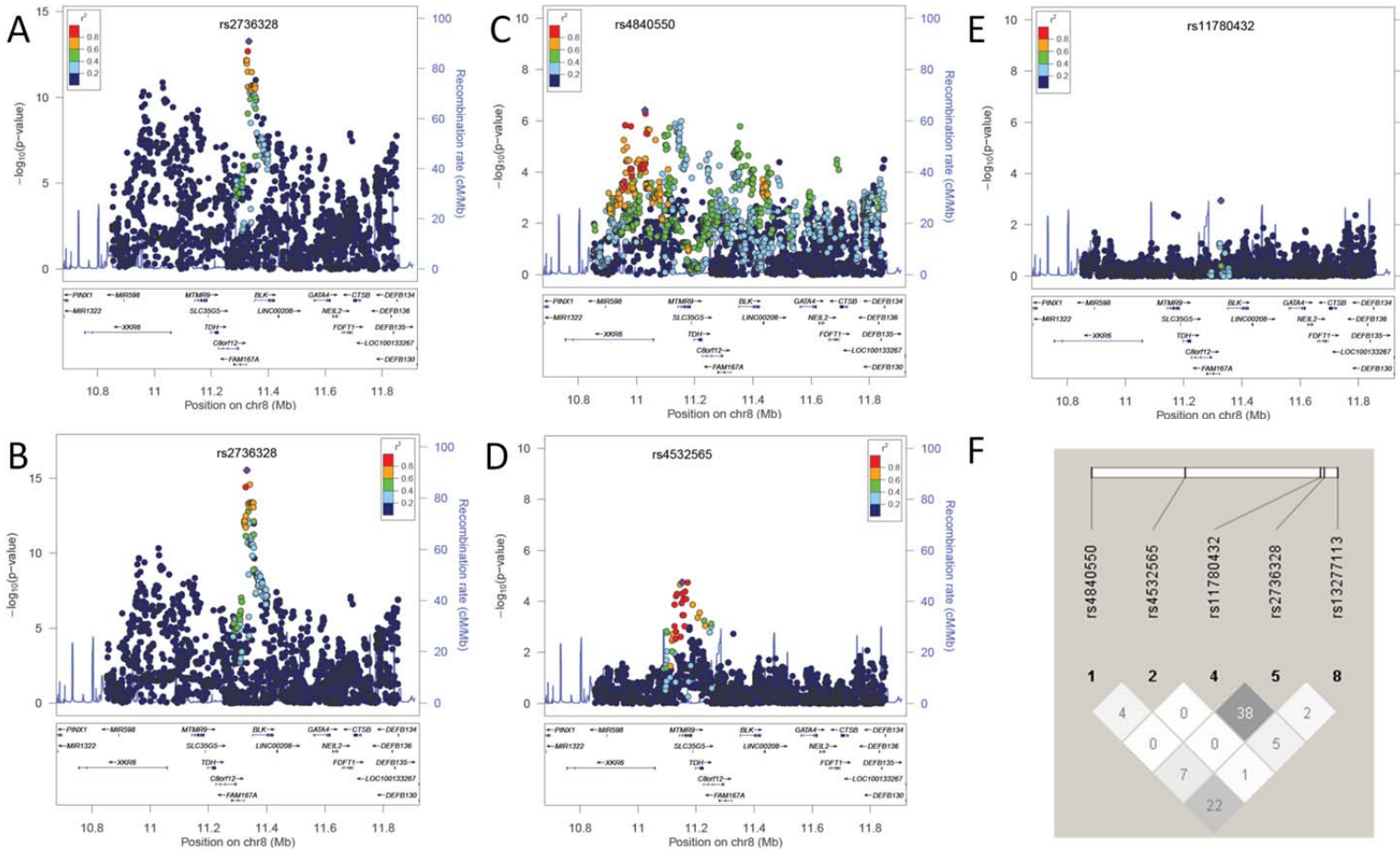


Figure 3.4. Conditional linear regression analysis of resting healthy control monocyte *C8orf13* eQTL. Genetic modulation of monocyte *C8orf13* expression shows evidence of 3 independent signals that are not related to the baseline joint damage associated SNP rs13277113. Linear regression models were sequentially conditioned on the lead SNP from the previous analysis following the primary unconditioned linear regression model. A – Primary eQTL analysis of *C8orf13* (ILMN_1687213) monocyte expression shows the strongest signal with rs2736328 intermediate to *C8orf13* and *BLK*. There are a large number of SNPs across the locus with small p-values. B – Including rs13277113 as a covariate in the linear regression analysis of monocyte *C8orf13* expression has no appreciable effect on the eQTL signal at this locus. C – Conditional linear regression controlling for the lead SNP rs2736328 shows the independent signal ~300kb centromeric to the *C8orf13* coding region. This signal of association also appears to explain the enrichment of small p-values seen in panel A as there is considerable locus-wide LD with this SNP. D – Conditional linear regression analysis controlled for rs2736328 and rs4840550 shows evidence of a third association signal with rs4532565 downstream of *C8orf13*. E – no further eQTL associations are observed following conditioning on the 3 lead SNPs in panels A, C and D. F – LD between monocyte *C8orf13* eSNPs and rs13277113 show independence of eSNPs from each other, as well as from rs13277113. *C8orf13* is also annotated *FAM167A*.

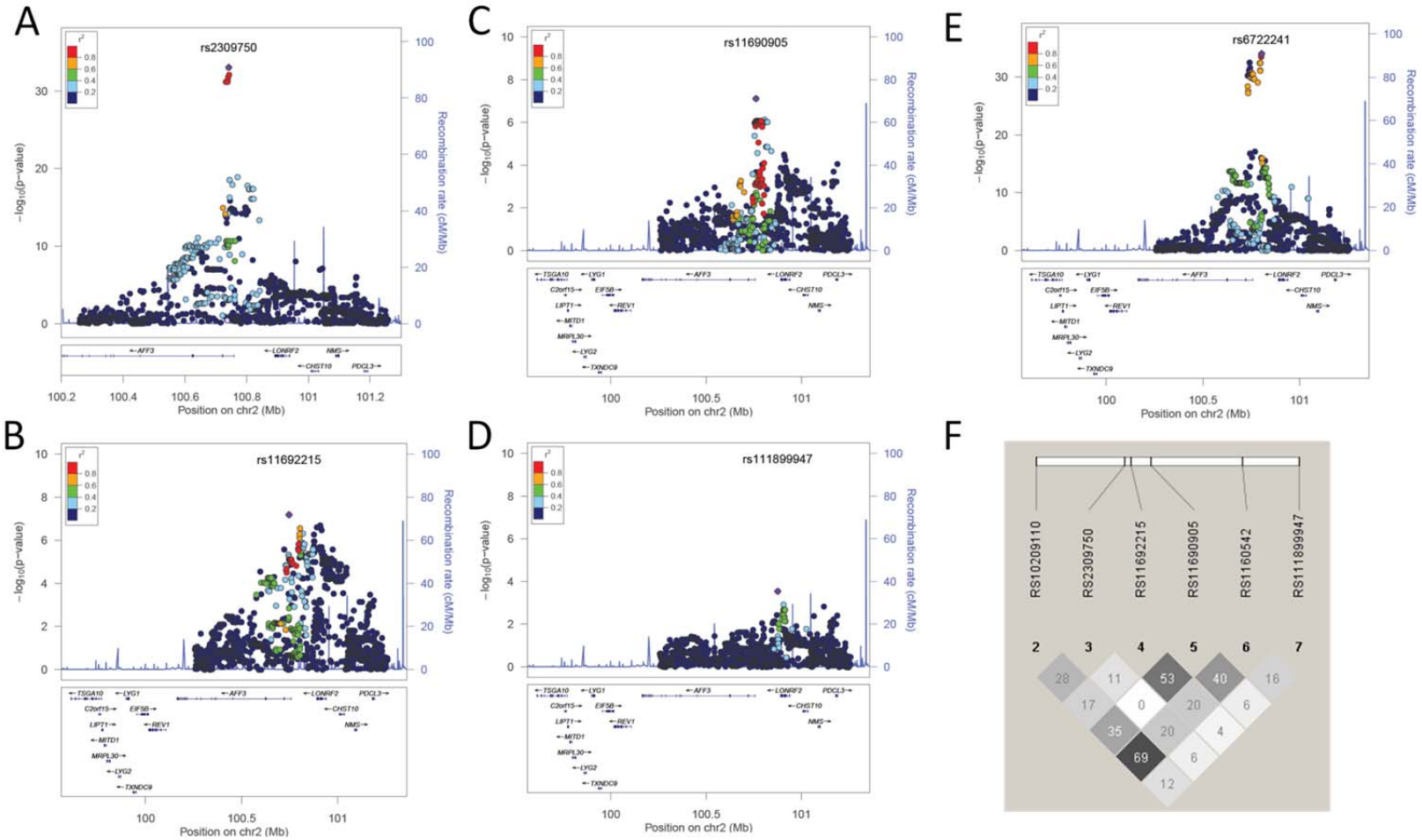


Figure 3.5. High resolution Manhattan plots of *AFF3* isoform 1 expression probe conditional eQTL analysis. $-\log_{10}$ p-values are plotted on the y-axis for each SNP included in the analysis after QC-filtering. SNP colouring is based on LD, measured by r^2, with the lead SNP. The LD key is in the top left of each high resolution Manhattan plot. The lead SNP for the association analysis is shown as a purple diamond and labelled with the appropriate rs# ID. The locus-wide recombination is plotted on the right-hand y-axis and is calculated from the HapMap CEU population. The bottom panel of each plot shows the refseq genes mapping to this particular genomic interval. A – B cell derived expression data from probe ILMN_1697308 maps to *AFF3* isoform 1. Evidence of an eQTL with this probe can be seen over the 2nd intron of *AFF3.* B – Conditional linear regression controlled for rs2309750 identified a second eQTL signal in the intergenic region between *AFF3* and *LONRF2*. C – Further conditional linear regression analysis on rs2309750 and rs11692215 show a third eQTL signal of association. D – No additional signals of association are observed following conditioning on the lead SNPs in panels A-C. E – Adjusting for the joint damage associated SNP rs1160542 has no appreciable impact on the primary association signal in panel A. F – LD patterns between the eSNPs and RA susceptibility SNPs rs1160542 and rs10209110 (recently identified in a large-scale fine-mapping study of autoimmune susceptibility loci [32] . Moderate LD exists between rs1160542 and two eSNPs rs11690905 and rs11692215.

In the MuTHER data three probes showed evidence of eQTLs in the region around rs2900180 at the *TRAF1/C5* locus (Table 3.32). The same eQTL signals were observed in both the monocyte and B cell expression data for *MEGF9* and *C5*, but only in the monocyte data for one of the probes measuring *GSN* expression. The linear regression models for each of these probes were conditioned on rs2900180 and the relevant lead SNP from the primary analysis (Figure 3.6). The eQTL signal with the monocytes *GSN* probe was unaltered by controlling for this trait associated SNP (Figure 3.6B); conditioning on the lead SNP rs4836848 accounted for the totality of the eQTL observed with this probe. In the B cell *MEGF9* (ILMN_2290118) eQTL conditional analysis on rs2900180 appeared to affect the strength of the association marginally with the lead SNP rs7044106 ($p=3.0x10^{-10}$ vs. $3.2x10^{-8}$) which suggested it may affect expression of this isoform of *MEGF9* in B cells (Figure 3.6C and 3.6D). However, conditioning on rs7044106 found a residual eQTL signal with rs4836827; neither SNP is in strong or moderate LD with rs2900180 (Figure 3.6F). The same effect of a weakening of association with the lead SNP when conditioned on rs2900180 was observed in the monocyte expression data. Again, no LD was observed between this SNP and rs10984990 (lead monocyte eSNP).

The publication of the primary B cell and monocyte data describes a strong eQTL with the expression probe for *C5* (Figure 3.7A and 3.7B) [277]. Conditional analysis on the lead SNP in both of these data did not reveal strong evidence for further eQTL effects with this probe (Figure 3.7C and 3.7D). Controlling for rs2900180, the trait-associated SNP in this context, did not alter the primary eQTL signal (Figure 3.7E and 3.7F). Weak LD exists between rs2900180 and the eQTL lead SNPs (r^2 0.06-0.10).

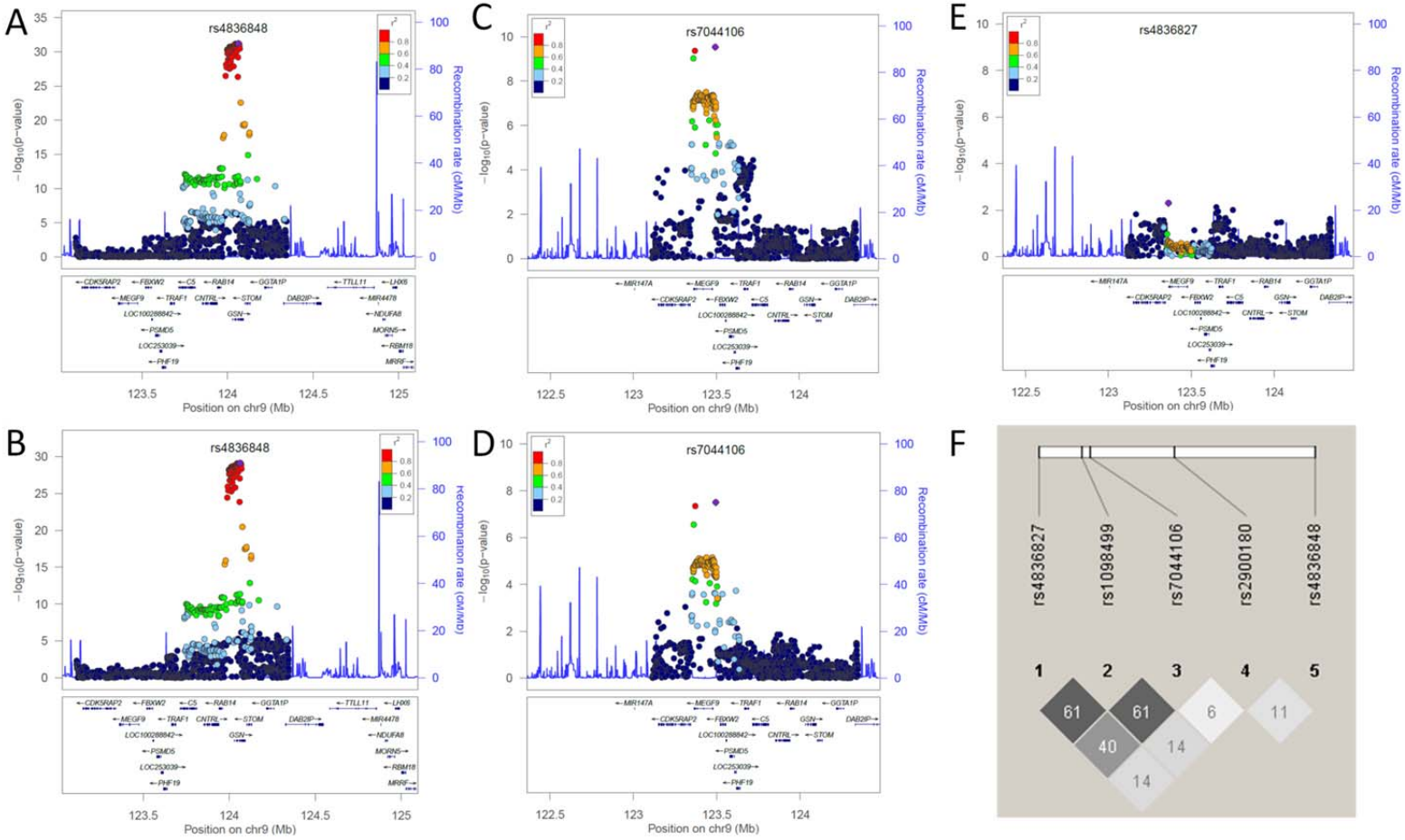


Figure 3.6. eQTL signals with *GSN* and *MEGF9* gene expression probes at the *TRAF1/C5* locus. High resolution Manhattan plots showing eQTL signals with monocyte *GSN* and B cell *MEGF9*. Each point represents a single genotyped SNP which are coloured by LD (r^2) with the annotated lead SNP in each plot. The right-hand x-axis shows the recombination rate estimated using HapMap populations. A – *GSN* probe ILMN_1801043 primary eQTL association in resting healthy control monocytes. B – conditional linear regression of monocyte *GSN* expression controlling for rs2900180 which appears to have no effect on this eQTL. C – B cell *MEGF9* (ILMN_2290118) eQTL association. D – B cell *MEGF9* conditioned on rs2900180 does not explain the eQTL with this expression probe. E – B cell *MEGF9* linear regression conditioned on rs7044106 appears to account for the signal observed at this locus. F – LD plot between eSNPs and the joint damage associated SNP rs2900180. Limited LD exists between these SNPs.

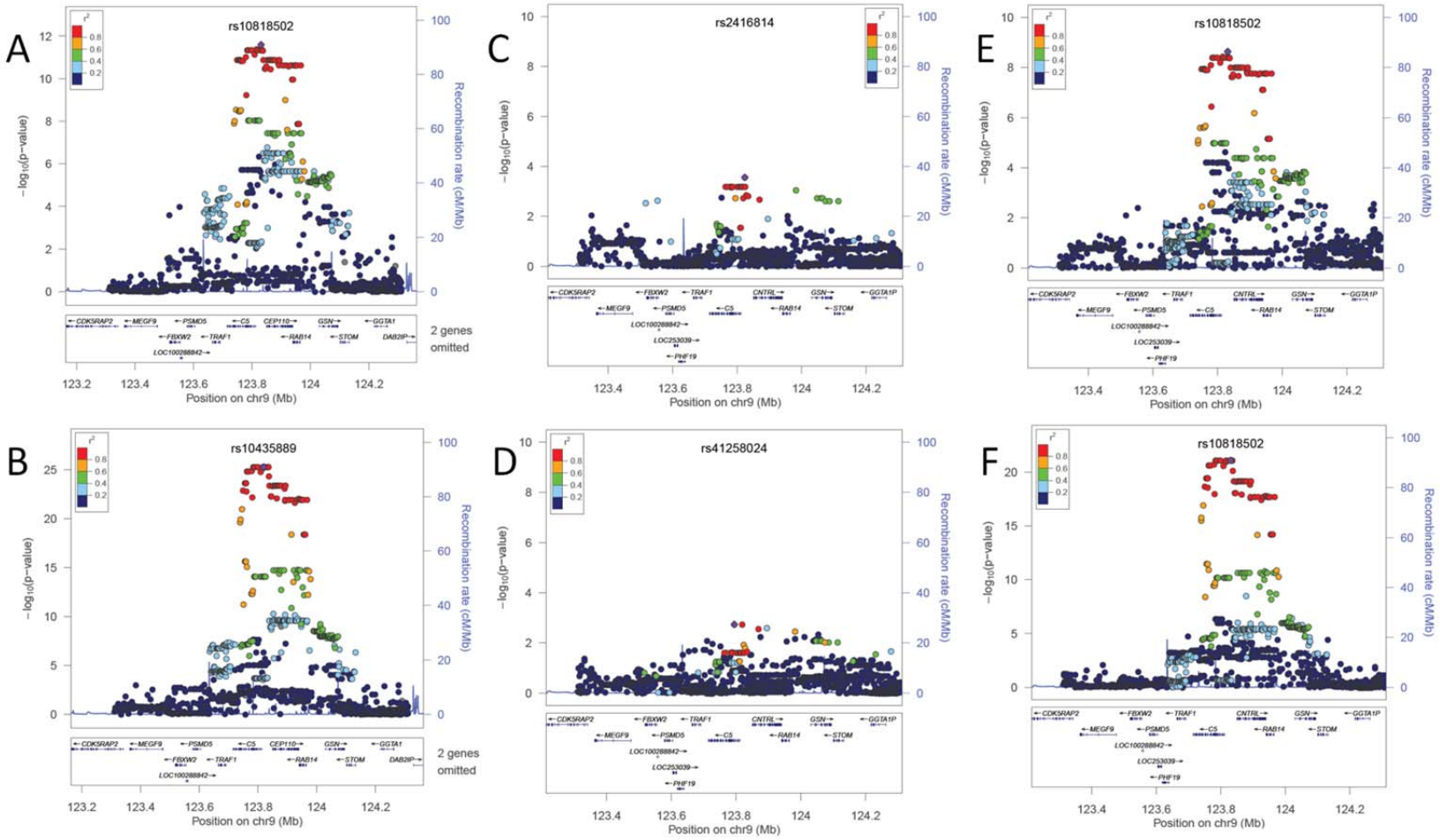


Figure 3.7. *cis*-eQTL signals with *C5* in resting healthy primary B cells and monocytes. A strong *cis*-eQTL is observed with the *C5* gene expression probe ILMN_1746819, which was reported in the primary literature [275]. A – B cell *C5 cis*-eQTL with SNPs across a large genomic interval covering several coding genes. B – The same *cis*-eQTL is observed in monocytes. The eSNPs rs10818502 and rs10435889 are in strong LD (r^2=0.98). C –B cell *C5 cis*-eQTL conditioned on rs10818502 accounts for the primary signal at this locus. Residual eSNPs do not reach the threshold for statistical significance. D – Monocyte *C5 cis*-eQTL conditioned on rs10435889 shows no further statistically significant *cis*-eQTL signals with SNPs at this locus. E – B cell *C5* expression conditioned on rs2900180, and F – monocyte *C5* expression. No effect is observed on the association signals seen with the expression of *C5* with the trait-associated SNP rs2900180.

A single *cis*-eQTL association signal was present in the MuTHER data in the region 0.5Mb of rs1678542. The association signal between this locus and RA susceptibility was refined by a high-density genotyping chip to a region 165Kb centromeric to rs1678542 [32]. This region is proximal to the coding region for *METTL21B*, identified as containing a *cis*-eQTL with an expression probe for this gene (ILMN_1723846) in the MuTHER data across all 3 tissue types (Table 3.32). A *cis*-eQTL was also observed in both the primary B cell and monocyte data for this same probe, which was reported in the results for the primary publication (Table 3.34 and Figure 3.8A and 3.8B) [277]. The trait-associated SNP rs1678452 was used in a conditional linear regression for both the B cell and monocyte *cis*-eQTL analysis. No effect was observed on the *cis*-eQTL when controlled for this SNP; conditional linear regression controlled for the lead SNP in the primary analysis appeared to account for this association signal in both the B cell and monocyte data (Figure 3.8C and 3.8D). Moderate LD exists between rs1678542 and the *METTL21B* eSNPs ($r^2 \approx 0.27$, Figure 3.8E). No genotype or LD information was available on the refined RA susceptibility SNP (rs10683701), thus its contribution to both the association with joint damage and any potential influence on the *METTL21B cis*-eQTL could not be tested formally.

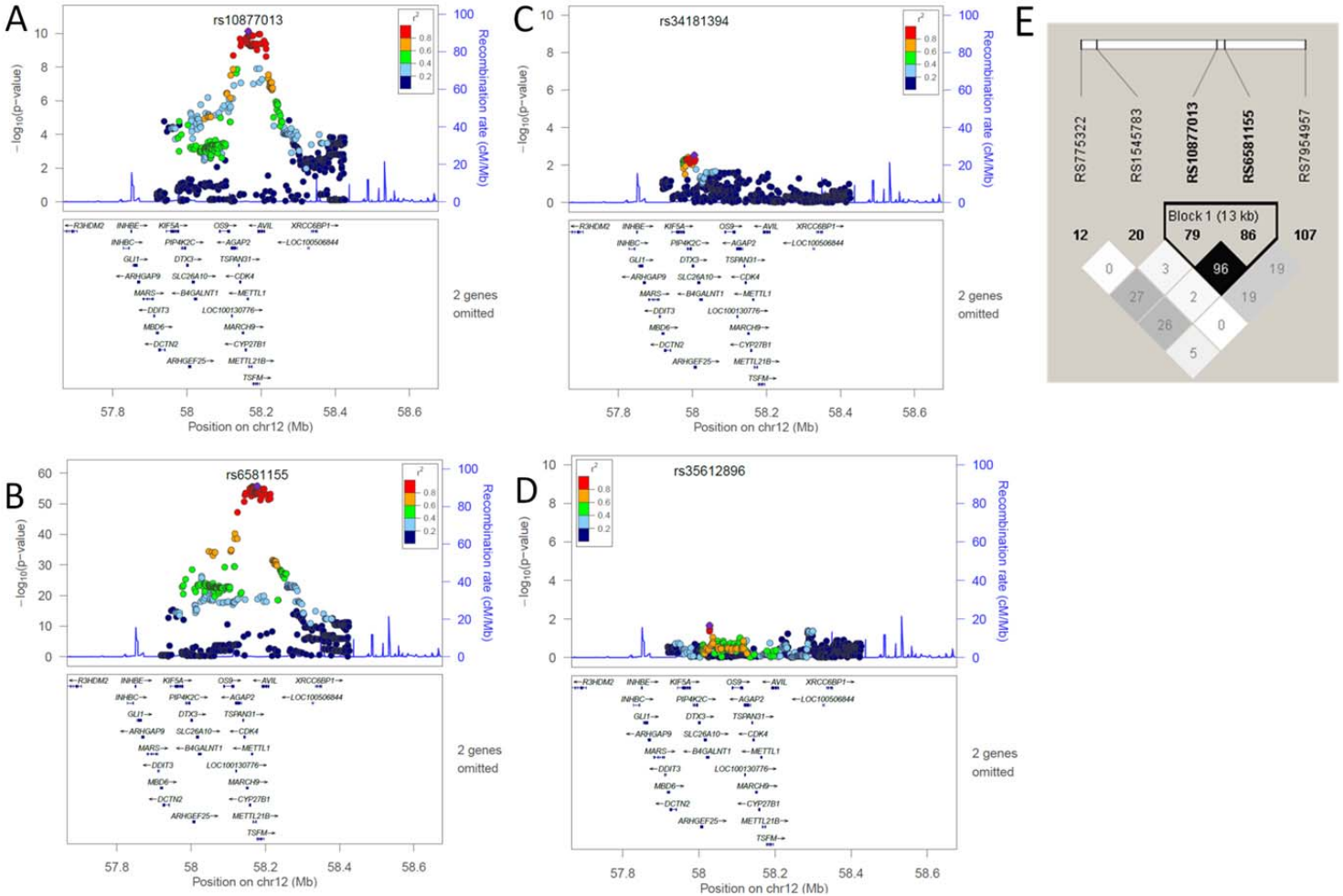


Figure 3.8. *cis*-eQTL with *METTL21B* global expression probe ILMN_1723846. High-resolution Manhattan plots of the *cis*-eQTL with *METTL21B* in primary B cells and monocytes. Strength of LD with the annotated lead SNP is illustrated by colouring shown in the key in each plot. A – B cell, and B – monocyte, *cis*-eQTL associations with *METTL21B* and surrounding SNPs. C and D – Conditional linear regression of *METTL21B* expression controlled for the lead SNPs labelled in panels A and B. No further associations are observed with this gene. E – Moderate LD exists between eSNPs and the proxy SNP for the trait-associated rs1678542.

3.3.4.2 Functional annotation of overlapping *cis*-expression quantitative trait loci and disease severity associated single nucleotide polymorphisms

SNPs in strong LD (r^2=0.9-1.0) with the eSNP for each observed *cis*-eQTL signal that were also in moderate LD ($r^r \geq 0.2$) with an RA susceptibility trait-associated SNP, including those from the conditional analyses, were input into RegulomeDB. Several loci displayed evidence of at least one SNP with annotated ENCODE and RoadMap Epigenome data that were likely to affect gene expression (Table 3.36). *AFF3* isoform 1 eSNP rs2309750, or any SNP in strong LD, did not show evidence of putative regulatory function, despite the strength of the association and estimated effect size (~40% variance in expression explained by this SNP). The strongest evidence for affecting gene expression regulation is for SNPs rs13397886 and rs62149357 which are predicted to affect a range of histone 3 lysine methylation modifications. Both SNPs are also predicted to overlap with an androgen receptor binding site detected from ChIP-seq experiments in a prostate cancer cell line (VCaP). The ENCODE project sampled a limited number of cell types and transcription factor binding sites; thus these SNPs may affect *AFF3* expression in resting B cells via regulatory factors not directly assayed in the ENCODE project. The RoadMap Epigenomics project data was accessed via the UCSC Genome Browser; rs62149357 showed limited evidence of overlap with a histone 3 lysine 36 trimethylation (H3K36me3) modification in CD19+ cells, the same population of cells this *AFF3* isoform *cis*-eQTL was detected in. Stronger evidence of regulatory function was found for SNPs in LD with rs11692215 and rs11690905 (Table 3.36). A SNP in LD with rs11692215 is predicted to affect a number of CD19 expressed transcriptional regulatory proteins, transcription factors and components of the transcriptional machinery; rs4850920 overlaps with multiple transcription factor binding regions as well as a binding region for a subunit of RNA polymerase II, the principle RNA polymerase responsible for transcription in eukaryote cells. This region is annotated as a weak promoter on the UCSC genome browser which

incorporates chromatin modifications, transcription factor ChIP-seq and RNA polymerase occupancy data into account. There is strong evidence that suggests rs13003982 as a functional candidate for the effect of rs11690905 on *AFF3* expression as it likewise overlaps with a large number of CD19+ expressed regulatory proteins. This SNP is also in stronger LD with rs1160542 than both rs2309750 and rs11692215 (r^2 0.25 vs. 0.43).

A number of SNPs in strong LD with rs13277113 overlap with annotated transcription factor binding regions (Table 3.36), including rs9693589 which has been described as disrupting RNA polymerase III binding *in vitro* [289]. Ge *et al* first described a *cis*-eQTL with *BLK* in LCLs from HapMap samples and showed direct functional evidence that a haplotype of SNPs overlapping the 5' untranslated region of the *BLK* mRNA were responsible for the differential expression of this gene in this cell type; strong LD exists between these promoter and 5' UTR SNPs.

In both the MuTHER and primary B cell data *C8orf13* is oppositely expressed to *BLK*; where SNPs are associated with lower levels of *BLK* transcripts, the same alleles are associated with higher *C8orf13* mRNA levels (Table 3.34). This may indicate a shared promoter region between these two genes that overlaps with the *cis*-eQTL association signals. These two B cell *cis*-eQTLs also share a number of eSNPs in common including the previously described rs9693589 which disrupts RNA polymerase III binding, in addition to a number of SNPs predicted to affect multiple transcription factor binding regions, regions of chromatin modification and annotated strong enhancer elements. The SNPs which affect the latter of these annotations may be strong functional candidates given the up-regulation of *C8orf13* transcripts associated with the minor alleles of these SNPs.

The eSNPs associated with *METTL21B* expression in both B cells and monocytes were in strong LD, as were 15 other SNPs, 7 of which are assigned strong evidence of affecting gene expression based on RegulomeDB annotation (Table 3.36). The majority of this evidence is

assigned based on their previous association with expression levels of a number of genes in the region in both LCLs and monocytes [290, 291]. There is a lack of data on CD14+ monocyte regulatory elements collected as part of the extant ENCODE and RoadMap Epigenome projects which hinders their use for this purpose. However, available annotation based on LCLs suggested rs2291617 as a strong functional candidate as it is predicted to affect an active promoter, and lies in the 5' UTR of the *METTL21B* gene.

Table 3.36. ENCODE annotation of candidate regulatory SNPs overlapping with RA disease severity associated SNPs

Trait-associated SNP	Cytoband	Gene (*cis*-eQTL)	Lead SNP	Candidate SNPs	Schaub *et al* RegulomeDB score	LD with trait-associated SNP	ENCODE annotation	RoadMap Epigenomics annotation
rs1160542	2q11.2	*AFF3* (isoform 1)	rs2309750	rs13397886	5	0.21	H3Kxmex, AR	NA
				rs62149357	5	0.21	H3Kxmex, AR	H3K36me3
			rs11692215	rs13019618	3a	0.25	H3Kxmex	H3K36me3
				rs13000759	4	0.25	H3Kxmex	H3K36me3
				rs4850920	4	0.25	Weak promoter, POL2RA, SPI1, IRF3, STAT3, USF2, H3Kxmex	HeK3me3, DNase HS
			rs11690905	rs13003982	2b	0.43	POL2RA, BHLH40, CHD2, EBF1, Mxi1, p300, WHIP, NFKB1, SP1, PAX5, POU2F2, Elk-1 – Footprint, DNase HS, H3Kxmex	DNase HS
				rs11123810	3a	0.43	weak promoter, NFKB1, BHLHE40, p300, SMC3 H3Kxmex	DNase HS
				rs4851252	4	0.43	weak promoter, POL2RA, SPI1, H3Kxmex	DNase HS, H3K4me3
				rs11900482	4	0.43	Strong enhancer H3Kxmex, H3K9/27Ac	H3K4me3
				rs2309756	4	0.42	H3K9/27me3	H3K9me3
rs13277113	8p23.1	*BLK*	rs1478901	rs2618476	3a	0.90	Active promoter, NFKB1, POL2RA, POU2F2, TBP, BHLH40E, EBF1, Mxi1, STAT1, DNase HS, HxKxmex	H3K4me3
				rs998683	4	0.91	Active promoter, POL2RA, NFKB1, EBF1, TBP, WHIP, HxKxmex	H3K4Me3

Trait-associated SNP	Cytoband	Gene (*cis*-eQTL)	Lead SNP	Candidate SNPs	Schaub *et al* RegulomeDB score	LD with trait-associated SNP	ENCODE annotation	RoadMap Epigenomics annotation
rs13277113	8p23.1	*BLK/C8orf13*	rs1478901/ rs2736332	rs13277113	3a	-	Strong enhancer, POL2RA, BHLH40E, Mxi1, p300, SMC3, STAT1, TBP, ZNF143, H3K27me3, H3K4me3	DNase HS
				rs2061831	2b	0.98	Strong enhancer, CTCF, SMC3, DNase HS, H3Kxmex	H3K9me3
				rs9693589	2b	0.99	Strong enhancer, NF-YB, H3Kxmex, Disrupts PolIII binding[a] [289]	-
				rs2736336	3a	0.97	Strong enhancer, BHLH40E, CHD2, EBF1, POL2RA, SMC3, TBP, ZNF143, DNase HS, H3Kxmex	-
				rs2736337	3a	0.97	Strong enhancer, BHLH40E, CHD2, EBF1, POL2RA, SMC3, TBP, ZNF143, DNase HS, H3Kxmex	-
				rs2736338	3a	0.97	Strong enhancer, BHLH40E, CHD2, EBF1, POL2RA, SMC3, TBP, ZNF143, DNase HS, H3Kxmex	-
rs13277113	8p23.1	*C8orf13*	rs2736332	rs4840568	2b	0.85	Strong enhancer, BHLHE40, CDH2, BRCA1, EBF1, IRF3, Mxi1, NF-YB, p300, STAT3, WHIP, ZNF143, POL2RA, NFKB1, TBP, RFX5, SPI1, JUND, ATF5 – Footprint, DNase HS, H3Kxmex	DNase HS, H3K4me3

Trait-associated SNP	Cytoband	Gene (*cis*-eQTL)	Lead SNP	Candidate SNPs	Schaub *et al* RegulomeDB score	LD with trait-associated SNP	ENCODE annotation	RoadMap Epigenomics annotation
rs13277113	8p23.1	*C8orf13*	rs922483	rs922483	4	0.80	Active promoter, BHLHE40, CHD2, POL2RA, CTCF, EBF1, Mxi1, ZNF143, WHIP, TAF1, TBP, DNase HS, H3Kxmex	DNase HS, H3K4me3
			rs4840550[b]	rs34069332	6	0.22	-	-
				rs13276836	6	0.21	H4K20me1	-
				rs4840550	6	0.22	H4K20me1	-
				rs7000132	6	0.21	H4K20me1	-
rs1678542	121q13.3	*METTL21B*	rs10877013/rs6581155	rs2069502	4	0.28	Weak promoter, POL2RA, WHIP, DNase HS, H3Kxmex,	MeDIP-seq, H3K36me3
				rs10877013	4	0.28	Weak promoter, H3Kxmex	H3K36me3
				rs2291617	4	0.27	Active promoter, BHLHE40, CHD2, Mxi1, POL2RA, ZNF143, H3Kxmex	PBMC RNA-seq, K3K4me1, K3K9Ac
				rs11172335	4	0.26	DNase HS, H3Kxmex	-
				rs1021469	4	0.26	Weak transcribed, H3K36me3, H4K20me1	PBMC – MeDIP-seq
				rs11172349	4	0.26	Weak transcribed, POL2RA, H3Kxmex,	PBMC – MeDIP-seq
				rs11172351	4	0.26	Weak transcribed, POL2RA, c-MYC-PWM, H4K20me1, H3Kxmex	PBMC – H3K36me3

Table 3.36. Candidate regulatory SNP annotations based on ENCODE and RoadMap Epigenome project data. Candidate SNPs were selected based on strong LD with eSNPs from both B cell and monocyte *cis*-eQTLs in Table 3.34, as well as conditional analyses as illustrated in Figures 3.3-3.8. SNP rs# IDs were input into RegulomeDB [280] where they are scored based on evidence of affecting regulatory elements using the amendment published by Schaub *et al* (Section 3.2.6: Table 3.4). SNPs were also visually inspected using the UCSC Genome Browser to identify evidence of affecting other regulatory elements measured in the RoadMap Epigenome and ENCODE project data. Annotations are recorded for the relevant cell type sampled, i.e. CD14+ monocytes, LCLs, CD19+ B cells and peripheral blood monocytes. HxKxmex – Histone x lysine x x-methylation (where x=any number), DNase HS – DNase hypersensitivity site, POL2RA – RNA polymerase III sub-unit 2, SPI1 – Transcription factor PU.1, IRF3 – Interferon regulatory factor 3, STAT1/3 – Signal transducer and activator of transcription 1/3, USF2 – Upstream transcription factor 2, BHLHE40 – Basic helix-loop-helix family, member e40, CHD2 – Chromodomain helicase DNA binding protein 2, EBF1 – Early B-cell factor 1, Mix1 – MAX interactor 1, dimerization protein, p300 – Histone acetyltransferase p300, WHIP – Werner helicase interacting protein 1, NFKB1 – Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1, SP1 – sp1 transcription factor, PAX5 – Paired box 5, POU2F2 – POU class 2 homeobox 2, Elk-1 – Elk1, member of ETS oncogene family, SMC3 – structural maintenance of chromosomes 3, TBP – TATA box binding protein, ZNF143 – zinc finger protein 143, CTCF – CCCTC-binding factor, NF-YB – Nuclear transcription factor Y, beta, RFX5 – Regulatory factor X, 5, JUND – Jun D proto oncogene, ATF5 – Activating transcription factor 5, TAF1 – RNA polymerase II, TBP-associated factor. a - Function assigned based on published literature, b – limited data collected on CD14+ monocytes in both ENCODE and RoadMap Epigenome projects – all SNPs with RegulomeDB score <7 were included for this eQTL rather than <5.

3.4 Discussion

3.4.1 Multiple testing burden – the elephant in the pharmacogenetics room

Testing multiple SNPs against one or more outcome measures raises the immediate issue of a burden of multiple testing. If one is to adopt a threshold for statistical significance to minimise the number of type I errors, it must account for how many times the dataset is tested. The simplest correction is a Bonferroni correction, which divides the type I error probability by the number of tests performed. For instance if the type I error probability is 5% and 10 tests are performed the threshold for statistical significance would $p<0.005$. A Bonferroni correction is conservative where tests are not statistically independent, as is often the case in genetic studies where genetic variants may be physically or statistically linked despite large genetic distances. The statistical power of a test is influenced by the p-value threshold, so an overcorrection for multiple testing comes at the cost of reduced statistical power. This is a particular concern for studies with small or modest sample sizes that attempt to test for association with common genetic variation that invariably has a small effect on the outcome of interest, be it disease susceptibility or treatment response. The logical solution is to increase the sample size as this allows a more robust approach. Statistical significance thresholds can be more stringent, multiple testing can be addressed appropriately and statistical power can be maintained. However, this has its own costs associated with it, namely logistical and financial. Recruiting large cohorts requires a population to sample from that is willing to participate in academic research, a clinical and informatics infrastructure that can deal with registration of patients and collection and storage of clinical and demographic data. For large data sets (1000+ patients) this may involve the collaboration of multiple centres or academic institutions. These are not insurmountable barriers and have been overcome by the genetic epidemiology field in the last decade, which has greatly facilitated the study of the genetic basis of complex diseases.

This approach should be, and is in some fields, echoed by research groups involved in pharmacogenetics. Though in pharmacogenetics, it is further complicated by heterogeneity between cohorts which can make combining them difficult and subsequently limits sample sizes.

These issues of heterogeneity, limited sample sizes and multiple testing burden are apparent in this doctoral work, as such results have been highlighted at $p \leq 0.01$, though this likely yields a number of false positive results. With this in mind the results from this chapter must be interpreted in light of the multiple testing burdens herein.

3.4.1 Rheumatoid arthritis susceptibility loci and disease modifying anti-rheumatic drug response

The aim of this doctoral work was to identify potential genetic biomarkers of treatment response in RA, as well as understand the impact of disease susceptibility loci on disease. Do the genetic loci that predispose to disease continue to act on disease prognosis? Do they directly impact on how and whether a patient will respond to their prescribed treatment?

The SJC28 counts how many joints are affected by inflammation-related swelling (i.e. synovitis), determined by a trained physician or research nurse, whilst the CRP level is a biochemical measure of the acute-phase response. These measure different aspects of RA pathology, and whilst they are often correlated, they may be subject to different influences and factors. In 1990 van der Heijde *et al* published their first paper on the DAS [246], which was later modified to the DAS28 based on clinical discussions of whether the foot joints should be included in clinical measures of disease activity [292]. Firstly this raises the issue of how treatment response is measured, and to what extent the current measures in the published literature are confounded by subjective components of composite disease activity scores. The DAS was originally developed to formalise physician treatment

decisions, such as when to prescribe a DMARD and when to remove them. The weighting attached to each component of the DAS, and by extension the DAS28, were determined from an analysis of "high" and "low" disease activity; categories which themselves were based on physicians' decisions to begin, continue or cease DMARD therapy. Thus the DAS28 includes subjective components (TJC28 and general health assessment), has the most weighting attached to subjective components (the TJC28 has the second strongest weighting) and is defined by subjective decision-making (all of the weightings are influenced by the original decisions of the treating physicians). This implies that the DAS28 is highly subjective and may not be appropriate as an outcome measure in pharmacogenetic association studies in RA.

The use of objective outcome measures is therefore paramount. In this doctoral work I have chosen to use the SJC28 and the CRP for this purpose. The DAS28 has been included as this is a familiar outcome measure and forms the basis for many of the treatment response categories used in the rheumatology literature.

The lack of consistent genetic associations between RA susceptibility SNPs and treatment response to combined DMARDs or either MTX or SSA monotherapy implies several possibilities:

- these SNPs do not individually have a clinically meaningful impact on treatment outcome
- treatment outcome is not measured objectively so any influence from genetic loci is masked by measurement error
- the individual influence of each SNP is small and requires larger, well powered cohorts to detect
- individual SNP effects are modulated by the therapy given, including concomitant non-DMARD therapies such as glucocorticoids which are unaccounted for

- unknown and unaccounted for heterogeneity in disease trajectory may mask the influence of individual SNPs

A biomarker by definition must be robust to differences between individuals, i.e. it must be generalizable across a population, it must predict or track a normal biological or pathological process and it must be easily measured. The RA susceptibility SNPs do not meet these basic criteria, and thus are not appropriate individually as biomarkers of treatment response in RA.

In the analysis of all DMARD treatment patients a SNP at the *TRAF6* locus was associated specifically with a lower improvement in SJC28 (Section 3.3.1.1; Table 3.6) and had no apparent effect on the improvement in CRP. No association was seen in the MTX or SSA monotherapy analyses either. This suggests this is a false positive association, or, that the analyses stratified by MTX and SSA were underpowered to detect this association. A larger cohort would be required to adequately test the effect of this SNP on treatment response.

The analysis of *HLA-DRB1* shared epitope alleles with DMARD response found a nominal association with a lower ΔDAS28 and ΔlnCRP. The observed effect was consistent in both the ΔDAS28 and ΔlnCRP analyses which suggest it may be a genuine effect as the SE alleles are also associated with progression of radiographic joint damage [293, 294] , though reports of an influence on treatment response to DMARDs and biologics conflict [218, 228, 295-297]. The effect of the SE alleles on treatment response was consistent in both the MTX and SSA monotherapy analysis which suggests that it has an effect through disease severity which may be related to its mechanism of effect on joint damage progression. Alternatively the observed association with joint damage is a result of treatment inefficacy in SE carriers as there is no effect observed with baseline joint damage in these data. Adjusting the analysis of joint damage progression for autoantibodies largely attenuated the effect of the SE alleles. This is consistent with reports that SE alleles are themselves

associated with susceptibility to ACPA rather than RA *per se* [27, 52]. There remained a small residual effect after this adjustment in the modified Poisson regression model. This might suggest an ancillary effect at the MHC not fully captured by the SE alleles [63]. The most recent association between RA susceptibility and the HLA region refined the association to amino acid positions 11, 13, 71 and 74 of *HLA-DRB1* and additional independent associations with *HLA-B* amino acids at position 9 and *HLA-DPB1* amino acids at position 9 [67]. It would be interesting to test whether the association with treatment response, in particular the improvement in CRP, is due to the same signals of association as that observed for disease susceptibility.

3.4.3 The effect of rheumatoid arthritis susceptibility loci on response to methotrexate and sulphasalazine monotherapy

Whilst MTX and SSA to an extent have overlapping mechanisms of action, the possibility exists of treatment-specific effects. This idea was tested using a subset of the YEAR cohort treated with MTX monotherapy and a cohort of patients from the placebo arms of several Roche-sponsored clinical trials. The only consistent result between the two cohorts was a nominal association with a SNP upstream of *FCGR2A* (Section 3.3.1.2; Table 3.9). This effect appeared to be confined to the ΔSJC28 in the YEAR data which may indicate either a lack of power in the Roche clinical trials data to detect its effect or a type I error. A fine-mapping and re-sequencing study of the low-affinity *FCGR* locus highlighted two non-synonymous polymorphisms in LD with this SNP [95]. Replication of this nominal association with MTX treatment response warrants pursuit as FcγRIIa binds CRP as well as IgG and may be directly involved in RA pathology as ACPA are able to activate macrophages in a FcγRIIa dependent manner [29]. There were no SNPs which were consistently associated with treatment response across all of the outcome measures in both the YEAR and Roche MTX data. Several SNPs showed high heterogeneity in the meta-analyses which

may be more indicative of the influence of large variation by chance in small data sets than genuine biological heterogeneity. This phenomenon was perhaps more striking in the SSA monotherapy analysis where several SNPs showed association in the ERAS data, but no effect in the YEAR data, and *vice versa* (Section 3.3.1.3; Tables 3.13-3.15). This includes the aforementioned SNP at the *TRAF6* locus which had opposite directions of effect between the YEAR and ERAS patients (lower Δ%DAS28 but a higher Δ%DAS). This once again highlights the need for large well powered cohorts to study pharmacogenetics of complex treatment response.

In the bullet points above the majority of problematic issues are well known, i.e. small datasets, small effect sizes, etc. However, there exists another issue, that of heterogeneity between datasets that cannot be accounted for using recorded clinical and demographic variables. These are issues which are not easy to overcome and are more likely facets of using observational cohorts compared to clinical trials data:

- not all patients will maintain therapy, particularly where iatrogenic side effects occur
- a patient may visit their GP and receive a glucocorticoid injection to alleviate inflammation and symptoms; this additional therapy may not be recorded and thus the patient may be perceived as “responding” better than they are in their next study visit
- co-morbidity or extra-articular manifestations may influence treatment response, or drug dosage, which are not recorded or accounted for in analyses
- “treatment response” is measured at specific time points and thus does not reflect the changes and fluctuations in pathology over time, only a snap-shot

- patient treatment and clinical decision making may differ from hospital to hospital, particularly if there is no standardised treatment protocol; this is not the case in the YEAR or Roche clinical trial cohorts, but may apply to the ERAS cohort
- drug toxicity may force a patient to withdraw from a specific therapy or drop out of a study

Of these issues only the last one could be accounted for in this study; patients in the monotherapy analyses were selected on the basis of their baseline treatment, whether they ceased it or not. This in itself may also introduce confounding as patients that either change their therapy or cease treatment are still included with patients that were treated for at least 6 months.

3.4.4 Rheumatoid arthritis susceptibility loci and radiographic joint damage

Erosive bone damage in RA is thought to be the most specific type of damage, i.e. it is partially driven by disease pathology. This is supported by observations of structural bone damage in ACPA positive healthy controls and arthralgia patients [39, 40]. The van der Heijde modification of the Sharp score (SHS) measures both erosions and joint space narrowing, thus allowing the investigation of both types of damage in unison or as separate entities.

The genetic basis of radiographic joint damage has been demonstrated in a cohort of RA patients from a homogenous island population; around 50% of joint damage progression in these patients was attributed to a heritable component [298]. The focus of genetic studies of joint damage in RA has been on progression of damage over several years rather than early joint damage. Whilst an analysis of joint damage progression risk has been performed in this doctoral work, the main focus is on the contribution of RA susceptibility loci to the baseline radiographic joint damage.

The distribution of joint damage in the YEAR cohort displayed considerable zero-inflation with a non-normal distribution of positive values. This precluded the use of a data transformation which has been commonly applied in the literature to allow the use of linear models for analysis. The zero-inflation in joint damage data has also been reported in the ERAS cohort with progression of Larsen score [294]. Using a ZINB regression model therefore allowed a model to be fit to the data that could be used to test each SNP against the SHS, as well as the separate damage components ERN and JSN. There were 4 SNPs associated with baseline SHS, two of which were in moderately strong LD (r^2>0.7). Based on the model estimates, these effects were limited to increases and decreases in SHS score rather than altering the odds of zero-inflation. In essence, these SNPs appear to affect the extent of joint damage, but not whether it occurs or not at this early point in the disease. Despite this it is worth noting that a biological or environmental factor that predisposes to joint damage may still affect the extent of joint damage. For example, a SNP may affect the severity of joint damage which can manifest as both a susceptibility to joint damage and extent of damage if they occur by the same or overlapping biological processes. A SNP of this nature would be most evident in an analysis that was suitably powered to test both the baseline and progression of joint damage. The joint damage progression analysis in this doctoral work was limited by the availability of data beyond two years and by the extent of missing data at this time point (baseline n=415, 2years n=272). No SNP associations were consistent between the baseline and risk of progression analyses, which suggests that these SNPs affect joint damage early in the disease process, are false positive associations, or that the progression analysis was underpowered to detect modest effects. In fact no single SNP was associated with the risk of radiographic joint damage (p<0.01), though a number of nominal associations were observed, including a SNP associated with worse MTX response at the *FCGR2A* locus (Section 3.3.1.2). The nominal association with progression risk was protective; this does not immediately lend itself to an intuitive

explanation and may indicate a false positive result. This is reinforced by the nominal association with the minor allele of rs2900180 at the *TRAF1/C5* locus and protection from ERN progression. An association has been reported with increased Larsen score up to 7.5 years of disease duration [294]. This disparity again suggests this is likely a false positive result.

One of the strengths of the SHS is that it can be separated into its component parts which measure ERN and JSN separately. The sub-analyses of the ERN found the majority of SNPs associated with SHS were also associated with ERN, and that larger effects were estimated in this sub-analysis. This was most evident for a SNP at the *C8orf13-BLK* locus (rs13277113) where a more than two-fold increase in baseline ERN score was estimated per-minor allele (Section 3.3.2.2; Table 3.19). The analyses of the JSN data did not see a comparable number of associations, though a borderline JSN-specific association was seen with the minor allele of a SNP at the *CDK6* locus. Together these indicate that the associations with baseline SHS are primarily driven by effects on the presence and extent of erosive joint damage in these data. Furthermore, this indicates a certain disconnection between ERN and JSN early in the disease, but with some overlap in processes. For instance, the association with the *C8orf13-BLK* SNP rs13277113 minor allele was notably stronger in the ERN analysis than the JSN, suggesting an effect on a biological process specific to erosive damage. The association with the minor alleles of linked SNPs at the *IL2RB* locus is suggestive of an effect on shared biology as the effects of these SNPs are comparable in both sub-analyses.

Regardless of the initial strength of a statistical association, replication is important in order to establish a robust association between an outcome and a predictor variable. Baseline joint damage data for a set of SNPs was kindly provided by Professor Annette van der Helm-van Mil (Leiden Medical Centre). No association was seen with the strongest

associated SNPs in the YEAR dataset for SHS, ERN or JSN. Rather a SNP on chromosome 12q13.3 which was nominally associated with baseline ERN was associated with baseline SHS and JSN in this Dutch dataset (Section 3.3.2.3 Table 3.22). The signal of association for this locus with RA susceptibility has been refined by fine-mapping to a region upstream of *OS9* [32]; fine-mapping this region may therefore yield a more accurate picture of the role of this locus in early radiographic joint damage. Interestingly, there was little evidence of zero-inflation in these replication data (SHS 11.6%). This raises several questions:

- Despite the strength of association in the YEAR data, are these false positive results?
- Despite comparable symptom duration (<24months) are there important disease-specific or clinical differences between the YEAR and Dutch cohorts that indicate a lack of genuine comparability for this type of analysis?
- Does the replication of a SNP minor allele with ERN in one dataset and JSN in another indicate a convergence of biological processes, or a shared one?

These questions cannot be answered directly from the data available; however, clues may be gathered from observation and investigation of correlated disease measures, such as inflammatory markers (CRP) or measures of synovitis (SJC28). These are proxy measures for biological processes associated with RA which may themselves be affected by genetic factors and impact on radiographic joint damage. SNP associations with multiple correlated outcomes may therefore give an indication of the general biological effect of a SNP on disease severity or implicate a particular biological pathway that would warrant further investigation.

3.4.5 Correlated disease activity and severity measures, joint damage, treatment response and rheumatoid arthritis susceptibility loci – exploratory analyses

The correlation of baseline clinical and disease activity measures showed a number of correlations that might be expected, for instance between CRP and SJC28, though the relative weakness of the correlation indicates some factors specific to each. This is reinforced by their individual weak correlations with other measures, for example the correlation between CRP and JSN (ρ=0.11); JSN was not correlated with SJC28 (ρ 0.07). Of particular interest was the lack of correlation between ERN and either CRP or SJC28, which suggests the early erosive damage in the YEAR cohort was not due to the untreated inflammatory burden in these patients, or at least not the inflammatory burden that could be measured by CRP or SJC28. Sub-clinical inflammation below the lower limit of detection for SJC28 and CRP measurement may yield to a more sensitive method of measurement. Thus the apparent discontinuity between inflammatory burden and joint damage may only be a technical artefact. SNPs at least nominally associated with baseline joint damage were tested in regression models against 5 disease outcome measures, including the ERN and JSN scores, where each model was adjusted for any measures correlated with it (Section 3.3.3.1; Figure 3.2A). No SNP was associated with additional outcome measures in this analysis. One conclusion is that early joint damage in the YEAR cohort is not driven by inflammation based on SJC28 and CRP. A more sensitive measure of synovitis such as ultrasound may be able to detect sub-clinical inflammation and resolve this apparent discrepancy between inflammation and joint damage. This is supported by the recapitulation of the relationship between synovitis and joint damage using MRI which is highly sensitive to inflammation undetected by either the SJC28 or CRP [13].

The correlation between disease outcome measures was extended to 1 year follow-up, which should incorporate treatment effects; the correlation between measures was

different compared to the baseline (Section 3.3.3.1; Figure 3.2B). ERN score was correlated with SJC28 at this time point (ρ=0.20), whilst CRP was no longer correlated with JSN (ρ=0.05). The central position of SJC28 in Figure 3.2B may indicate the central role of synovitis in disease and joint damage progression, rather than CRP which is a measure of the acute-phase response. Alternatively, it may also indicate the central role of synovitis and joint swelling in clinical decision making [246]. In essence, CRP is a biomarker of systemic inflammation whilst SJC28 is a measure of local RA-associated inflammation. One might expect genetic factors that influence synovitis (i.e. SJC28) to also affect joint damage due to the correlation of these two measures, including genetic factors associated with treatment response changes in SJC28.

A single SNP was nominally associated with more than one outcome measure at 1 year follow-up near a gene of unknown function (rs231707; *FAM193A*). The minor allele was associated curiously with a lower SJC28 but higher JSN. Firstly, there is a considerable burden of multiple testing in this analysis, so all results need to be interpreted in light of this. With that in mind SJC28 and JSN are moderately positively correlated at this time point (ρ=0.26) so one might expect a genetic effect on one to be congruent with the other. If not a false positive, this disparity may alternatively indicate a delayed treatment effect, i.e. SJC28 is reduced because of more aggressive treatment in carriers of the minor allele of this SNP, but any effect on JSN lags behind.

Two further observations merit a mention: a SNP at the *IL2RB* locus associated with lower baseline joint damage was nominally associated with a lower CRP at baseline (p=0.05) and 1 year AUC (p=0.02). This consistency across multiple time points and disease measures could be indicative of a genuine biological effect that warrants further investigation.

The second observation of interest is a SNP at the *TRAF6* locus associated with a worse improvement in SJC28; this SNP was associated with a higher SJC28 AUC (p=0.006). This

effect may have been driven by the effect observed after 6 months of treatment, or it may represent a genuine and consistent effect on synovitis.

3.4.6 *cis*-expression quantitative trait loci aid the interpretation of genetic associations

In an attempt to assign a functional role to the nominally associated SNPs from this chapter a search of publicly available gene expression data was performed. This found an overlap between 6 SNPs and *cis*-eQTLs within 0.5Mb. This particular range was selected based on observational evidence that regions within 0.5Mb of transcriptional start sites (TSS) are enriched for eSNPs, which rapidly drops off with greater genetic distance [277, 288, 290, 299]. There was very little strong LD between the trait associated SNPs and the eSNPs from the MuTHER expression data (Section 3.3.4; Table 3.33), with the exception of rs13277113 which has been associated with the expression of both *C8orf13* and *BLK* in multiple independent datasets of both LCLs and primary unstimulated cells [276, 277, 288, 290, 300].

The lack of strong LD between the majority of trait associated SNPs and lead eSNPs from the *cis*-eQTLs immediately suggests there is no relationship between these SNPs and the gene expression levels in question. However, a number of potentially confounding factors needs to be addressed:

- A SNP may affect the expression of a specific isoform that would not be detected by an analysis of global expression of the gene in question
- An eQTL of large effect may mask independent smaller effects on gene expression from SNPs related to the trait of interest or in LD with the trait associated SNP(s)
- An eQTL in one cell type does not preclude an effect in another cell or tissue type, though some eQTLs do exist across multiple tissue and cell types. This may include specific cell subsets, e.g. an eQTL may be present in CD4+ but not CD8+ T cells.

- A gene may require a stimulatory signal in order to be expressed; the majority of current expression data with available matched genotyping has been performed on unstimulated cells or tissue
- The context in which a SNP affects a trait may be specific to the conditions of that trait and this may not manifest within a non-trait/disease environment

The first point was observed with a *cis*-eQTL with *AFF3* which was specific to isoform 1 of this gene; no *cis*-eQTL was seen for isoform 2. The second of these was addressed using a conditional regression approach which adjusted the expression analysis for the strongest eSNP. In combination with the isoform specific expression of *AFF3* this revealed a complex signal of multiple independent and partially independent eSNP associations (Section 3.3.4; Figure 3.5). A number of SNPs that overlap with strong promoter elements and RNA polymerase II binding sites based on the ENCODE data annotations were in moderate LD with the baseline joint damage associated SNP (r^2=0.43); this may indicate a complex relationship between expression levels of this isoform and joint damage. Further investigation of the product of this isoform in multiple cell types is now warranted to establish whether or not it affects radiographic joint damage in early RA.

Further evidence of complex relationships between trait associated SNPs and eSNPs were evident at the *C8orf13-BLK* locus. The *cis*-eQTL with *BLK* in LCLs was functionally refined to a haplotype of SNPs in the 5' UTR of the *BLK* mRNA that is in strong LD with the baseline ERN score associated SNP rs13277113 [288]. However, the *cis*-eQTL with *C8orf13* was not simple and exhibited evidence of further independent eSNPs in a conditional linear regression analysis; the lead eSNPs from this and subsequent conditional analyses were in strong LD with rs13277113 which indicated a haplotype modulating the expression of this gene of unknown function in primary B cells (Section 3.3.4; Table 3.35) . Indeed, 7 SNPs in strong LD, including rs13277113, overlap with active promoter elements in the intergenic

region between *C8orf13* and *BLK* (Section 3.3.4.1; Table 3.36). Simpfendorfer *et al* found the B cell *cis*-eQTL was not evident in all B cell subsets they studied, but was confined to umbilical rather than adult B cells, particularly naïve and transitional B cells [301]. They also observed differential allelic expression in *BLK* mRNA in CD4+ and CD8+ T-cells which was not seen at the protein level.

C8orf13 is a gene of unknown function which encodes the protein FAM167A. The predicted amino acid sequence shows >90% homology to full length predicted proteins in 6 species within the primate clade, with >80% with a further 17 non-primate species from *Felis catus* (domestic cat) to *Cricetulus griseus* (Chinese hamster). This implies strong sequence conservation across mammalian species; the sequence is present in evolutionarily divergent species such as puffer fish (*Takifugu rubripes*) and zebra finch (*Taeniopygia guttata*) at >60% homology. The presence of a strong *cis*-eQTL that overlaps with a region of association with both RA susceptibility and possibly early radiographic damage suggests *C8orf13* as a region that requires further functional and genetic investigation to establish what role, if any, it has to play in the biology of these traits.

3.4.7 Rheumatoid arthritis, biological complexity and genetic associations

Figure 3.9 illustrates a simplified view of the relationship between a functional genetic variant (in this case a SNP) and the biological processes and pathways it impacts upon. This illustration assumes a 1:1 relationship between each step from functional SNP to gene/protein function to cellular response; a relationship which is unlikely to represent the normal state of genotype-phenotype relationships, and may only represent the “low hanging fruit”. Nevertheless, it provides a framework from which one can try to understand and interpret the findings of genetic associations with complex diseases and traits.

This chapter has investigated candidate SNPs based on their previous association with RA susceptibility. In essence it has attempted to fulfil two parallel aims: to find predictive biomarkers of treatment response and prognostic biomarkers of disease severity and, to understand the effect of these SNPs on RA pathology by investigating disease-specific phenotypes (i.e. radiographic joint damage, SJC28, CRP, etc.). This second aim represents the "physiological process" of Figure 3.9, whilst the *cis*-eQTL analysis represents the relationship between the "functional SNP" and "Gene/protein function". Further experimental investigation of "cellular response" and "tissue structure and function" could focus on the genes and proteins identified from this work. For instance characterising the tissue and cellular expression of FAM167A might highlight a role for this protein of unknown function in RA. *IL2RB* encodes the β chain of the IL-2 receptor (CD122) which is expressed on a number of RA relevant cell types including T-cell subsets, B cells and NK cells. It forms part of the IL-15 receptor required by NK and T-cells which may regulate their activation and development. To add further weight for the potential role of IL-15 signalling and function in joint damage pathology, Knevel *et al* report a signal of association with several SNPs in strong LD and increased rate of joint destruction at the *IL15* locus [302].

An attempt to replicate several baseline joint damage associated SNPs did not find any association with *IL2RB, AFF3* or *C8orf13-BLK*; this might be due to an oversimplification of how symptom duration is defined and recorded. The symptoms reported by patients in the YEAR cohort may have been caught much earlier than in the Dutch cohort because of the recruitment process. Patients were recruited from GP referrals across the Yorkshire region. This may go some way to explaining the zero-inflation compared with the second dataset, where patients may not have been referred until further in the disease process, giving them more time to accrue damage.

More in-depth phenotyping across larger cohorts will be required to properly test the effect of these SNPs on RA severity. For several loci multiple SNPs were tested for their association with the outcome measures in this chapter; however, for the majority only a single SNP was tested for association. Fine-mapping, particularly of the baseline joint damage associated loci may reveal a more refined picture of the genetic architecture of radiographic damage.

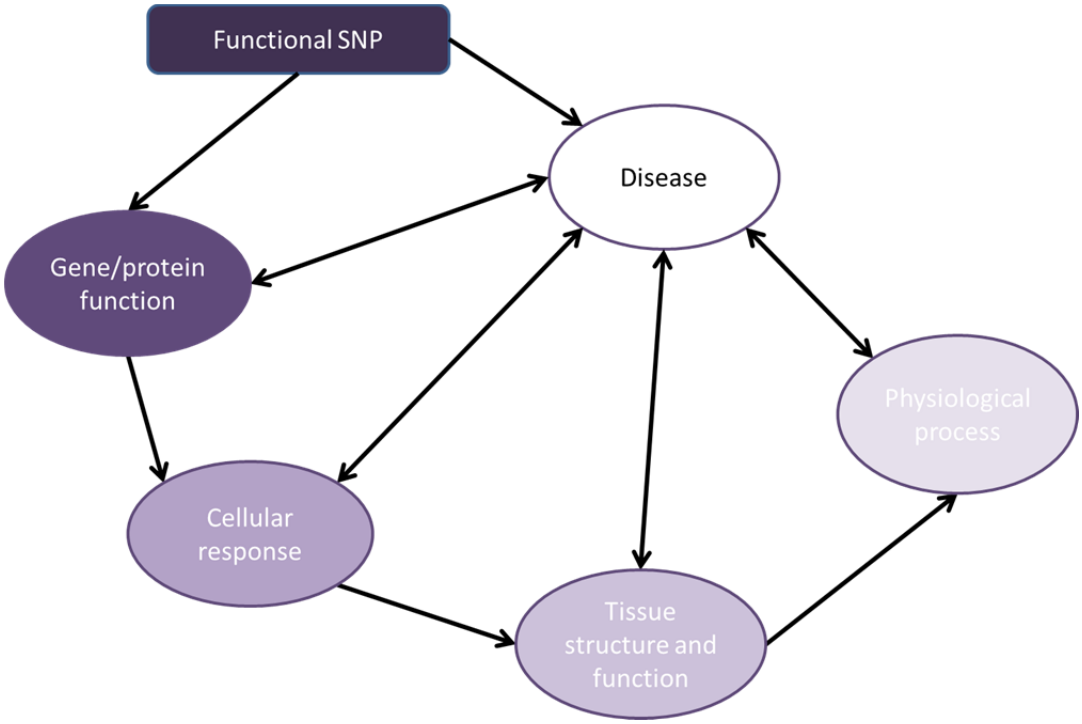


Figure 3.9. Genotype-phenotype relationships and the interpretation of genetic associations. The association between a SNP and trait or disease can be statistically robust, but does not necessarily lend itself to an immediate functional interpretation. Experiments that characterise the relationship between the SNP(s) of interest and biological processes related to the associated disease or trait may be subject to confounding from the trait itself. For instance differences in cellular response to a given stimulus may be due to the influence of the SNP in question or confounding from disease specific processes. One would expect that as an experimental output is further removed from the genetic level the potential for confounding factors would increase and the strength of association or correlation would diminish – this is represented by the gradient of dark to light purple; larger sample sizes are therefore needed to detect these subtle genetic effects. The two-way arrows illustrate the potential for disease-specific effects to confound experimental interpretation. The establishment of a relationship between genetic variants and a trait, as well as multiple related phenotypes can help in understanding of how a SNP predisposes to a disease or impacts upon a particular trait. The first stage in this process would be to look for a relationship between associated genetic variants and a simple molecular biological process such as gene expression quantity. Any correlation between gene expression and SNP allele could then be extended to further disease or trait related cellular processes.

Chapter 4 – Methotrexate pharmacogenetics: a candidate gene association study of methotrexate metabolism, transport and mechanism of action

4.1 Introduction

This chapter initially describes an association study and meta-analysis of two candidate polymorphisms previously associated with toxicity and sensitivity to MTX as a chemotherapeutic agent [170, 303]. A number of small studies and heterogeneous meta-analyses have been conducted [304, 305] using a mixture of response outcome measures in RA. The meta-analysis described herein was designed using a more stringent set of study inclusion criteria in mind to minimise sources of heterogeneity. Given the lack of association between these variants and MTX response a more detailed candidate gene study was undertaken.

The second part of this chapter describes this candidate gene study across regions that contain genes in the metabolism and proposed mechanism of action of MTX. Rather than focus on tagging SNPs or candidate non-synonymous changes as in previous studies [189] a more comprehensive approach was taken. This was in part driven by differences in genotyping platform within and between the two cohorts used in this analysis, but also in order to capture non-coding variation that may affect response to MTX.

4.2 Methods

4.2.1 Methylene tetrahydrofolates reductase single nucleotide polymorphism genotyping

Two SNPs were selected for genotyping in the *MTHFR* gene based upon their functional roles using a restriction fragment length polymorphism (RFLP) assay [172]. Specific sequence surrounding each SNP of interest, rs1801131 and rs1801133, were amplified using PCR following previously described protocols for each SNP [171, 172] (Figure 4.1). These protocols use two separate restriction endonucleases, *HinfI* and *MboII* (New England BioLabs Inc., Hitchin, Hertfordshire, UK) for rs1801133 and rs1801131 respectively, resulting in digested fragments on a gel indicative of genotype (Figure 4.2). These differences arise through the creation or abolition of recognition sites for these restriction enzymes specific to each allele of each SNP. The rs1801133 T allele creates a *HinfI* site which cleaves the 198bp amplicon into two fragments of 175 and 23bp.

The rs1801131 C allele abolishes the fourth *MboII* recognition site leaving four cleavage sites, as opposed to five for the wild-type allele. The expected fragments from this digest of the wild type amplicon would be 56, 31, 30, 28 and 18bp; whereas the C-allele amplicon generates fragments of 84, 31, 30 and 18bp. Fragments were resolved using 3.0% Tris borate EDTA (TBE) agarose gel electrophoresis for rs1801133 and 8% polyacrylamide microtitre array diagonal gel electrophoresis (MADGE) for rs1801131 [306]. Genotyping by this technique was carried out by Mr Stephen Martin, Mrs Nashwa El-Shaarawy and Mr Mike Morgan (the doctoral candidate).

rs1801133 RFLP

```
                              |HinfI
5'-TGAAGGAGAAGGTGTCTGCGGGAG[C/T]CGATTTCATCATCACGCAGCTTTTCTTTGAGGCTGACACATTCTTCCGCT
MTHFR C677T F>>                *rs1801133
TTGTGAAGGCATGCACCGACATGGGCATCACTTGCCCCATCGTCCCCGGGATCTTTCCCATCCAGGTGAGGGGCCCAGGAGAG

CCCATAAGCTCCCTCCACCCCACTCTCACCGCACCGTCCT-3'
                     <<MTHFR C677T R
```

rs1801131 RFLP

```
                            |MboII                          |MboII
5'-CTTTGGGGAGCTGAAGGACTACTACCTCTTCTACCTGAAGAGCAAGTCCCCCAAGGAGGAGCTGCTGAAGA
MTHFR A1298C F>>                                |MboII
TGTGGGGGGAGGAGCTGACCAGTGAAG[A/C]AAGTGTCTTTGAAGTCTTCGTTCTTTACCTCTCGGGAGAACCAAACC
         |MboII                *rs1801131 |MboII
GGAATGGTCACAAAGTG-3'
<<MTHFR A1298C R
```

Figure 4.1. Sequence context of the RFLPs for rs1801133 and rs1801131 using the restriction enzymes *HinfI* and *MboII* respectively. The forward primer sequence for each assay are shown higlighted in green and the reverse primer sequences are highlighted in red. Each box represents the restriction enzyme recognition site used for each SNP assay – *HinfI* for rs1801133 and *MboII* for rs1801131. The primer name and SNP sites are annotated under the corresponding sequence

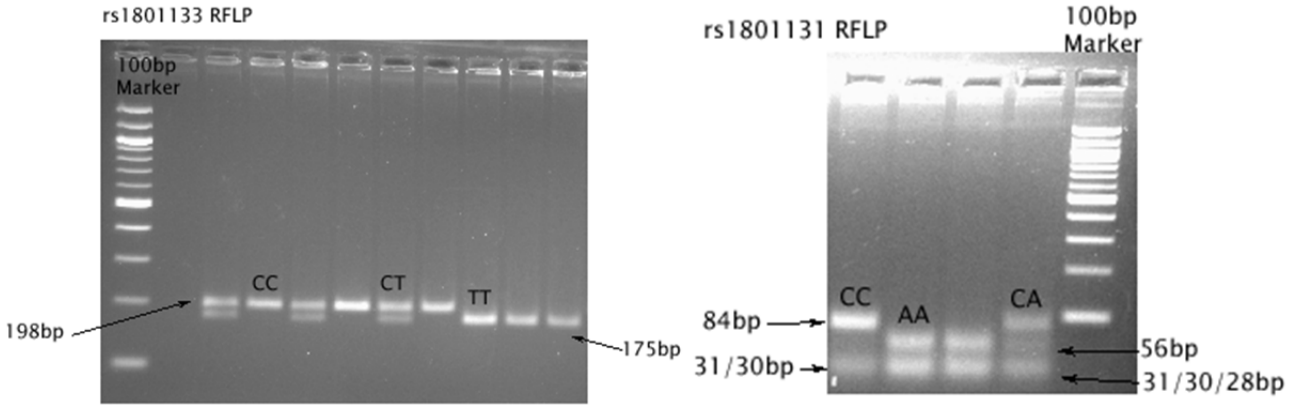


Figure 4.2. Example RFLP results for rs1801133 and rs1801131. RFLP band patterns are shown for each SNP genotype with the relevant sized bands as described in Figure 4.1; individual genotypes are annotated with fragment sizes. DNA ladder sizes correspond to a 100bp marker (New England BioLabs Inc., Hitchin, Herts, UK). A – RFLP band pattern for rs1801133. The DNA ladder shows the RFLP bands migrate further than the 200bp band. B – RFLP band pattern for rs1801131. The DNA ladder shows all of the RFLP bands migrate more than the 100bp band, the smallest band in this DNA ladder.

4.2.2 Meta-analysis

In order to test the association between *MTHFR* functional SNPs and MTX treatment response information from the literature was collated to perform a meta-analysis. Prior to, and during, the course of this doctoral work four meta-analyses were published attempting to answer the question of whether two functional *MTHFR* SNPs were associated with MTX outcomes [183, 304, 305, 307]. All four varied in the number of studies included in the final meta-analysis, largely based on the selection criteria for the studies. Prior to the commencement of this doctoral research program only a single meta-analysis had been published by Fisher *et al* [304]. They report heterogeneity within their meta-analysis of the C677T polymorphism and MTX toxicity, but not the metric used to quantify or identify it. They did not address the question of the impact of these SNPs on MTX response.

It was felt that by combining a set of rigorous study inclusion criteria for published studies with new primary data on *MTHFR* SNPs and MTX response in the YEAR and Roche clinical trials cohorts we would be able to better test the relationship between these two functional variants and MTX response. Original literature was identified from PubMed and ISI Web of Science searches using the terms "*MTHFR* methotrexate rheumatoid" and "methotrexate response rheumatoid genetic". The following criteria were then applied to exclude ineligible studies:

- studied in RA only (no juvenile arthritis or other autoimmune diseases)
- original research article (no reviews, comments or meta-analyses)
- abstract/full text article available
- address the questions of MTX response
- report OR/RR with 95% CI, regression coefficients and/or genotype counts
- minimum sample size n=100
- use of comparable response measures, i.e. DAS28, DAS, EULAR response

- MTX monotherapy dose 10-25 mg/week, either oral or subcutaneous

The meta-analysis was performed using a random effects model with the pooled effect size estimated using the method of DerSimonian and Laird [308]. Heterogeneity was quantified using the I^2 statistic [274]. Figure 4.3 illustrates the identification and exclusion criteria used in this systematic review and meta-analysis.

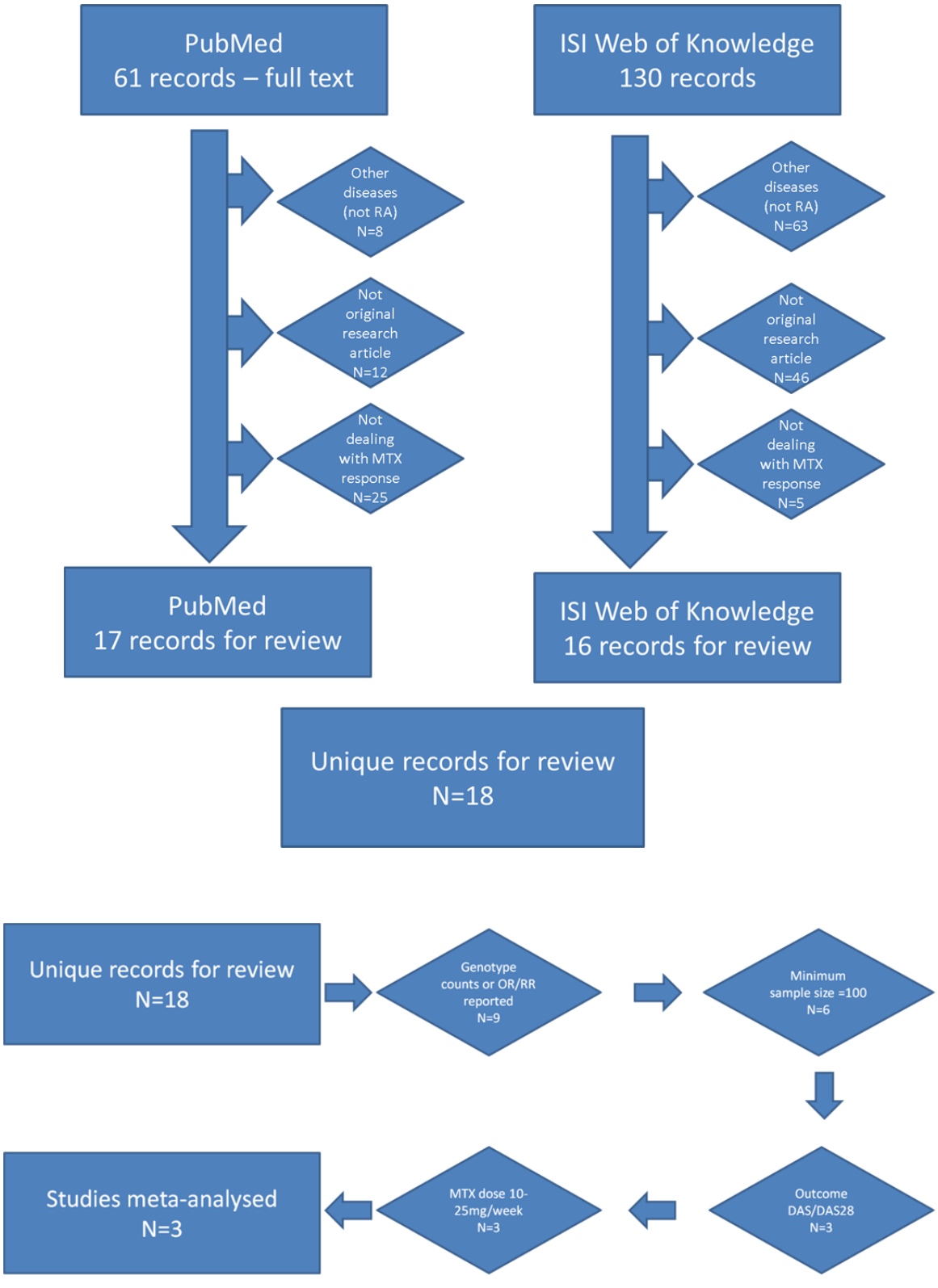


Figure 4.3. Systematic review and study exclusion criteria flowchart. PubMed and ISI Web of Knowledge were queried for original research articles investigating the role of *MTHFR* SNPs on MTX response in RA; 18 eligible articles were identified from the systematic review. Each was read in full, applying the exclusion criteria on reported data, minimum sample size, reported outcome measure and MTX dose.

4.2.3 Methotrexate metabolism candidate gene association study

4.2.3.1 Methotrexate metabolism candidate gene selection

Candidate genes were identified based upon their putative role in MTX metabolism and mechanism of action (Table 4.1 and Chapter 1.3.2.1 Figure 1.3). Pathways were identified in PharmGKB (http://www.pharmgkb.org/) using the search term "methotrexate". Non-redundant genes listed under the heading "Antimetabolite Pathway" and "Methotrexate Pathway (Cancer Cell), Pharmacodynamics" were listed as candidates. In addition, given that MTX is thought to work through adenosine accumulation and signalling, the adenosine receptor genes were also included. This list included several genes listed as candidates for MTX pharmacogenetics in a recent PharmGKB summary (e.g. *PPAT, ITPA*) [309].

4.2.3.2 Single nucleotide polymorphism genotyping quality control and imputation

Genomic intervals 500kb upstream and downstream of each gene (GRCh37/hg19) were selected for imputation from a set of reference haplotypes (1000 Genomes release March 2012) using IMPUTE2 into the YEAR and Roche clinical trials datasets (Chapter 2.1.1 and 2.1.3). Pre-imputation QC removed SNPs with MAF≤0.03, <98% genotype call rate and statistically significantly deviation from Hardy-Weinberg equilibrium (HWE, $p<0.001$). Sample QC was based on identity by descent (IBD) ≥0.25, genotyping call rate <98% and gender discordance between reported and X chromosome genotypes. Within-study population stratification was assessed using principal components analysis (PCA) compared to the HapMap phase III reference population samples. Imputation and genome-wide genotyping QC were carried out by Mr John Taylor.

4.2.3.3 Methotrexate metabolism candidate gene association analysis

In order to test the association between SNPs in MTX metabolism genes and outcome, linear regression was used to estimate effect sizes and calculate p-values with the ΔDAS28-CRP as the primary analysis dependent variable. The model was adjusted for the baseline

DAS28-CRP as described in Chapter2.3.4. The improvement in the natural log transformed CRP (ΔlnCRP) and the improvement in the SJC28 (ΔSJC28) variables were also used as dependent variables in order to test the association between these specific DAS28 components and MTX metabolism SNPs in secondary analyses. Imputed SNPs with an INFO score <0.8 were excluded from the analysis, and the YEAR imputed data were used in genotype dosage format, whilst the imputed Roche data used the best guess genotype. This was because specific SNP genotype data were required for the *MTHFR* meta-analysis and the investigation of RA susceptibility loci in Chapter 3. All association analyses were carried out using the linear regression function in PLINK and assumed an additive effect. Association p-values were plotted on high-resolution Manhattan plots using LocusZoom [267]. Genomic inflation of p-values was visualised by quantile-quantile (Q-Q) plots. A threshold for statistical significance was set at $p<5x10^{-5}$, with a second tier of nominal association highlighted at $p<5x10^{-4}$. The lower threshold for significance was based on 1000 independent tests at each locus (m) with a false positive rate of 5% (α) so: $\frac{\alpha}{m} = \frac{0.05}{1000} = 5x10^{-5}$.

4.2.3.4 Meta-analysis of metabolism association studies

The association results from the YEAR and Roche data were combined by meta-analysis for each of the three outcome measures. A random-effects model was used to combine the two data sets in PLINK. Heterogeneity was assessed using the I^2 statistic. Meta-analysis association p-values were assessed for inflation using Q-Q plots and plotted on high-resolution Manhattan plots using LocusZoom. Statistical significance threshold for association was set at $p<5x10^{-5}$, with nominal associations highlighted with $p<5x10^{-4}$.

4.2.3.5 Statistical power calculations

Statistical power for the MTX metabolism gene association analysis was calculated across a range of MAF (0.05-0.5) and effect sizes (β_G 0.3-2.4) for the primary outcome measure

ΔDAS28-CRP ($\bar{x}$ = 1.74, σ=1.78). Calculations were performed assuming an additive effect on the phenotype, a case-only design of unrelated individuals (n=478) with a continuous variable phenotype using Quanto v1.2.4 (http://hydra.usc.edu/gxe/).

Table 4.1. Candidate gene regions for association with MTX response

Gene	Cytoband	Pathway
MTHFR	1p36.22	Folate metabolism
AMPD1	1p13.2	Adenosine metabolism
ADORA3	1p13.2	Adenosine signalling
ADORA1	1q32.1	Adenosine signalling
MTR	1q43	Homocysteine metabolism
ATIC	2q35	Adenosine accumulation
PPAT	4q12	*De novo* purine synthesis
ABCG2	4q22.1	MTX transport
MTRR	5p15.31	Homocysteine metabolism
DHFR	5q14.1	Folate metabolism
TPMT	6p22.3	Purine metabolism
SLC29A1	6p21.1	Adenosine transport
NT5E	6q14.3	Adenosine metabolism
ABCB1	7q21.12	MTX transport
GGH	8q12.3	MTX metabolism
FPGS	9q34.11	MTX metabolism
ADK	10q22.2	Adenosine metabolism
ABCC2	10q24.2	MTX transport
MTHFD1	14q23.2	Folate metabolism
MTHFS	15q25.1	Folate metabolism
ADORA2B	17p12	Adenosine signalling
SHMT1	17p11.2	Folate metabolism
TYMS	18p11.32	Folate metabolism
ITPA	20p13	Purine metabolism
ADA	20q13.12	Adenosine accumulation
GART	21q22.11	Adenosine accumulation
SLC19A1	21q22.3	MTX transport
CBS	21q22.3	Homocysteine metabolism
ADORA2A	22q11.23	Adenosine signalling

4.3 Results - Methotrexate pharmacogenetics

4.3.1 Methylene tetrahydrofolate reductase functional polymorphisms and methotrexate response

Three eligible studies were identified from the literature search for use in a meta-analysis of the *MTHFR* functional SNPs rs1801131 and rs1801133 (Table 4.2). Genotyping of these two SNPs was performed by a RFLP assay on the YEAR cohort, and genotype data were extracted from the genome-wide genotyping of these SNPs in the Roche clinical trials patients described in Chapters 2.1.1 and 2.1.3. Genotypes for each response category were manually extracted from each published article for each SNP. Including the YEAR and Roche trials patients there were data available on 986 patients for rs1801133 and 866 patients for rs1801131 (ref [187] did not test rs1801131).

Table 4.2. Studies for *MTHFR* SNP meta-analysis – low-disease activity (DAS28-CRP (≤3.2) at 6months)

Study	Population	Outcome	MTX dose	SNP	MAF	N	OR	ref
YEAR 2013	UK	DAS28≤3.2	10-25mg/week	rs1801131	0.338	300	0.87	n/a
				rs1801133	0.298	302	0.67	
Roche Clinical Trials 2013	USA, Europe	DAS28≤3.2	10-25mg/week	rs1801131	0.336	174	0.84	n/a
				rs1801133	0.336	174	1.29	
Stamp *et al* 2010	New Zealand	DAS28≤3.2	15mg/week	rs1801131	0.312	187	0.78	[185]
				rs1801133	0.374	187	0.73	
Lee *et al* 2009	USA	DAS28≤3.2	13mg/week	rs1801133	0.351	118	0.90	[187]
Wessels *et al* 2006	The Netherlands	DAS44≤2.4	20.7mg/week	rs1801131	0.350	203	1.30	[180]
				rs1801133	0.317	205	1.39	

Table 4. 2 - Meta-analysis studies summary information. Summary data shown are mean MTX dose reported, SNP minor allele frequency and reported OR for each study. MAF; minor allele frequency

The MAFs of rs1801131 were broadly comparable across all 4 data sets, whilst the MAF of rs1801133 varied more widely. LD exists between these two SNPs (r^2=0.24 HapMap CEU Phase III/Rel27 population) which form three haplotypes of similar frequency (AT; freq=0.36, CC; freq=0.32, AC: freq=0.31).

In the meta-analysis of rs1801133 the combined ORs for the genotype effect relative to the wild type homozygote group show no evidence of association with MTX response (Table 4.3 and Figure 4.4A-B). There was low heterogeneity in the two analyses (I^2=16.5-24.5%). The lack of association was mirrored in the meta-analysis of rs1801131 and MTX response (Table 4.4 and Figure 4.4C-D).

Table 4.3. Meta-analysis of effect of rs1801133 on MTX response (DAS28≤3.2)							
Study	n	$OR_{CT\ vs.\ CC}$	95% $CI_{CT\ vs.\ CC}$	% weight	$OR_{TT\ vs.\ CC}$	95% $CI_{TT\ vs.\ CC}$	% weight
YEAR 2013	302	0.76	0.46-1.26	29.24	0.67	0.30-1.49	29.49
Roche Trials 2013	174	1.53	0.81-2.90	18.65	1.43	0.56-3.70	21.24
Stamp *et al* 2010	187	0.59	0.31-1.12	18.41	0.73	031-1.72	26.32
Lee *et al* 2009	118	1.20	0.52-2.75	11.09	0.17	0.02-1.38	4.27
Wessels *et al* 2006	205	0.98	0.55-1.76	22.61	1.39	0.50-3.81	18.68
Pooled OR	**986**	**0.92**	**0.70-1.21**	**p=0.55**	**0.87**	**0.56-1.35**	**p=0.53**

Table 4.3 - Meta-analysis results for the effect of rs1801133 on MTX response. OR are presented for comparison of each genotype against the wild type homozygote group. OR – odds ratio, 95% CI – 95% confidence interval, % weight – percentage weight attributed to each study in the meta-analysis. I^2 CT vs. CC = 24.5%, I^2 TT vs. CC = 16.5%

Table 4.4. Meta-analysis of effect of rs1801131 on MTX response (DAS28≤3.2)							
Study	n	$OR_{AC\ vs.\ AA}$	95% $CI_{AC\ vs.\ AA}$	% weight	$OR_{CC\ vs.\ AA}$	95% $CI_{CC\ vs.\ AA}$	% weight
YEAR 2013	300	1.34	0.81-2.22	33.23	0.99	0.45-2.17	34.05
Roche Trials 2013	174	1.25	0.65-2.39	20.09	2.13	0.85-2.38	24.99
Stamp *et al* 2010	187	0.75	0.41-1.38	22.76	1.04	0.33-3.27	16.21
Wessels *et al* 2006	205	0.63	0.35-1.14	23.91	1.30	0.51-3.29	24.75
Pooled OR	**866**	**0.97**	**0.72-1.29**	**p=0.81**	**1.29**	**0.81-2.05**	**p=0.28**

Table 4.4 - Meta-analysis results for the effect of rs1801131 on MTX response. ORs are presented for each genotype group compared against the wild-type homozygote group. OR – odds ratio, 95% CI – 95% confidence interval, % weight – the percentage weight attributed to each study in the meta-analysis. I^2 AC vs. AA = 38.2%, I^2 CC vs. AA = 0.0%

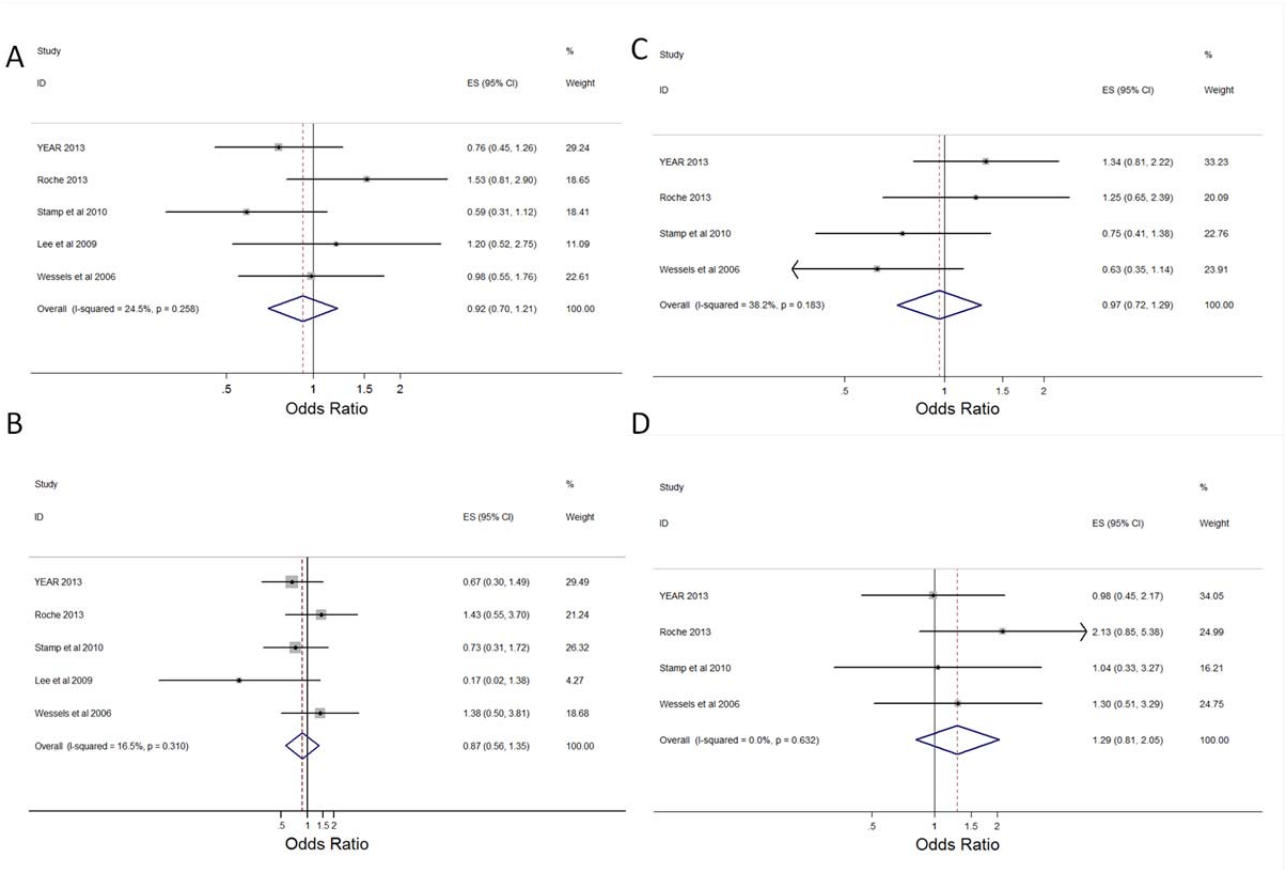


Figure 4.4. Forest plots of *MTHFR* SNP random effects meta-analyses for effect on DAS28/DAS low-disease activity threshold. The red dotted line represents the combined effect estimate from the DerSimonian and Laird random effects meta-analysis, and the horizontal width of the blue diamond represents the 95% confidence intervals. A – meta-analysis of rs1801133 CT vs. CC genotypes. The differences in odds ratio direction of effect is represented by the moderate I^2=24.5%. B – meta-analysis of rs1801133 TT vs. CC genotypes. C – meta-analysis of rs1801131 AC vs. AA genotypes. The odds ratios from these studies vary more in direction which is reflected by the higher I^2=38.2% and wider confidence intervals. D – meta-analysis of rs1801131 CC vs. AA genotypes. The odds ratios are in a more consistent direction of effect shown by the low I^2=0%. ES – effect size, 95% CI – 95% confidence interval.

4.3.2 Single nucleotide polymorphisms across methotrexate metabolism candidate genes and treatment response following methotrexate

4.3.2.1 Genotyping quality control and clinical data availability

Genome-wide genotyping was performed on two methotrexate-treated cohorts of RA patients (see Chapters 2.1.1, 2.1.3 and 2.2.3 for details). YEAR patients were selected for inclusion in the PhArmacogenomics of MEthotrexate in Rheumatoid Arthritis (PAMERA) consortium study based on a diagnosis of RA according to the 1987 or 2010 ACR criteria, symptom duration ≤12months at treatment initiation, MTX as first DMARD and clinical response data available at either 3 or 6 months. Only patients with treatment response data at 6 months were included in this doctoral work. A total of 380 samples were sent to consortium collaborators in the USA; 9 samples were not sent on to Japan for genotyping due to low concentration, 4 samples failed genotyping and were discounted and a further 4 were discordant for X chromosome SNP genotypes based on their reported gender and were removed prior to analyses. A total of 363 samples therefore had available genotyping data prior to QC which were assessed for IBD and sample genotype call rate. No patients were found to be closely related (IBD<0.25) and all had a genotyping call rate ≥98%. Patient ethnicity and population stratification were assessed using PCA. There were 21 individuals identified with non-European Caucasian ethnicity, the majority of which clustered with HapMap reference samples from South American and Indian sub-continent populations (Figure 4.5A). These samples were removed before analysis. A total of 349,566 SNPs were removed by QC prior to imputation based on call rate <98% (59,249), MAF<3% (284,233) and HWE p-value<0.001 (6,084). Many of these SNPs are low frequency non-synonymous SNPs which is a feature of the HumanOmniExpressExome BeadChip design.

Genome-wide genotyping data were available from 272 Roche clinical trial placebo arm patients (Chapter 2.1.3). PCA of these samples identified a large proportion of non-

European individuals, many of which matched their reported ancestry. Therefore 185 samples of European ancestry with matched demographic data were retained for analysis (Figures 4.5B-D). SNP QC was performed within each Roche clinical trial cohort prior to imputation; 229,268 removed in Ambition (3855 <98% call rate, 209,421 MAF<3%, 15,992 HWE p<0.001), 239,860 in FILM (1,308 <98% call rate, 211,916 MAF<3%, 26,636 HWE p<0.001) and 280,942 in IMAGE (1,598 <98% call rate, 202,821 MAF<3%, 76,523 HWE p<0.001).

Genotype data across the regions in Table 4.1 were extracted from each data set and imputed as described in Section 4.2.3.2.

Matched genotyping and clinical data for the primary analysis (ΔDAS28-CRP) were available on 304 YEAR patients and 174 Roche clinical trial patients. Following imputation a total of 132,770 and 96,215 SNP genotypes across the 29 genomic intervals were available in the YEAR and Roche data sets, respectively. SNPs with an INFO score <0.8 and MAF<0.05 were excluded from analysis, therefore a total of 52,803 and 71,104 SNPs were tested in the YEAR and Roche data for their effect on MTX response in the primary analysis. The study workflow is summarised in Figure 4.6.

A summary of the clinical and demographic data for the YEAR and Roche cohorts are shown in Table 4.5.

Table 4.5. YEAR and Roche MTX patient cohort summary and demographics				
Clinical/demographic variable	YEAR		Roche	
	Summary statistic	n	Summary statistic	n
Age at disease onset (years), mean (s.d.)	58.5 (13.2)	379	51.7 (12.7)	185
Disease duration (months), mean (s.d.)	8.5 (10.8)	356	4.6 (7.4)	185
Baseline DAS28-CRP, mean (s.d.)	5.29 (1.35)	369	5.45 (1.19)	185
Baseline SJC28, median (IQR)	7 (24)	377	13 (21)	185
Baseline CRP, median (range)	17 (0.1-339)	371	3.39 (0.4-119)	185
ΔDAS28-CRP, mean (s.d.)	1.44 (1.75)	304	2.26 (1.71)	174
ΔSJC28, median (IQR)	4 (45)	335	6.5 (36)	176
ΔlnCRP, median (range)	5.3 (-84-198)	314	2.1 (-38-112)	178
Gender, % female	68.9%	379	72.4%	185
Seropositive, %	75.9%	348	98.6%	142
ACPA, %	67.7%	263	94.6%	92
RF, %	69.0%	332	67.0%	185

4.3.2.2 Statistical power of association testing with the primary outcome measure

Statistical power calculations based on the primary outcome measure, using a sample size n=478, indicate there is >70% power to detect an additive effect of a 1.2 units of ΔDAS28-CRP at a MAF 0.05 and $p \leq 5.0 \times 10^{-5}$. An estimated β_G of this magnitude would explain ~4% of the variation in ΔDAS28-CRP. A SNP allele with a more moderate effect size, as is most often seen in complex traits, shows testing in a cohort of this size would be much lower powered. For example a test of a SNP minor allele with a frequency of 5% that explains just 1% of the phenotypic variance at a $p \leq 5 \times 10^{-5}$ would have less than 4% power. Therefore the association testing of SNPs in the genomic intervals around the MTX metabolism genes in Table 4.1 would only be adequately powered to detect the largest effect sizes of clinical relevance at the lower end of the allele frequency spectrum. Tests of more common SNPs (MAF 0.3-0.5) would have ≥70% power to detect an effect of 0.6 ΔDAS28-CRP units (Table 4.6).

Table 4.6. Statistical power for the primary outcome analysis			
MAF	β_G	R^2	Power
0.05	0.3	<0.01	<1%
	0.6	0.01	3.8%
	1.2	0.04	70.5%
	2.4	0.17	99.9%
0.10	0.3	<0.01	<1%
	0.6	0.02	18.1%
	1.2	0.08	99.0%
	2.4	0.33	99.9%
0.30	0.3	0.01	4.8%
	0.6	0.05	78.2%
	1.2	0.19	99.9%
	2.4	0.76	99.9%
0.50	0.3	0.01	7.5%
	0.6	0.06	89.1%
	1.2	0.23	99.9%
	2.4	0.91	99.9%

Table 4.6. Statistical power calculations for the primary outcome analysis. Power was calculated in Quanto v.1.2.4 assuming an additive model across a range of MAF and effect sizes. Calculations assumed a $\bar{x}$=1.74 and σ=1.78 for the ΔDAS28-CRP (Table 4.5), n=478, $p<5x10^{-5}$. MAF – minor allele frequency, β_G – effect size in units of ΔDAS28-CRP, R^2 – phenotype variance explained by the genetic model

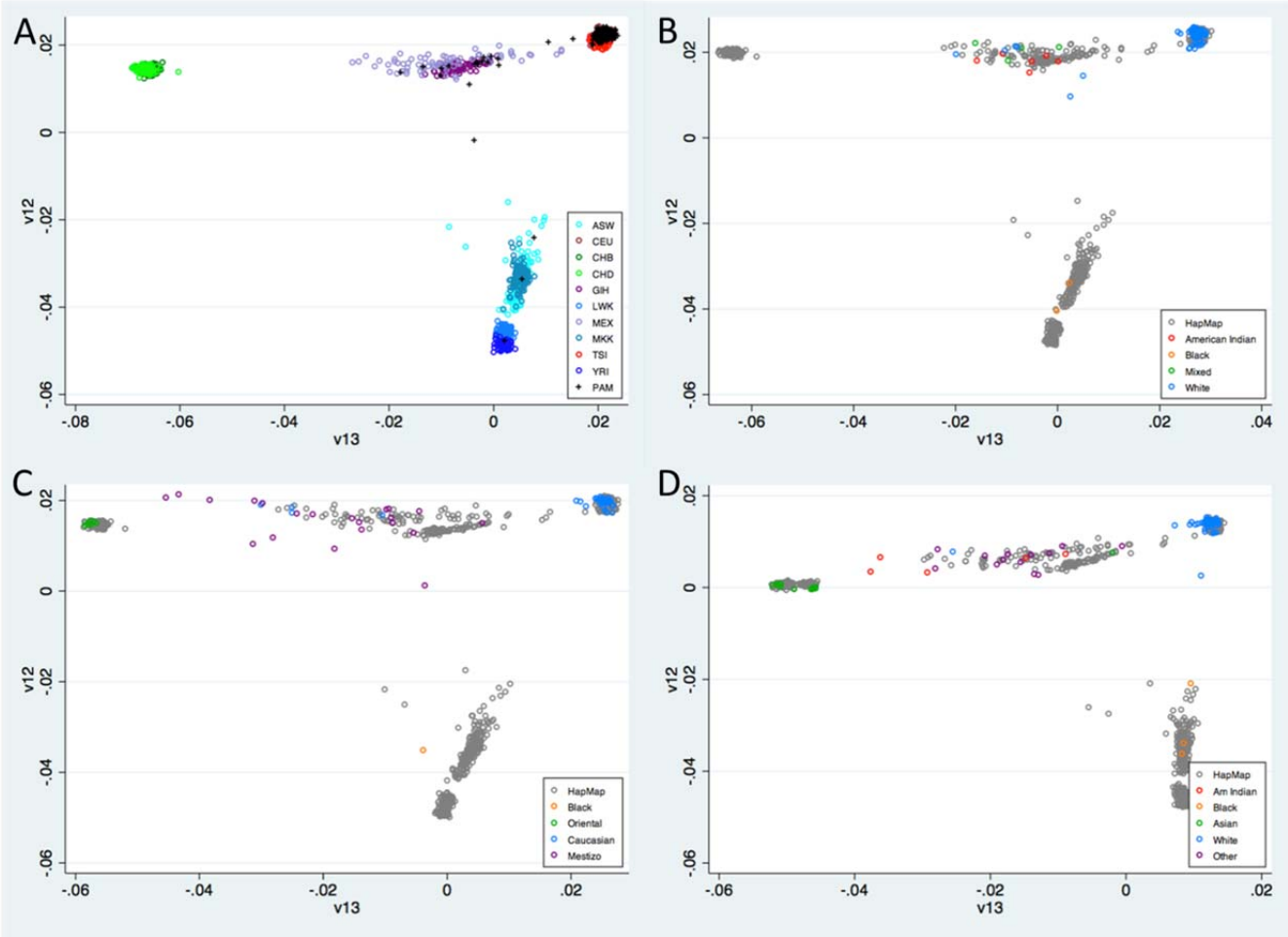


Figure 4.5. Principal components analysis of YEAR and Roche samples detects population stratification. The first two principle components from the PCA are plotted for each cohort: YEAR, Ambition, FILM and IMAGE against the HapMap phase III population reference samples. A – YEAR genotyped patients primarily cluster with the TSI (Toscani) and CEU (Northern European) samples. Individuals are colour coded according to their reference population (bottom right corner) and the YEAR samples are plotted as black crosses. B-D – Roche clinical trials (B – Ambition, C – FILM, D – IMAGE) are coloured according to their self-reported ancestry and the HapMap reference population samples are coloured in grey according to the key (bottom right corner of each plot). Only individuals that cluster with the CEU and TSI populations (top right corner of each plot) were retained for analysis.

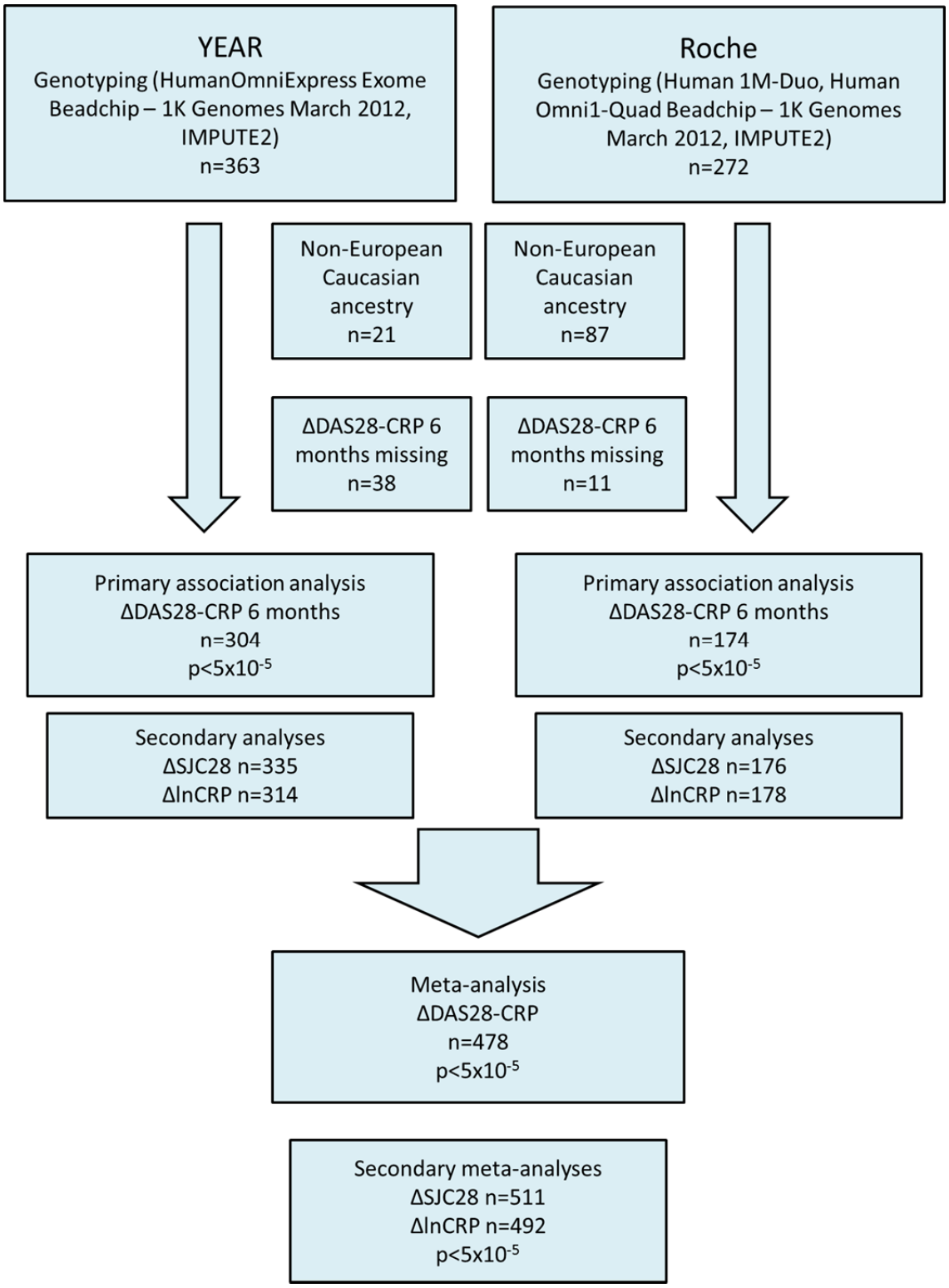


Figure 4.6. MTX metabolism candidate gene study design and workflow. The YEAR and Roche patient samples were genotyped across 3 different Illumina whole genome genotyping chips. Imputation of genotypes into each dataset was performed from reference haplotypes in the 1000 Genomes March 2012 release in IMPUTE2. Sample QC excluded 59 samples from YEAR and 98 samples from the Roche data. Association testing of the primary and secondary outcome measures was performed in PLINK prior to a random effects meta-analysis. Statistical significance was place at a threshold $p<5x10^{-5}$ for both primary analyses and the meta-analysis. YEAR – Yorkshire Early Arthritis Register, ΔDAS28-CRP – improvement in disease activity score with 28 joints, ΔSJC28 – improvement in swollen joint count of 28 joints, ΔlnCRP – improvement in natural log transformed C-reactive protein.

4.3.2.3 Single nucleotide polymorphism association with primary outcome in the Yorkshire Early Arthritis Register and Roche clinical trial cohorts

No SNP reached the threshold for statistical significance in either the YEAR or Roche datasets ($p \leq 5x10^{-5}$) for the primary analysis of ΔDAS28-CRP. A total of 99 SNPs reached the threshold for nominal significant association ($p \leq 5x10^{-4}$) in the YEAR data and 56 SNPs in the Roche data. The lead SNPs (smallest p-value) for all 29 genomic intervals are shown in Table 4.7. A meta-analysis of the YEAR and Roche data sets combined 38,115 SNPs across the 29 genomic intervals. The Q-Q plot of expected versus observed p-values from the meta-analysis shows deviation from the expected proportion of p-values at smaller values (Figure 4.7A). 15 SNPs reached the threshold for statistical significance within the *MTHFS* candidate genomic interval on 15q25.1 (Figure 4.7B). The lead SNP at this locus is associated with a lower ΔDAS28-CRP (β -0.72) and lies ~170kb downstream of the *MTHFS* coding region. The high-resolution Manhattan plot shows a second group of SNPs in weak LD with the lead SNP that are associated with a higher ΔDAS28-CRP (Figure 4.7B). These SNPs lie within the second intron of the *MTHFS* gene and may constitute a second independent signal of association at this locus. A summary of the lead SNP for each nominally associated locus in the meta-analysis is shown in Table 4.8, as well as whether they are imputed or genotyped and the effect estimates from the separate YEAR and Roche association analysis.

A further 21 SNPs across 3 genomic intervals reached the threshold for nominally significant association, the lead SNPs for which are highlighted in Table 4.7. There was no obvious enrichment of associations within any of the pathways that the candidate genes were attributed to.

Table 4.7. Candidate gene regions for association with MTX response ΔDAS28

Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFR	**1p36.22**	**Folate metabolism**	**rs4846014**	**$3.3x10^{-3}$**	**174**	**rs2745273**	**$5.8x10^{-3}$**	**304**	**rs2745273**	**0.465**	**$3.9x10^{-4}$**	**478**	**0**
AMPD1	1p13.2	Adenosine metabolism	rs2938323	0.01	174	rs191762325	0.01	304	rs4240537	0.114	$8.6x10^{-3}$	478	0
ADORA3	**1p13.2**	**Adenosine signalling**	**rs77779101**	**0.02**	**174**	**rs12123824**	**$5.4x10^{-5}$**	**304**	**rs4839129**	**0.450**	**$3.9x10^{-4}$**	**478**	**0**
ADORA1	1q32.1	Adenosine signalling	rs2486963	$1.5x10^{-3}$	174	rs9787266	$6.9x10^{-3}$	304	rs10800924	0.229	$1.9x10^{-3}$	478	0
MTR	1q43	Homocysteine metabolism	rs72764017	$2.4x10^{-3}$	174	rs78539606	$8.2x10^{-4}$	304	rs57657252	0.295	0.01	478	0
ATIC	2q35	Adenosine accumulation	rs62181409	0.02	174	rs12997148	0.01	304	rs4624380	0.082	0.01	478	0
PPAT	4q12	*De novo* purine synthesis	rs62310574	$2.3x10^{-3}$	174	rs14590771	$4.0x10^{-3}$	304	rs73240555	0.123	0.01	478	0
ABCG2	**4q22.1**	**MTX transport**	**rs146970867**	**$9.6x10^{-5}$**	**174**	**rs6817223**	**$5.3x10^{-4}$**	**304**	**rs6817223**	**0.326**	**$5.6x10^{-4}$**	**478**	**12.9**
MTRR	5p15.31	Homocysteine metabolism	rs6555490	$5.3x10^{-5}$	174	rs12520142	$4.4x10^{-3}$	304	rs1840159	0.330	$8.1x10^{-3}$	478	0
DHFR	5q14.1	Folate metabolism	rs71636231	$6.7x10^{-3}$	174	rs79043242	$3.5x10^{-3}$	304	rs76469466	0.115	0.02	478	0
TPMT	6p22.3	Purine metabolism	rs58347115	$7.4x10^{-3}$	174	rs7761743	$6.1x10^{-4}$	304	rs12192017	0.068	$1.2x10^{-3}$	478	0
SLC29A1	**6p21.1**	**Adenosine transport**	**rs7767854**	**$2.1x10^{-4}$**	**174**	**rs79343816**	**$1.6x10^{-3}$**	**304**	**rs2096199**	**0.204**	**$3.2x10^{-4}$**	**478**	**0**
NT5E	6q14.3	Adenosine metabolism	rs990274	$1.6x10^{-4}$	174	rs1744121	$5.7x10^{-3}$	304	rs653241	0.09	$6.7x10^{-3}$	478	0
ABCB1	7q21.12	MTX transport	rs2373363	$2.1x10^{-3}$	174	rs149494767	$3.5x10^{-4}$	304	rs112854881	0.100	$5.5x10^{-3}$	478	10.3
GGH	8q12.3	MTX metabolism	rs1455578	0.02	174	rs113485429	$7.5x10^{-3}$	304	rs72653281	0.086	0.02	478	0
FPGS	9q34.11	MTX metabolism	rs4837216	$4.3x10^{-3}$	174	rs77333490	$3.5x10^{-3}$	304	rs10156580	0.156	$3.3x10^{-3}$	478	0.6
ADK	10q22.2	Adenosine metabolism	rs10430522	$3.2x10^{-3}$	174	rs60602315	0.05	304	rs58588373	0.160	0.04	478	0
ABCC2	10q24.2	MTX transport	rs142354158	$8.6x10^{-3}$	174	rs4919367	0.02	304	rs873441	0.218	$8.4x10^{-3}$	478	0

Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFD1	14q23.2	Folate metabolism	rs1152587	0.03	174	rs3853411	0.01	304	rs1256063	0.059	$7.9x10^{-3}$	478	0
MTHFS	**15q25.1**	**Folate metabolism**	**rs186864369**	**$2.5x10^{-3}$**	**174**	**rs185888913**	**$3.9x10^{-4}$**	**304**	**rs185888913**	**0.098**	**$2.4x10^{-5}$**	**478**	**0**
ADORA2B	17p12	Adenosine signalling	rs11657857	0.01	174	rs9906205	$2.2x10^{-3}$	304	rs2530062	0.098	$7.8x10^{-3}$	478	0
SHMT1	17p11.2	Folate metabolism	rs76089176	$6.7x10^{-3}$	174	rs2975002	$5.8x10^{-4}$	304	rs854788	0.390	0.02	478	0
TYMS	18p11.32	Folate metabolism	rs12956482	$3.7x10^{-4}$	174	rs556277	$5.5x10^{-4}$	304	rs2606241	0.267	$5.1x10^{-4}$	478	0
ITPA	20p13	Purine metabolism	rs3761243	$9.2x10^{-3}$	174	rs6051655	0.03	304	rs11697157	0.156	0.02	478	0
ADA	20q13.12	Adenosine accumulation	rs1155566	$5.6x10^{-3}$	174	rs244078	$6.2x10^{-5}$	304	rs12481475	0.114	$6.0x10^{-3}$	478	0
GART	21q22.11	Adenosine accumulation	rs192088586	$5.8x10^{-3}$	174	rs184228857	$8.9x10^{-3}$	304	rs184228857	0.143	$6.7x10^{-3}$	478	0
SLC19A1	21q22.3	MTX transport	rs28432869	$1.5x10^{-4}$	174	rs56359647	$5.1x10^{-4}$	304	rs78757161	0.103	$3.3x10^{-3}$	478	0
CBS	21q22.3	Homocysteine metabolism	rs2839600	$8.5x10^{-3}$	174	rs8128259	$2.2x10^{-3}$	304	rs1788467	0.218	$6.0x10^{-3}$	478	0
ADORA2A	22q11.23	Adenosine signalling	rs2877206	0.02	174	rs9608321	0.01	304	rs8135864	0.337	0.02	478	0

Table 4.7. Methotrexate metabolism candidate gene association results with ΔDAS28-CRP. The lead SNP at each locus is listed with its respective p-value from the association testing in each cohort and the meta-analysis. SNPs are highlighted in bold with $p<5x10^{-4}$ from the DerSimonian and Laird random effects meta-analysis.

Table 4.8. Summary details of meta-analysis nominally associated MTX metabolism loci with ΔDAS28-CRP													
Candidate Gene	Cytoband	Lead SNP	Genotype/Imputed	MAF	Roche clinical trials MTX			YEAR MTX			Meta-analysis		
					β	p-value	n	β	p-value	n	β	p-value	n
MTHFR	1p36.22	rs2745273	Y=I, Amb=I, IM/FI=I	0.465	0.36	0.03	174	0.44	$5.8x10^{-3}$	304	0.40	$3.9x10^{-4}$	478
ADORA3	1p13.2	rs4839129	Y=I, Amb=I, IM/FI=I	0.450	-0.37	0.03	174	-0.39	$5.1x10^{-3}$	304	-0.38	$3.9x10^{-4}$	478
ABCG2	4q22.1	rs6817223	Y=I, Amb=I, IM/FI=I	0.326	-0.28	0.10	174	-0.52	$5.3x10^{-4}$	304	-0.42	$5.6x10^{-4}$	478
SLC29A1	6p21.1	rs2096199	Y=I, Amb=I, IM/FI=I	0.204	0.47	0.01	174	0.42	0.01	304	0.44	$3.2x10^{-4}$	478
MTHFS	15q25.1	rs185888913	Y=I, Amb=I, IM/FI=I	0.098	-0.63	0.03	174	-0.77	$3.9x10^{-4}$	304	-0.72	$2.4x10^{-5}$	478
		rs34903151	Y=I, Amb=I, IM/FI=I	0.101	0.62	0.03	174	0.75	$4.0x10^{-4}$	304	0.70	$2.7x10^{-5}$	478

Table 4.8. ΔDAS28-CRP meta-analysis associated and nominally associated SNPs. Individual study regression estimates and p-values are shown for the lead SNP from the meta-analysis nominally associated loci. All of the SNPs are imputed in all of the studies with and INFO≥0.8. In the Genotype/Imputed column Y = YEAR data, Amb = Roche Ambition data and IM/FI = Roche FILM and IMAGE data. I - Imputed, G - genotyped. Two SNPs are listed for the *MTHFS* locus which represents the two possibly independent signals of association (Figure 4.7B). MAF is in 1000 Genomes GBR population (March 2012).

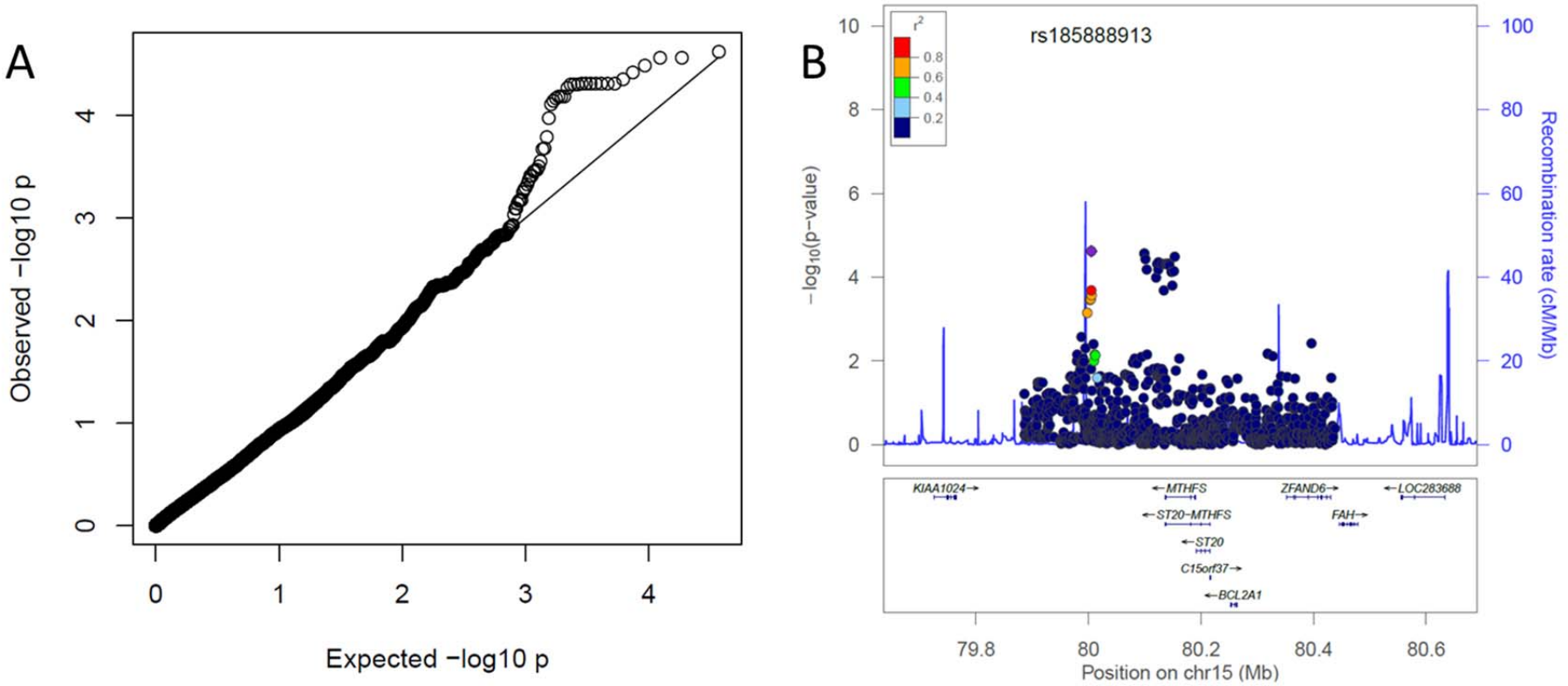


Figure 4.7. Meta-analysis of ΔDAS28-CRP YEAR and Roche association results. A – Q-Q plot of expected vs. observed $-\log_{10}$ p-values from the meta-analysis of ΔDAS28-CRP YEAR and Roche association studies shows an enrichment of smaller p-values and deviation from the expected proportion below $p=10^{-3}$. B - High resolution Manhattan plot of meta-analysis p-values across the genomic interval containing the candidate gene *MTHFS*. SNPs are plotted along the x-axis by base pair position on chromosome 15 with refseq genes shown below. –log10 p-values are plotted on the left y-axis and recombination rate on the right hand y-axis. LD with the lead SNP at the locus (rs185888913) is shown in colour (key top left corner). Two signals of association are observed at this locus within the 2nd intron of the *MTHFS* gene and ~170kb downstream close to a large recombination peak. These two groups of associated SNPs are in weak LD and may represent two independent signals of association with ΔDAS28-CRP.

4.3.2.4 Secondary outcome association analysis with single nucleotide polymorphisms in the Yorkshire Early Arthritis Register and Roche clinical trials cohorts

The secondary analyses used the ΔSJC28 and ΔlnCRP as outcome measures in both the YEAR and Roche cohorts. Following QC 71,103 and 52,806 SNPs were analysed in the Roche and YEAR data sets respectively for ΔSJC28, and 71,103 and 52,924 SNPs respectively for ΔlnCRP. A single SNP on 18p11.32 was statistically significantly associated with an increased ΔSJC28 in the YEAR cohort (rs556277 - imputed, MAF 0.12, β 3.9 95% CI [2.3-5.5], $p=1.4x10^{-6}$) (Table 4.9). This SNP lies in the 6th intron of the *CLUL1* gene which encodes clusterin-like 1, a retinal expressed protein and is ~26kb upstream of the *TYMS* transcriptional start site. A further 10 SNPs reach the threshold for nominal significance in the YEAR dataset (Table 4.9) whilst 32 SNPs in the Roche dataset reach this threshold. A random-effects meta-analysis of 37,039 shared SNPs between the Roche and YEAR association results found 2 SNPs at the *MTRR* genomic interval statistically significantly associated with a lower ΔSJC28 (Table 4.9 and Figure 4.8A). These two SNPs lie within the 8th and 15th introns of *ADCY2*, a brain tissue expressed adenylate cyclase which catalyses the formation of the intracellular second messenger cyclic adenosine monophosphate (cAMP). A further 15 SNPs across the *MTRR* and 3 other genomic intervals were nominally associated with ΔSJC28 (Figure 4.8B-D); no statistically significant signals of association were observed at the *TYMS* locus in the meta-analysis. Regression effect estimates and p-values are shown for each study in Table 4.10. All except one SNP was imputed in all of the genotyping data (rs6555490).

The 4 SNPs that show nominal association in the ΔSJC28 meta-analysis at the *MTHFS* locus were also at least nominally associated in the ΔDAS28 meta-analysis. This may add weight to the evidence that these SNPs have, or are in LD with SNPs that have genuine effects on treatment outcome. One might expect an association with ΔDAS28-CRP that is not

observed with any of the components to represent a false positive association, whilst those also associated with the DAS28 components may suggest a consistent effect. These 4 SNPs map ~17-60kb downstream of the *MTHFS* coding region and are in perfect LD (r^2=1.0) with a non-synonymous SNP, rs8923. This SNP results in an alanine to threonine substitution at position 202 of the methylene tetrahydrofolate synthetase protein.

No SNPs reached the threshold for statistical significance in the analysis of ΔlnCRP in either the Roche or YEAR cohorts. Nominal association was observed for 2 SNPs at 2 different loci in the YEAR data and 10 SNPs in the Roche data across 6 loci. In the meta-analysis a total of 37,097 SNPs were shared between the Roche and YEAR data, of which none reached the threshold for statistically significant association with ΔlnCRP. Nominal association was observed with 3 SNPs on 1p36.22 ~200kb from the 3' untranslated region of *MTHFR* (Table 4.11). These SNPs are not in strong LD with either rs1801131 or rs1801133 investigated in Section 4.3.1 (r^2<0.01). The lack of statistically significant associations in the meta-analysis of ΔlnCRP is reflected in the Q-Q plot of p-values which show a deficit of smaller p-values (Figure 4.9B). This is in contrast to the equivalent Q-Q plot for the ΔSJC28 (Figure 4.9A) and ΔDAS28-CRP (Figure 4.9B) meta-analysis results which show a slight enrichment of smaller p-values.

Table 4.9. Candidate gene regions for association with MTX response ΔSJC28

Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFR	1p36.22	Folate metabolism	rs56182507	$1.1x10^{-2}$	176	rs12403639	$1.0x10^{-4}$	335	rs2745273	0.472	$9.6x10^{-4}$	511	0
AMPD1	1p13.2	Adenosine metabolism	rs2878760	$6.6x10^{-2}$	176	rs117608626	$5.3x10^{-3}$	335	rs80081023	0.050	$3.4x10^{-3}$	511	0
ADORA3	1p13.2	Adenosine signalling	rs17655441	$5.7x10^{-3}$	176	rs1264879	$1.1x10^{-3}$	335	rs4839129	0.451	$3.1x10^{-3}$	511	0
ADORA1	1q32.1	Adenosine signalling	rs73072180	$6.6x10^{-3}$	176	rs6671900	$1.2x10^{-3}$	335	rs10800922	0.423	$1.0x10^{-3}$	511	0
MTR	1q43	Homocysteine metabolism	rs34647600	$3.4x10^{-4}$	176	rs677628	$1.7x10^{-3}$	335	rs41421448	0.094	$2.4x10^{-3}$	511	0
ATIC	2q35	Adenosine accumulation	rs17519171	$2.1x10^{-3}$	176	rs12997148	$7.1x10^{-4}$	335	rs142741707	0.078	$8.3x10^{-3}$	511	0
PPAT	4q12	*De novo* purine synthesis	rs56073671	$1.6x10^{-2}$	176	rs114763852	0.02	335	rs73240555	0.122	$4.7x10^{-3}$	511	0
ABCG2	**4q22.1**	**MTX transport**	**rs76634520**	$\mathbf{8.9x10^{-4}}$	**176**	**rs6817223**	$\mathbf{8.4x10^{-4}}$	335	**rs6817223**	0.32	$\mathbf{1.5x10^{-4}}$	511	**0**
MTRR	**5p15.31**	**Homocysteine metabolism**	**rs16879202**	$\mathbf{3.5x10^{-4}}$	**176**	**rs62342476**	$\mathbf{3.3x10^{-4}}$	335	**rs6555490**	0.129	$\mathbf{7.9x10^{-6}}$	511	**0**
DHFR	5q14.1	Folate metabolism	rs10073069	$6.0x10^{-3}$	176	rs79043242	$7.0x10^{-3}$	335	rs13187593	0.164	$4.0x10^{-3}$	511	0
TPMT	6p22.3	Purine metabolism	rs214590	$1.8x10^{-3}$	176	rs2842937	$7.7x10^{-3}$	335	rs11970773	0.240	$7.2x10^{-3}$	511	0
SLC29A1	6p21.1	Adenosine transport	rs9472255	$6.8x10^{-4}$	176	rs1409112	$5.0x10^{-4}$	335	rs1409112	0.095	$7.8x10^{-4}$	511	6.3
NT5E	6q14.3	Adenosine metabolism	rs78557317	$1.4x10^{-2}$	176	rs1744121	0.02	335	rs2225546	0.140	0.01	511	0
ABCB1	7q21.12	MTX transport	rs2157930	$5.6x10^{-4}$	176	rs55737734	$6.2x10^{-3}$	335	rs1358064	0.155	0.02	511	0
GGH	**8q12.3**	**MTX metabolism**	**rs17265947**	$\mathbf{2.2x10^{-3}}$	**176**	**rs113485429**	$\mathbf{2.0x10^{-3}}$	335	**rs72653281**	0.087	$\mathbf{3.9x10^{-4}}$	511	**0**
FPGS	9q34.11	MTX metabolism	rs12551781	$5.1x10^{-2}$	176	rs55996464	$3.4x10^{-4}$	335	rs72752823	0.127	$1.1x10^{-3}$	511	12.8
ADK	10q22.2	Adenosine metabolism	rs16931023	$4.8x10^{-3}$	176	rs187333189	0.06	335	rs4745757	0.078	0.05	511	0
ABCC2	10q24.2	MTX transport	rs2111326	$2.4x10^{-2}$	176	rs6584287	$5.0x10^{-3}$	335	rs6584287	0.086	$1.4x10^{-3}$	511	0

Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFD1	14q23.2	Folate metabolism	rs17824591	$1.7x10^{-2}$	176	rs2045023	0.01	335	rs1499520	0.256	0.01	511	0
MTHFS	**15q25.1**	**Folate metabolism**	**rs75958782**	**$7.9x10^{-3}$**	**176**	**rs1781652**	**$5.1x10^{-4}$**	335	**rs16971397**	0.104	**$3.5x10^{-4}$**	511	**0**
ADORA2B	17p12	Adenosine signalling	rs146778604	$9.5x10^{-3}$	176	rs2779205	0.01	335	rs4286144	0.086	$5.5x10^{-3}$	511	0
SHMT1	17p11.2	Folate metabolism	rs854811	$9.0x10^{-4}$	176	rs145835293	0.02	335	rs77284899	0.258	0.03	511	0
TYMS	18p11.32	Folate metabolism	rs492347	$1.0x10^{-4}$	176	rs556277	$1.4x10^{-6}$	335	rs2846786	0.379	$9.2x10^{-4}$	511	0
ITPA	20p13	Purine metabolism	rs79355864	$7.4x10^{-3}$	176	rs6084297	$8.4x10^{-3}$	335	rs117725082	0.058	$5.2x10^{-3}$	511	0
ADA	20q13.12	Adenosine accumulation	rs6017418	$5.3x10^{-3}$	176	rs6093976	$3.1x10^{-3}$	335	rs59952658	0.243	0.01	511	0
GART	21q22.11	Adenosine accumulation	rs8126739	$2.4x10^{-3}$	176	rs9984273	$5.5x10^{-3}$	335	rs4816464	0.429	$5.3x10^{-3}$	511	0
SLC19A1	21q22.3	MTX transport	rs9637201	$1.2x10^{-3}$	176	rs56359647	$9.3x10^{-5}$	335	rs2150444	0.156	$1.4x10^{-3}$	511	0
CBS	21q22.3	Homocysteine metabolism	rs3972	$1.5x10^{-2}$	176	rs1788467	$5.0x10^{-3}$	335	rs1788467	0.218	$1.4x10^{-3}$	511	0
ADORA2A	22q11.23	Adenosine signalling	rs5751902	$3.8x10^{-3}$	176	rs149546420	0.01	335	rs189665441	0.090	$7.2x10^{-3}$	511	0

Table 4.9. Methotrexate metabolism candidate gene association secondary analysis of ΔSJC28. The lead SNP for each locus in each cohort is shown with its respective p-value. SNPs highlighted in bold are those that meet the threshold for nominal statistical significance in the random effects meta-analysis only ($p<5x10^{-4}$).

Table 4.10. Summary details of meta-analysis nominally associated MTX metabolism loci with ΔSJC28													
Candidate Gene	Cytoband	Lead SNP	Genotype/Imputed	MAF	Roche clinical trials MTX			YEAR MTX			Meta-analysis		
					β	p-value	n	β	p-value	n	β	p-value	n
ABCG2	4q22.1	rs6817223	Y=I, Amb=I, IM/FI=I	0.32	-1.08	0.07	176	-1.44	$8.4x10^{-4}$	335	-1.32	$1.5x10^{-4}$	511
MTRR	5p15.31	rs6555490	Y=I, Amb=G, IM/FI=I	0.129	-2.67	$5.2x10^{-3}$	176	-1.94	$5.0x10^{-4}$	335	-2.13	$7.9x10^{-6}$	511
GGH	8q12.3	rs72653281	Y=I, Amb=I, IM/FI=I	0.087	-2.50	$9.4x10^{-3}$	176	-1.81	0.01	335	-2.07	$3.9x10^{-4}$	511
MTHFS	15q25.1	rs16971397	Y=I, Amb=I, IM/FI=I	0.104	2.07	0.03	176	1.65	$5.3x10^{-3}$	335	1.77	$3.5x10^{-4}$	511

Table 4.10. ΔSJC28 meta-analysis nominally associated and associated loci with MTX response. The lead SNP for each locus at least nominally associated in the ΔSJC28 meta-analysis with the effect estimates and p-values from the separate YEAR and Roche association analyses. The MAF is from the 1000 Genomes GBR population (March 2012 release). All except one SNP is imputed in all studies. Only rs6555490 is genotyped directly in the Ambition study. The INFO score for all of the other SNPs are ≥0.8. Y – YEAR, Amb – Ambition, IM/FI – IMAGE and FILM, I – imputed, G – genotyped.

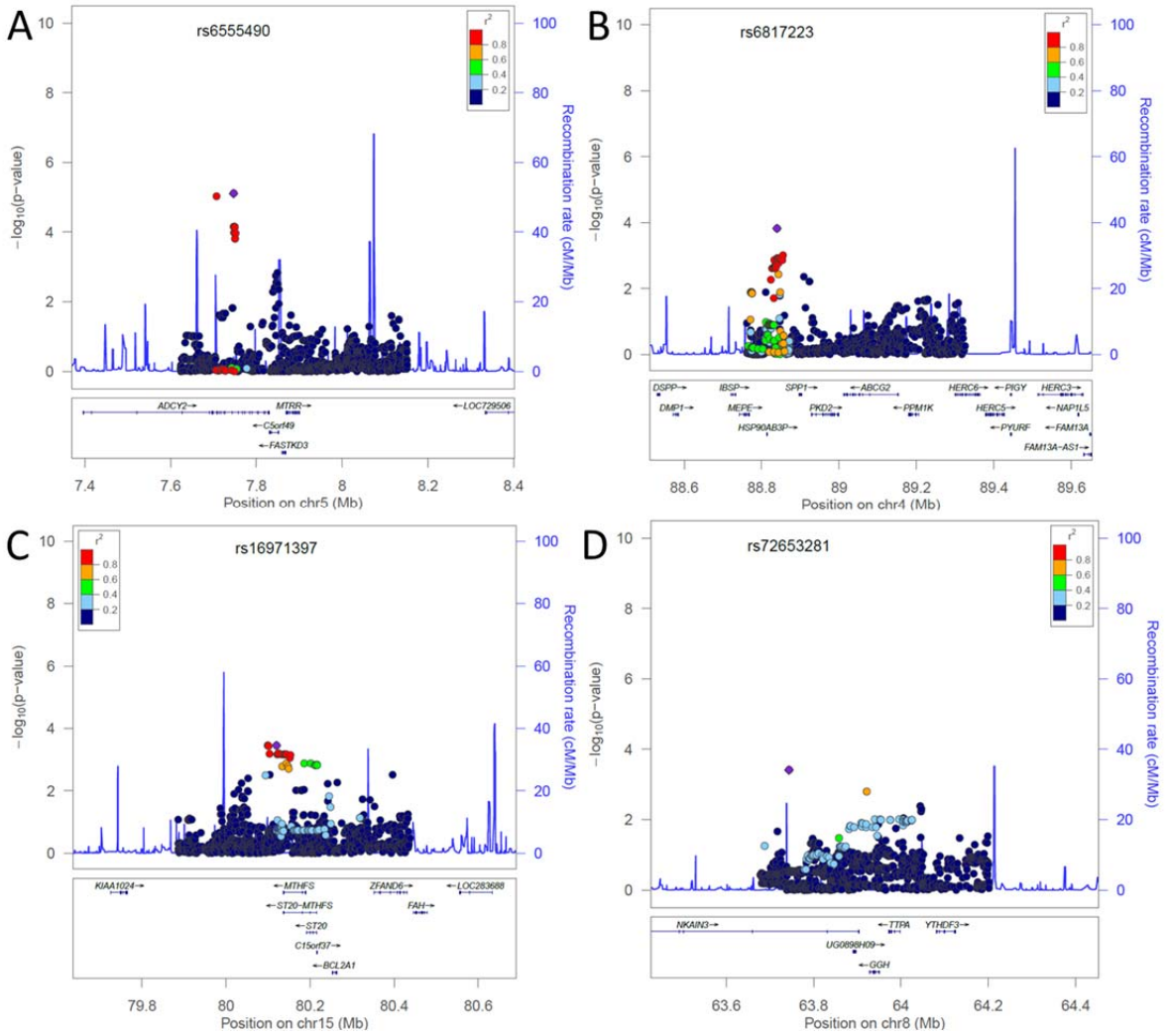


Figure 4.8. Meta-analysis results of ΔSJC28 association testing. High resolution Manhattan plots are shown of the signals of association at the A-*MTRR* genomic interval and the nominally associated SNPs across the B-*ABCG2*, C-*MTHFS* and D-*GGH* genomic intervals. The lead SNP from the analysis is labelled in each plot and the colouring represents LD with the lead SNP based on the colour code key.

Table 4.11. Candidate gene regions for association with MTX response ΔlnCRP													
Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFR	**1p36.22**	**Folate metabolism**	**rs11554507**	**$2.4x10^{-3}$**	**178**	**rs2745273**	**$1.8x10^{-3}$**	**314**	**rs11587747**	**0.325**	**$1.0x10^{-4}$**	**492**	**0**
AMPD1	1p13.2	Adenosine metabolism	rs10157284	$1.3x10^{-2}$	178	rs113511534	$9.6x10^{-3}$	314	rs115391531	0.071	$8.8x10^{-3}$	492	0
ADORA3	1p13.2	Adenosine signalling	rs2252740	$3.3x10^{-3}$	178	rs12735640	$3.6x10^{-3}$	314	rs6673839	0.296	$2.2x10^{-3}$	492	0
ADORA1	1q32.1	Adenosine signalling	rs4950894	$2.8x10^{-6}$	178	rs55768624	$2.5x10^{-3}$	314	rs58402552	0.054	$6.0x10^{-3}$	492	0
MTR	1q43	Homocysteine metabolism	rs4659764	$4.4x10^{-3}$	178	rs138482850	$2.4x10^{-3}$	314	rs6656267	0.084	$3.5x10^{-3}$	492	0
ATIC	2q35	Adenosine accumulation	rs1250255	$2.2x10^{-2}$	178	rs12694351	0.01	314	rs12694351	0.376	$8.3x10^{-3}$	492	0
PPAT	4q12	*De novo* purine synthesis	rs56394067	$5.0x10^{-3}$	178	rs9312681	0.01	314	rs10017460	0.112	0.01	492	0
ABCG2	4q22.1	MTX transport	rs13120254	$4.8x10^{-5}$	178	rs62310036	$2.4x10^{-3}$	314	rs34056272	0.226	$2.0x10^{-3}$	492	0
MTRR	5p15.31	Homocysteine metabolism	rs75271545	$4.0x10^{-4}$	178	rs162031	0.01	314	rs11134286	0.289	0.02	492	7.5
DHFR	5q14.1	Folate metabolism	rs6894961	$1.1x10^{-3}$	178	rs1677658	$7.5x10^{-3}$	314	rs420522	0.278	$1.5x10^{-3}$	492	0
TPMT	6p22.3	Purine metabolism	rs10807621	$1.5x10^{-3}$	178	rs34969716	$4.8x10^{-4}$	314	rs34991140	0.392	$1.7x10^{-3}$	492	0
SLC29A1	6p21.1	Adenosine transport	rs57904877	$8.3x10^{-3}$	178	rs60987108	$4.7x10^{-3}$	314	rs4711769	0.485	$3.4x10^{-3}$	492	0
NT5E	6q14.3	Adenosine metabolism	rs149019033	$7.7x10^{-4}$	178	rs9362179	0.03	314	rs9362179	0.187	0.03	492	0
ABCB1	7q21.12	MTX transport	rs5014436	$6.3x10^{-3}$	178	rs802102	$2.7x10^{-3}$	314	rs802062	0.420	$3.7x10^{-3}$	492	0
GGH	8q12.3	MTX metabolism	rs11785601	$6.6x10^{-4}$	178	rs17265947	0.03	314	rs4297033	0.093	0.02	492	0
FPGS	9q34.11	MTX metabolism	rs11787917	$2.7x10^{-5}$	178	rs35000896	$9.7x10^{-3}$	314	rs10124356	0.061	0.04	492	0
ADK	10q22.2	Adenosine metabolism	rs61865576	$2.5x10^{-2}$	178	rs10824232	$3.3x10^{-3}$	314	rs11000915	0.076	0.01	492	0
ABCC2	10q24.2	MTX transport	rs142212934	$5.1x10^{-3}$	178	rs186059830	$9.6x10^{-3}$	314	rs75483041	0.061	0.01	492	0

Candidate Gene	Cytoband	Pathway	Roche clinical trials MTX			YEAR MTX			Meta-analysis				
			Lead SNP	p-value	n	Lead SNP	p-value	n	Lead SNP	MAF	p-value	n	I^2
MTHFD1	14q23.2	Folate metabolism	rs13329053	$3.7x10^{-3}$	178	rs56021161	0.04	314	rs80192041	0.066	0.03	492	0
MTHFS	15q25.1	Folate metabolism	rs7183862	$7.4x10^{-3}$	178	rs73487353	$1.4x10^{-3}$	314	rs3893384	0.424	$9.6x10^{-4}$	492	0
ADORA2B	17p12	Adenosine signalling	rs178656	$1.3x10^{-2}$	178	rs9906205	$1.5x10^{-3}$	314	rs2324082	0.472	$7.6x10^{-3}$	492	0
SHMT1	17p11.2	Folate metabolism	rs112441014	$6.4x10^{-3}$	178	rs17436616	$4.4x10^{-3}$	314	rs651495	0.363	$5.7x10^{-3}$	492	0
TYMS	18p11.32	Folate metabolism	rs61423706	$6.2x10^{-3}$	178	rs11081364	$1.9x10^{-3}$	314	rs974071	0.459	$1.5x10^{-3}$	492	0
ITPA	20p13	Purine metabolism	rs2004688	$3.3x10^{-3}$	178	rs6076470	0.02	314	rs144102485	0.073	$5.9x10^{-3}$	492	0
ADA	20q13.12	Adenosine accumulation	rs191957620	$3.4x10^{-2}$	178	rs6017342	$1.8x10^{-3}$	314	rs2425640	0.343	$4.0x10^{-3}$	492	0
GART	21q22.11	Adenosine accumulation	rs8126534	$1.9x10^{-3}$	178	rs73199881	$5.4x10^{-3}$	314	rs139056856	0.069	$1.4x10^{-3}$	492	0
SLC19A1	21q22.3	MTX transport	rs7279250	$2.0x10^{-3}$	178	rs4819100	0.01	314	rs56770975	0.052	$5.2x10^{-4}$	492	0
CBS	21q22.3	Homocysteine metabolism	rs2096853	$9.5x10^{-4}$	178	rs9976191	$3.6x10^{-3}$	314	rs11910684	0.107	$7.4x10^{-3}$	492	0
ADORA2A	22q11.23	Adenosine signalling	rs4822528	$9.7x10^{-4}$	178	rs6004167	$1.6x10^{-3}$	314	rs2330772	0.484	$4.1x10^{-3}$	492	0

Table 4.11. Meta-analysis of MTX metabolism candidate gene loci SNPs with ΔlnCRP. The lead SNP for each locus is listed with its respective p-value for the separate Roche and YEAR cohort association analyses, as well as the meta-analysis of these two data sets. Only SNPs that reach the threshold for nominal statistical significance in the meta-analysis are highlighted in bold.

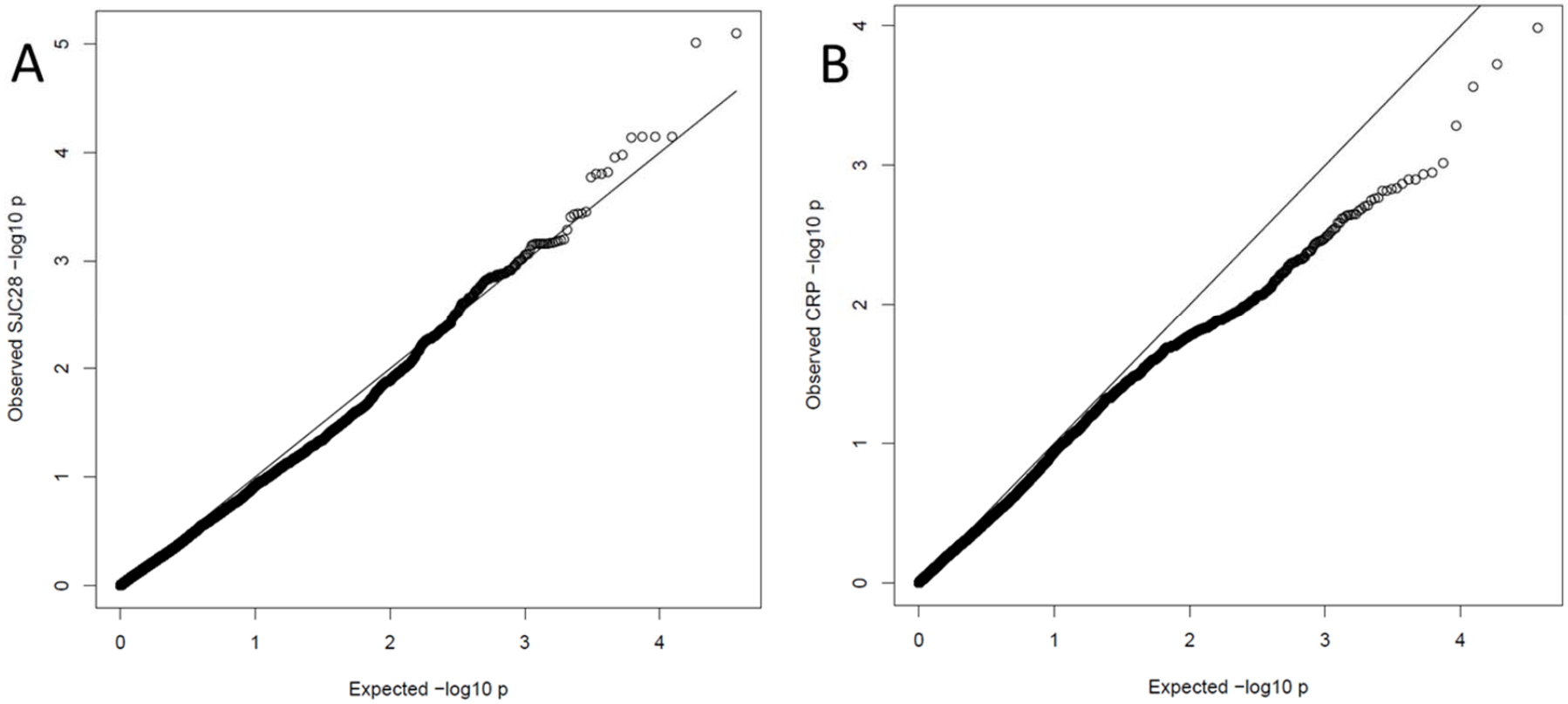


Figure 4.9. Quantile-quantile plots of meta-analysis $-\log_{10}$ p-values from the secondary outcome analyses. A - An enrichment of smaller p-values is observed in the ΔSJC28 meta-analysis above the expected proportion. B – A deficit of smaller p-values is observed in the ΔlnCRP meta-analysis results demonstrated by the distribution of p-values that falls below the expected proportion.

4.3.2.5 Replication analysis of previously associated methotrexate pharmacogenetic markers

During the analysis of these MTX metabolism pathway genes, SNPs previously associated with MTX efficacy have been tested for their association with response. The results for each SNP are reported in Table 4.12. Of the 14 SNPs in MTX metabolism genes, excluding *MTHFR* (Chapter 1.4.3; Table 1.5), one at the *ADA* locus (rs244076, β 0.55 95% CI [0.21-0.89], $p=1.7x10^{-3}$) demonstrated an association $p<0.05$ in the YEAR primary outcome analysis. This effect was not observed in the Roche data or meta-analysis. A second SNP, rs2853539, at the *TYMS* locus demonstrated an association with $p<0.05$ with all three outcome measures in the Roche data, but was not observed in the YEAR or meta-analysis results. None of the other previously reported SNPs examined were associated with MTX response in any of the analyses conducted in this study.

Table 4.12. Replication analyses of previous MTX pharmacogenetic associations												
SNP	Candidate Gene	Cytoband	YEAR			Roche			Meta-analysis			
			ΔDAS28-CRP p-value	ΔSJC28 p-value	ΔlnCRP p-value	ΔDAS28-CRP p-value	ΔSJC28 p-value	ΔlnCRP p-value	ΔDAS28-CRP p-value	ΔSJC28 p-value	ΔlnCRP p-value	
rs17602729	*AMPD1*	1p13.2	0.38	0.83	0.50	0.69	0.91	0.82	0.70	0.81	0.52	
rs4673993	*ATIC*	2q35	0.80	0.96	0.57	0.19	0.48	0.43	0.56	0.68	0.35	
rs2372536	*ATIC*	2q35	0.82	0.92	0.44	0.32	0.55	0.72	0.66	0.82	0.41	
rs3821353	*ATIC*	2q35	0.35	0.28	0.67	0.14	0.22	0.96	0.80	0.92	0.73	
rs1045642	*ABCB1*	7q21.12	0.55	0.24	0.63	0.58	0.36	0.25	0.93	0.13	0.26	
rs12681874	*GGH*	8q12.3	0.74	0.19	0.41	0.99	0.68	0.63	0.79	0.18	0.35	
rs1544105	*FPGS*	9q34.11	0.68	0.32	0.24	0.71	0.58	0.68	0.58	0.26	0.59	
rs2236225	*MTHFD1*	14q23.2	0.45	0.20	0.13	0.70	0.55	0.62	0.41	0.67	0.61	
rs2853539	*TYMS*	18p11.32	0.24	0.53	0.36	0.018	0.018	0.010	0.72	0.53	0.63	
rs1127354	*ITPA*	20p13	0.87	0.83	0.77	0.58	0.22	0.98	0.64	0.39	0.84	
rs244076	*ADA*	20q13.12	$1.7x10^{-3}$	0.06	0.29	0.39	0.16	0.30	0.59	0.97	0.99	
rs2274808	*SLC19A1*	21q22.3	0.57	0.22	0.22	0.94	0.36	0.35	0.70	0.89	0.12	
rs1051266	*SLC19A1*	21q22.3	0.90	0.45	0.69	0.70	0.89	0.63	0.72	0.60	0.53	
rs5751876	*ADORA2A*	22q11.23	0.14	0.57	0.81	0.51	0.70	0.37	0.21	0.82	0.43	

Table 4.12. Association p-values of previously associated SNPs with MTX efficacy in the published literature. A single SNP previously associated with MTX demonstrated a $p<0.05$ in this association analysis in the YEAR primary analysis. This association was not observed in the Roche data, or in the meta-analysis of the two datasets.

4.4 Discussion

4.4.1 Methotrexate pharmacogenetics and *MTHFR*

The *MTHFR* gene has been the subject of many genetic association studies. SNPs within the coding region of the gene have been associated with a range of phenotypes affected by folate metabolism including susceptibility to vascular disease and colorectal cancer, response and adverse reactions to chemotherapeutic doses of MTX, plasma homocysteine and serum folate levels [171, 310-315]. As a single gene it has received unparalleled attention in the field of MTX pharmacogenetics, particularly within RA. The findings from these studies have been confusing and contradictory. Primarily this has arisen due to small sample sizes where random variation can have much larger effects than in larger cohorts. Another confounding factor is the use of different treatment response measures, ranging from the dose of MTX administered to definitions of clinical response based on composite measures of disease activity. In the previous results chapter discussion (Chapter 3.4.1-3.4.2) issues within the field of pharmacogenetics in MTX were highlighted. Many of these apply to previous association analyses of *MTHFR* SNPs with MTX response. Addressing some of these issues in the meta-analysis described in Section 4.3.1, it is noticeable that there remains no association between these two SNPs and response to MTX using DAS28-based measures of treatment response. The sample sizes (n=986 for rs1801133 and n=866 for rs1801131) are adequately powered to detect the sort of effect sizes that are thought to be of clinical utility (OR>2.0). If these two SNPs do affect MTX response in RA, it would require even larger (assuming OR 0.87 n>3000) and more extensively phenotyped cohorts than those described in this doctoral thesis to adequately test this.

4.4.2 Methotrexate pharmacogenetic candidate genes and methotrexate treatment response

The aim of the work described in this chapter was to test the hypothesis that genetic variation within and around genes encoding proteins in MTX metabolism and transport pathways are associated with response to this drug. The design of the study was such that it maximised the capture of common genetic variation (MAF≥5%) within 0.5Mb of the transcriptional start site of each gene. Previous studies have focussed on non-synonymous SNPs as candidates or have used a tagging approach to capture much of the common variation at a locus [189]. Owen *et al* captured variation within the exonic and intronic regions of several MTX candidate genes, and within a 10kb flanking region. This particular approach does not necessarily capture genetic variants that affect transcriptional variation that lie outside of this region. A number of eQTL studies have demonstrated the enrichment of genetic variants affecting gene expression within 0.5Mb of the TSS of most genes [290, 300]. Therefore this distance was used to capture potential *cis* regulatory effects missed by focussing on proximal regions. This is illustrated by the presence of published *cis*-eQTLs for 17 of the 29 genes investigated in this doctoral work [276, 277], including *MTHFS* and *MTRR* which both demonstrate nominal associations with MTX treatment response.

The finding of a group of SNPs at least nominally associated with ΔDAS28-CRP and ΔSJC28 indicates that those SNPs close to *MTHFS* may have a role in determining treatment response. The appearance of two independent association signals, based on limited LD between groups of associated SNPs (Figure 4.6A) may be a genuine finding. However, the most consistent finding was with the group of SNPs in strong LD (r^2>0.9 GBR 1K Genomes) with a non-synonymous change in the 3rd exon of *MTHFS*. This alanine to threonine mutation is predicted by Polyphen to be benign and lies within a region of low conservation

between evolutionary divergent species. No description exists of the functional consequences of this non-synonymous change to date (2013). The most strongly associated SNPs with the published *cis*-eQTL with *MTHFS* lie within the 1st intron and are in weak LD with the SNPs associated with ΔDAS28-CRP in this doctoral work. Therefore, they are unlikely to account for this signal of association.

The statistically significant association with ΔSJC28 in the meta-analysis of the YEAR and Roche data is focussed in the 15th intron of *ADCY2*, which lies 70kb telomeric to *MTRR*. *MTRR* has been a target of candidate gene association studies in the past, as it lies within the pathway responsible for catalysing homocysteine to methionine (Chapter 1.3.2.1; Figure 1.3). A non-synonymous SNP has been the focus of investigation in RA, though no consistent association has been described with treatment response [176, 185]. Stamp *et al* found that the G allele of the 66A>G SNP (rs1801394) was associated with lower red blood cell folate concentrations [185], decreased levels of which have been correlated with greater decreases in DAS28 in response to MTX [196]. This SNP was not associated with treatment response in any of the meta-analyses. Stamp *et al* also found expression of *MTRR* transcripts increased in synovial tissue compared to patients not receiving MTX treatment [316]. A *cis*-eQTL has been reported at this locus in skin biopsy tissue and lymphoblastoid cell lines in a large TwinsUK study [276]. The SNPs associated with lower ΔSJC28 are not, however, associated with transcript levels of *MTRR*. A second peak of SNPs with small p-values ($p<10^{-3}$) can be seen in Figure 4.8A that overlap with the TSS of *MTRR*. Whether these constitute a more modest effect on MTX response will require a larger cohort, with more power to detect smaller effect sizes, to test robustly.

4.4.3 Methotrexate pharmacogenetics: pitfalls, caveats and robust study design

4.4.3.1 Measuring treatment response in rheumatoid arthritis: is there an objective marker?

In the discussion of Chapter 3.4.1 a number of issues were raised pertaining to the use of composite measures of treatment response, particularly those influenced and defined by subjective decision making. Currently there does not exist a single consistent robust objective measure of treatment response in RA, analogous to that used in Warfarin dose optimisation. A panel of 12 serum biomarkers has been described [317] and is currently marketed under the name "Vectra DA" by Crescendo Bioscience (San Francisco, California, USA) [318, 319]. Serum measurements of vascular cell adhesion molecule-1 (VCAM-1), epidermal growth factor (EGF), vascular-endothelial growth factor A (VEGF-A), IL-6, TNF receptor, type 1 (TNFR1), MMP-1 and -3, human cartilage glycoprotein-39 (YLK-40), leptin, resistin, serum amyloid A (SAA) and CRP are combined to predict separate DAS components [317]. This generates a score on a scale 1-100 which is correlated with several measures of disease activity, including the DAS28-CRP, CDAI and SDAI. The development of this multi-marker assay involved comparisons with imaging markers and was correlated with synovial thickening and joint vascularity measured by power Doppler ultrasound [319]. This suggests that this assay may represent a useful biomarker for assessing both clinical disease activity and treatment response. Ergo it may also represent, in conjunction with ultrasound imaging, an appropriately objective measure of treatment response for pharmacogenetic studies in RA. A composite measure that uses multiple markers may be under heterogenous genetic influences. However, if the effect on treatment response is mediated through one or more of these serum markers one might expect to observe an association. Given the genetic and phenotypic heterogeneity of RA one might also expect a similar heterogeneity at the level of treatment response. Therefore any set of genetic predictors of treatment response may only be of use in a sub-set of patients which appear

to respond in the same fashion due to a convergence of genetic effects at the phenotype level (Figure 4.10). This suggests it is important to identify genetic biomarkers of response, and then further understand how they influence response to maximise their clinical utility.

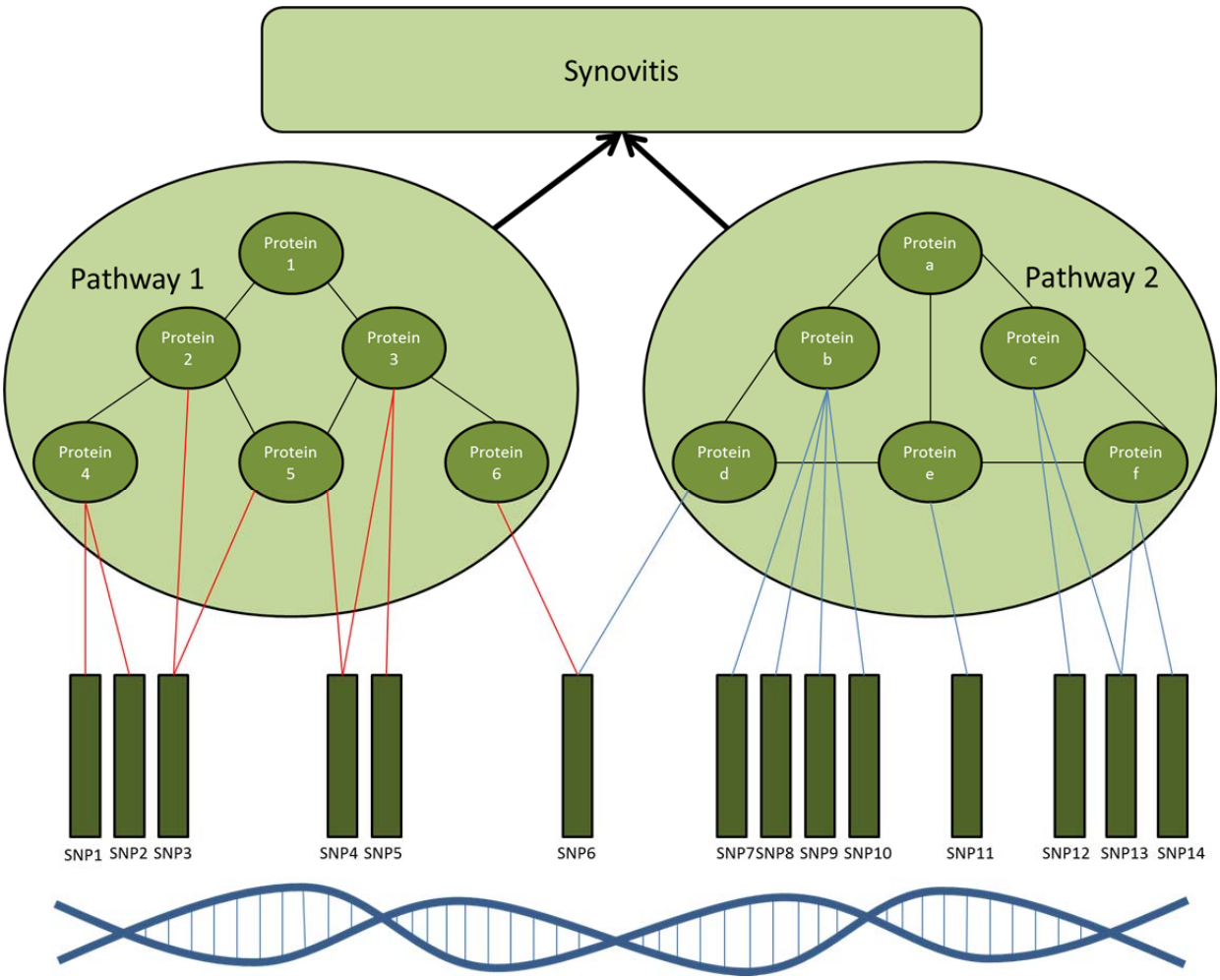


Figure 4.10. Genetic biomarkers of response using a multi-marker assay of serum proteins may converge at the phenotype level. Each of the 14 SNPs are associated with treatment response using a semi-objective measure of synovitis such as the SJC28. Individually SNPs may affect transcript regulation (regulatory SNPs), protein structure and/or function (non-synonymous SNPs) or transcript structure (splice site SNPs). Some SNPs act upon transcription factors that in turn affect regulatory networks, detected as *trans*-eQTL associations. Pathway 1 represents one pathway that contributes to synovitis whilst pathway 2 represents another; the two are independent. Both converge at the phenotype of "synovitis" despite different genetic influences. It is therefore important to understand how genetic variation affects treatment response. If Protein b in Pathway 2 is secreted in response to an inflammatory stimulus it can be influenced by many SNPs independently because of both signalling pathways that induce its expression as well effects on the proteins that regulate its transcription. All of these SNPs would potentially manifest as biomarkers of response using a measure of synovitis, but only Pathway 2 if using objective serum markers alone. Likewise Protein 5 is affected by two independent SNPs but contributes to Pathway 1. The serum markers measured in the Vectra DA assay are analogous to the proteins in Pathway 1 and 2. Thus one might expect genetic associations to be specific to each protein. However the assay markers are integrated to predict DAS28 components, including the SJC28, then used to generate a composite disease activity score which could be used as the outcome measure for association testing. The assay score therefore reflects serum changes of markers which track synovitis but do not rely on subjective decision making.

4.4.3.2 Inconsistency and failed replications in rheumatoid arthritis pharmacogenetics – the importance of study design

The candidate gene association study described in this chapter is the largest investigation of MTX pharmacogenetics in RA, to date. At the time of writing a larger GWAS of MTX response is underway, which includes the YEAR patients described herein. Previous pharmacogenetic studies of MTX focussed on candidate genes and candidate SNPs, which were also tested for association with MTX response in this study (Table 4.12). A single SNP in the *ADA* gene that encodes adenosine deaminase demonstrated an association $p<0.05$. This failure to replicate the associations of previous studies highlights the caveats and pitfalls that plague pharmacogenetics of RA, and more specifically MTX response. Studies have been small and underpowered to detect modest effect sizes (n=108-281), have used heterogeneous definitions and measures of treatment response and largely failed to account for multiple testing burden. In order to address some of these issues a more stringent statistical significance threshold has been adopted (driven by the large number of SNPs tested in this chapter), a larger sample size used (though still underpowered to detect modest effects) and more objective outcome measures defined (ΔSJC28 and ΔlnCRP). CRP is measured in the Vectra DA test described above, which suggests it is a useful marker of treatment response, but does not adequately capture the joint-specific aspect of RA. In order to achieve this, the ΔSJC28 was used and has been shown to correlate well with ultrasound measurements in a small cohort [248].

However, using multiple outcome measures increases the multiple testing burden which demands appropriate adjustment of test statistics or p-value thresholds to account for this. One approach would be to use a permutation procedure to control the type I error rate. For instance the phenotype data (including baseline measurement) could be held and each genotype for each individual permuted to generate a set of null distribution test statistics

for each SNP. The test statistic for each SNP using the non-permuted genotypes could then be compared to this null distribution.

Effective sample size can be increased by recruiting more patients into a cohort or by combining existing studies. The latter of these is on first inspection an attractive option as it circumvents the need to spend more time recruiting patients into a research study which is often a long process (potentially taking many years). An important limitation when combining cohorts is that of between-study heterogeneity. In genetic epidemiological studies of disease this is often not a primary concern where there are defined diagnostic criteria. In pharmacogenetic studies treatment protocols outside of clinical trials are rarely standardised (the YEAR cohort provides a possible exception to this). This is a major source of heterogeneity, particularly where administration of glucocorticoid steroids are not recorded. Rather than directly combining cohorts, test statistics from each cohort can instead be combined by meta-analysis and can account for between-study heterogeneity, using random effects models such as those described by DerSimonian and Laird [308]. This can also help to account for non-overlapping SNPs that arise when different genotyping chips are used and require imputation to overcome this. Rather than directly combining genotyped and imputed SNPs, meta-analysis allows the combination of studies in these circumstances.

Despite the larger sample size in this study (n=478) compared to those conducted previously there was limited power to detect modest effect sizes (Section 4.3.2.2: Table 4.6). When taken together with the association findings from the meta-analysis of the YEAR and Roche datasets (for all three outcome measures) it becomes evident that pharmacogenetic studies require careful design to: 1) minimise intra- and inter-cohort heterogeneity, 2) maximise statistical power, 3) test for associations in a statistically robust manner and 4) incorporate both discovery and replication phases.

Chapter 5 – Fc γ receptors, copy number variation and response to biologics in rheumatoid arthritis

5.1 Introduction

Several biologic treatments that target molecules and cells important in RA pathogenesis and pathology are either monoclonal antibodies or recombinant molecules that contain the IgG Fc region. These antibodies can bind to the cognate IgG receptors (FcγR) on cells of the immune system and elicit their therapeutic potential through these cell-surface receptors. As a result, variation of the genes encoding these receptors (*FCGR*) are candidates for pharmacogenetic investigation.

5.1.1 Cell surface receptors for immunoglobulin G

Immunoglobulins are able to induce cell activation, inhibition and effector functions that do not rely on the antigen specificity of the antibody, i.e. Fab-independent. Each monomeric antibody molecule contains two variable antigen binding regions that recognise specific epitopes, the fragment antigen-binding (Fab) region. In addition these molecules contain an invariant region that does not recognise antigen, the fragment crystallisable (Fc) region. Antigen-agnostic cell signalling is mediated by a class of receptors that bind the Fc region (Fc Receptors) that recognise specific immunoglobulin classes (i.e. IgG, IgE and IgA are bound by FcγR, FcεR and FcαR, respectively). Immunoreceptors for IgG (FcγR) are expressed on the cell surface of certain immune system cells, including macrophages, monocytes, B cells, NK cells and neutrophils. This implies they have important roles in the operation and regulation of the immune system. Human and murine FcγR are classified according to their affinity for IgG and whether they induce an activation or inhibitory signal. There are 3 classes of human FcγRs that bind IgG with high (FcγRI) and low (FcγRII and FcγRIII) affinity. The FcγRII and FcγRIII are sub-classified into activatory receptors, FcγRIIa and FcγRIIc, that contain an integral immunoreceptor tyrosine-based activatory motif (ITAM), and FcγRIIIa, which associates with the FcR γ or ζ chain, which also contain ITAMs. The inhibitory FcγRIIb contains an immunoreceptor tyrosine-based inhibitory motif (ITIM).

A fifth low affinity FcγR (FcγRIIIb) exists with expression restricted to neutrophils and a subset of eosinophils that lacks an integral intracellular signalling domain and is glycosyl phoshatidylinositol (GPI) anchored in the cell membrane. A summary of the FcγRs are shown in Table 5.1.

IgG molecules are grouped into sub-classes based on their relative abundance in plasma and are defined by differences in the constant heavy chain regions in their Fc region. FcγRI consistently binds monomeric IgG subclasses, whereas only FcγRIIIa 158V allotype of the low affinity receptors can bind monomeric IgG3. All FcγRs and their allotypes have been shown to bind to IgG-containing immune complexes with different affinities [320, 321] (Table 5.1).

Table 5.1. High and low affinity FcγR summary						
CD marker	CD16a	CD16b	CD32a	CD32b	CD32c	CD64
Protein	FcγRIIIa	FcγRIIIb	FcγRIIa	FcγRIIb	FcγRIIc	FcγRI
Gene	*FCGR3A*	*FCGR3B*	*FCGR2A*	*FCGR2B*	*FCGR2C*	*FCGR1A/B/C*
mRNA isoforms	a1	b1	a1, a2	b1-b3	c1-c5	a, b, c
Cell expression	Monocytes, macrophages, NK cells	Neutrophils, eosinophils	Myeloid lineage cells, platelets, endothelial cells	B cells, monocytes, macrophages	NK cells, neutrophils	Neutrophils, monocytes, macrophages
Signalling motif	FcR γ ITAM	GPI anchor	ITAM	ITIM	ITAM	FcR γ ITAM
Non-synonymous genetic variation	F158V L48H/R	hNA haplotype SH antigen	Q27W R131H	I232T	Q57X	V39I, R92x, Q224X, I301M, I338T
IgG subclass binding preferences	1,3>>>2,4	3>>1>>>2,4 (hNA1a) 3,1>>>2,4 (hNA1b) 3≥1>>>2,4 (hNA1c)	1≥3>>>2,4 (131H) 3>1>>2,4 (131R)	3≥1>4>>>2	1≥3>>>2,4	1≥3>>4>>>2

Table 5.1. Summary of the high and low affinity FcγRs. The 5 low affinity FcγR proteins are highly homologous within each class, and show specific patterns of cell surface expression. FcγRIIIa, FcγRIIa, and FcγRIIc all contain ITAMs in their cytoplasmic signalling tails whilst FcγRIIb contains an ITIM, and is the only inhibitory FcγR. A number of mRNA isoforms exist for the FcγRII molecules which, for FcγRIIb, encode differences in the cytoplasmic signalling tail. Non-synonymous variation exists in all of the FcγRs, some of which affect IgG affinity (158 V/F). The Q57X variant of FcγRIIc generates a premature stop codon which is present in ≥60% of Caucasian individuals. IgG subclass binding preferences are shown in order from [320]. The affinities of the FcγR for each IgG isotype are altered by the polymorphic allotypes (see text for details). CD – cluster of differentiation, ITAM – immunoreceptor tyrosine-based activatory motif, ITIM – immunoreceptor tyrosine-based inhibitory motif, GPI – glycosyl phosphatidylinositol. ≥ - more than or equal binding affinity, > - binding affinity greater than.

5.1.2 Fc γ receptors and tumour necrosis factor antagonist therapy

Cross-linking of FcγRIIIa and FcγRIIa on monocytes and macrophages by IgG containing immune complexes induces receptor internalisation, phagocytosis, superoxide generation and TNF release [322-324]. Ligand binding to FcγRIIIa on NK cells is capable of inducing ADCC of antibody coated target cells, one of the principal mechanisms of tumour killing induced by monoclonal antibody therapy [325], as well as B cell depletion in RA by the anti-CD20 biologic rituximab [148]. Expression of a transfected human mutant membrane-bound TNF (mTNF) on murine myeloma cells was used to investigate the mechanism of action of infliximab, an IgG1 chimeric anti-TNF. Scallon *et al* demonstrated that these mTNF expressing cells were killed by both ADCC and complement dependent cytotoxicity (CDC) when incubated with peripheral blood lymphocytes [326]. Whilst this study did not highlight a specific cell type responsible these findings suggest a potential role for FcγRIIIa and/or FcγRIIc expressed on NK cells in mediating depletion of mTNF-expressing cells by monoclonal antibodies (mAbs) directed at TNF. Monocyte apoptosis has been observed within 4 hours of infliximab infusion in patients diagnosed with Crohn's disease [327]. However, this induction of apoptosis through mTNF was also evident in monocytes incubated with an infliximab Fab indicating this specific mechanism is not Fc mediated and potentially secondary to reverse signalling. Induction of apoptosis in synovial tissue was not observed from a small cohort (n=24) of RA patients 48 hours post treatment with infliximab, despite decreased synovial cell numbers of macrophages, T-cells and plasma cells [328].

Etanercept is a recombinant fusion molecule of human TNF receptor 2 to human IgG1 hinge, CH2 and CH3 regions. This structural difference between etanercept and mAb TNF antagonists may affect their binding affinity to FcγR.

The low affinity FcγRs are hypothesised to be the main effector molecules for autoantibody-mediated biological effects relevant to RA pathogenesis [29, 329]. These IgG receptors may play an important role in the clearance of immune complexes, thus generating complex associations across this receptor family specific for each autoantibody and type of immune complex [29, 330]. TNF antagonists form immune complexes with their cognate soluble TNF molecules, indicating a role for FcγRs in the clearance of these TNF-containing immune complexes in the therapeutic efficacy of these drugs [331]. RFs have the potential to increase the size and modify the structure of circulating biologic-TNF immune complexes, potentially increasing the diversity of FcRs bound, depending on the RF subclass present within the immune complex (e.g. IgM RF binding to Fcα/μR).

Neutrophils express FcγRIIa, the only low affinity FcγR to bind strongly to IgG2, and FcγRIIIb. Upon activation, neutrophils demonstrate a respiratory burst; release of H_2O_2 and reactive oxygen species (ROS) [332, 333]. Both of these molecules are essential for host defence against invading pathogens; they also have the potential to cause tissue damage, particularly where there is a sustained inappropriate inflammatory response as present in RA [334]. However, polymorphonuclear leukocytes (PMNs) incubated with synovial fluid of RA patients or large insoluble immune complexes do not induce this respiratory burst [333]. Instead these 'primed' neutrophils, and collectively PMNs, are responsible for immune complex clearance via FcγRIIIb [335], whilst serum-opsonised particle induced phagocytosis is largely mediated by FcγRIIa [336]. The bivalency of mAb TNF antagonists may allow the formation of large immune complexes which can be cleared by FcγR-mediated phagocytosis. Etanercept is monovalent and therefore unlikely to form large immune complexes with sTNF. This would possibly prevent cross-linking of cell surface FcγR and reduce subsequent immune complex clearance.

5.1.3 Fc γ receptor biology guided analysis

The investigation of genetic variation at the low affinity *FCGR* locus is based upon these IgG receptors known biology and proposed mechanisms of effect of TNF antagonist therapy. The following hypotheses are proposed to guide statistical analyses:

1. TNF antagonists bind to mTNF on TNF producing cells resulting in cross-linking of cell surface FcγRs that induce downstream effector functions that are both FcγR and cell type specific. These mechanisms are postulated to be altered for the two mAbs (infliximab and adalimumab) as opposed to the recombinant TNFR (etanercept). Furthermore, both ADCC (FcγRIIIa and FcγRIIc on NK cells) and/or phagocytosis (FcγRIIa and FcγRIIIa on macrophages and FcγRIIa and FcγRIIIb on neutrophils) of mTNF expressing cells may be enhanced in the presence of RF, with IgG RFs also being capable of binding to FcγRs
    a. *FCGR3A* CN and/or allelic variation associated with FcγRIIIa expression or IgG binding affinity leads to improved responses to TNF antagonists that are increased in the subgroup receiving mAbs and those individuals with circulating RFs.
    b. Genetic variation that results in the expression of FcγRIIc on NK cells, including both allelic variation and *FCGR2C* CN states are associated with improved responses, particularly within the mAb TNF antagonist drug class.
    c. *FCGR2A* allelic variants that alter IgG binding affinity leads to more effective phagocytosis of mTNF expressing cells and improved responses to TNF antagonists, particularly for the monoclonal antibodies in the RF positive subgroup.
2. TNF antagonists form circulating immune complexes with sTNF which are cleared by either the mononuclear phagocyte system (FcγRIIa, FcγRIIIa and FcγRIIc) or circulating neutrophils (FcγRIIa and FcγRIIIb). It is proposed that improved

outcomes are seen with variants that increase cell surface expression (increased CN) or those that enhance IgG binding.

The exact cellular and molecular functions of TNF antagonists cannot be defined from a single genetic association study, however, these data can be used to refine the hypotheses described above and direct focussed functional experiments to answer them. An additional layer of complexity not described here, but mentioned in Chapter 1.1.3, is the pathogenic role of autoantibodies. Therefore, whilst these hypotheses are based on improvements in clinical response there exists the possibility of autoantibody-mediated pathogenicity directed through these FcγRs.

5.1.4 Fc γ receptors, genomic organization and genetic variation

5.1.4.1 Fc γ receptor genomic organization and copy number variation

The genes encoding the low affinity FcγRs are located on human chromosome 1q23.3. High sequence homology exists between the two class III Fc γ receptor genes (*FCGR3A* and *FCGR3B*) and between the three class II genes (*FCGR2A, FCGR2B* and *FCGR2C*). This high homology arose via a segmental duplication (SD) of a genomic region originally containing a single *FCGR3* gene and two homologous *FCGR2* genes (Figure 5.1). An unequal crossover event resulted in the generation of a second *FCGR3A* gene, which later diverged into *FCGR3B* by accumulated point mutations, and a hybrid gene of *FCGR2A* and *FCGR2B* (*FCGR2C)* [337]. SDs are structural genomic features that occur at more than one site in a genome which are defined by their size (1-400kb) and homology (>90%) [338]. SDs are predicted to affect ~5% of the human genome sequence and occur in regions of high gene density, with genes related to immune system function being just one ontology enriched within these genomic features [338, 339]. Highly homologous repeats often flank SDs which may predispose to non-allelic homologous recombination and the generation of copy number variable regions (CNVRs); there is a 4-5 fold enrichment of CNVRs in regions

overlapping with SDs [339]. Copy number variation (CNV) can vary in size from 1kb to several megabases and encompasses duplications and deletions relative to a reference genome. Single nucleotide differences between SD homologs may be used to uniquely identify specific paralogs, called paralogous sequence variants (PSVs). Genotyping SNPs that lie within SDs may be subject to confounding from both PSVs (when two alleles of a SNP are actually two different paralogous nucleotides) and complex single nucleotide variants called multisite variants [340]. CNV may represent simple duplications or deletions of a region that is polymorphic within a population. Complex CNV has been described for some loci where both duplications and deletions are documented, whilst others show tandem CNVs that may act independently; these complex CNVs are particularly enriched within regions of SD [341]. The genomic structure of the *FCGR* locus and high sequence homology precludes its in-depth investigation using genotyping technologies that rely on short probe hybridisation or short sequencing reads (<100bp).

Deletions of the genomic region containing *FCGR3B* and *FCGR2C* had been reported in a number of families prior to association between glomerulonephritis and loss of *FCGR3B* in both rats and humans [342-344]; this was attributed to the loss of *FCGR3B* and concomitant loss of neutrophil FcγRIIIb expression and thus impaired immune complex clearance. Multiple CNVs have been recorded at the *FCGR* locus, with both deletions and duplications reported to affect *FCGR3A, FCGR3B* and *FCGR2C*, but there are no reports for *FCGR2A* or *FCGR2B* [341, 345-347]. *FCGR3B* deletions have been associated with susceptibility to SLE and RA in humans [330, 348-350], whilst *FCGR2C* alleles carrying an active form of this receptor, with an intact open reading frame (ORF), have shown association with an increased risk of idiopathic thrombocytopenia purpura [347]. CNV of *FCGR2C* does not appear to be independent of the *FCGR3A* and *FCGR3B* CNVRs [345], which are themselves independent of each other.

The breakpoints and limits of the *FCGR* CNVRs have not been fully resolved, though the positions of multiple putative breakpoints indicate these coincide with the boundaries of the SD [351, 352]. The presence of multiple breakpoints may also indicate that deletions and duplications of both CNVRs arise independently and are not associated with a specific genetic background.

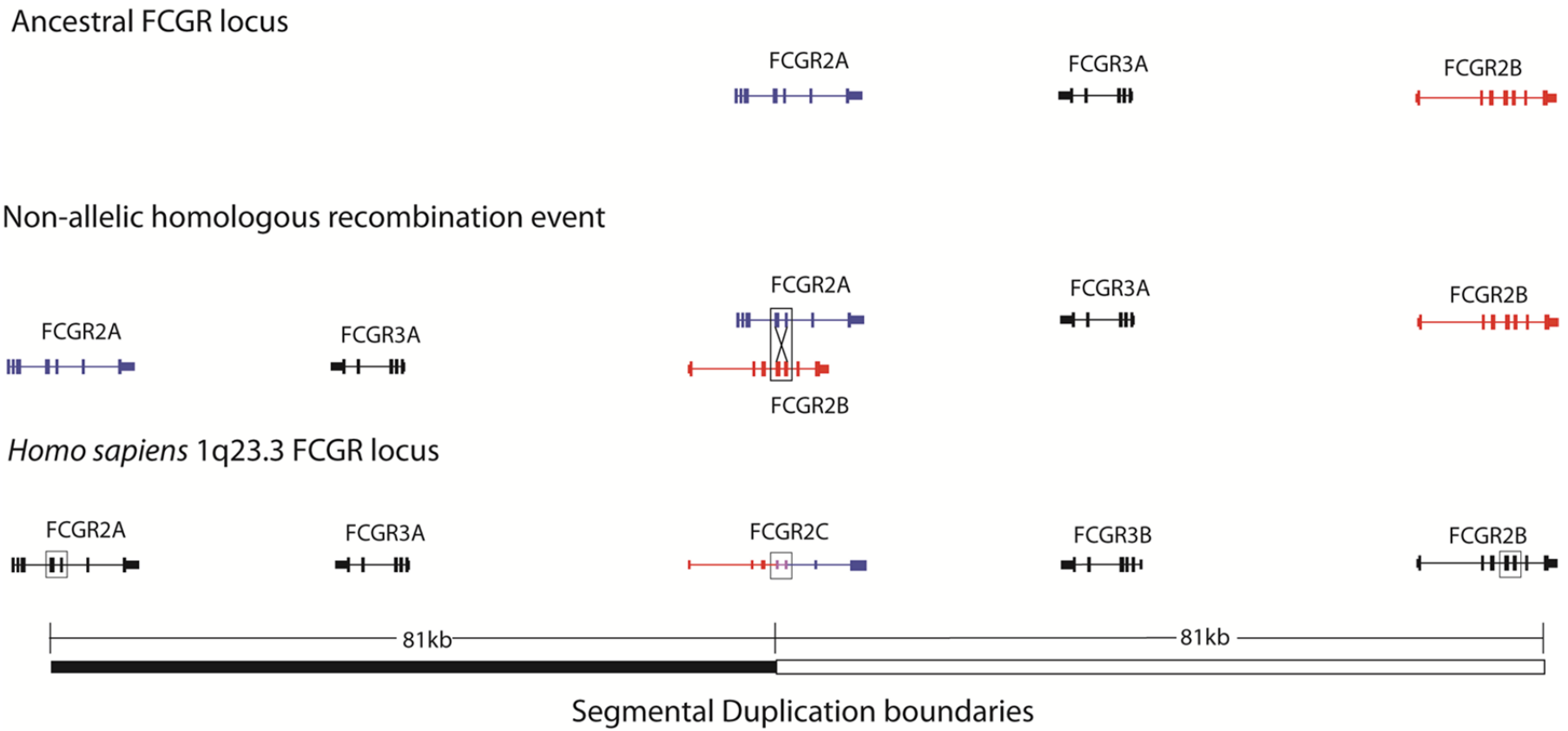


Figure 5.1. A model of the evolution of the *Homo sapiens* chromosome 1q23.3 low-affinity *FCGR* locus. The extant *FCGR* locus is thought to have arisen due to an unequal cross-over event, or non-allelic homologous recombination, between a highly homologous region between the ancestral *FCGR2A* (blue) and *FCGR2B* (red) genes. This recombination event resulted in a duplicated locus, or segmental duplication, flanked by *FCGR2A* at the centromeric end, and *FCGR2B* at the telomeric end. A new hybrid gene, *FCGR2C*, was created by this recombination and a duplicated *FCGR3A* gene, now known as *FCGR3B* which has functionally diverged from its ancestral precursor. The highly homologous region within *FCGR2A, FCGR2B* and *FCGR2C* is shown by the open box. The boundaries of the SD are defined by the limits of homology. CNV at this locus may have arisen as a direct result of the SD as the reported breakpoint regions correspond closely to these SD boundaries.

5.1.4.2 Measuring copy number variation

Genome-wide detection methods for CNV have utilised both high-throughput SNP and custom tiling path arrays (array comparative genomic hybridisation (aCGH)) that utilise the bacterial artificial chromosomes used to sequence and construct the human genome reference sequence [338, 341, 353, 354]. Single locus CNV measurement has also used quantitative PCR and probe hybridisation based multiplex ligation dependent probe amplification (MLPA) [330, 345]. Probe hybridisation based methods of CNV quantification are confounded by the presence of homologous sequences unrelated to the CNVR of interest, such as simple repeat elements or homologs of a SD. The low affinity FcγR locus displays two independent CNVRs which relate to the SD at this locus; hybridisation of short probes designed for one CNVR are therefore confounded by the presence of a highly homologous sequence that can skew copy number (CN) quantification. Alternative chemistries exist that utilise paralogous sequence amplification, but which contain indel PSVs that can differentiate between the two paralogs on the basis of size by capillary gel electrophoresis (paralog ratio test (PRT)) [355]. Where a single CNV and single CN invariant paralog are co-amplified, absolute quantification can be achieved. Where both paralogs are subject to CNV, a third homologous invariant region must also be co-amplified. Total CN of both variable regions can be obtained using this approach but individual CNVR measurement cannot. A second assay is required that can differentiate the total signal from the first assay into the relative proportions from each CNVR [348, 349]. The design and genomic position of the invariant reference primers for these PCR-based assays must take into consideration the position of the CNVR breakpoints in order to avoid confounding from dynamic breakpoints, i.e. those that vary between individuals. If the reference primers overlap with a region of CNVR breakpoints then the reference assay may be subject to confounding from the CNVR it is attempting to measure (Figure 5.2).

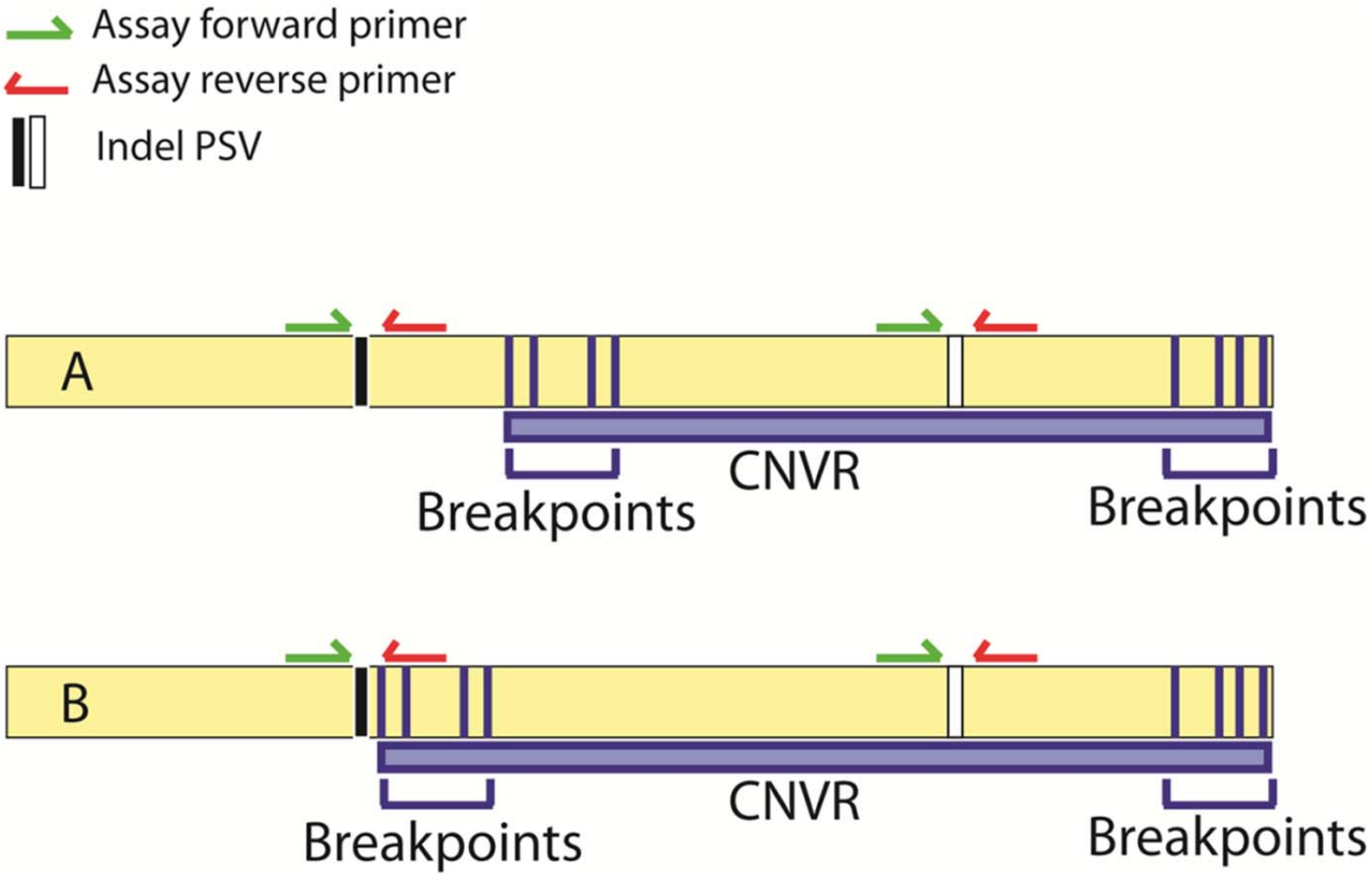


Figure 5.2. CNV assays are subject to confounding from dynamic breakpoints. The design and positioning of primers for CNV assays such as the PRT are critical and need to take into account the location of dynamic breakpoints. A – The PRT assay infers relative CN of the locus of interest using an indel PSV (filled/open box). The fluorescently labelled primers co-amplify the homologous region and the PCR products are separated by capillary gel electrophoresis. The signal intensity corresponds to the relative proportion of each product, i.e. a paralog ratio of 1 (1:1) represents equal CN, whilst 0.5 (1:2) represents a loss relative to the reference sequence. B – The reference primers overlap with the breakpoint region of this CNVR. Accurate measurement of this CNV will only be achieved for the CNVRs that have breakpoints that do not overlap with the reverse primer. Those that do overlap may give spurious ratios introducing bias into downstream analyses. The positioning of the invariant reference sequence need not be in the same region of the genome.

5.1.4.3 Non-synonymous single nucleotide polymorphisms

Several common non-synonymous SNPs have been described which affect FcγR function, including those described in Table 5.1.

Two non-synonymous genetic variants have been described in the gene encoding FcγRIIa (*FCGR2A*). The 131H>R variant (rs1801274) lies in a putative IgG binding region of the FcγRIIa ectodomain [356]. FcγRIIa is the most widely expressed FcγR; allotypes of the 131H>R show differences in IgG2 binding [357], and IgG1 and IgG3 opsonized bacteria induced phagocytosis [358]. PMNs from 131H allele homozygotes were able to phagocytose IgG2 opsonised encapsulated streptococcus bacteria, whilst 131R homozygotes lacked the ability to phagocytose these bacteria in an IgG2 dependent manner [359]. The crystal structure of FcγRIIa indicates different dimer formations between these two allotypes [356], a difference that may account for their differential binding of CRP; 131R allotypes bind CRP stronger than 131H allotypes [360]. The second variant is formed by a dinucleotide change that results in a glutamine to tryptophan change at amino position 27 [95, 361]. The functional impact of this variant has not been published to date (2013).

FcγRIIb is the only inhibitory IgG receptor and is expressed on B cells, and some monocytes and macrophages. A low-frequency non-synonymous polymorphism (in Caucasian populations) exists within the transmembrane domain of the FcγRIIb protein (I232T); the isoleucine allele has a reduced inhibitory capacity on B cell receptor signalling [362]. Japanese individuals homozygous for the T allele are at an increased risk of SLE [363]. The minor allele of this SNP has a lower frequency in populations of European and Caucasian ancestry (Caucasian MAF=0.10 vs. African MAF=0.25).

FcγRIIc is expressed on NK cells and can also mediate ADCC upon receptor ligation like FcγRIIIa, though it signals through its cytoplasmic ITAM, whereas FcγIIIa signals through the common γ chain (FcRγ) [364]. The gene encoding FcγRIIc (*FCGR2C*) arose through an unequal crossover event between *FCGR2A* and *FCGR2B* and contains a common SNP that results in a missense change at codon 57 [337, 365]. Expression of this FcγR is therefore determined by the alleles of this SNP.

The phenylalanine to valine polymorphism (amino acid 158) in the second extracellular domain of FcγRIIIa affects IgG isotype binding preferences, affinity and effector cell function [148, 366]. Healthy control donors homozygous for the V allele bind more IgG1 and IgG3 mAb than donors homozygous for the F allele, and show low levels of binding of IgG4, which the FF donors are unable to bind. Wu *et al* showed that this resulted in 3-fold higher internal calcium mobilization and CD25 expression in NK cells when stimulated with aggregated human IgG [366]. They also noted an association between the F allele of this variant and susceptibility to SLE, which has since been replicated in both European and Asian populations [367], whilst the higher affinity V allele was associated with RA, which may be specific to autoantibody-positive males [241, 368, 369].

The neutrophil-specific FcγRIIIb is characterised by the human neutrophil antigen system (hNA), a haplotype of 4 non-synonymous and 1 synonymous common SNPs that affect neutrophil function. Investigation of the functional consequences of the hNA1a and hNA1b allotypes found a lower phagocytic capacity of neutrophils from donors homozygous for hNA1b compared to those homozygous for hNA1a [370]. Salmon *et al* demonstrated this to be due to carbohydrate:non-IgG ligand interactions between allotypes, which is concordant with observed differences in the number and nature of glycosylation sites between hNA1a and hNA1b homozygous donor neutrophils [371]. Bredius *et al* reported differences between hNA allotypes specific to IgG1 and IgG3 opsonised whole bacteria, but

not IgG2 [358]. A third allele (SH; hNA1c) of the hNA system has also been described on the hNA1b genetic background, though the functional consequences are unknown [372].

5.2 Methods

5.2.1 *FCGR2C* quantitative sequence variant assay: concept, background and development

CNV at the *FCGR2C* locus is less well characterised as it lies intermediate to the CN variable genes *FCGR3A* and *FCGR3B* and may be influenced by each of these CNVRs. Previous unpublished and extensive re-sequencing of the region, by Dr Jim Robinson, uncovered numerous PSVs that were able to differentiate *FCGR3A* from *FCGR3B*, as well as PSVs that were specific to the homologous regions of *FCGR2A*, *FCGR2B* and *FCGR2C* [95]. These included a single tri-parallelic PSV that was able to discriminate between all three *FCGR2* genes. A published PRT assay that measured the relative CN of a homologous sequence proximal to *FCGR2A* and *FCGR2C* displayed large variation in CN estimates [348]. This may be related to the design of the assay primers in a region that may harbour CNVR breakpoints or overlap with SNPs (Section 5.1.4.2, Figure 5.2). This chapter describes the development and optimisation of a novel *FCGR2C* CNV measurement assay. All of the optimisation experiments were carried out using a panel of Northern European ancestry reference samples (HapMap CEU), which have been characterised with a number of CNV detection and measurement assays, both in the literature [341] and in-house [95].

5.2.1.1 The quantitative sequence variant method

The quantitative sequence variant (QSV) method measures the relative contribution of a particular nucleotide within a homologous sequence to the total electropherogram peak height for that position (Figure 5.3). For instance, in a pool of homologous sequences a difference may exist at a single nucleotide that can differentiate one sequence from all the others (e.g. a PSV). All of these sequences are the same length, and the only difference is the nucleotide at this position. Sanger sequencing and measurement of electropherogram peak heights make it possible to infer the relative contribution of this sequence to the total

pool of sequences [373]. A copy number proportion (CNP) score is calculated that is normalised to the sequence context of the PSV and the background signal. When extended to the inference of CN at a given locus, homologous sequences are co-amplified by common PCR primers. The peak heights of a known PSV can be measured from sequencing electropherograms, and the relative contribution of one nucleotide to its homolog obtained. This approach has been used, in conjunction with a PRT assay, to measure CNV of both *FCGR3A* and *FCGR3B* [349].

The QSV method also allows the determination of SNP genotypes in a quantitative manner in regions of high homology without relying on gene-specific amplification. This is achieved by generating two reference sequences for each SNP, one for each homozygous state of the two SNP alleles. These reference sequences define the upper and lower limit of the peak heights possible at this nucleotide position. QSV Analyser therefore generates CNP scores based on these two maximum values (Figure 5.3) [373].

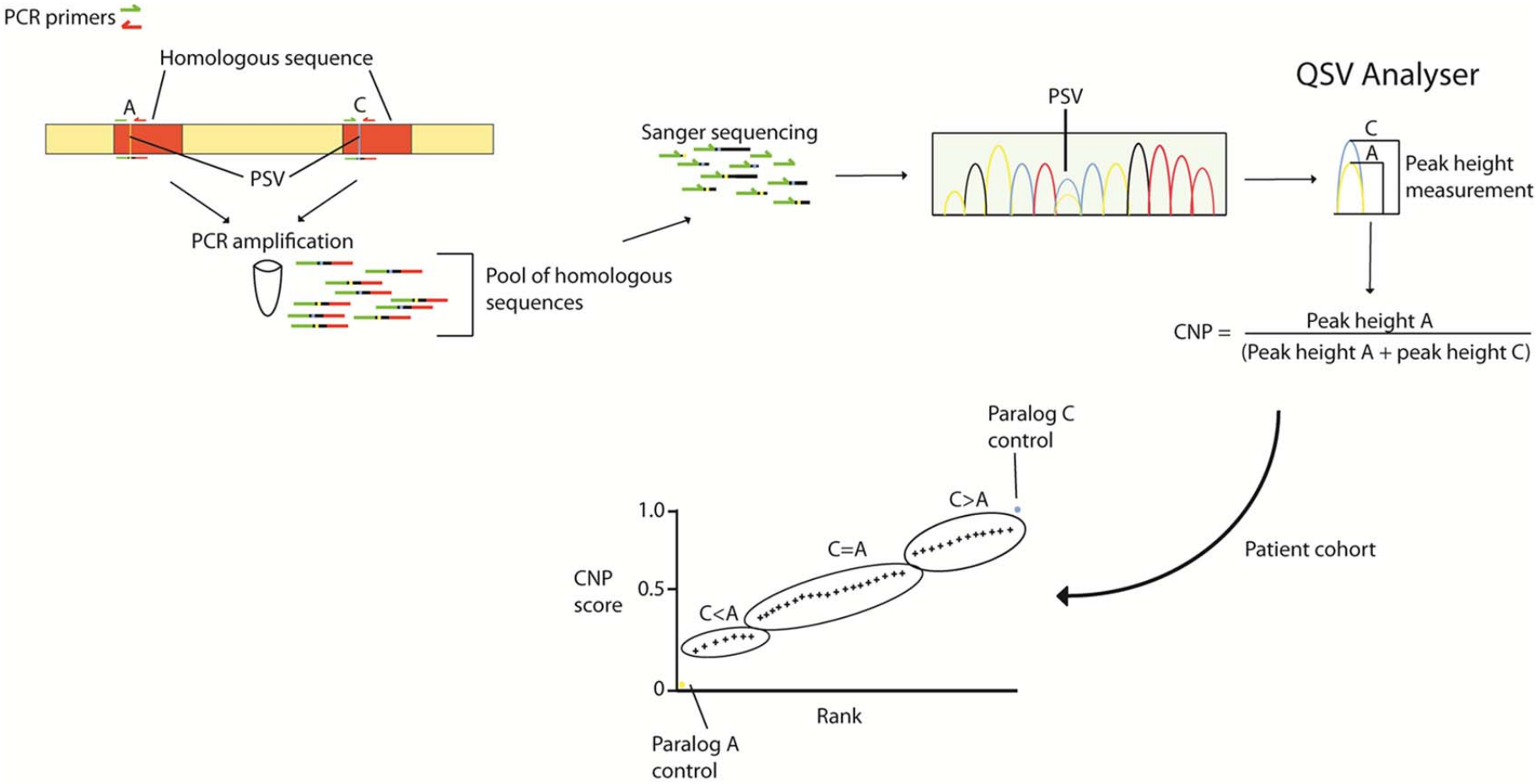


Figure 5.3. The QSV method of CNV measurement. Homologous sequences that contain PSVs are first co-amplified in a PCR reaction followed by Sanger sequencing of the amplicons. Electropherograms are processed and the peak heights at the PSV nucleotides are measured. QSV Analyser normalises the peak heights measurement to the background sequencing trace signal and the sequence context of the PSV, both of which affect individual peak heights. The CNP score is the proportion of the test nucleotide (peak height A) relative to the total signal at that position (peak height A + peak height C). By running this method on a cohort of DNA samples, such as those from a patient cohort, relative CN can be inferred by plotting all of the ranked CNP scores. Controls sequencing electropherograms are included which provided maximum and minimum peak heights for QSV Analyser. This method can be extended to included SNP genotyping where the two homozygous states form the upper and lower limits of each nucleotide peak height (analogous to the paralog controls).

5.2.1.2 A tri-parallelic sequence variant differentiates between three homologous genes

The PSV that differentiates between the 3 *FCGR2* genes was investigated as a potential target for measuring CN of *FCGR2C* (herein referred to as "TriPSV"). Redon *et al* developed a custom whole genome tiling path (WGTP) array; a 150kb probe mapped to the *FCGR* locus (8H4), for which data on the HapMap CEU population was publicly available (http://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-107/) [341, 353] (Figure 5.4A). The $\log_2$ probe intensity of 8H4 was plotted directly against the *FCGR2C* nucleotide peak height measured by QSV Analyser [373] for the TriPSV on these samples. Clusters indicative of CN states were well delineated on both axes with a broadly linear relationship (Figure 5.4B). The observation of 4 clusters suggests this single nucleotide may discriminate between different CN states, and appears comparable to a 150kb tiling path probe. The 90 HapMap CEU individuals (30 trios) were used to develop and validate a CN assay based on the TriPSV.

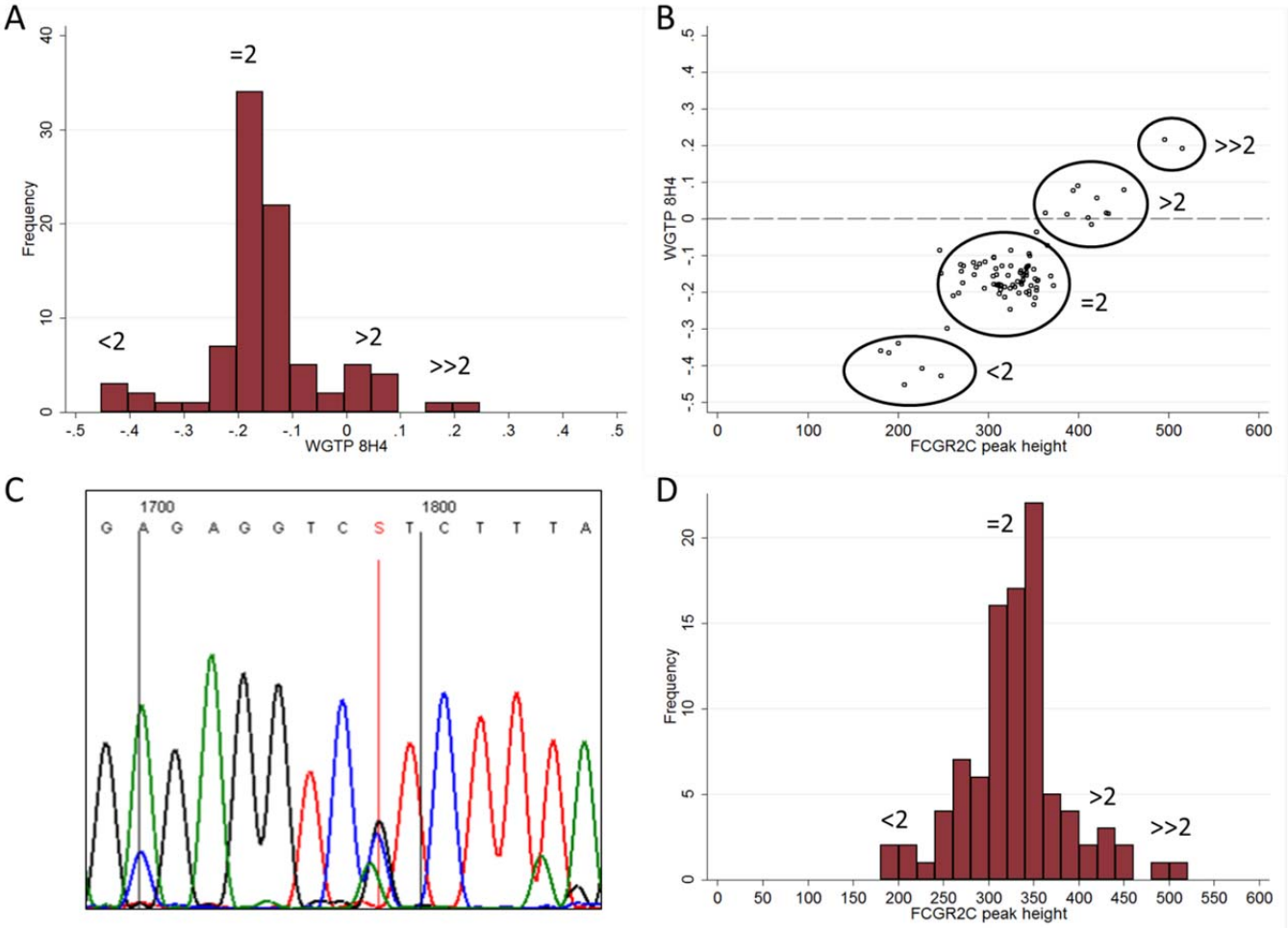


Figure 5.4 – The detection of CNV by a single nucleotide variant within the FCGR2C gene. A – The distribution of WGTP probe 8H4 $\log_2$ intensity indicates CNV groups at the *FCGR* locus. B- The tri-allelic PSV is comparable to the 150kb WGTP array probe (8H4) in discriminating between CN states across the locus seen by the presence of 4 clusters C) Snap shot from QSV Analyser of the tri-allelic PSV sequence context. The green peak (adenine) represents the *FCGR2C* peak in this individual with a deletion of *FCGR2C*. QSV Analyser measures the peak height of each nucleotide signal at a single site, normalized to the background and surrounding sequence context [372]. D – The distribution of *FCGR2C* normalised peak heights measured using QSV Analyser shows deviations that may represent CN states of the genomic interval containing this nucleotide.

5.2.1.3 TriPSV assay optimisation

A single PCR reaction that co-amplifies a homologous region of the 3 *FCGR2* genes, followed by Sanger sequencing allows the quantification of electropherogram peak heights. A number of CNV assays are sensitive to subtle differences in processing and chemistry, which can affect their ability to discriminate between CN states as well as introducing systematic bias into downstream statistical analysis [374, 375]. The chemistry of the TriPSV assay is PCR-based, thus it relies on the efficient amplification of specific templates in both an initial exponential amplification reaction, and a subsequent linear amplification Sanger sequencing reaction. Technical variation and confounding in assay performance can be introduced at multiple stages, requiring the optimisation of these steps in order to minimise both batch effects and technical bias before quality control filters are imposed. The first step of any PCR reaction is to optimise the annealing temperature of the primers, which is sequence and salt concentration dependent. The estimated melting temperature [376] of the primers was used as a starting point for an annealing temperature gradient PCR reaction to select the optimum annealing temperature. The optimum annealing temperature for this primer pair was 55°C. Other reaction conditions that may affect assay performance and CN discrimination are Mg^{2+} concentration, primer:template ratio, PCR reaction thermocycling conditions, Sanger sequencing thermocycling conditions, capillary electrophoresis running and injection voltage, and Sanger sequencing reaction salt concentration. These conditions were optimised for discrimination of CN clusters using a panel of 31 HapMap CEU reference individuals with known CN states across the *FCGR* locus. For instance, Figure 5.5C shows scatter plots of the *FCGR2A*:*FCGR2C* and *FCGR2B*:*FCGR2C* CNP scores across a number of primer concentrations. The higher primer concentrations (e.g. 15µM) are better able to discriminate the *FCGR2C* CN<2 states from the rest of the samples, whilst the 7.5µM concentration appeared to discriminate well between all three groups.

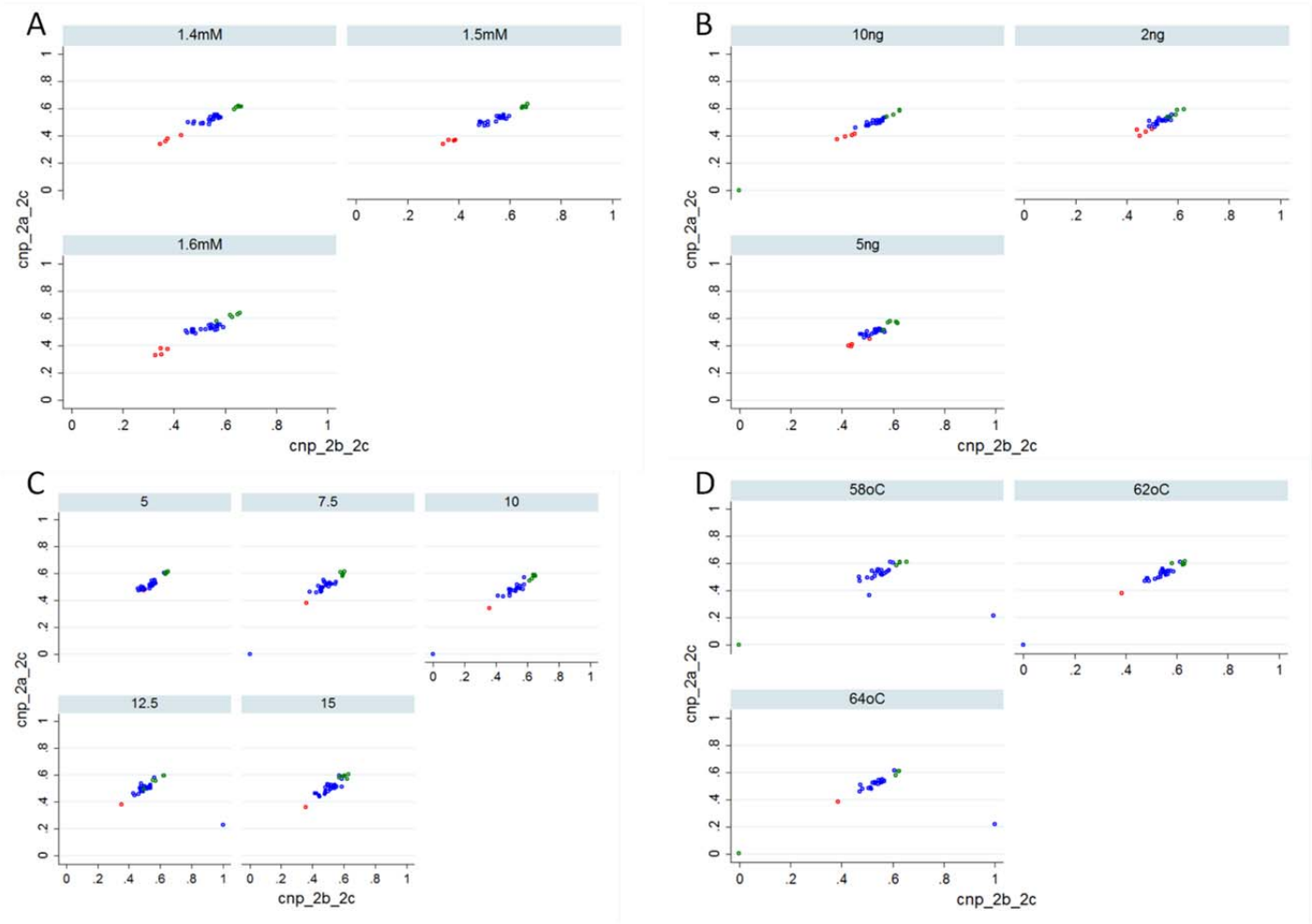


Figure 5.5 – *FCGR2C* CN assay optimisation. CN cluster discrimination is dependent on the reaction conditions in both the PCR and Sanger sequencing amplification reactions. Red circles – CN<2, blue circles – CN=2, green circles – CN>2. A – The availability of Mg^{2+} ions affects amplification efficiency. However, it does not appear to have an appreciable impact on discriminating between CN clusters. B – The ratio of primer to template in the initial PCR reaction can also affect amplification efficiency at low concentrations of template DNA. The best discrimination between the CN clusters is observed in the 5ng reactions, but does not have a major impact above 10ng. C – Template to primer ratio can also be optimised by altering the primer concentration. Discrimination of both deletion and duplication CN individuals is optimal around 7.5µM. D – Sanger sequencing reaction kinetics are altered by the thermocycling conditions. The recommended extension temperature for the Big Dye 3.1 reagent is 60°C. However better discrimination of CN clusters is observed between 62 and 64°C.

Optimisation of the other reaction conditions listed above was carried out in order to improve the accuracy of the CN calls. For the initial PCR reactions, $MgCl_2$ concentration, initial genomic DNA concentration and reaction buffer were optimised; the final reaction conditions are listed in Table 5.2. Altering the $MgCl_2$ concentration in ±0.1mM increments per reaction did not alter the assay results, and so it was maintained at 1.5mM (Figure 5.5A). Genomic DNA concentration did not appreciably alter the discrimination of CN clusters in the range 20-100ng; however, differences were observed at concentration <10ng, thus all further reactions used DNA above 20ng (Figure 5.5B). The Sanger sequencing reaction and capillary electrophoresis conditions were optimised by altering the reaction extension temperature ±2°C between 56°C and 64°C. Multiple sample sequencing reactions failed to adequately amplify in the 56°C and 58°C extension temperature experiments. The optimal condition was at 62-64°C, despite the manufacturer's recommended extension temperature of 60°C (Figure 5.5D); 62°C was selected to minimise thermal denaturing and degradation of the polymerase in the Big Dye 3.1 reagent. In order to measure the maximum electropherogram QSV peak height, QSV Analyser looks for the highest value for each peak within a 10 nucleotide range. The number of measurements per peak can be increased by slowing down the passage of each nucleotide past the detector cell. This was achieved by reducing the running voltage from the default 8kV. By increasing the number of measurements within that range a more accurate quantification of peak height may be obtained. The capillary electrophoresis running voltage was altered to increase the number of scans per nucleotide by running at 3kV and 5kV. Whilst more measurements per peak were obtained at 3kV the total running time increased, resulting in truncated sequencing products. Using a running voltage at 5kV increased the number of scans per nucleotide, but the differences in peak height measurements were negligible so the running voltage was maintained at 8kV.

Table 5.2. TriPSV PCR reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	40.5µl
10X Mg+ buffer	1.5mM $MgCl_2$	5.0µl
dNTPs	200µM each dNTP	1.0µl
TriPSV F (7.5µM)	150nM	1.0µl
TriPSV R (7.5µM)	150nM	1.0µl
Taq polymerase	1U	0.5µl
gDNA	20ng	1.0µl
Total		50µl

The incorporation efficiency of a dideoxy nucleotide triphosphate (ddNTP) into a growing nucleotide chain in a sequencing reaction is influenced by the preceding sequence, and thus the peak height of any given nucleotide is context-dependent on the sequence preceding it. This phenomenon has been documented previously and informed the development and optimisation of the Big Dye terminator dye by ABI [377]. It is not known whether Applied Biosystems include MnCit in their 5x sequencing buffer, as they do not reveal the constituents. MnCit concentration per sequencing reaction was titrated in 0.1mM increments from 0-1.0mM per reaction. The results based on variance in peak height and CN state discrimination suggested that inclusion of MnCit at any concentration did not improve the accuracy of the CN assay.

5.2.1.4 TriPSV peak height measurement and copy number determination

QSV Analyser is able to output both normalised peak heights and CNP scores [373]. Figure 5.4B shows a scatter plot of the WGTP log_2 probe intensity versus the normalised peak height for the *FCGR2C* paralog which was used to determine the CN states of the HapMap

CEU individuals. In order to calculate the CNP, QSV Analyser requires a reference sequence for each nucleotide to provide a theoretical maximum value for each pair-wise comparison; that is *FCGR2A* vs. *FCGR2C*, *FCGR2B* vs. *FCGR2C* and *FCGR2A* vs. *FCGR2B*. Reference sequences were amplified in a gene-specific manner using a long-PCR approach employing gene-specific primer pairs by Dr Jim Robinson [95]. A nested PCR using the TriPSV primer pair, followed by Sanger sequencing with the TriPSV reverse primer, generated paralog-specific electropherograms as reference sequences. Reference sequencing electropherograms were visually checked for the presence of PSVs that would indicate a non-specific sequence. The TriPSV sequence context from these reference traces can be seen in Figure 5.6 where combinations of PSVs, including the TriPSV, indicate the specificity of each reference sequence.

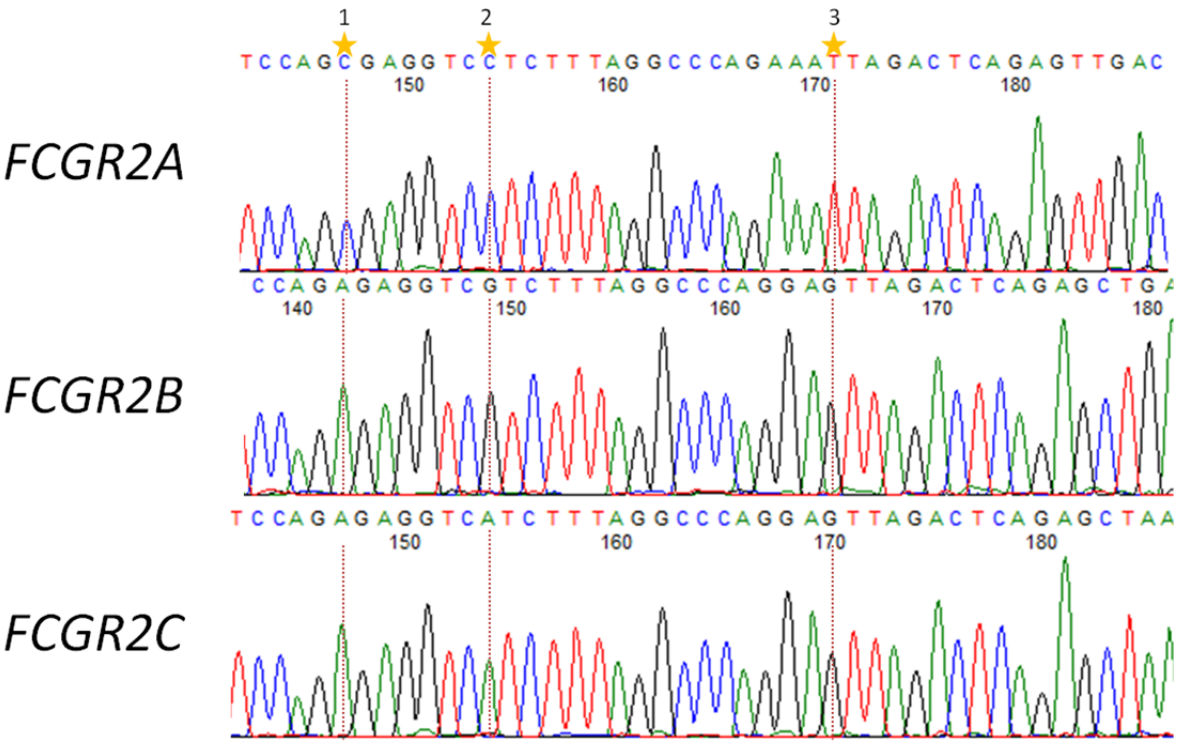


Figure 5.6. *FCGR2* paralog-specific reference sequences show the TriPSV context. The sequence surrounding the TriPSV nucleotide shows the gene-specific reference sequences used to calculate pair-wise CNP by QSV Analyser. The orange stars denote PSVs. Homology between *FCGR2C* and *FCGR2B* is higher than with *FCGR2A* in this region, but they can be distinguished by the nucleotide present at PSV position 2; the TriPSV.

QSV Analyser calculates the relative proportion of the test nucleotide to the reference at the position specified (Figure 5.4C). QSV Analyser can only calculate a CNP score for two paralogs at a single position; therefore three separate CNP scores were calculated for each pair-wise combination of nucleotides (*FCGR2A:FCGR2C*, *FCGR2B:FCGR2C, FCGR2A:FCGR2B*). The performance of several scores for determining CN were assessed: the individual CNP scores calculated by QSV Analyser, a relative peak height (RPH) ratio of *FCGR2C* to *FCGR2A* and *FCGR2B* to *FCRG2C* individually and averaged scores for both the CNP and RPH (Table 5.3). The principal difference between the CNP and the RPH scores is that the RPH is the proportion of the test peak height relative to the total signal for all 3 peaks at the TriPSV position, whilst the CNP is only proportional to the sum of the test and reference peaks in question. Figure 5.7 shows a comparison of scatter plots of CNP and RPH scores. The RPH

data are more dispersed, both within and between CN clusters, but do not show any improvement over the CNP score in ability to discriminate between CN states. This is reflected in the spread of the RPH scores around the mean in Table 5.3; the standard deviation (σ) is higher for the RPH than the CNP for all 3 CN states. The 3 CNP scores are all comparable with similar mean and σ values within each CN states. Figure 5.8 shows a histogram of the distribution of the average of the *FCGR2A:FCGR2C* and *FCGR2B:FCGR2C* scores plotted in the scatter plot in Figure 5.7 for the CNP and RPH. Theoretically the scores for *FCGR2A:FCGR2C* and *FCGR2B:FCGR2C* should be nearly equal under the assumption that neither *FCGR2A* nor *FCGR2B* vary in CN. The averaged CNP score shows a prominent peak in the expected region for individuals with CN<2 (~0.3), where the averaged RPH is more dispersed. Based on these data and comparisons made here, the CNP score was chosen over the RPH.

Table 5.3. Comparison of QSV scores for *FCGR2C* CN quantification				
TriPSV score	Equation	Mean (σ)		
		CN<2 n=8	CN=2 n=69	CN>2 n=12
FCGR2A:FCGR2C CNP	$CNP_{FCGR2A} = \frac{r_{FCGR2C}}{r_{FCGR2A} + r_{FCGR2C}}$	0.38 (0.07)	0.50 (0.03)	0.60 (0.03)
FCGR2B:FCGR2C CNP	$CNP_{FCGR2B} = \frac{r_{FCGR2C}}{r_{FCGR2B} + r_{FCGR2C}}$	0.36 (0.06)	0.49 (0.04)	0.60 (0.03)
Averaged CNP	$CNP = \frac{(CNP_{FCGR2A} + CNP_{FCGR2B})}{2}$	0.37 (0.06)	0.50 (0.03)	0.60 (0.03)
FCGR2A:FCGR2C RPH	$RPH_{FCGR2A} = \frac{\left[\frac{NPH_{FCGR2C}}{\sum NPH_{FCGR2}}\right]}{\left[\frac{NPH_{FCGR2A}}{\sum NPH_{FCGR2}}\right]}$	0.56 (0.16)	0.87 (0.10)	1.27 (0.13)
FCGR2B:FCGR2C RPH	$RPH_{FCGR2B} = \frac{\left[\frac{NPH_{FCGR2C}}{\sum NPH_{FCGR2}}\right]}{\left[\frac{NPH_{FCGR2B}}{\sum NPH_{FCGR2}}\right]}$	0.47 (0.10)	0.76 (0.11)	1.16 (0.16)
Averaged RPH	$RPH = \frac{(RPH_{FCGR2A} + RPH_{FCGR2B})}{2}$	0.51 (0.12)	0.82 (0.10)	1.21 (0.14)

Table 5.3. Comparison of QSV scores for *FCGR2C* CN quantification. Two different scores were calculated for the TriPSV QSV to measure relative CN of *FCGR2C* to *FCGR2A* and *FCGR2B*. QSV Analyser by default calculates the CNP score, a pair-wise score for the test (*FCGR2C*) and reference (*FCGR2A* or *FCGR2B*) nucleotides. The RPH score is the proportion of the test peak relative to the sum of all 3 peaks. Data are from the HapMap CEU samples used to develop and validate the TriPSV assay. The mean score for each CN state is an indication of the expected value for each CN group. The standard deviation (σ) shows the spread of the data for the RPH score is much broader than for the CNP score. σ – standard deviation, CNP – copy number proportion score, RPH – relative peak height score, NPH – normalised peak height, CN – copy number, QSV – quantitative sequence variant, TriPSV – tri-allelic paralogous sequence variant.

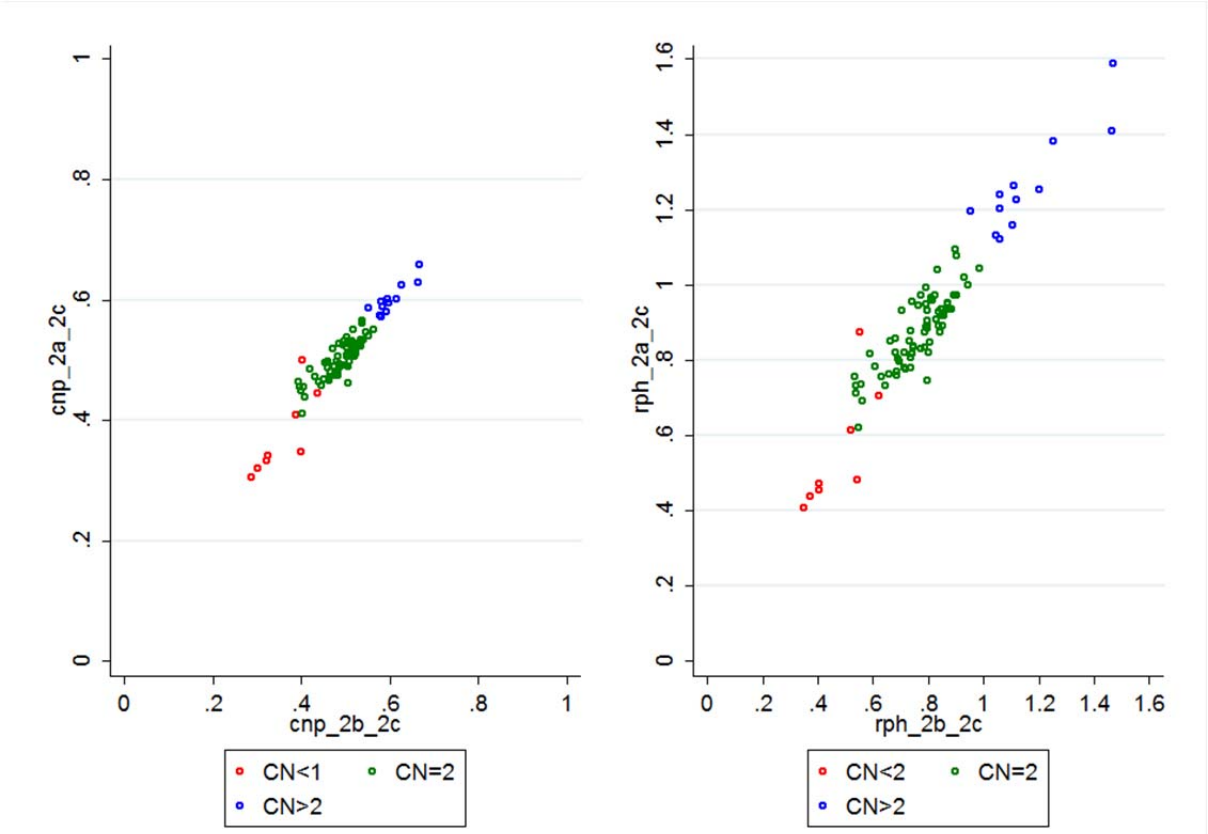


Figure 5.7. Comparison of CNP and RPH scores for *FCGR2C* CN determination. Left panel - *FCGR2A:FCGR2C* CNP plotted against the *FCGR2B:FCGR2C* CNP. *FCGR2C* deletion individuals can be delineated with low CNP values (<0.4), duplication carriers are less well defined by comparing CNP scores. Right panel – *FCGR2A:FCGR2C* RPH plotted against *FCGR2B:FCGR2C* RPH. *FCGR2C* duplication and deletion individuals are more distinct, however, there is increased variation in the data and data points within each CN group are more dispersed. CN states as deviation from 2n are coloured where red represents individuals carrying a deletion of the *FCGR2C* region, green two copies and duplications (3 or more) are in blue.

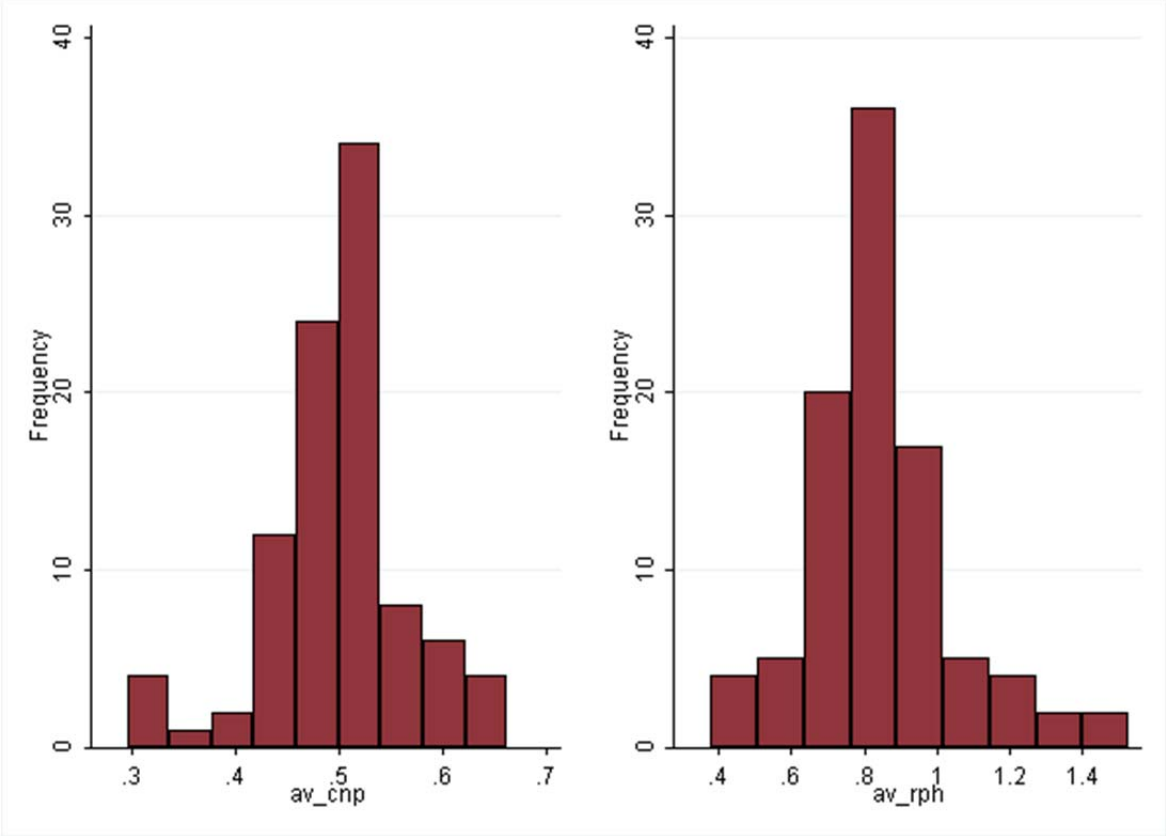


Figure 5.8. Comparison of averaged CNP and RPH scores for determining CN states in the CEU HapMap population. The average of the FCGR2A:FCGR2C and FCGR2B:FCGR2C scores for both the CNP and RPH were calculated and plotted as histograms. The left panel shows the distribution of the averaged CNP score whilst the right panel shows the averaged RPH score. The CN<2 group appear more distinct when using the averaged CNP score, while the groups in the averaged RPH are more dispersed which mirrors their distribution in Figure 5.7.

5.2.1.5 The tri-parallelic sequence variant includes an intrinsic quality control measure

By measuring the TriPSV with the QSV method, a third CNP score is calculated: the relative proportion of *FCGR2A* to *FCGR2B*. No CNV has been reported for these two genes, nor has any indication of CNV been observed across these loci in a panel of 32 reference samples used to re-sequence and fine-map all 5 *FCGR* genes at this locus [95]. The assumption of this CN invariant nature is utilised to infer CN states (CN<2, CN=2, CN>2) for *FCGR2C*. The mean CNP score for the CN=2 state, for both the *FCGR2A:FCGR2C* and *FCGR2B:FCGR2C* comparisons, is 0.5. The assumption of *FCGR2A* and *FCGR2B* CN invariance can therefore be checked by plotting the distribution of the *FCGR2A:FCGR2B* CNP. The expected distribution of values would have a mean of 0.5 and small standard deviation without

ancillary peaks. Assuming a normal distribution of values, >99% of the data would fall within ±3xσ of a mean value of 0.5 (Figure 5.9). A single sample (NA12750) falls outside of the upper limit. This sample has a deletion of the *FCGR3A* CNVR with a concomitant deletion of the overlapping portion of *FCGR2C* and TriPSV, which could indicate a loss of this homologous sequence in *FCGR2A*. However, the daughter of individual NA12750 has inherited this CNVR, but does not show any deviation from the expected proportion for the *FCGR2A:FCGR2B* score (0.52 against the expected 0.5). This indicates that it is unlikely that the TriPSV position in *FCGR2A* is affected by CNV in this family and that large deviation from the expected proportion is probably due to technical variation. This CNP score therefore gives an indication of the quality of the sequence electropherograms used to calculate CNP scores, and thus may be used as a quality control measure.

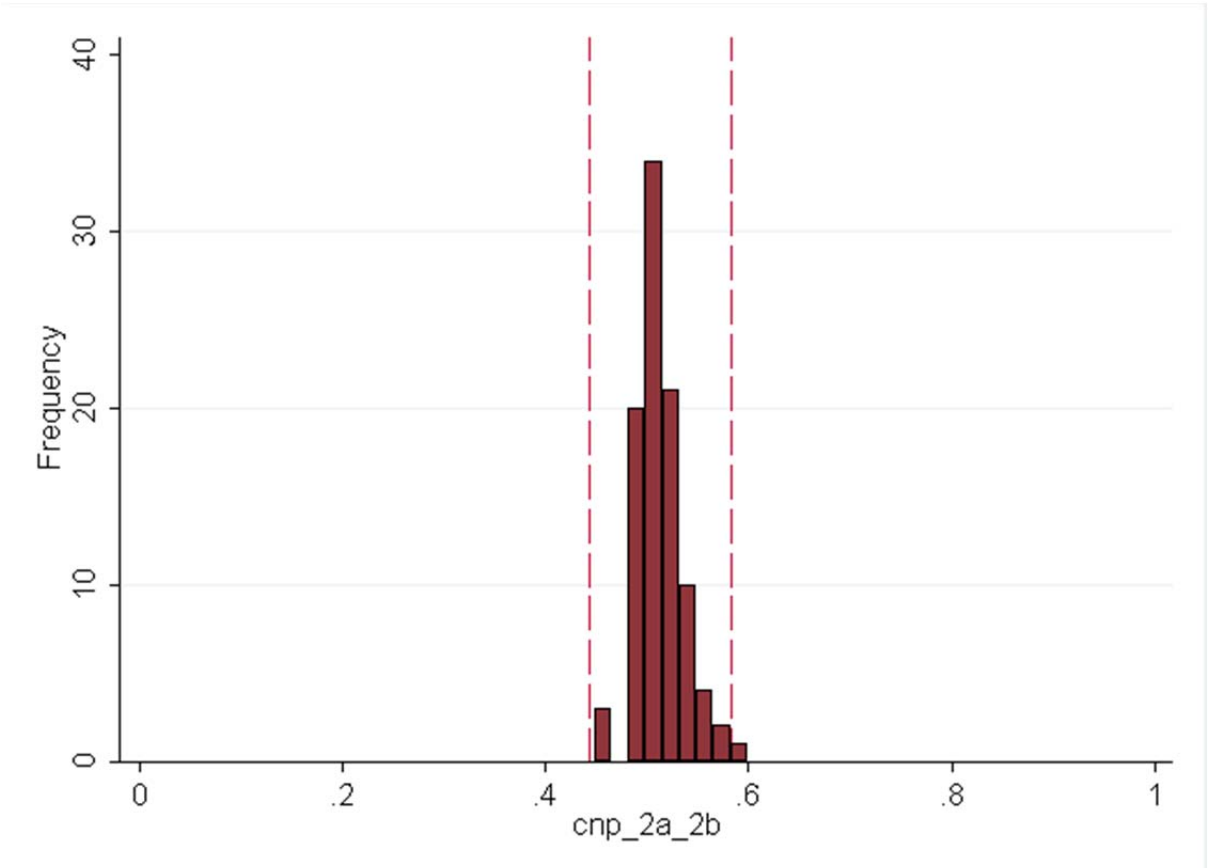


Figure 5.9. A histogram of *FCGR2A:FCGR2B* CNP values for the HapMap CEU population. Based on the assumption that *FCGR2A* and *FCGR2B* are CN invariant CNP scores are expected to approximately fit a normal distribution with >99% of values within ±3xσ. The upper and lower limits are shown by the vertical red dash lines (lower=0.44, upper=0.58). A single value falls outside of this range in the upper portion of the distribution which harbours a deletion of the *FCGR3A* CNVR, with subsequent loss of a copy of the portion of *FCGR2C* overlapping with the TriPSV.

5.2.2 Patient cohort description

RA case DNA samples were provided by the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate (BRAGGSS). Patients were invited to register into BRAGGSS following recruitment onto the British Society of Rheumatology Biologics Register (BSRBR). Inclusion criteria for BRAGGSS were physician diagnosis of RA, receiving one of three TNF antagonist biologics (etanercept, infliximab or adalimumab) and Caucasian ancestry. Blood for DNA extraction was donated at a single time point and processed in multiple batches; these are referred to as BRAGGSS1 and BRAGGSS4; high quality DNA from BRAGGSS2 and BRAGGSS3 were not available at the time of this doctoral work. Healthy control DNA samples from the Genetics of Rheumatoid Arthritis (GORA) study of European ancestry were donated by Prof Anthony G. Wilson at the University of Sheffield.

5.2.3 *FCGR2C* rs10917661 (Q57X) genotyping

A QSV method-based (Section 5.2.1.1) genotyping assay was developed on genomic DNA for the Q57X SNP (rs10917661) that determines the presence or absence of an ORF of *FCGR2C*. Primers were designed to anneal to a homologous region of *FCGR2B* and *FCGR2C* and were used to amplify this region in a panel of samples derived from cohorts of healthy control volunteers (n=9) and early RA patients (n=37). These samples had already been characterised for a number of flow cytometry markers, including CD32 (Clone 3D3 BD Biosciences), on CD3- CD56+ (CD3; Clone UCHT1, CD56; Clone B159, BD Biosciences, Oxford, UK) NK cells and had RNA extracted from whole blood. CD3-CD56+ NK cells express FcγRIIc, but not FcγRIIa or FcγRIIb therefore they were used to evaluate the translation of RNA transcripts arising from *FCGR2C,* since there are no FcγRIIc-specific antibodies. The *FCGR2C* cDNA from these samples had also previously been characterised for alternative splice forms of *FCGR2C* and for allele-specific expression of rs10917661 (unpublished experiments performed by Dr Dawn Cooper and Mrs Lubna Haroon-Rashid). Gene-specific

amplification of *FCGR2C* is not possible across the region containing rs10917661 using genomic DNA (gDNA); thus these data were used to validate the genomic assay for this SNP. *FCGR2C* isoform 1 was previously reported to correspond to the ORF allele of rs10917661, whilst isoform 3 corresponds to the stop codon allele [365]. A concordance of 100% was observed between the allele observed in the sequence electropherogram derived from the cDNA, the predicted splice isoform and the genomic assay for rs10917661.

A ~350bp region was amplified using the reaction conditions in Table 5.4 and the following thermocycling conditions: denature at 95°C for 30secs, followed by 32 cycles of 95°C for 15secs, primer annealing at 62°C for 15secs and chain elongation at 72°C for 1mins, followed by a final 5mins extension phase at 72°C. Primer sequences are listed in Appendix I: Table 1.

Table 5.4. rs10917661 polymerase chain reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	40.5µl
10X Mg+ buffer	1.5mM $MgCl_2$	5.0µl
dNTPs (10mM)	200µM each dNTP	1.0µl
2C ORF QSV 2 F (10µM)	150nM	1.0µl
2C ORF QSV 2 R (10µM)	150nM	1.0µl
Taq polymerase	1U	0.5µl
gDNA	50ng	1.0µl
Total		50µl

Amplicons were purified and Sanger sequenced with the forward primer using the method described in Chapter 2.2.2.2-2.2.2.3. QSV Analyser was used to measure the ratio of C:T peaks at the position for rs10917661 and calculate a QSV score. QSV score concordance was checked on 24 patient samples.

5.2.4 *FCGR3A* rs396991 genotyping

The *FCGR3A* non-synonymous SNP rs396991 (158F>V) was previously genotyped by Sanger sequencing by Dr Jim Robinson and Mr Stephen Martin as part of an association study of this polymorphism with RA susceptibility [241]. DNA from the BRAGGSS cohort was used as the template for PCR amplification with the conditions in Table 5.5 using the following thermocycling conditions: denature at 95°C for 30s, then 36 cycles of denature at 95°C for 30s, anneal at 52°C and elongation at 72°C for 1mins, followed by a final 72°C for 5mins. PCR reactions were checked for successful amplification by agarose gel electrophoresis. Amplification products were purified using the Charge Switch clean-up kit described in Chapter 2.2.3.2 prior to Sanger sequencing using the conditions described in Chapter 2.2.3.3 and the IIIA F primer.

Table 5.5. rs396991 polymerase chain reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	15.5µl
10X Mg+ buffer	1.5mM $MgCl_2$	2.0µl
dNTPs (10mM)	200µM each dNTP	0.4µl
IIIA F (10µM)	150nM	0.3µl
IIIA R (10µM)	150nM	0.3µl
Taq polymerase	1U	0.2µl
gDNA	20ng	1.0µl
Total		20µl

QSV Analyser was used in the current study to measure peak heights and calculate a CNP score for each sample. CNP scores for SNPs represent the contribution of each allele to a total maximum defined by the two homozygous states. Heterozygote individuals would expect to fall in a region intermediate to these two groups. For instance, if the T allele is

the test allele and the G is the reference a CNP score of 0 would represent the GG homozygotes, a score of 1 the TT homozygotes and score ~0.5 the heterozygotes.

5.2.5 *FCGR2A* non-synonymous single nucleotide polymorphism genotyping

5.2.5.1 rs9427399 genotyping

Two consecutive SNPs (rs9427397 and rs9427398) generate a dinucleotide polymorphism in *FCGR2A* that generates a previously described glutamine to tryptophan change at amino acid 27 of the full length receptor [361]. Two haplotypes of these two SNPs exist in Caucasian populations, though a third exists in populations of African descent that generates a third amino acid (arginine) change at position 27 of FcγRIIa (1000 Genomes African populations, March 2012 release). Several tag SNPs exist for the rs9427397/rs9427398 dinucleotide change which tag the two alleles together in Caucasian populations (D'=1.0), but individually are in moderate LD (r^2=0.67) with either of these two consecutive SNPs. One of the tag SNPs was selected based on its proximity (~300bp) and successful conversion into a Sequenom MassARRAY genotyping assay (rs9427399). Therefore this tag SNP is able to act as a proxy for the two haplotypes of rs9427397 and rs9427398 in populations of Caucasian ancestry. YEAR patients were genotyped on the Sequenom MassARRAY (n=817) or the Illumina Immunochip platform (n=179). A subset of the BRAGGSS cohort were genotyped for this proxy SNP on the Illumina Immunochip platform (n=1443). Patient DNA samples not genotyped on this platform were genotyped directly using the QSV method for rs9427399 by Mrs Lubna Haroon-Rashid. Template DNA from BRAGGSS patients was used in a PCR reaction that amplified a region around rs9427399 using the following thermocycling conditions: initial denature at 95°C for 30s, then 32 cycles of denaturing at 95° for 15s, primer annealing at 64°C and extension at 72°C for 1mins. A final extension at 72°C for 5mins was performed. Samples were checked for successful amplification by agarose (2%) gel electrophoresis stained with EtBr and

visualised using a BioRad GelDoc (BioRad, Hercules, California, USA). The reaction conditions are shown in Table 5.6 and the primer sequences are listed in Appendix 1:Table2. Amplicons were purified and Sanger sequenced as described in Chapters 2.2.2.2 and 2.2.2.3. QSV Analyser measured peak heights and calculated QSV scores which were used to manually genotype individuals based on the ranked QSV score.

Table 5.6. rs9427399 polymerase chain reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	15.6µl
10X Mg+ buffer	1.5mM $MgCl_2$	2.0µl
dNTPs (10mM)	200µM each dNTP	0.4µl
rs9427399F (10µM)	200nM	0.4µl
rs9427399R (10µM)	200nM	0.4µl
Taq polymerase	1U	0.2µl
gDNA	20ng	1.0µl
Total		20µl

5.2.5.2 rs1801274 (H131R) genotyping

A second non-synonymous SNP exists within the coding region of *FCGR2A* and was genotyped directly by either the custom Illumina Beadchip (Immunochip) or by the QSV method (by Mrs Lubna Haroon-Rashid), where data were not available from Immunochip typing. Template DNA from BRAGGSS patients were amplified in a PCR reaction designed to amplify the sequence around rs1801274 (Table 5.7) using the following thermocycling conditions: denature at 95°C for 30s, then 35 cycles of 95°C, 30s at 56.5°C and 72°C for 1min. A final extension step was performed at 72°C for 5mins. Reactions were checked for successful amplification by 2% agarose gel electrophoresis stained with EtBr prior to amplicon purification and Sanger sequencing (Chapters 2.2.2.2 and 2.2.2.3). QSV Analyser was used to measure peak heights and calculate QSV scores, which were used to manually genotype samples based on their ranked score.

Table 5.7. rs1801274 polymerase chain reaction conditions

Reagent	Reaction Concentration	Volume
dH_2O	n/a	15.76
10X Mg+ buffer	1.5mM $MgCl_2$	2.0µl
dNTPs (10mM)	200µM each dNTP	0.4µl
2A PCR1 (10µM)	160nM	0.32µl
2A R2 (10µM)	160nM	0.32µl
Taq polymerase	1U	0.2µl
gDNA	20ng	1.0µl
Total		20µl

Data were available for rs1801274 on the YEAR patients genotyped on the Immunochip platform. A tag SNP for rs1801274 (rs12139150; 1000 Genomes GBR r^2=0.91) was genotyped on the Sequenom MassARRAY (n=810) platform, which was also available on the YEAR patients genotyped on the Immunochip platform (n=179).

5.2.6 *FCGR2C* copy number variation and association testing

CNV at the FCGR locus has been genotyped previously for *FCGR3A* and *FCGR3B* using a combined QSV and PRT method and tested for association with RA susceptibility [349]. Patients carrying at least one deletion of the CNVR containing the *FCGR3B* gene were found to have an increased odds of RA, which was not seen with the CNVR containing *FCGR3A*, although the latter occurred at a lower frequency. CNV of *FCGR2C* is not independent of CNV in *FCGR3A* and *FCGR3B* as it spans the SD that is thought to overlap with the boundaries of the two CNVRs. However, the effect of *FCGR2C* CNV has not been tested for its impact on RA susceptibility, to date. A case-control association study of *FCGR2C* CNV was carried out using a set of UK healthy controls and a subset of the RA cases used in the *FCGR3* study above. In addition, *FCGR2C* CNV was tested in the cohort of RA patients who were treated with TNF antagonists for its effect on response to these therapeutics.

5.2.6.1 TriPSV assay and quality control

The TriPSV assay was performed by the doctoral candidate on all available DNA samples from the patient and control cohorts described in 5.2.2, using the PCR and sequencing reaction conditions described in 5.2.1.3. QC was performed on all successfully base-called sequencing electropherograms based on the CNP score for *FCGR2A:FCGR2B*; these two genes are not subject to CNV thus the CNP score for these two nucleotides should remain stable (CNP~0.5). Deviation from a predetermined range (± 3xσ) may indicate assay failure; samples with CNP for *FCGR2A:FCGR2B* outside of this range were therefore deemed to have failed QC. CNP scores were used for both *FCGR2A:FCGR2C* and *FCGR2B:FCGR2C* to measure *FCGR2C* CNV as described in Section 5.2.1.4.

5.2.6.2 CNVtools and *FCGR2C* CNV disease susceptibility association testing

CNV testing is susceptible to confounding variation from sample handling differences and batch effects that may arise from this [354]. If not accounted for appropriately during statistical analysis the type I error rate may be inflated [374]. Estimation of discrete CN states from continuous data such as the CNP score for the TriPSV can lead to incorrect assignment of CN where there is uncertainty, i.e. where there is no clear separation between CN states, and the most likely CN state is assumed. Allowing for uncertainty in statistical analyses can account for this. CNVtools, run in the R environment (R version 2.15.0), can incorporate both known batch effects and uncertainty in CN calling from a continuous CN measure [378]. CNVtools fits a finite mixture model of Gaussian distributions to the data, with the number of components defined by the ability to discriminate between the CN states; each distribution thereby representing a CN state. For instance, if CNVtools is able to fit a model with 3 components to the TriPSV data, this implies it is able to discriminate the 3 CN states CN=2, CN<2 and CN>2. In addition the mean of each distribution can vary based on known batch effects, such as known sample handling differences between cohorts. It is unable to account for hidden or unknown batch

effects. Formal association testing in a case-control design study is carried out by fitting the distributions under the null and alternate hypothesis that CN depends on case-control status, and performing a likelihood-ratio (LR) test on the model log-likelihoods. Association between CN and case-control status can then be tested with a 2 degrees of freedom χ^2 test. Using this method, individuals are not assigned to discrete CN states, but a posterior probability for each CN assignment can be derived from the model. ORs for each comparison were estimated using the *pD* output from the model, which is the estimated probability of disease given the CN state.

The TriPSV data were rescaled to mean = 0 and standard deviation =1 in order to fit Gaussian distributions to the data. Known batch effect was accounted for by defining each of the two case cohorts and one control cohort as batches in the model. The number of components in the model was set at 3 representing the states of CN=2 (wild type), CN<2 (deletions) and CN>2 (duplications). This represented the most parsimonious model based on the observed Bayesian information criterion (BIC) values across a number of models with different number of components (Figure 5.10). The model variance was constrained to vary only within each combination of batch and CN state.

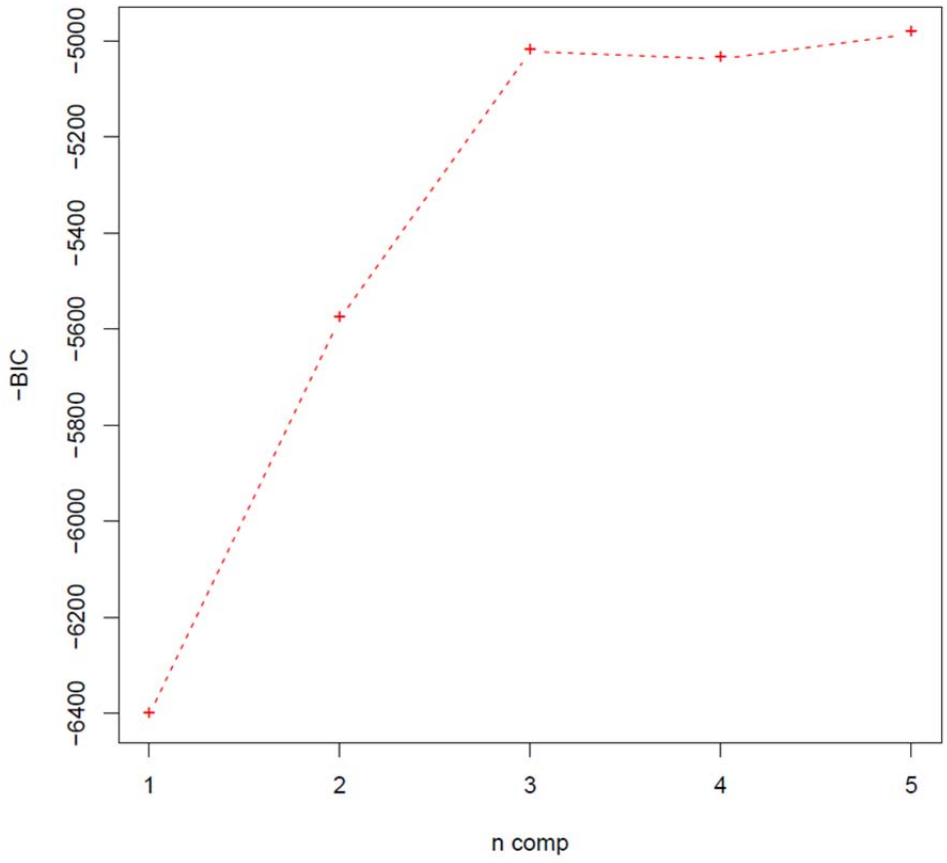


Figure 5.10. CNVtools model selection using the TriPSV *FCGR2A:FCGR2C* CNP score with varying number of components. Model variance was constrained to being proportional to CN and batch. Model selection was run over 6 iterations under the null hypothesis of no association between CN and case-control status. –BIC is plotted on the y-axis for each model with differing number of components (x-axis). The –BIC is slightly larger for the 5 component model compared to the 3 component model, however the added complexity in the model did not allow the EM algorithm to converge properly and CNVtools was unable to assign posterior probabilities appropriately.

Three tests of association were specified for the primary analysis:

1. Disease risk differs between *FCGR2C* CN states (2 d.f.)
2. *FCGR2C* deletions confer risk of disease compared to wild type and duplications combined (1 d.f.).
3. *FCGR2C* duplications confer risk of disease compared to wild type and deletions combined (1 d.f.)

5.2.6.3 *FCGR2C* copy number variation and response to tumour necrosis factor antagonist biologics

Patients from the BRAGGSS cohort were treated with TNF antagonist biologics, including the mAbs infliximab and adalimumab, and the TNF receptor fusion molecule etanercept. The effect of *FCGR2C* CNV was tested on the response to TNF antagonist therapy in this cohort using the ΔDAS28-CRP as the primary measure of response. Sub-analyses were performed using the ΔSJC28 and ΔlnCRP, as these are the most objective components of the DAS28-CRP. A secondary analysis was performed using the EULAR response categories as the outcome measure, comparing EULAR good and moderate responders vs. non-responders.

CNVtools is able to test whether a quantitative trait varies by CN. The ΔDAS28-CRP, ΔSJC28 and ΔlnCRP are measures over time and are therefore not independent of their baseline level [379]. Association testing with these outcome measures requires the baseline measurement as a covariate, which cannot be easily accommodated in CNVtools. Instead the primary analysis used linear regression to regress the ΔDAS28-CRP, adjusting for the baseline DAS28-CRP, on the posterior probability of a particular CN assignment (from CNVtools) in three separate analyses. Thus three regression models were run:

1. ΔDAS28-CRP regressed on the probability that *FCGR2C* CN<2
2. ΔDAS28-CRP regressed on the probability that *FCGR2C* CN=2
3. ΔDAS28-CRP regressed on the probability that *FCGR2C* CN>2

The sub-analyses using the ΔSJC28 and ΔlnCRP were analysed, adjusting for their respective baseline measurements, in similar models. The threshold for nominal association was set at $p<0.05/3$ ($p<0.016$) as each outcome measure was tested 3 times for association with *FCGR2C* CN.

The EULAR response categories were modelled as a binary variable and *FCGR2C* CN was tested for association in CNVtools as described for the disease susceptibility association (Section 5.2.6.2). However, the model variances were constrained to vary for each CN state only. Association testing was performed under three alternate hypotheses:

1. EULAR response category varies dependent on *FCGR2C* CNV, d.f.=2
2. EULAR response category is associated with *FCGR2C* CN<2 relative to CN=2 and CN>2 combined, d.f.=1
3. EULAR response category is associated with *FCGR2C* CN>2 relative to CN=2 and CN<2 combined, d.f.=1

Analysis of *FCGR3A* and *FCGR3B* duplications with TNF antagonist response was performed as for *FCGR2C* using the posterior probabilities from CNVtools (data published previously [349]). A threshold for nominal association was the same as for *FCGR2C* for consistency ($p<0.016$).

The stratified analysis by TNF antagonist drug class was performed using these same models for ΔDAS28-CRP, ΔSJC28 and ΔlnCRP regressed on CN posterior probability, adjusting for their respective baseline measurement. A multivariate regression model for each of these outcome measures was also adjusted for RFs (presence/absence) based on the hypotheses described in Section 5.1.3. A modified Poisson regression model was used to estimate the relative risk of EULAR response (Good/moderate vs. poor) in a drug class-stratified analysis, including a multivariate model adjusting for presence/absence of RFs.

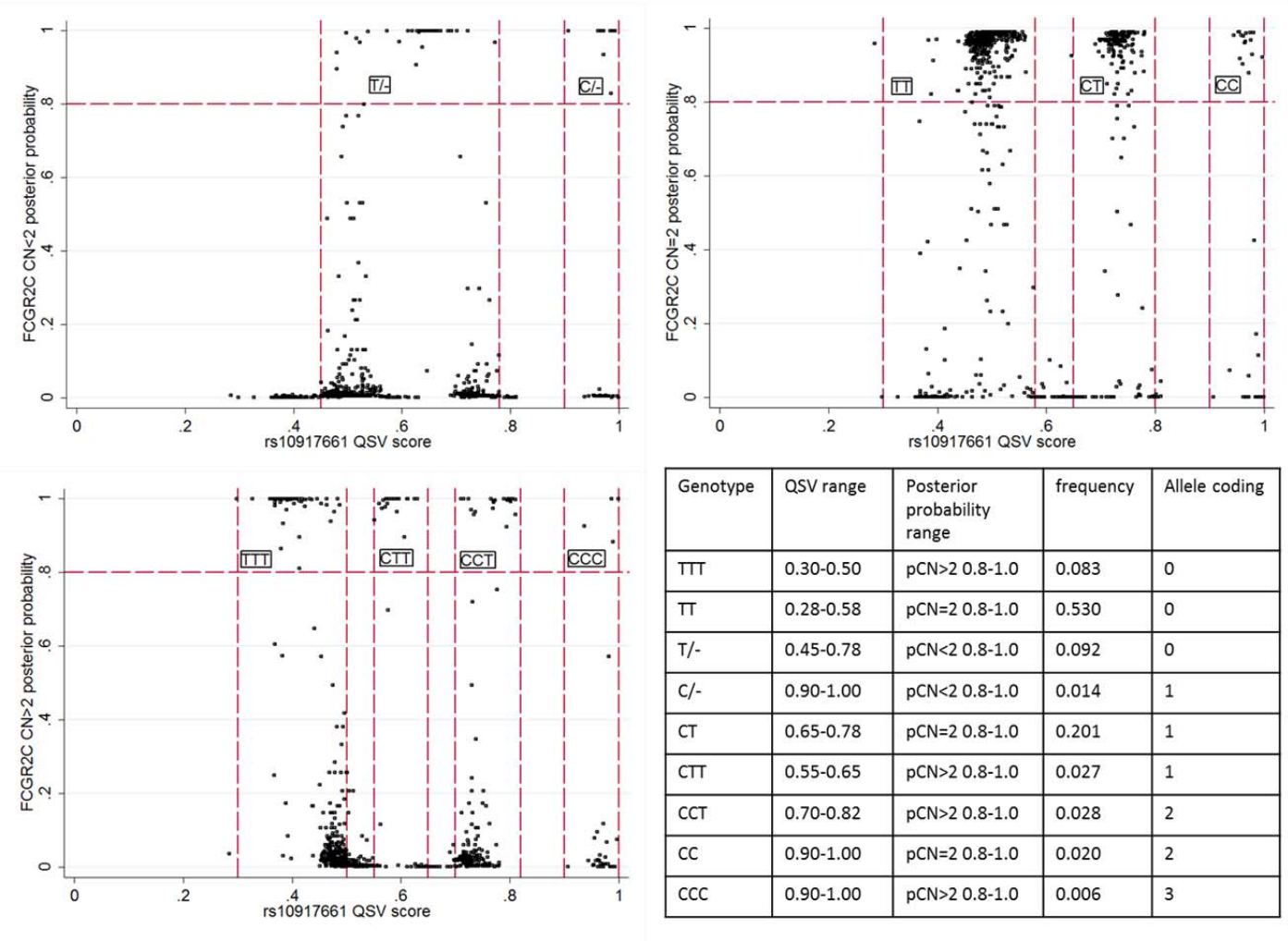
5.2.7 Tumour necrosis factor antagonist biologic response and copy number integrated functional *FCGR2C* allele association testing

Association testing with CNV is with gain and loss of genomic regions relative to a reference sequence, but does not account for allelic variation that is affected by this CNV. Allelic

variation within *FCGR3A, FCGR3B* and *FCGR2C* is affected by CNV; thus any statistical test of the effect of allelic variation must take into account CN changes. The integration of CNV and allelic variation is a complex challenge. The effect of the rs10917661 C (ORF) allele on response to TNF antagonist biologics was tested by integrating the QSV score for this SNP with the posterior probability for *FCGR2C* CNV from CNVtools to generate an integrated rs10917661/CNV genotype, i.e. the number of copies of a functional/ translated allele. This was achieved by plotting the posterior probability for each *FCGR2C* CN (i.e. 1, 2 or 3) against the QSV score for rs10917661. In order to maintain a high probability of correct CN assignment, a cut-off was placed on the CN posterior probability of 0.8 to call integrated genotypes. The scatter plots of rs10917661 QSV score versus *FCGR2C* CN posterior probability and thresholds used to call integrated genotypes are shown in Figure 5.11. The reference lines on each plot represent the boundaries of the genotype calls.

Association testing between CN integrated rs10917661 was performed for the primary outcome measure ΔDAS28-CRP and the sub-analysis of ΔlnCRP and ΔSJC28 using linear regression and the allele coding shown in Figure 5.11. This coding tests the hypothesis that increasing numbers of the ORF allele of *FCGR2C* (C allele) are associated with response to TNF antagonist therapy. The secondary outcome analysis used a modified Poisson regression model to estimate the relative risk of good EULAR response associated with the number of copies of the C allele of rs10917661.

Stratified analyses by drug class (mAbs and etanercept) were carried out using these same outcome measures adjusting for their relevant baseline measurement. A modified Poisson regression model was used to estimate the relative risk of achieving a EULAR response (Good/moderate vs. poor) in these stratified analyses. Based on the hypotheses defined in Section 5.1.3 RFs (presence/absence) were also entered into each linear regression and modified Poisson regression model as a covariate in these stratified analyses.



Genotype	QSV range	Posterior probability range	frequency	Allele coding
TTT	0.30-0.50	pCN>2 0.8-1.0	0.083	0
TT	0.28-0.58	pCN=2 0.8-1.0	0.530	0
T/-	0.45-0.78	pCN<2 0.8-1.0	0.092	0
C/-	0.90-1.00	pCN<2 0.8-1.0	0.014	1
CT	0.65-0.78	pCN=2 0.8-1.0	0.201	1
CTT	0.55-0.65	pCN>2 0.8-1.0	0.027	1
CCT	0.70-0.82	pCN>2 0.8-1.0	0.028	2
CC	0.90-1.00	pCN=2 0.8-1.0	0.020	2
CCC	0.90-1.00	pCN>2 0.8-1.0	0.006	3

Figure 5.11. Integrated CN and genotype calling of rs10917661 in *FCGR2C*. The posterior probability for each CN state from CNVtools using the TriPSV assay was used to integrated rs10917661 genotype calls and CNV. The three scatter plots show the three CN posterior probabilities against the rs10917661 QSV score from QSV Analyser. The red reference lines show the boundaries of the genotype groups and are labelled accordingly within each cluster. The above Table shows the ranges of the QSV and posterior probabilities used to call the CN integrated genotypes and their respective frequencies in the successfully genotyped BRAGGSS patient DNA samples.

5.2.8 Fc γ receptor single nucleotide polymorphism association testing with tumour necrosis factor antagonist response

The 3 non-synonymous SNPs in *FCGR2A* and *FCGR3A* were entered into the appropriate regression model for each outcome measure. The primary outcome measure was the ΔDAS28-CRP, with sub-analyses performed using the ΔSJC28 and ΔlnCRP. The secondary outcome measure was the EULAR response classification. A modified Poisson regression model with robust error variance was used to estimate the relative risk of good or moderate response vs. poor response. SNPs were entered into each regression equation assuming an additive effect on the outcome. A threshold for statistical significance was set at a $p \leq 0.01$.

Stratified analyses by drug class were carried out for each of these SNPs in regression models on ΔDAS28-CRP, ΔSJC28 and ΔlnCRP, adjusting for their respective baseline measurement. A modified Poisson regression model was used to estimate the relative risk of SNP allele on EULAR response (good/moderate vs. poor). Multivariate models for each outcome measure included RFs (presence/absence) based on the hypotheses described in Section 5.1.3.

5.3 FCGR Results

5.3.1 A comparison of *FCGR2C* copy number variation assays

Neiderer *et al* have published a PRT assay that is designed to measure the relative CN of *FCGR2C* relative to *FCGR2A* [348]. This assay was performed on the same 90 HapMap CEU DNA samples to check concordance with the TriPSV and the WGTP probe 8H4 $\log_2$ intensity (Figure 5.4). There was poor concordance between the TriPSV and PRT and between the WGTP $\log_2$ probe intensity and PRT results (Figure 5.12). The reference sequence for the PRT, reported to be copy number invariant, was located within the SD, whilst the TriPSV reference sequence was located outside of the SD (Figure 5.12 lower panel). Several HapMap individuals displayed large deviations in their PRT results (5.12 middle and right panels). These individuals carried CNV of *FCGR3A*, the CNVR of which may extend over the reference sequence for the PRT.

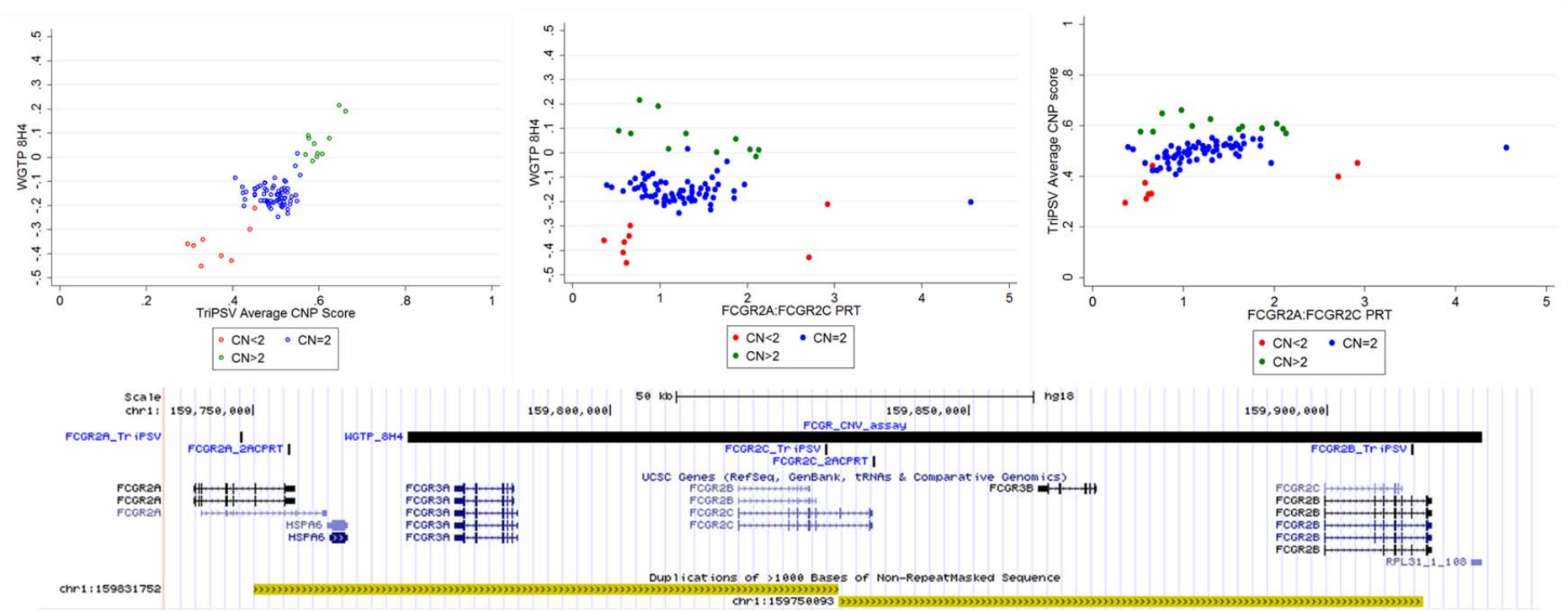


Figure 5.12. Comparison of the WGTP, TriPSV and PRT for measuring *FCGR2C* CN. The left panel shows the WGTP probe 8H4 plotted against the average TriPSV CNP score for the 90 CEU HapMap individuals. The middle panel shows the same WGTP array plotted against the PRT using the method described by Neiderer *et al* [347]. The right panel shows the same PRT plotted against the TriPSV average CNP score. The PRT shows a wide distribution of values that correspond poorly to the WGTP and TriPSV which are in good agreement. Three outliers are obvious in the PRT data, all of which carry deletions of the *FCGR3A* CNVR. The panel below the scatter plots shows the hg18 mapping positions of the assay primers and probes in the top 3 plots across the low affinity *FCGR* locus. Below the UCSC genes is shown the SD extent and boundaries, thought to form the limits of the CNVR breakpoints. The primers for the PRT reference sequence (FCGR2A_2ACPRT) lie within the SD and may therefore be affected by this CNVR in a subset of individuals. The TriPSV equivalent reference primers lie outside of the SD (FCGR2A_TriPSV). Colour coding in the scatter plots is by *FCGR2C* CN according to the key below each plot. CN states are derived from a comparison of the WGTP and *FCGR2C* TriPSV nucleotide peak height shown in Figure 5.4.

5.3.2 *FCGR2C* copy number variation and rheumatoid arthritis susceptibility

The TriPSV assay was performed on the GORA control and BRAGGSS RA case cohorts; PCR and Sanger sequencing reactions were performed on a total of 2289 samples. Sequencing reactions failed in 19 samples and could not be base-called by ABI software Sequence Analysis 5.2. CNP scores were calculated for *FCGR2A:FCGR2C, FCGR2B:FCGR2C* and *FCGR2A:FCGR2B* in QSV Analyser on the remaining 2270 sequences. Following QC of CNP scores 13 samples were deemed to have failed based on *FCGR2A:FCGR2B* CNP score<0.41 or >0.64 (±3xσ); this was due to poor starting DNA quality. A total of 2257 samples were therefore included in the analysis (genotyping success rate 98.6%). Histograms of each *FCGR2C* CNP score are shown in Figure 5.13, split by cohort, as this is often a major source of batch effect in CNV analyses. There appeared to be subtle differences between the GORA control and two BRAGGSS case cohorts in the *FCGR2A:FCGR2C* data (Figure 5.13A). These differences were more pronounced in the data for the *FCGR2B:FCGR2C* CNP data. In addition, there was a marked technical effect within the BRAGGSS 4 data (Figure 5.13B). Upon further investigation the majority of this effect appeared to occur across 3 sample plates (96 samples per plate), although it was not confined to these plates. CNVtools can account for known batch effects where each sample can be assigned to a specific batch. The exact source of the effect could not be defined within BRAGGSS 4, therefore the *FCGR2B:FCGR2C* CNP score could not be used for this analysis and the *FCGR2A:FCGR2C* score that was not subject to such a marked batch effect was used instead.

A single RA patient had a very low peak height measurement for the *FCGR2C* nucleotide at the TriPSV position. Inspection of the electropherogram revealed this individual did not have a *FCGR2C* peak at this position implying a complete deletion of this region in this individual. This individual was included in the analysis with a value assigned equal to the minimum value of the distribution of the remaining samples.

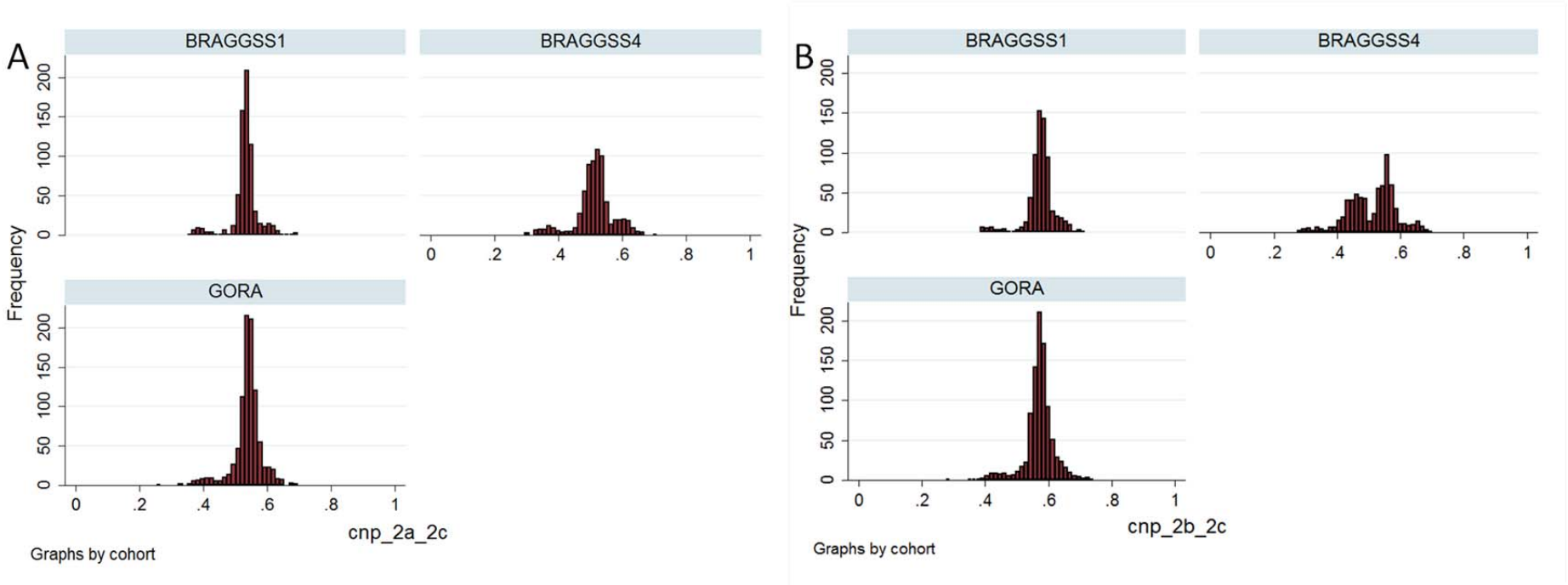


Figure 5.13. Histograms of TriPSV CNP scores plotted for each cohort. A – *FCGR2A:FCGR2C* CNP score for each cohort. BRAGGSS1 and BRAGGSS4 represent the two batches of the cases cohort. Differences in the distributions of each CN state between the three cohorts can be seen, particularly for the BRAGGSS4 data. B – *FCGR2B:FCGR2C* CNP score for each cohort. There is evidence of considerable batch effect within the BRAGGSS4 data which cannot be identified directly – this precludes using this score as batch effect can only be accounted for where it is known which batch samples are assigned to.

CNVtools was used to fit a mixture of Gaussian distributions to the data and perform a case-control association study using the three tests described in Section 5.2.6.2. The model parameters and test statistics for each test are described in Table 5.8 and represent the mean Z-score as it varies by batch (μ_i), the proportion of cases and controls in each CN state (α), model variance (σ^2) and the probability of disease given the CN state defined by the specified test (pD). The data distribution, model selection and posterior probability for each CN state by batch can be seen in Figure 5.14.

A three component model tested the hypothesis that case-control status was dependent on CN state (Table 5.8 – Test 1). No association between *FCGR2C* CN and RA susceptibility was found under this model. Two further tests were performed, comparing *FCGR2C* CN<2 (deletions – Test 2) against all other CN states and CN>2 against all other CN states (duplications – Test 3). No associations were found between either deletions or duplications and RA susceptibility.

Table 5.8. *FCGR2C* CNV disease susceptibility association parameters and test statistics

CN state	α	μ_1	μ_2	μ_3	σ^2	pD	OR	lnL H_A	χ^2	p
Test 1 – 2 d.f.										
CN<2	0.11	-2.31	-2.92	-1.96	1.06	0.56	0.91	-3989.4	0.40	0.82
CN=2	0.76	-0.34	-0.72	-0.17	0.13	0.59	ref			
CN>2	0.14	0.76	0.83	0.87	0.42	0.60	0.95			
Test 2 – 1 d.f.										
CN<2	0.11	-2.31	-2.91	-1.96	1.06	0.56	0.90	-3989.4	0.33	0.56
CN=2	0.75	-0.34	-0.72	-0.17	0.13	0.59	ref			
CN>2	0.14	0.75	0.82	0.83	0.43					
Test 3 – 1 d.f.										
CN<2	0.10	-2.30	-2.91	-2.00	1.05	0.58	ref	-3989.5	0.11	0.74
CN=2	0.76	-0.34	-0.72	-0.17	0.13					
CN>2	0.14	0.77	0.83	0.87	0.42	0.60	1.08			

Table 5.8. *FCGR2C* CNV disease susceptibility association CNVtools parameters and test statistics. Model parameters are described in the text and represent the proportion of CN states in the total data set, the mean Z-score for each batch, model variance and proportion of cases within each CN state given the defined test. For instance in test 1 there are three categories defined in the disease model, whereas for tests 2 and 3 there are two categories which represent the deletions vs. all others, and duplications vs. all others. The log likelihood under the alternate hypothesis ($\ln LH_A$) and χ^2 test statistics are presented.

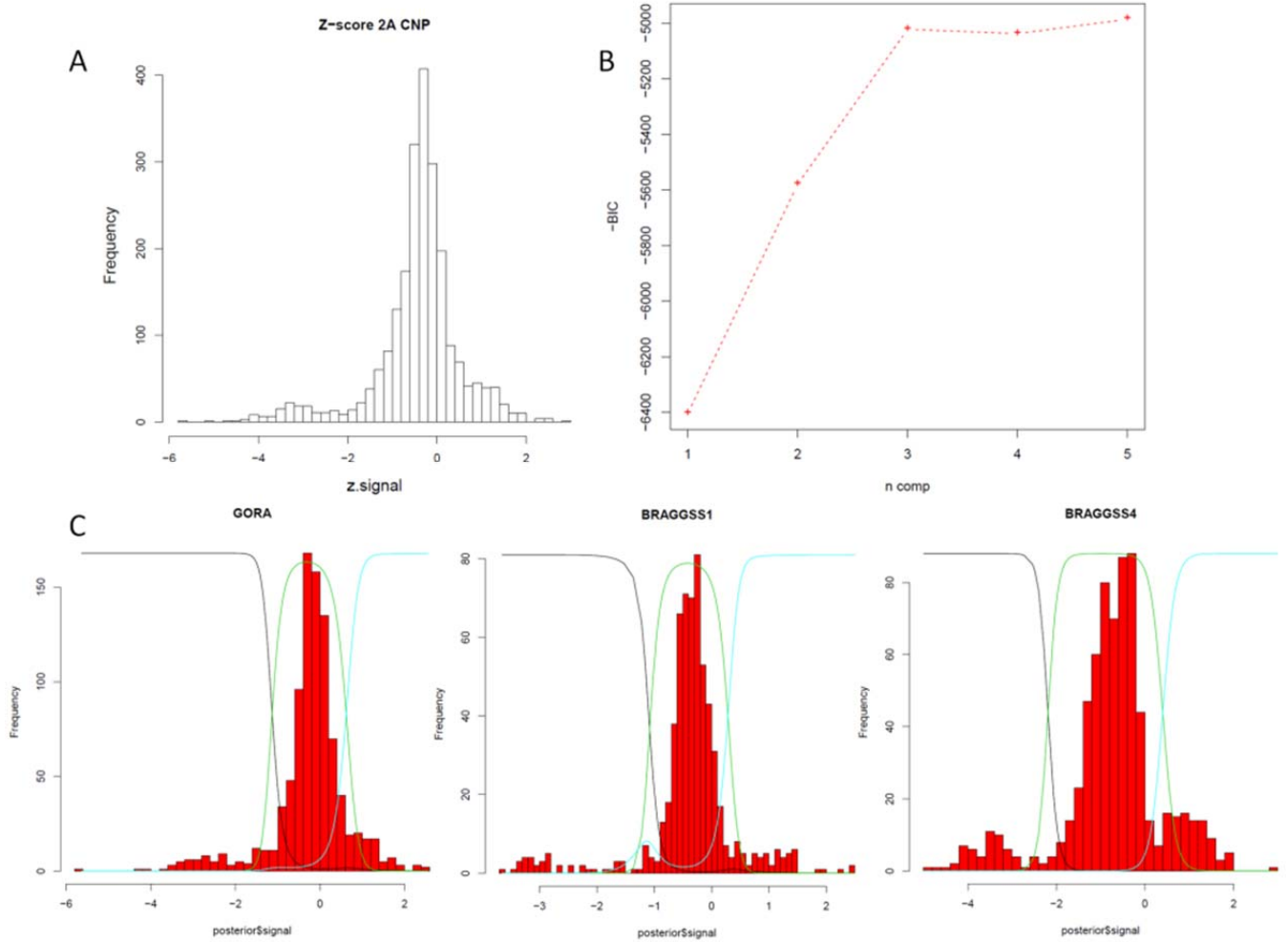


Figure 5.14. Rheumatoid arthritis susceptibility association testing with *FCGR2C* CNV and CNVtools. The *FCGR2A:FCGR2C* CNP standardised Z-scores were used to test the association between *FCGR2C* CN and RA susceptibility. A – Histogram of Z-scores showing the distribution and peaks representing CN states. The right hand shoulder represents CN>2 and is not easily discriminated from the CN=2 peak. B – CNVtools model selection function fits mixtures of Gaussian distributions to the Z-score data and selects the most likely number of components using the lowest Bayesian information criteria (BIC). The number of component (n comp) is representative of the number of CN states. A model with 3 components best fits the Z-score data. C – CNVtools accounts for between-batch effect in association testing, in this case the cohort from which the samples were taken were treated as different batches (see Figure 5.13). These CNV plots overlay the distribution of Z-scores with the fitted posterior probabilities from CNVtools under the null hypothesis of association. The black line is the posterior probability of CN<2, the green line CN=2 and the blue line CN>2 based on the interpretation of the CNP score.

5.3.3 Tumour necrosis factor antagonist biologic treated patients

CNVtools assigns a posterior probability to each sample for each CN state, the distributions of which are shown in Figure 5.14. The effect of *FCGR2C* CN on response to TNF antagonist biologic treatment was tested in the primary analysis by regressing ΔDAS28-CRP on these probabilities, adjusted for baseline DAS28-CRP. Matched clinical and CN data were available on up to 1277 patients with EULAR response category, 828 with matched baseline and 6month DAS28-CRP, 815 with matched baseline and 6month SJC28 and 462 with complete baseline and 6month CRP. Patients had evidence of active disease at baseline (mean DAS28-CRP =6.42), which is concordant with the NICE requirements for TNF antagonist biologic prescription of DAS28≥5.1 (Table 5.9). A high proportion of patients were seropositive for either ACPA (81.1%) or RF (78.6%), and many patients received concomitant DMARD therapy (74.6%). The study design for the subsequent analyses and the relevant hypothesis (described in Section 5.1.3) of genetic variation at the low-affinity *FCGR* locus is shown in Figure 5.15.

Table 5.9. BRAGGSS patient cohort summary and demographics			
Clinical/demographic variable	Summary statistic	n	p
Age at disease onset (years), mean (s.d.)	42.8 (13.3)	819	0.80
Disease duration (years), mean (s.d.)	13.5 (9.7)	820	0.005
Baseline DAS28-CRP, mean (s.d.)	6.42 (1.13)	828	8.5×10^{-25}
Baseline SJC28, median (IQR)	11 (28)	821	0.66
Baseline CRP, median (range)	27 (1-264)	546	0.009
ΔDAS28-CRP, mean (s.d.)	2.45 (1.63)	828	-
ΔSJC28, median (IQR)	7 (50)	815	-
ΔlnCRP, median (range)	14 (-140-212)	462	-
Gender, % female	74.2%	828	0.007
Seropositive, %	92.6%	767	0.008
ACPA, %	81.1%	645	0.061
RF, %	78.6%	827	0.034
Concurrent DMARD, %	74.6%	827	1.9×10^{-4}
Etanercept, n %	320, 38.7%	828	-
Infliximab, n %	354, 18.6%	828	-
Adalimumab, n %	154, 42.8%	828	-

Table 5.9. BRAGGSS patient cohort summary and demographics. Summary statistics are shown for the BRAGGSS patients with matched *FCGR2C* CN data. Data are presented as mean and standard deviation (s.d.) for normally distributed variables, and median and range or interquartile range (IQR) for non-normally distributed variables. Categorical variables are presented as percentages. P-values are from linear regression of ΔDAS28-CRP on these variables, adjusting for baseline DAS28-CRP.

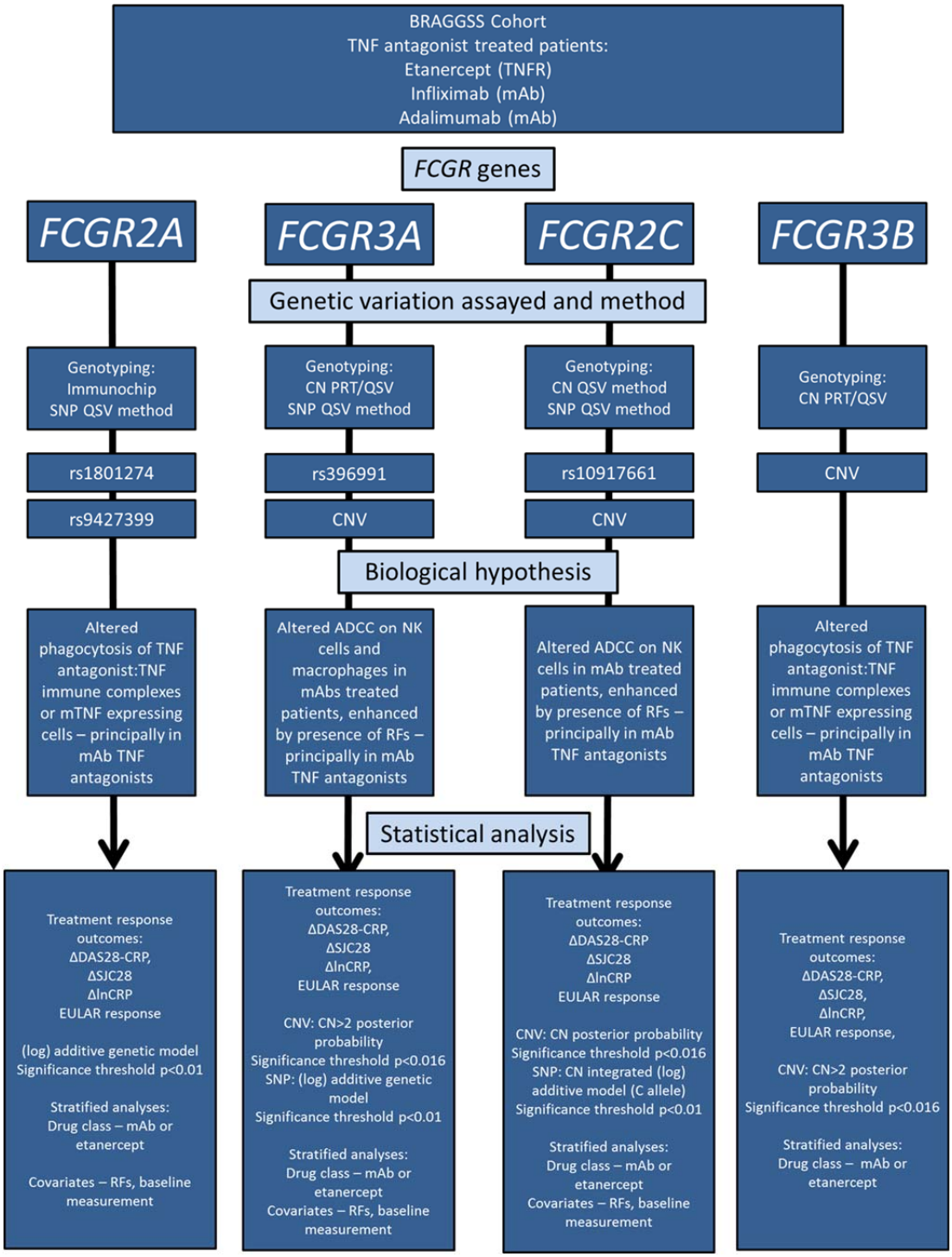


Figure 5.15. Study design of *FCGR* genetic variation association testing with response to TNF antagonist therapy in the BRAGGSS cohort. A combination of DNA sample and clinical data availability dictate the sample size for each analysis – these are shown in the relevant results tables in the following sections. A combination of methods were used to genotype SNPs and CNV at this locus. High-density SNP chip (Immunochip) data was available for two SNPs outside of the SD (rs1801274 and rs9427399) the remaining samples were genotyped by Mrs Lubna Haroon-Rashid and genotypes called using QSV Analyser by the doctoral candidate. CN posterior probability data generated by a combination of PRT and QSV assays were available for *FCGR3A* and *FCGR3B* from a previously published study by Dr Jim Robinson. *FCGR3A* SNP genotyping was performed by Mr Stephen Martin and Dr Jim Robinson and genotypes were called using the QSV method by the doctoral candidate. *FCGR2C* CN and SNP data were generated by the doctoral candidate using the QSV method developed and described in this chapter.

5.3.4 *FCGR2C* copy number variation and tumour necrosis factor antagonist biologic response

The effect of CNV on treatment response was tested for each CN state. A nominally statistically significant association was observed between *FCGR2C* CN>2 and an increased ΔDAS28-CRP (β 0.45 95% CI [0.09-0.8], p=0.015) relative to the combined CN<2 and CN=2 states (Table 5.10). The sub-analysis of ΔSJC28 and ΔlnCRP found this effect was due to a higher ΔlnCRP in the *FCGR2C* duplication group (β 0.33 95% CI [0.08-0.58], p=0.01).

Table 5.10. *FCGR2C* CNV linear regression models of biologic treatment response						
CN state	freq.	Outcome	β	95% CI	p	n
CN<2	0.08	ΔDAS28*	-0.14	-0.55-0.27	0.492	828
CN=2	0.81		-0.22	-0.52-0.07	0.137	
CN>2	0.11		**0.45**	**0.09-0.80**	**0.015**	
CN<2	0.08	ΔSJC28*	-1.06	-2.36-0.24	0.111	815
CN=2	0.81		0.04	-0.89-0.98	0.930	
CN>2	0.11		0.76	-0.39-1.90	0.196	
CN<2	0.09	ΔlnCRP*	-0.07	-0.36-0.22	0.618	462
CN=2	0.77		-0.19	-0.40-0.02	0.076	
CN>2	0.13		**0.33**	**0.08-0.58**	**0.010**	

Table 5.10. *FCGR2C* CNV linear regression models of anti-TNF biologic treatment response. Effect estimates from the linear regression models of primary treatment response outcome measures ΔDAS28-CRP and secondary ΔSJC28 and ΔlnCRP on each CN test are presented. 95% confidence intervals (95% CI) and p-values are shown for each test. Nominally statistically significant results with adjusted p<0.016 (α/3) are highlighted in bold. *analyses were adjusted for the baseline measurement, CN<2 – *FCGR2C* deletions, CN=2 – *FCGR2C* wild type, CN>2 – F*CGR2C* duplications.

The secondary analysis tested the hypothesis that EULAR response category was dependent on *FCGR2C* CNV. CNVtools was used to model the CNV data described in the methods under 3 alternate hypotheses. EULAR response was found to vary dependent on *FCGR2C* CN state in an analysis of good and moderate responders vs. poor-responders (Table 5.11). A higher relative risk (RR) of EULAR response was associated with *FCGR2C* duplications relative to combined wild type and deletions (RR 1.13, p=0.007). No association was seen with EULAR response for the *FCGR2C* deletions relative to wild type and duplications (Test 2). This association with increased RR of EULAR response is concordant with the association with higher ΔDAS28-CRP and ΔlnCRP (Table 5.10).

Table 5.11. *FCGR2C* EULAR Good/moderate vs. non response CNV association parameters and test statistics										
CN state	α	μ_1	μ_2	σ^2	pD	RR	lnL H1	χ^2	p	n
Test 1 – 2 d.f.										
CN<2	0.08	-2.68	-3.19	0.59	0.70	0.89	**-2081.6**	**8.74**	**0.013**	**1277**
CN=2	0.78	-0.34	-0.73	0.15	0.79	ref.				
CN>2	0.13	0.86	0.88	0.39	0.86	1.09				
Test 2 – 1 d.f.										
CN<2	0.08	-2.68	-3.19	0.58	0.72	0.90	-2084.3	3.22	0.072	1277
CN=2	0.79	-0.34	-0.72	0.15	0.80	ref.				
CN>2	0.13	0.94	0.91	0.35						
Test 3 – 1 d.f.										
CN<2	0.08	-2.67	-3.18	0.60	0.78	ref.	**-2082.3**	**7.29**	**0.007**	**1277**
CN=2	0.77	-0.35	-0.73	0.14						
CN>2	0.14	0.73	0.82	0.45	0.87	1.13				

Table 5.11 *FCGR2C* EULAR Good/moderate vs. poor response CNV association parameters and test statistics. CNVtools parameters and test statistics are shown for each CN association test defined in Section 5.2.6.3. CN – copy number, CN<2 – deletions, CN=2 – wild type, CN>2 – duplications. A – CN state frequency, μ – batch mean TriPSV Z-score, where batch is BRAGGSS cohort, σ^2 – within CN state variance, *pD* – proportion of cases given the CN state, RR – risk ratio, lnL H1 – log likelihood under the alternate hypothesis of CN state association with EULAR Good/moderate response, χ^2 – test statistic from the likelihood ratio test, p – p-value from the likelihood ratio test, d.f. – degrees of freedom in each test.

Data on prescribed biologic were available (etanercept n=320, infliximab and adalimumab n=508). Stratified analyses found the direction of effect in the ΔDAS28-CRP analysis, i.e. linear regression β estimates, were consistent across both classes (Table 5.12; etanercept β 0.36, p=0.23; mAb β 0.50, p=0.031). Adjustment for RFs in these regression models did not alter these findings. A modified Poisson regression was used to estimate the RR of good/moderate EULAR response in this stratified analysis. The observation of a statistically significant association was maintained in the etanercept treated patients (Table 5.12; RR 1.21 95% CI [1.13-1.31] $p=1.23x10^{-7}$), but not the mAb treated patients (RR 1.10 95% CI [0.99-1.22], p=0.08). These findings were not altered by entering RFs into the modified Poisson regression models as a covariate.

A statistically significant association was observed between EULAR response and *FCGR2C* duplications in the analysis stratified by drug class. Curiously this was not observed in the mAb treated subset of patients. An exploratory *ad hoc* analysis of this effect stratified by

case cohort (BRAGGSS1 or BRAGGSS4) within each TNF antagonist (etanercept, infliximab or adalimumab) indicated a consistent result across cohorts within the etanercept (BRAGGSS1 RR 1.21 p=0.002; BRAGGSS4 RR 1.22 $p=9.0x10^{-7}$) and infliximab (BRAGGSS1 RR 1.08 p=0.60; BRAGGSS4 RR 1.11 p=0.28) treated patients. Patients treated with adalimumab differed between the BRAGGSS1 and BRAGGSS4 cohorts (BRAGGSS1 RR 1.22 p=0.0005 n=61; BRAGGSS4 RR 1.06 p=0.46 n=277). The proportions of female patients, mean age at disease onset and symptom duration at baseline were similar across both cohorts. The concomitant DMARDs administered differed between cohorts (BRAGGSS1 68.3% vs. BRAGGSS4 74.0%), as did the proportion of RF positive patients (BRAGGSS1 90.2% vs. BRAGGSS4 54.9%). Whilst entering RFs as a covariate into the modified Poisson regression model did not alter the effects of *FCGR2C* duplications, the effect of RFs were different between the etanercept and mAb TNF antagonist patients (etanercept RR 0.99 p=0.84; mAb RR 0.91 p=0.013). This drug class-specific effect of RFs on treatment response was strongest in the BRAGGSS1 patients (RR 0.85 p=0.015 n=354) than the BRAGGSS4 patients (RR 0.93 p=0.13 n=475). This is likely due to the higher frequency of RF positive patients in the BRAGGSS1 case cohort.

CNV posterior probability data for *FCGR3A* and *FCGR3B* CN were available on a subset of the BRAGGSS1 patients (n=567) from a previous association study with RA susceptibility [349]. Duplications of *FCGR3A* and *FCGR3B* were tested for association with ΔDAS28-CRP, ΔSJC28 and ΔlnCRP in separate linear regression models in the subset of patients for which *FCGR2C* CN data were also available. There were no statistically significant (p<0.016) associations between TNF antagonist treatment response and either *FCGR3A* or *FCGR3B* duplications with either the primary or secondary outcome measures (Table 5.12).

Table 5.12. Refinement of *FCGR2C* duplications association with anti-TNF biologics response																
	FCGR3A CN>2				*FCGR3B* CN>2				Etanercept				mAbs (infliximab + adalimumab)			
Outcome	β	95% CI	p-value	n	β	95% CI	p-value	n	β	95% CI	p-value	n	β	95% CI	p-value	n
ΔDAS28	-0.28	-0.76-0.19	0.24	567	-0.20	-0.61-0.21	0.34	567	0.36	-0.23-0.95	0.23	320	0.50	0.05-0.96	0.031	508
ΔSJC28	-1.48	-2.95--0.02	0.05	591	-0.83	-2.08-0.42	0.19	591	0.98	-0.89-2.85	0.30	317	0.71	-0.75-2.17	0.34	498
ΔlnCRP	0.13	-0.28-0.56	0.52	241	0.05	-0.31-0.42	0.77	241	0.41	-0.04-0.85	0.07	166	0.30	0.00-0.61	0.052	296
Outcome	RR	95% CI	p	n	RR	95% CI	p	n	RR	95% CI	p	n	RR	95% CI	p	n
EULAR response	0.96	0.82-1.13	0.63	613	1.09	0.98-1.22	0.12	613	1.21	1.13-1.31	$1.2x10^{-7}$	447	1.10	0.99-1.22	0.08	830

Table 5.12. Refinement of *FCGR2C* duplication association and stratified analysis by TNF antagonist drug class. Treatment response outcomes were regressed on *FCGR3A* and *FCGR3B* duplications adjusting for the relevant baseline measure in each model, i.e. DAS28-CRP, SJC28 or lnCRP. A stratified analysis of *FCGR2C* duplication posterior probability of TNF antagonist response by drug class shows a consistent effect estimate across both drug classes, with a borderline significant effect in the mAbs treated patients. There is a strong statistically significant association with EULAR Good/moderate response and *FCGR2C* duplications in the etanercept treated patients.

5.3.5 Copy number integrated *FCGR2C* rs10917661 and tumour necrosis factor antagonist biologics response

A total of 960 BRAGGSS patient samples were genotyped for rs10917661 by the QSV method described in 5.2.3. Of these, 70 sequencing reactions failed to generate high quality sequencing electropherograms that could be base-called by ABI Sequence Analysis 5.2. Duplicate samples of 24 patients were included to check QSV score concordance for this assay, 3 of which failed to generate a base-callable sequence (included in the 70 described above). QSV score concordance was high (Spearman's ρ=0.96) for the remaining 21 samples that generated sequence electropherograms (Figure 5.16). Therefore 890 non-duplicate QSV scores were available to call rs10917661 genotypes. Matching *FCGR2C* CN posterior probabilities from CNVtools were available on 847 of these samples. CN integrated genotypes were generated for this SNP, as described in 5.2.7, with 787 genotypes called (call rate 92.9%). The frequency of each genotype is shown in Figure 5.8 for these 787 patients. Matched ΔDAS28-CRP and rs10917661 genotype data were available on 423 samples, ΔSJC28 on 420 and ΔlnCRP on 295. Matched EULAR response and CN integrated genotypes were available on up to 780 patients.

The effect of the number of copies of the C allele of rs10917661 was tested on the primary outcome measure ΔDAS28-CRP, and sub-analysis of ΔSJC28 and ΔlnCRP, using linear regression. No association was observed between increasing numbers of copies of rs10917661 C allele and ΔDAS28-CRP, ΔSJC28 or ΔlnCRP (Table 5.13) in all TNF antagonist treated patients. The secondary analysis, which tested the effect of rs10917661 C allele copies on EULAR response, found no association between good/moderate vs. poor response.

The hypotheses described in Section 5.1.3 state a proposed effect of *FCGR2C* genetic variants in the mAb treated patients only. A stratified analysis by drug class did not find

any association between rs10917661 C allele copies and mAb response ΔDAS28-CRP or ΔSJC28 (Table 5.13; ΔDAS28-CRP p=0.44, ΔSJC28 p=0.85). However, a nominal association was observed using the ΔlnCRP as the outcome measure (Table 5.13; ΔlnCRP β 0.25 95% CI [0.05-0.44], p=0.014, n=197). The direction of this effect is concordant with that observed with *FCGR2C* duplications (Table 5.10). Adjusting the linear regression model of ΔlnCRP on rs10917661 C allele copies for RFs did not alter this effect (Table 5.13). No effect was observed on EULAR response in this stratified analysis by drug class.

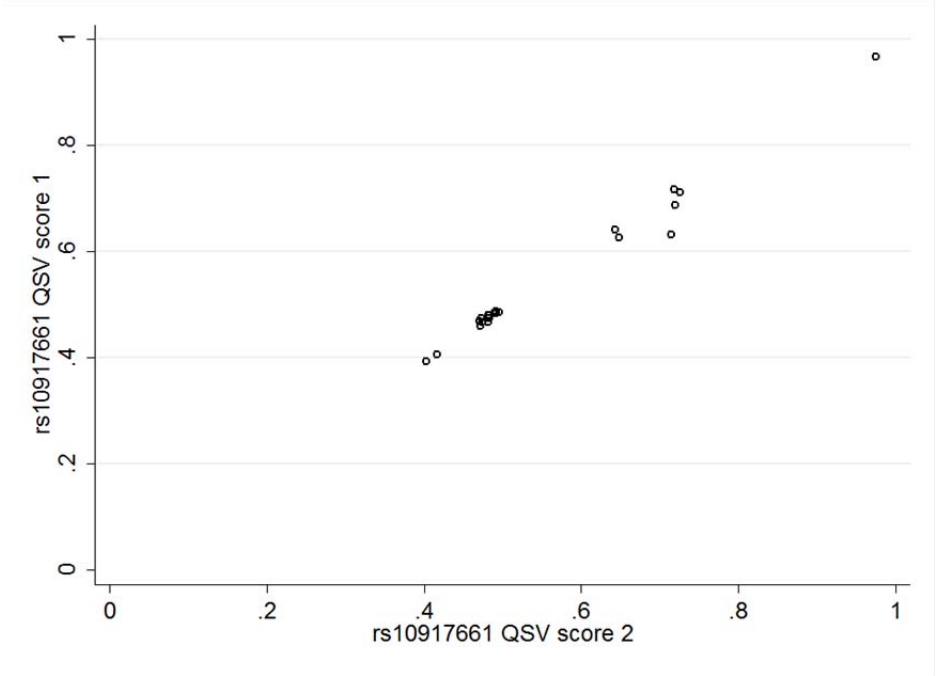


Figure 5.16. Concordance of rs10917661 QSV scores. Two independent replicates of the rs10917661 QSV genotyping assay were performed on 24 BRAGGSS samples, 21 of which generated base called electropherograms. The concordance between scores is high (Spearman's ρ=0.96) which suggests the assay is robust.

Table 5.13. Association test of CN integrated rs10917661 with TNF antagonist treatment response, and stratified analysis by drug class														
			All TNF antagonists				Etanercept				mAbs (infliximab + adalimumab)			
Outcome	MAF	Allele	β	95% CI	p-value	n	β	95% CI	p-value	n	β	95% CI	p-value	n
ΔDAS28-CRP	0.179	C	-0.02	-0.26-0.23	0.90	423	-0.26	-0.64-0.12	0.18	149	0.13	-0.19-0.45	0.44	274
ΔSJC28			0.00	-0.80-0.81	0.98	420	0.26	-0.64-1.16	0.57	243	-0.07	-1.13-0.99	0.89	272
ΔCRP			0.09	-0.07-0.25	0.27	295	-0.23	-0.52-0.06	0.12	98	0.25	0.05-0.44	0.014*	197
Outcome	MAF	Allele	RR	95% CI	p-value	n	RR	95% CI	p-value	n	RR	95% CI	p-value	n
EULAR Good/moderate vs. poor	0.179	C	0.99	0.94-1.06	0.86	780	1.04	0.97-1.11	0.27	249	0.96	0.88-1.05	0.38	531

Table 5.13. Association test of CN integrated rs10917661 with TNF antagonist treatment response. Linear regression estimates (β), 95% CI's and p-values are presented for the primary outcome measure ΔDAS28-CRP, adjusted for baseline DAS28-CRP. Sub-analyses for ΔSJC28 and ΔlnCRP were adjusted for their respective baseline measure. Modified Poisson regression was used to estimate the RR of good/moderate EULAR response vs. poor response. Stratified analyses by TNF antagonist drug class are also presented. * adjusted for RFs (presence/absence) β 0.26 95% CI [0.06-0.45], p=0.012. The minor C allele represents the ORF of *FCGR2C*.

5.3.6 *FCGR3A* rs396991 and TNF antagonist biologic therapeutic response

Allotypes of FcγRIIIa affect the binding affinity of different IgG isotypes. The Fc-containing TNF antagonist biologics are IgG1 based, and thus their therapeutic effect may be altered by these differences in binding affinity. The SNP rs396991 was tested for association with response to TNF antagonist biologic treatment in the BRAGGSS cohort described in Section 5.2.2. Matched rs396991 genotyping and clinical data were available on 1604 patients. There was no observed association between the rs396991 minor allele, assuming an additive model, and ΔDAS28-CRP in the primary analysis (p=0.61). The sub-analysis of ΔSJC28 found a trend towards a borderline association with the T allele (F allotype) of this SNP and worse response (p=0.06). There was no association observed with ΔlnCRP (p=0.62), though there was a large proportion of patients with missing data for ΔlnCRP (~55%). In the secondary analysis the relative risk of EULAR response was estimated using a modified Poisson regression model (see Chapter 2.3.3). No association was observed between good and moderate response relative to poor response (Table 5.14).

The hypothesis that *FCGR3A* allelic variation would alter response to TNF antagonists is based on binding of IgG1 mAb therapies to FcγRIIIa, with a possible enhanced cross-linking by RFs (Section 5.1.3). Therefore a stratified analysis on drug class was conducted using the same outcome measures. No drug class specific association was observed in this stratified analysis (Table 5.14). These findings were not altered when RFs were entered into each regression model.

Table 5.14. Association test results of *FCGR3A* rs396991 with TNF antagonist treatment response														
			All patients				Etanercept				mAbs ((infliximab + adalimumab)			
Outcome	MAF	Allele	β	95% CI	p-value	n	β	95% CI	p-value	n	β	95% CI	p-value	n
ΔDAS28	0.336	T	0.03	-0.08-0.14	0.61	1604	0.00	-0.17-0.18	0.97	605	0.04	-0.10-0.19	0.54	999
ΔSJC28			-0.60	-1.23-0.03	0.06	1562	-0.75	-1.63-0.13	0.10	597	-0.44	-1.29-0.41	0.31	965
ΔlnCRP			0.03	-0.07-0.13	0.62	725	0.00	-0.17-0.17	0.98	265	0.04	-0.09-0.16	0.55	460
Outcome	MAF	Allele	RR	95% CI	p-value	n	RR	95% CI	p-value	n	RR	95% CI	p-value	n
EULAR Good/moderate vs. poor	0.336	T	1.00	0.97-1.04	0.88	2063	1.02	0.97-1.07	0.50	743	0.99	0.95-1.04	0.77	1320

Table 5.14 Association test results of *FCGR3A* rs396991 with TNF antagonist treatment response. The primary outcome measure linear regression model of ΔDAS28-CRP on rs396991 minor allele was adjusted for baseline DAS28-CRP. The sub-analyses of ΔSJC28 and ΔlnCRP were adjusted for their baseline measure. The RR of rs396991 minor allele on EULAR Good/moderate vs. poor response was estimated using a univariate modified Poisson regression model. The minor T allele of rs396991 represents the F allotype of FcγRIIIa.

5.3.7 *FCGR2A* non-synonymous single nucleotide polymorphism association testing with tumour necrosis factor antagonist therapeutic response

5.3.7.1 rs9427399 association testing with tumour necrosis factor antagonist treatment response

The proxy SNP for the Q27W (rs9427399, r^2=0.91 1K Genomes GBR population) non-synonymous change in *FCGR2A* was tested for its effect on response to TNF antagonist biologic therapy in the BRAGGSS cohort. Genotyping was performed by the QSV method on 403 samples, of which 110 failed to generate a sequence electropherogram, due to poor DNA quality. Replicate samples were not available prior to submission of this doctoral thesis. Combined with the data generated by the Immunochip genotyping, a total of 1873 genotypes were available for this SNP. Matched ΔDAS28-CRP data were available on 1446 patients, whilst EULAR response category data were available on 1873.

The minor allele (C) of rs9427399 (27W allele) was tested for its effect on treatment response against the primary outcome measure ΔDAS28-CRP and the sub-analysis outcome measures ΔSJC28 and ΔlnCRP. No association was observed between the minor allele of this SNP and ΔDAS28-CRP, ΔSJC28 or ΔlnCRP (Table 5.15). In the secondary analysis there was no effect observed on the relative risk of achieving a good or moderate EULAR response (Table 5.15). The stratified analysis of this SNP by TNF antagonist drug class did not find any association with treatment response using any outcome measure (Table 5.15).

5.3.7.2 rs1801274 (H131R) association testing with response to tumour necrosis factor antagonist treatment

Genotyping data were available on 1745 BRAGGSS patient samples from the Immunochip genotyping and samples genotyped by the QSV assay described in the methods (Section 5.2.5.2) for rs1801274 (H131R). Matched primary outcome and genotyping data were

available on 1032 patients, and 1445 with EULAR response category data. The minor allele of rs1801274 (A) was tested for association with ΔDAS28-CRP in these patients. No effect was observed for the minor allele on treatment response in this primary outcome analysis or either sub-analyses with ΔSJC28 or ΔlnCRP (Table 5.15). The secondary analysis tested the effect of the minor allele on good/moderate vs. poor EULAR response; no association was observed between the minor allele of this SNP and response to TNF antagonist biologic therapy with this outcome measure (Table 5.15). The stratified analysis of rs1801274 found a statistically significant association between the A allele (131H) lower ΔDAS28-CRP (p=0.005) in the etanercept treated patient and a borderline higher ΔDAS28-CRP (p=0.038) in the mAb treated patients (Table 5.15). The differences in direction of effect between these two classes of TNF antagonist biologic are consistently mirrored in the non-significant linear regression effect estimates on the ΔlnCRP, but not the ΔSJC28 (Table 5.15). A borderline association was also observed with a lower RR of EULAR response in the etanercept treated patient group (RR 0.93 95% CI [0.88-0.99] p=0.034). The effect in the etanercept treated patients was consistent with an additive effect on ΔDAS28-CRP (AG β -0.36, AA β -0.60). Given the multiple testing burden and lower statistical power in these stratified analyses these findings should be interpreted with caution.

The observed association between the rs1801274-A allele and lower ΔDAS28-CRP is concordant with the hypothesis that this 'high affinity' allele would produce a stronger pro-inflammatory signal in the presence of autoantibodies. Genotyping data were available for a tag SNP of rs1801274 (rs12131950) on the DMARD treated YEAR patients described in Chapter 2.1.1. Matched ΔDAS28-CRP and genotyping data were available for 545 patients, 383 of which were positive for RF and/or ACPA. Linear regression of ΔDAS28-CRP on rs12139150-G allele (tag for rs1801274-A), adjusted for baseline DAS28-CRP, found a borderline nominal association with lower ΔDAS28-CRP regardless of autoantibody status (β -0.18 95% CI [-0.36-0.00], p=0.056). A stratified analysis in the autoantibody positive

patients found a nominally statistically significant association with lower ΔDAS28-CRP (β -0.27 95% CI [-0.48--0.05], p=0.014). These data therefore provide statistical evidence for an influence of *FCGR2A* non-synonymous genetic variation on RA disease pathology in individuals seropositive for autoantibodies.

Table 5.15. Association test results of *FCGR2A* rs9427399 and rs1801274 with anti-TNF treatment response												
	rs9427399						rs1801274					
Outcome	Allele	MAF	β	95% CI	p-value	n	Allele	MAF	β	95% CI	p-value	n
ΔDAS28	C	0.131	0.01	-0.15-0.17	0.89	1446	A	0.479	0.00	-0.12-0.13	0.96	1032
ΔSJC28			-0.10	-0.84-0.63	0.78	1815			-0.26	-0.85-0.34	0.39	1407
ΔlnCRP			0.07	-0.07-0.20	0.32	686			0.02	-0.09-0.13	0.68	505
Outcome	Allele	MAF	RR	95% CI	p	n	Allele	MAF	RR	95% CI	p	n
EULAR Good/mod. vs. poor	C	0.131	0.97	0.95-1.05	0.91	1873	A	0.479	0.99	0.96-1.03	0.75	1445
Stratified analysis by TNF antagonist drug class												
Etanercept												
rs9427399							rs1801274					
Outcome	Allele	MAF	β	95% CI	p-value	n	Allele	MAF	β	95% CI	p-value	n
ΔDAS28	C	0.131	-0.19	-0.45-0.05	0.13	548	A	0.479	-0.30	-0.51--0.09	0.005	362
ΔSJC28			-0.45	-1.70-0.80	0.48	541			-0.27	-1.40-0.85	0.63	358
ΔlnCRP			-0.09	-0.31-0.14	0.45	253			-0.14	-0.33-0.06	0.16	171
Outcome	Allele	MAF	RR	95% CI	p	n	Allele	MAF	RR	95% CI	p	n
EULAR Good/mod. vs. poor	C	0.131	0.97	0.88-1.06	0.45	666	A	0.479	0.93	0.88-0.99	0.034	479
mAbs (infliximab + adalimumab)												
rs9427399							rs1801274					
Outcome	Allele	MAF	β	95% CI	p-value	n	Allele	MAF	β	95% CI	p-value	n
ΔDAS28	C	0.131	0.13	-0.07-0.34	0.20	898	A	0.479	0.17	0.01-0.32	0.038	670
ΔSJC28			0.04	-1.17-1.25	0.95	869			-0.37	-1.39-0.64	0.47	648
ΔlnCRP			0.16	-0.01-0.32	0.07	433			0.10	-0.03-0.23	0.13	334
Outcome	Allele	MAF	RR	95% CI	p	n	Allele	MAF	RR	95% CI	p	n
EULAR Good/mod. vs. poor	C	0.131	1.01	0.95-1.08	0.66	1207	A	0.479	1.03	0.98-1.08	0.28	966

Table 5.15. Association testing results of rs9427399 and rs1801274 with TNF antagonist treatment response. Regression effect size estimates, 95% CI and p-values are shown for the primary and sub-analyses of ΔDAS28-CRP, ΔSJC28 and ΔlnCRP as well as the secondary outcome measure EULAR response. Linear regression models of ΔDAS28-CRP, ΔSJC28 and ΔlnCRP were adjusted for their respective baseline measure. MAF – minor allele frequency, RR – risk ratio, mAbs – monoclonal antibodies

5.4 Discussion

5.4.1 Complex genomic arrangements, segmental duplications and technical challenges for genetic analyses

Genome-wide genotyping chips have shown limited ability to interrogate the low affinity *FCGR* locus in human genetic association studies. As described in the introduction of this chapter, (Section 5.1.4.1) the high sequence homology between the two paralogs of the SD confounded gene-specific genotyping using probe-hybridisation platforms. In addition, individuals may be incorrectly assigned to genotypes where they display CNV, particularly for heterozygous individuals where imbalances are likely to generate additional genotype clusters intermediate to the expected heterozygote and homozygote clusters. To this end several approaches to measuring both CN and allelic variation at the *FCGR* locus have used more focussed approaches (with the exception of aCGH) in order to resolve the issues of genotyping SNPs (without the confounding of PSVs), quantify CNV and map CNV breakpoints. The PRT method first described by Hollox and Armour [355], and subsequent modification [349], is able to quantify the *FCGR3* regions of the two CNVRs at this locus. A PRT assay that measures the CNV of *FCGR2C* has also been described [348]; however, there were a significant number of individuals excluded from their analysis, which is a limitation of this assay. This PRT assay has been performed on the HapMap CEU reference samples and compared directly to the TriPSV described in this chapter (Figure 5.12). There was poor agreement between the results of this PRT assay and both the WGTP array probe 8H4 and the TriPSV average CNP score. A potential reason for this problem is the positioning of the primers required for the PRT assay as there is significant overlap with the SD (>4.9Kb from the boundary), and potentially with CNVR breakpoints which are dynamic within a population, i.e. multiple CNV breakpoints exist. To this end an alternative assay that measures the peak height of a nucleotide at three homologous positions in *FCGR2A,*

FCGR2B and *FCGR2C* was developed and is described herein. This assay does not appear to be confounded by the CNVR breakpoints as the reference nucleotide in *FCGR2A* lies >1.5kb centromeric to the boundary of the SD, which suggest it is a more robust alternative to the *FCGR2* PRT assay described by Neiderer *et al* [348]. In addition, it appears to be equivalent in CN detection to an aCGH probe of 185kb that spans the *FCGR* locus (Section 5.2.1.2, Figure 5.4 and Figure 5.12).

Sequencing approaches for defining CNVR breakpoints have pinpointed several regions, but have yet been unable to resolve breakpoints at base-pair resolution due to stretches of near 100% homology between the SD paralogs [351, 352]. This may also limit the ability to resolve the genomic structure of these CNVRs without resorting to low-throughput and highly technically challenging approaches such as fluorescent *in situ* hybridisation (FISH), which allow the direct visualisation of genome structure [380, 381]. This is of particular relevance when interpreting findings from CNV genetic association studies at this locus. As with the interpretation of any genetic association, functional experiments need to be invoked, or at the very least carefully designed in order to answer the specific question of what effect that particular CNV arrangement has on human biology. In the case of deletions of the CNVR containing *FCGR3B* and a fragment of *FCGR2C*, this appears to create a gene that is indistinguishable from *FCGR2B*, but which is expressed on CD56+ NK cells [352, 382]. Stewart-Akers *et al* described a lower frequency of CD32+ CD56+ NK cells in RA patients carrying the nonsense allele of rs10917661 which expressed FcγRIIb [383]. This phenotype is consistent with that described for individuals with deletions of the *FCGR3B* CNVR [352, 382]. However, no association was found between RA susceptibility and deletions of *FCGR2C* itself.

Defining absolute CN states from CN quantification assays is a major challenge, but it is by no means the only challenge that requires addressing in any CNV-based analysis. The

sensitivity of specific assays and CNV-typing platforms to experimental variation is a recognised issue in the field. This can in part be minimised by stringent and systematic methodology; for instance the particular DNA extraction method can affect assay performance [95], as well as the reaction conditions of the assay. The TriPSV assay described in this thesis was subject to extensive experimental and analytical optimisation in order to minimise this experimental variation (Section 5.2.1.3 and 5.2.1.4). However, between-cohort or between-sample batch variation can still introduce confounding into the statistical analysis of CNVs, and was notably present in the TriPSV data (Section 5.3.2 Figure 5.13). An additional source of confounding or heterogeneity can come from the inability to robustly segregate all samples into discrete CN states. Neiderer *et al* did not address this issue with their PRT assay and consequently discarded data points that could not be clustered. This runs the risk of drastically reducing statistical power by lowering the effective sample size of the cohort. Additionally, and arguably more importantly, if the sample exclusion is due to particular CN states, then this may introduce bias into the analysis.

CNVtools models continuous CN measurements as a mixture of Gaussian distributions and is able to incorporate between-batch variation into analyses. Rather than assigning discreet CN states to each individual, a posterior probability is calculated for each component of the mixture model, each of which relates to a particular CN state. Therefore the principal sources of confounding and heterogeneity in CNV association studies can be accounted for. CNVtools in its current version is limited by an inability to incorporate covariates into analyses; therefore it could not be applied directly for all of the analysis of CNV in this chapter. However, the posterior probabilities it assigns to each model component can be used to model the three CN states of deletion, wild type and duplication, which can then be used in appropriate regression models.

5.4.2 *FCGR2C* allelic and copy number variation and tumour necrosis factor antagonist treatment response

This doctoral work has found a statistical association between duplications of a region of *FCGR2C* that contained the tri-parallelic PSV used to measure CNV of this gene and treatment response to therapeutic TNF antagonists. This is the first known report of association between *FCGR2C* CNV and TNF antagonist therapy. As with any statistically significant ($p<0.016$) genetic association, independent replication is required to confirm this CN state as a predictor of TNF antagonist biologic response. This association was observed in the ΔDAS28-CRP, ΔlnCRP and EULAR response category analysis. The consistency across response measures with direction of effect lends some weight to this being a genuine effect. CRP is a component of the ΔDAS28-CRP, so one might expect an association to appear in both of these measures; however, CRP is only a minor component of ΔDAS28-CRP.

There are three immediate questions that arise from this association with *FCGR2C* duplications:

1. Do duplications of *FCGR2C* relate directly to increased amounts of FcγRIIc on affected cells?
2. Is the association with *FCGR2C* due to a specific CNVR that contains either *FCGR3A* or *FCGR3B* or a specific gene-rearrangement?
3. Is the association with *FCGR2C* duplication a false positive generated by the combined CNV of both *FCGR3A* and *FCGR3B* CNVRs, or a simple type I error itself?

The first of these questions can be addressed indirectly by further genetic analyses. The derived allele of a SNP that lies in the third exon of *FCGR2C* generates a premature stop codon that prevents cell surface expression of FcγRIIc. This nonsense allele has a high frequency in the Caucasian populations it has been studied in to date (derived allele

frequency ~60%) though in the BRAGGSS patients studied here it was surprisingly as high as 82%. This is the largest population for which an estimate of the frequency of this allele has been made, to date. The reason for the higher frequency of this allele in this patient cohort is unknown, but may warrant further investigation in the context of disease susceptibility. Allelic information for this SNP can therefore be integrated with CNV information to test the number of ORF alleles (i.e. those predicted to be expressed on the cell surface). This analysis was performed and no association was found between the number of copies of this ORF allele and response to TNF antagonist therapy.

The hypotheses in Section 5.1.3 are based on a predicted greater improvement in mAb TNF antagonist therapeutic response in individuals carrying increasing numbers of copies of the C allele of rs10917661. To this end a stratified analysis by drug class (i.e. etanercept and mAbs) was performed. Stratification reduces the effective sample size and thus the statistical power of these association tests. However, a combined analysis of all TNF antagonists may mask effects on one specific class of drug. A nominal association of increasing number of copies of the rs10917661-C allele were found with greater improvement in ΔlnCRP (Table 5.13). This effect is consistent with the results in Section 5.3.4 and the proposed mechanism in Section 5.1.3, but was not affected by including RFs in the regression model. Given the low statistical power (22%) and small sample size (n=197) this result must be interpreted with caution. This finding requires corroboration in this cohort once additional DNA samples become available. If this association with ΔlnCRP holds up, which may also be detectable in the ΔDAS28-CRP with increased numbers and better signal to noise ratio, then it will require replication in an independent cohort of mAb TNF antagonist treated patients. Given the observed effect estimate, distribution of ΔlnCRP and a type I error rate of 1% this replication cohort would require n=708 to achieve 80% power. Given these limitations this nominal association suggests a role for FcγRIIc,

possibly through NK cell mediated ADCC, in the therapeutic response to mAb TNF antagonist biologics.

An association was observed with higher RR of EULAR response in the analysis of *FCGR2C* duplications in the etanercept treated patients. This effect was not observed as strongly in the mAb treated patients, though a difference in effect was observed between adalimumab patients in BRAGGSS1 and BRAGGSS4 case cohorts. There was also a difference in the frequency of RF positive patients between these two cohorts. Whilst RF did not affect the regression estimates of EULAR response regressed on *FCGR2C* duplication posterior probability the estimates of RF were different between BRAGGSS1 and BRAGGSS4 concordant with an effect observed in BRAGGSS1 where they were more frequent. RFs are autoantibodies directed against self IgG epitopes and therefore capable of recognising humanised IgG antibody biologic therapies. Infliximab is a human-murine chimeric IgG1 antibody whilst adalimumab is fully humanised. Etanercept contains the human IgG1 Fc hinge, CH2 and CH3 regions. RFs epitopes recognise the CH2 and CH3 regions of human IgG [384], therefore their interaction with etanercept, infliximab and adalimumab may differ and ultimately alter their therapeutic efficacy mediated by binding to FcγRs.

A complimentary QSV assay exists for measuring the relative CN of *FCGR3A* and *FCGR3B* which requires de-convoluting with a second assay that measures the combined CN of both genes [349]. With this additional data available on a subset of individuals, it was also possible to investigate the second question. There was no observed association between duplications of either *FCGR3A* or *FCGR3B* with the same treatment outcome measures used to analyse *FCGR2C* duplication association with TNF antagonist response. These data were restricted to the BRAGGSS1 samples at the time of this doctoral work. A complete analysis of all BRAGGSS samples with these CNV assays will be a priority in order to understand the effect of CNV at this locus on TNF antagonist therapeutic response.

5.4.3 FcγRIIIa, macrophages, natural killer cells and therapeutic response to TNF antagonist therapy

Allotypes of FcγRIIIa have differential IgG isotype binding [320], which could affect binding affinity of TNF antagonist biologics if they mediate their effects through this IgG receptor. Association testing of rs396991 (F158V) did not find a detectable effect on response to TNF antagonism (Section 5.3.5), which suggests binding affinity differences to this receptor do not mediate differences in therapeutic response between individuals. This initial analysis did not take into account differences between the mAbs infliximab and adalimumab and the recombinant TNFR2 fusion protein etanercept. To test the hypotheses described in Section 5.1.3 this analysis was performed stratified by prescribed mAb TNF antagonist biologic, and each model was adjusted for the presence of RFs. There was no effect observed in this stratified analysis which indicates that this allelic variant is unlikely to alter response to these biologic drugs.

5.4.4 FcγRIIa, antibody binding, C-reactive protein and phagocytosis

IgG binding differences and functional consequences, such as phagocytosis of opsonised bacteria, are affected by allelic variation of FcγRIIa [320, 358]. Two non-synonymous changes exist in the *FCGR2A* gene, one of which is caused by a dinucleotide variant, i.e. two consecutive SNPs which together generate a missense substitution [95, 361]. TNF antagonists bind sTNF forming soluble immune complexes, which are able to bind and cross-link FcγR such as FcγRIIa on macrophages and neutrophils [358]. This may induce phagocytosis of these immune complexes or mTNF-expressing cells and mediate the therapeutic effects of TNF antagonist biologic therapy. No association was found between either minor alleles of rs1801274 or rs9427399 with response to these combined biologics (Table 5.15).

As per the hypotheses outlined in Section 5.1.3, the stratified analysis by TNF antagonist drug were performed for these two non-synonymous variants. No statistically significant association was observed with the minor allele of rs9427399 and response to either mAb TNF therapy or etanercept. The linear regression of rs1801274-A allele (131H) on ΔDAS28-CRP found a statistically significant association with lower improvement (Table 5.15; p=0.005). However, this association was not observed with either the ΔSJC28 or ΔlnCRP, though the direction of effect was consistent across all three outcome measures. A nominal association with a lower RR of EULAR response was also observed, consistent with this association of a worse outcome in individuals carrying this allele. Conversely there was a nominal association with higher ΔDAS28-CRP in the mAb TNF antagonist treated patients, which was also observed with ΔlnCRP albeit with a larger p-value (p=0.13). Binding affinity of monoclonal IgG1 to FcγRIIa allotypes of this SNP are very similar [320] despite this variant mapping to the IgG binding region of this Fc receptor [356]. The observation of a worse outcome in the rs1801274-A allele carriers is discordant with a small published study of etanercept treated psoriatic arthritis patients (n=55) [385]; the rs1801274-AA+AG patients had a higher OR of achieving a EULAR response.

FcγRIIa is able to bind both CRP and IgG [360], the former of which is also affected by the allotypes of this molecule defined by the alleles of rs1801274; 131HH donor neutrophils and monocytes bind CRP at a much lower affinity than 131RR donor cells. This was evident in the Ca^{2+} signalling induced in neutrophils from 131RR individuals but not 131HH. These two allotypes form different dimer structures which may account for this difference in CRP binding ability [356]. The opposite directions of effect based on TNF antagonist treatment suggest differential roles for FcγRIIa depending on which drug is administered. The improved response in the mAb treated 131H allele (rs1801274-A) carriers may be due to a lack of competition between CRP and IgG anti-TNF molecules. This may allow more efficient clearance of anti-TNF:sTNF immune complexes or improved phagocytosis of

mTNF-expressing cells concordant with the hypothesis in Section 5.1.3, particularly in RF positive patients.

The association between rs1801274-A alleles and lower ΔDAS28-CRP in the etanercept treated cohort may reflect the effect of underlying pathology mediated by autoantibodies through FcγRIIa rather than a specific difference in therapeutic effect. Autoantibodies may mediate their pathological effects through promotion of pro-inflammatory effects on neutrophils and macrophages that express FcγRIIa. In this scenario the higher affinity allele (rs1801274-A; 131H) would bind IgG autoantibodies leading to increased activation of neutrophils and macrophages, including release of IL-6 and TNF [29, 53]. Therefore individuals carrying this allele would be predicted to experience a worse clinical response relative to those carrying the low affinity allotype of FcγRIIa (rs1801274-G; 131R). This effect, if genuine, would also be detectable in other independent patient cohorts treated with other non-mAb therapies. The YEAR cohort of DMARD treated early RA patients provides such a cohort. A nominal association was indeed observed between the tag allele for 131H with lower ΔDAS28-CRP, thus providing independent replication of this statistical association. This provides the first known observation of a replicated statistical association between *FCGR2A* genetic variation and disease pathology measured by clinical response in RA.

These opposing effects for FcγR allotypes highlight the considerable biological complexity involved in mediating therapeutic response in RA. In particular this indicates the need for biologically driven analyses in sufficiently powered cohorts. Given the multiple testing burden and low statistical power in these analyses of *FCGR2A* SNPs these must be interpreted with caution. Further independent replication cohorts of etanercept and mAb TNF antagonist treated patients will be required to corroborate these findings. Based on the observed effect sizes and a 1% type 1 error rate n=642 would be required to achieve

80% statistical power for the association with etanercept and n=2015 for the mAbs using the ΔDAS28-CRP.

5.4.5 Study design, statistical power and stratified analyses

The study design in Section 5.3.3; Figure 5.15 shows the complex analysis carried out in the context of *FCGR* genetic variation and response to TNF antagonist biologics. This design has involved analysis of combined biologics and specific drug classes based on biological hypotheses of different effects. This latter approach using analyses stratified by drug class have found nominal associations with allelic variation in two *FCGR* genes, one of which is affected by CNV. Stratification of this cohort reduces the effective sample size and thus statistical power at the cost of increasing the signal to noise ratio present in the data. This is particularly evident in the analysis of rs1801274 alleles which appear to show differential effects between etanercept and mAb anti-TNFs. This study design also increases the multiple testing burden; each genetic variant is tested for its effect on 4 outcome measures, increasing the probability of generating a type 1 error. A type I error rate of 1% was selected to reduce the number of false positive associations for the allelic tests. Given the number of genetic variants tested (m=9; both CN and allelic), the number of outcome measures and stratified analyses, this data set has been tested 48 times for allelic variants and 28 times for CNV.

The hypotheses used to inform these analyses are based on enhanced effects for patient positive for RF. RF presence/absence was included as a covariate in regression models to account for any effect mediated through these autoantibodies. This is not the same, however, as analysing whether an effect is present only in either the positive or negative group of patients. This requires further stratification which diminishes statistical power further and is therefore unlikely to be an accurate result where there is a low frequency of RF negative patients (i.e. an increased chance of a type 2 error).

Replication and corroboration of the nominal associations observed here in will be required to confirm the functional interpretation of *FCGR* non-synonymous genetic variation with response to TNF antagonist biologic therapy. The effects and biological mechanisms responsible for these effects can then be refined by focussed functional experiments at the molecular and cellular level.

Chapter 6 – General Discussion, Conclusion and Future Work:

Rheumatoid arthritis, genetic variation and treatment response

6.1 Biomarkers of response to therapy

GWAS have been phenomenally successful at identifying common genetic variation associated with susceptibility to common complex diseases. This has in part confirmed one aspect of the "common variant common disease" hypothesis; complex diseases are underpinned by tens, hundreds or even thousands of common alleles each of small effect. Issues of adequate statistical power indicate GWAS are ideally placed to identify common variants, lower frequency variation of similar or larger magnitude of effect requires much larger samples sizes [386] or an alternative study design [387]. The magnitude of treatment response for any individual patient for any disease with a complex genetic aetiology is also likely to be under the genetic influence of multiple alleles. Investigation of the genetic architecture of treatment response therefore requires a careful, rigorous and robust approach in order to realise the goal of a health care system integrated with personalised medicine. The apparent phenotypic, and probable genetic, heterogeneity of disease pathology is also manifest in treatment response as multiple factors influence both biological and clinical response to medications. Patients may miss medications, or ingest chemical substances that interact with their prescribed treatment (such as coffee, alcohol or grapefruit juice). Outside of the rigorously and tightly controlled environment of a clinical trial it is difficult to control all of these factors directly, particularly with real world observational cohorts of patients and their various co-morbidities and idiosyncrasies. How then are we to meet the goal of a world of personalised medicine? A start can be made by performing statistically robust well designed studies that aim to both identify genetic associations with treatment response and understand their effect on objectively measurable biological (and clinical) phenotypes. These analyses may be guided by knowledge of molecular and cellular pathways and processes implicated in disease pathology.

6.1.1 Identification of genetic biomarkers of response to conventional therapies in rheumatoid arthritis

The first two results chapters of this doctoral thesis describe an attempt to identify genetic variation that is associated with treatment response to DMARDS, including MTX, in patients with diagnosed RA. Known disease susceptibility variants have been investigated for their impact on treatment response; the result here is largely one of inconsistency due to low statistical power and a previously unappreciated level of heterogeneity within and between observational cohorts. This highlights several issues of between-study variation: definitions of disease and treatment response, measures of treatment response, prescribed therapeutic regimen and progressive changes in treatment goals with the increasingly widespread adoption of the treat to target approach in routine clinical practise. In any investigation of predictive biomarkers one might expect those that are consistent and robust to manifest regardless of between-study variation or subtleties in treatment response definition. However, even large effects may be masked by considerable differences in study design and patient treatment, more so for complex traits such as treatment response in RA.

In these circumstances one must make an attempt to understand this heterogeneity in order to account for it either in statistical analyses or future study design. It is apparent that the nature of RA pathology changes over time and in response to treatment. For instance the clinical variables correlated with measures of disease activity at baseline study visit in the YEAR cohort are different at the 1 year follow-up visit. There is also a need for objective measures of treatment response that are able to capture the concept of “disease activity” adequately; measures that may then be applied to pharmacogenetics research. These measures must fit certain criteria:

- Measurements must be independent of the individual doing the measuring – subjective influence will introduce noise and potential bias that can lead to false positive or false negative findings. Measurements must also be reproducible regardless of the laboratory and environment in which they are made.
- Measurements must track disease pathology including disease response to therapeutic treatment
- Measurements must have a clinical and biological interpretation that can be used to guide treatment decision making and not be dictated and defined by physician judgement

Patients prescribed different drugs, including one-off treatments, such as glucocorticoid injections, should have all information recorded. Furthermore, an objective measurement of treatment response should be recorded pre- and post-injection with steroids, to understand the immediate effects on biomarkers and their effect on long term treatment response.

Once a genetic biomarker has been identified it is not sufficient to simply start testing its clinical utility, particularly for traits that display genetic heterogeneity. Genetic variants that are associated with a complex trait are also likely to have a complex relationship with that trait. Therefore, it is important to understand how genetic variation is associated with a trait in order to understand this genetic heterogeneity. This can be achieved to an extent by estimating the heritable component of the treatment response measured using genome-wide genotyping data in large cohorts; GCTA is one method that has been utilised for complex traits [156]. A large majority of genetic variants associated with complex disease in GWAS are located outside of protein coding regions of the genome [281], some of which have been further related to changes in gene transcript levels with proximal genes [276, 277, 300]. Recently (2013) a large meta-analysis of peripheral blood eQTL studies

found *trans*-regulatory SNPs using a statistically robust study design, though lacking adequately powered replication analysis [388]. The authors of this study related their findings to the known disease pathology of type 1 diabetes (T1D) and observations of convergence with a disease phenotype. They noted two particular SNP alleles (rs3184504-T and rs4788084-C) were associated in *trans* with increased expression of multiple interferon-γ response genes as a result of different *cis* effects on the expression of nearby genes. Thus, independent SNP alleles with different molecular effects may converge on a particular phenotype. In these circumstances a genetic variant may therefore affect treatment response in one patient, but not another. Alternatively, this calls into question the clinical utility of genetic variants for these complex traits, and suggests soluble or cellular biomarkers of response that directly measure pathology or disease relevant molecules may be more promising.

As mentioned above, addressing these issues will require robust objective measures of response, careful collection of clinical data including all prescribed medications, *a priori* sample size calculations based on small effect sizes (for instance effect sizes estimated in Chapter 4.3.2.3) and similarly powered (or better powered; to allow for any instances of winner's curse) replication cohorts.

6.1.2 Identification of genetic biomarkers of response to biologic therapy in rheumatoid arthritis

The field of TNF antagonist biologic therapy has started to address many of the issues outlined above by the use of large multi-centre cohorts and replication analyses [226, 227, 232, 233, 389]. There still remains a lack of objective treatment response measure, which may be why there are few consistently replicated associations (Chapter 1.4.5).

This doctoral thesis analysed genetic variation at the low affinity *FCGR* locus and response to TNF antagonist biologic therapy based on known and predicted biological phenomenon.

To this end, the statistical analyses were undertaken in a manner that aimed to account for some of the biological complexity of these drugs and receptors using a stratified approach. These analyses were stratified according to drug class (etanercept or mAb) as these are predicted to mediate their therapeutic efficacy through different molecular and cellular mechanisms, particularly in the context of the low affinity FcγRs. Nominal associations were observed with variation predicted to affect ADCC mediated by NK cells and immune complex clearance and/or phagocytosis by macrophages and PMNs. As an indication of the biological complexity nominal associations with allotypes of FcγRIIa displayed differential (i.e. opposite directions of effect) dependent on TNF antagonist drug class. The interpretation present herein is that the effect observed in the mAb treated patients is due to the mechanism of immune complex clearance and/or phagocytosis. Patients with the high affinity allele (131H) are able to bind IgG without competition from CRP [360]and thus benefit from greater phagocytosis of mTNF expressing cells or clearance of large immune complexes formed between anti-TNF mAbs and sTNF. This effect is predicted to be enhanced in patients positive for RF; these autoantibodies are raised against self IgG which might include IgG1 mAb therapies. The analysis presented in this thesis did not stratify by RF positivity due to further loss of statistical power, however, it might be predicted that RF positive patients receiving infliximab or adalimumab would see a greater benefit than those treated with etanercept. Likewise an association between worse response and this same high affinity allele was observed in patients treated with etanercept. The binding of etanercept to FcγRs is unknown, however, it is not predicted to induce effector cell mechanisms via these receptors. The association with worse response is instead concordant with the predicted effect of FcγRIIa 131H allotypes on disease pathology. Autoantibodies are predicted to be pathological in RA, which may be mediated through the FcγRs [29]. FcγRIIa 131H allotypes would bind IgG autoantibodies with greater affinity thus inducing greater pro-inflammatory activation of macrophages and PMNs. This would

manifest as more severe disease and diminished therapeutic efficacy. If this interpretation is accurate autoantibody RA patients treated with non-mAb therapy (e.g. synthetic DMARDs) would experience more severe disease and/or diminished treatment response. Indeed, a nominal association was observed between the minor allele of a tag SNP for rs1801274 (rs12139150; 131H allele) and lower ΔDAS28-CRP, adjusted for baseline DAS28-CRP, in autoantibody positive patients in the YEAR cohort. This provides statistical evidence in favour of a pathological role of autoantibodies through FcγRIIa, which is modulated by allelic variation.

An association was observed with *FCGR2C* CNV duplications and improved response to TNF antagonist therapy. In a stratified analysis by drug class, an association was found to correspond to increasing number of copies of a translated allele of FcγRIIc. Whether these represent the same statistical and biological effect is currently unknown. Future stratified analyses will require larger sample sizes to address this question; based on the observed association in Chapters 5.3.5 and summary statistics of the BRAGGSS cohort this would require n=708 for a test of 80% power (α=0.01, $\bar{x}$=0.78, σ=1.06, β_G=0.25).

There are several further issues relevant to replication beyond those sources of heterogeneity discussed in the section above, notably of an experimental or statistical nature. CNV association studies have also been plagued by reports of false positive associations with complex traits that arise due to inadequacies in statistical power and CNV measurement assays. This doctoral thesis has described a CNV assay that measures a single nucleotide in three homologous genes (*FCGR2A, FCGR2B* and *FCGR2C*). Ideally any replication analysis would utilise this assay or an appropriately robust assay not confounded by CNVR breakpoints. These results attest to the considerable biological complexity inherent in therapeutic response in RA.

6.1.3 Interpreting genetic association signals

The interpretation of genetic association signals, whether they arise from GWAS of disease susceptibility or candidate gene studies of pharmacogenetics, is not easy. An understanding of the influence of genetic variation on human biology will require the integration of multiple sources of data at the molecular and cellular level. This would likely involve experimental investigations of gene transcript levels, mRNA splicing, non-coding RNA expression, and mRNA transcript structure in cells under both resting and stimulated states. The exact stimulation and environment would need to be guided by the cell type, the disease and specific hypotheses about the effect of the genetic variant on known disease pathology. At the protein level the study of non-synonymous variation will require the integration of biochemical approaches to understand protein structure and the involvement of proteins in interacting biological networks. All of these approaches ultimately rely on the generation of large experimental data sets followed by their integration and biological interpretation. This would require the collaboration of multiple research groups and institutes with different specialities and their integration with a sufficient informatics infrastructure to handle and store large quantities of data (petabytes).

A focussed approach can be taken for any given genetic loci or genetic variant, but it must ultimately be interpreted in the context of its environment and interacting partners to appreciate its biological effect.

6.2 Future Work – replication and interpretation

6.2.1 *HLA-DRB1* and disease-modifying anti-rheumatic drug response

Analysis of *HLA-DRB1* SE alleles in a cohort of DMARD treated YEAR patients found a nominal association with worse treatment response (Chapter 3.3.1.1) and increased risk of

joint damage progression (Chapter 3.3.2.4). The latter of these was largely attributed to the presence of autoantibodies, though a residual effect not captured by the SE allele coding may exist. Raychaudhuri *et al* have refined the association between the MHC and RA susceptibility to three loci: *HLA-DRB1, HLA-DP* and *HLA-B* [67]. These nominal associations with *HLA-DRB1* alleles therefore require both independent replication and refinement. Amino acids 11 and 13 identified by Raychaudhuri *et al* have not been investigated or incorporated into risk models of *HLA-DRB1* SE. Therefore, the effect of these two amino acid positions, and those of *HLA-B* and *HLA-DP* warrant investigation for their effect on treatment response in RA.

There appeared to be a residual effect in the analysis of *HLA-DRB1* on joint damage progression (Chapter 3.3.2.4, Table 3.24) after adjustment for autoantibodies. This suggests that there remains an effect from these SE alleles that warrants further investigation. A larger cohort than that described here will be required to elucidate this residual effect. Incorporation of *HLA-DRB1* amino acid positions 11 and 13, *HLA-DP* and *HLA-B* alleles may aid in this resolution of association signal with joint damage progression.

6.2.2 Methotrexate pharmacogenetics

This doctoral work described nominal associations between the minor alleles of SNPs at the *MTHFS* locus and improvements in DAS28-CRP and SJC28. These associations require replication in independent cohorts treated with MTX using these same treatment response measures. A GWAS of MTX response is currently underway that includes the YEAR samples described in this thesis (PAMERA), it will be of considerable interest to see which, if any, genetic variants are identified from their analysis. The PAMERA study aims to perform an initial GWAS with ~1000 MTX treated early RA patients to identify common genetic variants associated with response within the first 3-6months of treatment. Individual variants are then to be replicated in independent cohorts of MTX treated early RA patients. Confirmed

variants associated with MTX may also provide candidates for association testing with radiographic joint damage. These variants may shed light on what extent treatment failure influences the development and progression of joint damage in these individuals.

6.2.3 *FCGR* and tumour necrosis factor antagonist therapeutic response

A combined QSV and PRT assay to measure the CN of *FCGR3A* and *FCGR3B* has been published [349]. These assays can be used to investigate the role of further *FCGR* CNV in TNF antagonist response and test whether the association with *FCGR2C* duplication arises by a technical artefact, due to CNV of a specific gene or CNVR, or is an independent effect. A replication analysis should be designed with an appropriately powered cohort treated with mAb TNF antagonist therapy stratified by RF, and incorporate allelic variation known to define cell surface expressed FcγRIIc. This may also include a SNP predicted to alter an intron splice site and generate a downstream nonsense mutation in *FCGR2C* [382]. These specific replication analyses should also address the question of *FCGR2A* non-synonymous variation on mAb TNF antagonist therapy, and whether the effect of rs1801274-A (131H) alleles reflects autoantibody driven disease pathology in patients treated with non-mAb therapies. Based on the proposed hypotheses in Chapter 5.1.3 and the nominal associations observed in Chapter 5 several experiments are proposed to test the functional significance of these findings and refine their interpretation (Figure 6.1).

6.2.3.1 FcγRIIc and NK cell mediated antibody-dependent cell-mediated cytotoxicity

Increasing number of copies of FcγRIIc on the surface of CD3-CD56+ NK cells is predicted to correlate with increasing copies of the rs10917661-C allele, accounting for CNV. This increase in cell surface FcγRIIc can manifest as increased NK cell mediated ADCC of mTNF expressing cells (Figure 6.1A). Cellular systems exist for testing the effects of biological variables on ADCC [390]. Target cells expressing mTNF would be incubated with different TNF antagonist biologics (in separate experiments) and incubated with CD3-CD56+ NK cells

from patients with increasing number of copies of rs10917661-C allele (wild type TT individuals as controls) and target cell death measured using flow cytometry. The effect of RF can also be incorporated into this system by incubating the mTNF expressing target cells with RFs and repeating the experiments. Confirmation of FcγRIIc mediated ADCC can be achieved by incubating cells with a Fab fragment directed against FcγRII (no FcγRIIc-specific antibody exists) to block it's binding capacity.

6.2.3.2 FcγRIIa, immune complex clearance, phagocytosis and autoantibodies

Two hypotheses are proposed based on the results in Chapter 5.3.7.2 (Figure 6.1B):

1. Autoantibodies in RA mediate pathology via FcγRIIa which is altered by non-synonymous variation that affects IgG2 binding affinity leading to promotion of pro-inflammatory activation of macrophages and PMNs.
    a. Patients with the high affinity 131H allele of this receptor are predicted to have more severe disease and experience worse clinical response to non-mAb therapy.
    b. PMNs and/or macrophages from 131H and 131R (rs1801274-A and G respectively) allele donors are predicted to display altered pro-inflammatory effector function as a direct result of autoantibody and autoantibody-containing immune complexes binding to FcγRIIa.
2. IgG1 mAb TNF antagonists compete for binding with CRP to FcγRIIa which is altered by 131H>R (rs1801274A>G) allotypes leading to differences in therapeutic efficacy.
    a. 131R (rs1801274-G) allotype donor PMNs and/or macrophage internalisation of TNF:anti-TNF immune complexes, enlarged in the presence of RFs, is reduced in the presence of CRP. 131H (rs1801274-A) donor cells are unaffected by the presence of CRP.

b. 131R (rs1801274-G) allotype donor PMN and/or macrophage phagocytosis of mTNF-expressing (apoptotic) cells, enhanced in the presence of RFs, is reduced in the presence of competing CRP. 131H (rs1801274-A) donor cells are unaffected by the presence of CRP.

The first hypothesis would require the measurement of pro-inflammatory cytokines associated with PMN and macrophage activation through FcγRIIa, such as TNF, IL-6 and IL-1β. These could be measured from donor cells from healthy individuals of differing rs1801274 genotypes (to avoid confounding from disease associated cellular changes) incubated with ACPA and/or RF. Confirmation that this effect is mediated through FcγRIIa would be achieved by pre-incubating cells with an anti-FcγRIIa Fab agnostic to allotype.

The effect of rs1801274 alleles on mAb therapeutic efficacy via immune complex clearance could be tested in a similar system. IgG opsonised latex bead uptake could be measured with or without the presence of CRP from healthy control donor cells of differing genotypes. The effect of RF could also be tested in this system by pre-incubating the IgG opsonised beads with this autoantibody then adding the donor cells. Whether mAb TNF antagonists also mediate their therapeutic effect through apoptosis of mTNF-expressing cells *in vivo* in RA is unknown, though it has been observed in Crohn's disease [327]. These apoptotic cells, or somatic cells expressing mTNF, may then be cleared by PMNs and/or macrophages which may be altered by FcγRIIa allotypes. Phagocytosis of mTNF expressing cells, pre-incubated with mAb to induce apoptosis, by PMNs and macrophages can be measured by flow cytometry [391]. Both of these systems, immune complex clearance and mTNF expressing cell phagocytosis could then be incubated with CRP to assess the effect of this FcγRIIa binding competitor between receptor allotypes.

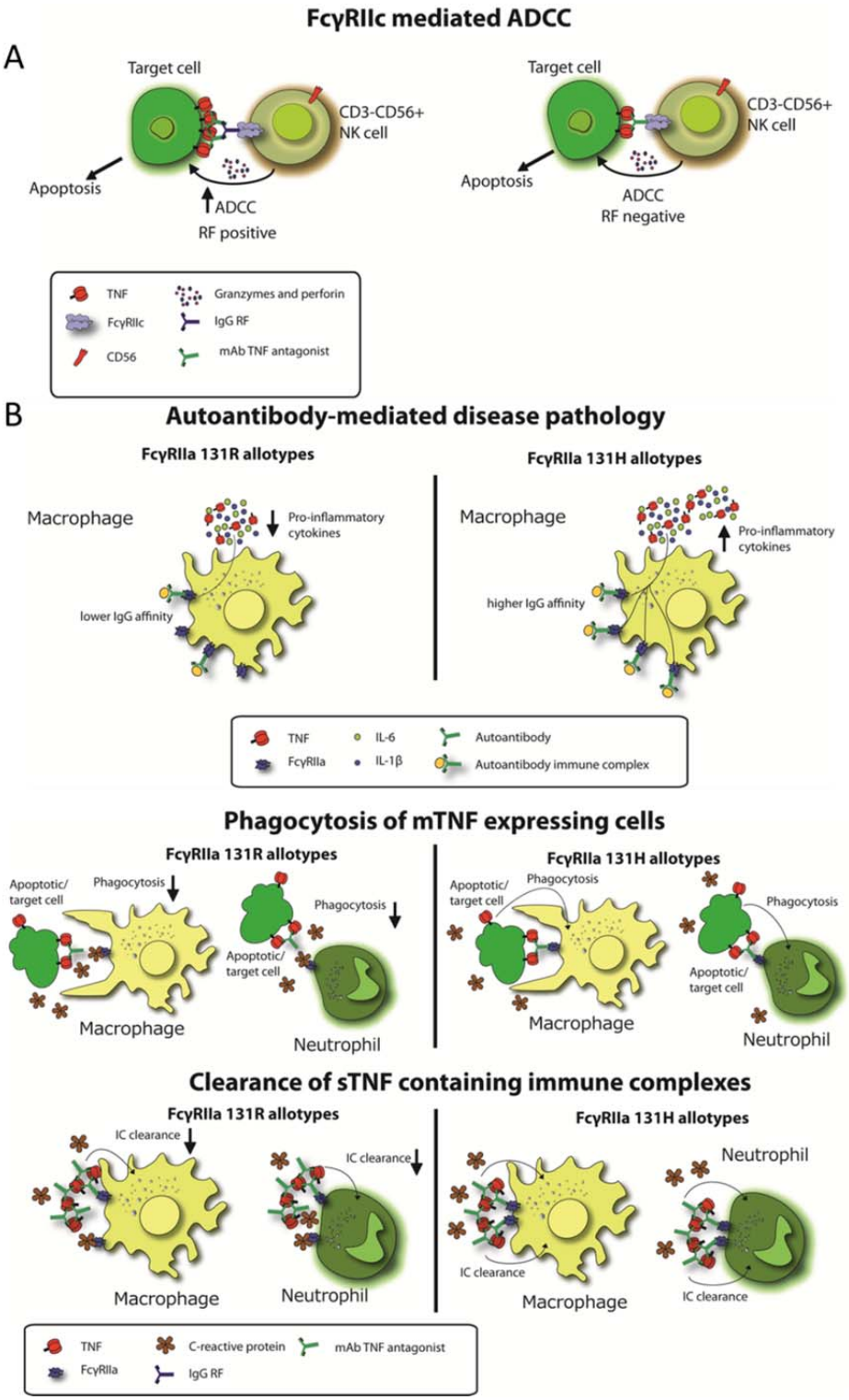


Figure 6.1. Hypotheses of FcγR genetic variation effects on RA disease pathology and treatment response to TNF antagonist therapy. A – FcγRIIc mediates mAb TNF antagonist induced ADCC by CD3-CD56+ NK cells, an effect enhanced by the presence of RFs which allow more cross-linking of FcγRIIc molecules. Increasing copies of the rs10917661 C allele corresponds to increased cell surfaced expression of FcγRIIc leading to increased ADCC of anti-TNF mAb coated cells. The presence of IgG RFs enhances ADCC by enabling further cross-linking of FcγRIIc. Increased ADCC of mTNF expressing cells leads to greater clinical improvement. B – In the absence of mAb TNF antagonist therapy autoantibody binding to FcγRIIa high affinity allotypes (131H) display increased activation and pro-inflammatory cytokine release manifested as a worse clinical response. The presence of mAb TNF antagonist biologics alters this relationship. Immune complex clearance of sTNF, or phagocytosis of mTNF expressing target of apoptotic cells, is reduced in 131R allotype carriers as CRP competes with IgG mAb TNF antagonist for binding to FcγRIIa. FcγRIIa 131H allotypes do not bind CRP so clear relatively more sTNF containing immune complexes, which manifests as a relatively improved clinical response.

6.2.4 Interpreting genetic association signals – *C8orf13-BLK, AFF3* and *KIF5A-PIP4K2C* loci

The finding of an isoform specific eQTL with *AFF3* gene transcripts warrants further investigation, both as a confirmed disease susceptibility locus as well as a nominally associated joint damage one. The function of the protein encoded by this gene in the context of B cell biology is of particular interest. Likewise the tissue expression pattern of *AFF3* in healthy and disease resting and stimulated cells should be investigated in order to gauge its impact on RA pathology. *C8orf13* has remained neglected compared to its neighbour *BLK*; the role of this gene of unknown function requires elucidating, especially given its apparent transcript regulation by a known RA susceptibility variant (rs13277113) and cross-species nucleotide sequence conservation (Chapter 3.4.6). The refinement and fine-mapping of these loci with early joint damage in multiple cohorts may yield insights into their effects on disease pathology. This includes the signal of association with joint damage at the *KIF5A-PIP4K2C* locus, given its recent refinement with RA susceptibility to the region proximal to *OS9* and *METTL21B* [32]; the transcription of the latter is observed to be under genetic regulation [276, 277].

6.3 Concluding remarks

There are significant problems in the field of RA pharmacogenetics that require resolution if personalised and stratified medicine are to be implemented in the rheumatology clinic. There already exist many cohorts internationally of RA patients treated with both conventional DMARDs and biologic therapies. These cohorts require standardisation across data collection and treatment response measures; their use requires careful study design to achieve these goals. Given the genetic and phenotypic heterogeneity of RA pathology and treatment response there exist the distinct possibility that hundreds, if not thousands,

of genetic variants underlie these traits. If this is the case, is pharmacogenetics in rheumatology clinically useless? A possibility that must always be entertained.

References

1. Symmons, D., et al., *The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century.* Rheumatology (Oxford), 2002. **41**: p. 793 - 800.
2. Young, A. and G. Koduri, *Extra-articular manifestations and complications of rheumatoid arthritis.* Best Practice & Research Clinical Rheumatology, 2007. **21**(5): p. 907-927.
3. Haagsma, C.J., et al., *Combination of sulphasalazine and methotrexate versus the single components in early rheumatoid arthritis: a randomized, controlled, double-blind, 52 week clinical trial.* Rheumatology, 1997. **36**(10): p. 1082-1088.
4. Weinblatt, M.E., et al., *Efficacy of Low-Dose Methotrexate in Rheumatoid Arthritis.* New England Journal of Medicine, 1985. **312**(13): p. 818-822.
5. Shmerling, R.H. and T.L. Delbanco, *The rheumatoid factor: An analysis of clinical utility.* The American Journal of Medicine, 1991. **91**(5): p. 528-534.
6. Miller, F.W., et al., *Epidemiology of environmental exposures and human autoimmune diseases: Findings from a National Institute of Environmental Health Sciences Expert Panel Workshop.* Journal of Autoimmunity, 2012. **39**(4): p. 259-271.
7. Heberden, W., *Commentaries on the History and Cure of Diseases*. 1818, Boston: Wells and Lilly.
8. Lawrence, J.S., *Heberden Oration, 1969 Rheumatoid Arthritis - Nature or Nurture?* Ann Rheum Dis, 1970. **29**: p. 357-379.
9. Bellamy, N., et al., *Rheumatoid arthritis in twins: a study of aetiopathogenesis based on the Australian Twin Registry.* Annals of the Rheumatic Diseases, 1992. **51**(5): p. 588-593.
10. Silman, A.J., et al., *Twin Concordance Rates for Rheumatoid Arthritis: Results from a Nationwide Study.* Rheumatology, 1993. **32**(10): p. 903-907.
11. MacGregor, A.J., et al., *Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins.* Arthritis & Rheumatism, 2000. **43**(1): p. 30-37.
12. Ohrndorf, S. and M. Backhaus, *Musculoskeletal ultrasonography in patients with rheumatoid arthritis.* Nat Rev Rheumatol, 2013. **9**(7): p. 433-437.
13. Conaghan, P.G., et al., *Elucidation of the relationship between synovitis and bone damage: A randomized magnetic resonance imaging study of individual joints in*

patients with early rheumatoid arthritis. Arthritis & Rheumatism, 2003. **48**(1): p. 64-71.

14. Ranganath, V.K., et al., *The utility of magnetic resonance imaging for assessing structural damage in randomized controlled trials in rheumatoid arthritis: Report from the imaging group of the American College of Rheumatology RA clinical trials task force.* Arthritis & Rheumatism, 2013. **65**(10): p. 2513-23.
15. Randen, I., et al., *The Identification of Germinal Centres and Follicular Dendritic Cell Networks in Rheumatoid Synovial Tissue.* Scandinavian Journal of Immunology, 1995. **41**(5): p. 481-486.
16. Young, A., et al., *How does functional disability in early rheumatoid arthritis (RA) affect patients and their lives? Results of 5 years of follow-up in 732 patients from the Early RA Study (ERAS).* Rheumatology, 2000. **39**(6): p. 603-611.
17. Carmona, L., et al., *Rheumatoid arthritis in Spain: occurrence of extra-articular manifestations and estimates of disease severity.* Annals of the Rheumatic Diseases, 2003. **62**(9): p. 897-900.
18. Hollan, I., et al., *Cardiovascular disease in autoimmune rheumatic diseases.* Autoimmunity Reviews. **12**(10): p. 1004-15.
19. Maradit-Kremers, H., et al., *Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: A population-based cohort study.* Arthritis & Rheumatism, 2005. **52**(2): p. 402-411.
20. van der Heijde, D., et al., *EULAR definition of erosive disease in light of the 2010 ACR/EULAR rheumatoid arthritis classification criteria.* Annals of the Rheumatic Diseases, 2013. **72**(4): p. 479-481.
21. Brown, A.K., et al., *An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis.* Arthritis & Rheumatism, 2008. **58**(10): p. 2958-2967.
22. Imafuku, Y., H. Yoshida, and Y. Yamada, *Reactivity of agalactosyl IgG with rheumatoid factor.* Clinica Chimica Acta, 2003. **334**(1–2): p. 217-223.
23. Young, A., et al., *Agalactosyl IgG: An aid to differential diagnosis in early synovitis.* Arthritis & Rheumatism, 1991. **34**(11): p. 1425-1429.
24. Parekh, R., et al., *A comparative analysis of disease-associated changes in the galactosylation of serum IgG.* Journal of Autoimmunity, 1989. **2**(2): p. 101-114.

25. Schellekens, G.A., et al., *The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide.* Arthritis & Rheumatism, 2000. **43**(1): p. 155-163.

26. Vossenaar, E., et al., *Rheumatoid arthritis specific anti-Sa antibodies target citrullinated vimentin.* Arthritis Res Ther, 2004. **6**(2): p. R142 - R150.

27. Klareskog, L., et al., *A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination.* Arthritis & Rheumatism, 2006. **54**(1): p. 38-46.

28. Trouw, L.A., et al., *Anti–cyclic citrullinated peptide antibodies from rheumatoid arthritis patients activate complement via both the classical and alternative pathways.* Arthritis & Rheumatism, 2009. **60**(7): p. 1923-1931.

29. Clavel, C., et al., *Induction of macrophage secretion of tumor necrosis factor α through Fcγ receptor IIa engagement by rheumatoid arthritis–specific autoantibodies to citrullinated proteins complexed with fibrinogen.* Arthritis & Rheumatism, 2008. **58**(3): p. 678-688.

30. Sokolove, J., et al., *Autoantibody Epitope Spreading in the Pre-Clinical Phase Predicts Progression to Rheumatoid Arthritis.* PLoS ONE, 2012. **7**(5): p. e35296.

31. Trynka, G., et al., *Chromatin marks identify critical cell types for fine mapping complex trait variants.* Nat Genet, 2013. **45**(2): p. 124-130.

32. Eyre, S., et al., *High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis.* Nat Genet, 2012. **44**(12): p. 1336-1340.

33. Dai, X., et al., *A disease-associated PTPN22 variant promotes systemic autoimmunity in murine models.* The Journal of Clinical Investigation, 2013. **123**(5): p. 2024-2036.

34. Zhang, J., et al., *The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness.* Nat Genet, 2011. **43**(9): p. 902-907.

35. van de Stadt, L.A., et al., *Development of the anti–citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis.* Arthritis & Rheumatism, 2011. **63**(11): p. 3226-3233.

36. Jordan, M.S., et al., *Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide.* Nat Immunol, 2001. **2**(4): p. 301-306.

37. Barra, L., et al., *Anti–Citrullinated Protein Antibodies in Unaffected First-Degree Relatives of Rheumatoid Arthritis Patients.* Arthritis & Rheumatism, 2013. **65**(6): p. 1439-1447.

38. Harre, U., et al., *Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin.* The Journal of Clinical Investigation, 2012. **122**(5): p. 1791-1802.

39. Kleyer, A., et al., *Bone loss before the clinical onset of rheumatoid arthritis in subjects with anticitrullinated protein antibodies.* Annals of the Rheumatic Diseases, 2013.

40. Krabben, A., et al., *MRI of hand and foot joints of patients with anticitrullinated peptide antibody positive arthralgia without clinical arthritis.* Annals of the Rheumatic Diseases, 2013.

41. Bizzaro, N., et al., *Anti-cyclic citrullinated peptide antibody titer predicts time to rheumatoid arthritis onset in patients with undifferentiated arthritis: results from a 2-year prospective study.* Arthritis Research & Therapy, 2013. **15**(1): p. R16.

42. Schneider, H., et al., *Reversal of the TCR Stop Signal by CTLA-4.* Science, 2006. **313**(5795): p. 1972-1975.

43. Qureshi, O.S., et al., *Trans-Endocytosis of CD80 and CD86: A Molecular Basis for the Cell-Extrinsic Function of CTLA-4.* Science, 2011. **332**(6029): p. 600-603.

44. Costantino, C.M., C.M. Baecher-Allan, and D.A. Hafler, *Human regulatory T cells and autoimmunity.* European Journal of Immunology, 2008. **38**(4): p. 921-924.

45. Sakaguchi, S., K. Wing, and M. Miyara, *Regulatory T cells – a brief history and perspective.* European Journal of Immunology, 2007. **37**(S1): p. S116-S123.

46. Ehrenstein, M.R., et al., *Compromised Function of Regulatory T Cells in Rheumatoid Arthritis and Reversal by Anti-TNFα Therapy.* The Journal of Experimental Medicine, 2004. **200**(3): p. 277-285.

47. Kong, Y.-Y., et al., *Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand.* Nature, 1999. **402**(6759): p. 304-309.

48. Glimcher, L.H., et al., *Ia antigen-bearing B cell tumor lines can present protein antigen and alloantigen in a major histocompatibility complex-restricted fashion to antigen-reactive T cells.* The Journal of Experimental Medicine, 1982. **155**(2): p. 445-459.

49. Julius, M.H., H. von Boehmer, and C.L. Sidman, *Dissociation of two signals required for activation of resting B cells.* Proceedings of the National Academy of Sciences, 1982. **79**(6): p. 1989-1993.

50. Fehr, K., et al., *Production of agglutinators and rheumatoid factors in plasma cells of rheumatoid and nonrheumatoid synovial tissues.* Arthritis & Rheumatism, 1981. **24**(3): p. 510-519.

51. Reparon-Schuijt, C.C., et al., *Secretion of anti–citrulline-containing peptide antibody by B lymphocytes in rheumatoid arthritis.* Arthritis & Rheumatism, 2001. **44**(1): p. 41-47.

52. Van Gaalen, F.A., et al., *Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis.* Arthritis & Rheumatism, 2004. **50**(7): p. 2113-2121.

53. Skokowa, J., et al., *Macrophages Induce the Inflammatory Response in the Pulmonary Arthus Reaction through Gαi2 Activation That Controls C5aR and Fc Receptor Cooperation.* The Journal of Immunology, 2005. **174**(5): p. 3041-3050.

54. Cheon, H., et al., *Increased expression of pro-inflammatory cytokines and metalloproteinase-1 by TGF-β1 in synovial fibroblasts from rheumatoid arthritis and normal individuals.* Clinical & Experimental Immunology, 2002. **127**(3): p. 547-552.

55. Kumkumian, G.K., et al., *Platelet-derived growth factor and IL-1 interactions in rheumatoid arthritis. Regulation of synoviocyte proliferation, prostaglandin production, and collagenase transcription.* The Journal of Immunology, 1989. **143**(3): p. 833-7.

56. Green, M.J., et al., *Serum MMP-3 and MMP-1 and progression of joint damage in early rheumatoid arthritis.* Rheumatology, 2003. **42**(1): p. 83-88.

57. Stastny, P., *Mixed lymphocyte cultures in rheumatoid arthritis.* The Journal of Clinical Investigation, 1976. **57**(5): p. 1148-1157.

58. Stastny, P., *Association of the B-Cell Alloantigen DRw4 with Rheumatoid Arthritis.* New England Journal of Medicine, 1978. **298**(16): p. 869-871.

59. Gregersen, P.K., J. Silver, and R.J. Winchester, *The shared epitope hypothesis. an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis.* Arthritis & Rheumatism, 1987. **30**(11): p. 1205-1213.

60. Ding, B., et al., *Different patterns of associations with anti–citrullinated protein antibody–positive and anti–citrullinated protein antibody–negative rheumatoid*

arthritis in the extended major histocompatibility complex region. Arthritis & Rheumatism, 2009. **60**(1): p. 30-38.

61. Verpoort, K.N., et al., *Association of HLA–DR3 with anti–cyclic citrullinated peptide antibody–negative rheumatoid arthritis.* Arthritis & Rheumatism, 2005. **52**(10): p. 3058-3062.
62. Jawaheer, D., et al., *Dissecting the Genetic Complexity of the Association between Human Leukocyte Antigens and Rheumatoid Arthritis.* The American Journal of Human Genetics, 2002. **71**(3): p. 585-594.
63. Vignal, C., et al., *Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci.* Arthritis & Rheumatism, 2009. **60**(1): p. 53-62.
64. Mackie, S.L., et al., *A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population.* Genes Immun, 2012. **13**(2): p. 120-128.
65. du Montcel, S.T., et al., *New classification of HLA–DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility.* Arthritis & Rheumatism, 2005. **52**(4): p. 1063-1068.
66. de Vries, N., et al., *Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67–74 of the HLA–DRB1 molecule.* Arthritis & Rheumatism, 2002. **46**(4): p. 921-928.
67. Raychaudhuri, S., et al., *Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis.* Nat Genet, 2012. **44**(3): p. 291-296.
68. Lee, H., Lee, A.T., Criswell, L.A., Seldin, M.F., Amos, C.I., Carulli, J.P., Navarrete, C., Remmers E.F., Kastner, D.L., Plenge, R.M., Li, W., & Gregersen, P.K.,, *Several region in the major histocompatibility complex confer risk for anti-CCP antibody positive rheumatoid arthritis, independent of the DRB1 locus.* Mol Med, 2007. **14**(5-6): p. 293-300.
69. Cornélis, F., et al., *New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study.* Proceedings of the National Academy of Sciences, 1998. **95**(18): p. 10746-10750.
70. Bali D., G.S., Kostyu DD., Goel N., Bruce I., Bell A., Walker DJ., Tran K., Zhu DK., Costello TJ., Amos CI., Seldin MF.,, *Genetic analysis of multiplex rheumatoid arthritis families.* Genes Immun, 1999. **1**(1): p. 28-36.

71. Jawaheer, D., et al., *A Genomewide Screen in Multiplex Rheumatoid Arthritis Families Suggests Genetic Overlap with Other Autoimmune Diseases.* The American Journal of Human Genetics, 2001. **68**(4): p. 927-936.

72. John, S., et al., *Linkage of cytokine genes to rheumatoid arthritis. Evidence of genetic heterogeneity.* Annals of the Rheumatic Diseases, 1998. **57**(6): p. 361-365.

73. Myerscough, A., et al., *Linkage of rheumatoid arthritis to insulin-dependent diabetes mellitus loci: Evidence supporting a hypothesis for the existence of common autoimmune susceptibility loci.* Arthritis & Rheumatism, 2000. **43**(12): p. 2771-2775.

74. Barton, A., et al., *High resolution linkage and association mapping identifies a novel rheumatoid arthritis susceptibility locus homologous to one linked to two rat models of inflammatory arthritis.* Human Molecular Genetics, 2001. **10**(18): p. 1901-1906.

75. Han, S., et al., *Meta-analysis of the association of CTLA-4 exon-1 +49A/G polymorphism with rheumatoid arthritis.* Human Genetics, 2005. **118**(1): p. 123-132.

76. Plenge, R.M., et al., *Replication of Putative Candidate-Gene Associations with Rheumatoid Arthritis in >4,000 Samples from North America and Sweden: Association of Susceptibility with PTPN22, CTLA4, and PADI4.* The American Journal of Human Genetics, 2005. **77**(6): p. 1044-1060.

77. Suzuki, A., et al., *Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis.* Nat Genet, 2003. **34**(4): p. 395-402.

78. Kurreeman, F., et al., *A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis.* PLoS Med, 2007. **4**: p. e278.

79. Consortium, W.T.C.C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-678.

80. Plenge, R.M., et al., *Two independent alleles at 6q23 associated with risk of rheumatoid arthritis.* Nat Genet, 2007. **39**(12): p. 1477-1482.

81. Thomson, W., et al., *Rheumatoid arthritis association at 6q23.* Nat Genet, 2007. **39**(12): p. 1431-1433.

82. Lee, E.G., et al., *Failure to Regulate TNF-Induced NF-κB and Cell Death Responses in A20-Deficient Mice.* Science, 2000. **289**(5488): p. 2350-2354.

83. Dieguez-Gonzalez, R., et al., *Analysis of TNFAIP3, a feedback inhibitor of nuclear factor-kappaB and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility.* Arthritis Research & Therapy, 2009. **11**(2): p. R42.

84. Orozco, G., et al., *Combined effects of three independent SNPs greatly increase the risk estimate for RA at 6q23.* Human Molecular Genetics, 2009. **18**(14): p. 2693-2699.

85. Freudenberg, J., et al., *Genome-wide association study of rheumatoid arthritis in Koreans: Population-specific loci as well as overlap with European susceptibility loci.* Arthritis & Rheumatism, 2011. **63**(4): p. 884-893.

86. Gregersen, P.K., et al., *REL, encoding a member of the NF-[kappa]B family of transcription factors, is a newly defined risk locus for rheumatoid arthritis.* Nat Genet, 2009. **41**(7): p. 820-823.

87. Kochi, Y., et al., *A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility.* Nat Genet, 2010. **42**(6): p. 515-519.

88. Okada, Y., et al., *Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population.* Nat Genet, 2012. **44**(5): p. 511-516.

89. Padyukov, L., et al., *A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2011. **70**(2): p. 259-265.

90. Raychaudhuri, S., et al., *Common variants at CD40 and other loci confer risk of rheumatoid arthritis.* Nat Genet, 2008. **40**(10): p. 1216-1223.

91. Stahl, E.A., et al., *Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci.* Nat Genet, 2010. **42**(6): p. 508-514.

92. Zhernakova, A., et al., *Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci.* PLoS Genet, 2011. **7**(2): p. e1002004.

93. Cobb, J.E., et al., *Identification of the Tyrosine-Protein Phosphatase Non-Receptor Type 2 as a Rheumatoid Arthritis Susceptibility Locus in Europeans.* PLoS ONE, 2013. **8**(6): p. e66456.

94. Wiede, F., et al., *T cell protein tyrosine phosphatase attenuates T cell signaling to maintain tolerance in mice.* The Journal of Clinical Investigation, 2011. **121**(12): p. 4758-4774.

95. Robinson, J.I., *A fine mapping study of the FCGR locus in rheumatoid arthritis*, in *Institute of Molecule Medicine, School of Medicine*. 2011, University of Leeds: Leeds. p. 332.

96. Karlson, E.W., et al., *Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk.* Annals of the Rheumatic Diseases, 2010. **69**(6): p. 1077-1085.

97. Raychaudhuri, S., et al., *Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk.* Nat Genet, 2009. **41**(12): p. 1313-1318.

98. Barton, A., et al., *Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13.* Nat Genet, 2008. **40**(10): p. 1156-1159.

99. Eyre, S., et al., *Confirmation of association of the REL locus with rheumatoid arthritis susceptibility in the UK population.* Annals of the Rheumatic Diseases, 2010. **69**(8): p. 1572-1573.

100. Barton, A., et al., *Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes.* Human Molecular Genetics, 2009. **18**(13): p. 2518-2522.

101. Remmers, E., et al., *STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus.* N Engl J Med, 2007. **357**: p. 977 - 986.

102. Barton, A., et al., *Re-evaluation of putative rheumatoid arthritis susceptibility genes in the post-genome wide association study era and hypothesis of a key pathway underlying susceptibility.* Hum Mol Genet, 2008. **17**: p. 2274 - 2279.

103. Zhernakova, A., et al., *Novel Association in Chromosome 4q27 Region with Rheumatoid Arthritis and Confirmation of Type 1 Diabetes Point to a General Risk Locus for Autoimmune Diseases.* The American Journal of Human Genetics, 2007. **81**(6): p. 1284-1288.

104. Shimane, K., et al., *The association of a nonsynonymous single-nucleotide polymorphism in TNFAIP3 with systemic lupus erythematosus and rheumatoid arthritis in the Japanese population.* Arthritis & Rheumatism, 2010. **62**(2): p. 574-579.

105. Plenge, R., et al., *TRAF1-C5 as a risk locus for rheumatoid arthritis - a genomewide study.* N Engl J Med, 2007. **357**: p. 1199 - 1209.

106. Chang, M., et al., *A Large-Scale Rheumatoid Arthritis Genetic Study Identifies Association at Chromosome 9q33.2.* PLoS Genet, 2008. **4**(6): p. e1000107.

107. Orozco, G., et al., *Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study.* Annals of the Rheumatic Diseases, 2010. **69**(5): p. 813-816.

108. Hu, X., et al., *Integrating Autoimmune Risk Loci with Gene-Expression Data Identifies Specific Pathogenic Immune Cell Subsets.* The American Journal of Human Genetics, 2011. **89**(4): p. 496-506.

109. Emery, P., *Therapeutic Approaches for Early Rheumatoid Arthritis. How Early? How Aggressive?* Rheumatology, 1995. **XXXIV**(suppl 2): p. 87-90.

110. Emery, P. and M. Salmon, *Early rheumatoid arthritis: time to aim for remission?* Annals of the Rheumatic Diseases, 1995. **54**(12): p. 944-947.

111. van Nies, J.A.B., et al., *What is the evidence for the presence of a therapeutic window of opportunity in rheumatoid arthritis? A systematic literature review.* Annals of the Rheumatic Diseases, 2013: p. [Epub ahead of print].

112. Raza, K., et al., *Early rheumatoid arthritis is characterized by a distinct and transient synovial fluid cytokine profile of T cell and stromal cell origin.* Arthritis Research & Therapy, 2005. **7**(4): p. R784 - R795.

113. Cronstein, B.N., *Low-Dose Methotrexate: A Mainstay in the Treatment of Rheumatoid Arthritis.* Pharmacological Reviews, 2005. **57**(2): p. 163-172.

114. Baugh, C.M., C.L. Krumdieck, and M.G. Nair, *Polygammaglutamyl metabolites of methotrexate.* Biochemical and Biophysical Research Communications, 1973. **52**(1): p. 27-34.

115. Rosowsky, A., et al., *Biochemical and Biological Studies on 2-Desamino-2-methylaminopterin, an Antifolate the Polyglutamates of Which Are More Potent Than the Monoglutamate against Three Key Enzymes of Folate Metabolism.* Cancer Research, 1992. **52**(8): p. 2148-2155.

116. Zeng, H., et al., *Transport of Methotrexate (MTX) and Folates by Multidrug Resistance Protein (MRP) 3 and MRP1.* Cancer Research, 2001. **61**(19): p. 7225-7232.

117. Flescher, E., et al., *Increased polyamines may downregulate interleukin 2 production in rheumatoid arthritis.* The Journal of Clinical Investigation, 1989. **83**(4): p. 1356-1362.

118. Nesher, G. and T.L. Moore, *The in vitro effects of methotrexate on peripheral blood mononuclear cells: Modulation by methyl donors and spermidine.* Arthritis & Rheumatism, 1990. **33**(7): p. 954-959.

119. Smith, D.M., J.A. Johnson, and R.A. Turner, *Biochemical perturbations of BW 91Y(3-deazaadenosine) on human neutrophil chemotactic potential and lipid metabolism.* Int J Tissue React, 1991. **13**: p. 1-18.

120. Furst, D.E. and J.M. Kremer, *Methotrexate in rheumatoid arthritis.* Arthritis & Rheumatism, 1988. **31**(3): p. 305-314.

121. Allegra, C.J., et al., *Inhibition of phosphoribosylaminoimidazolecarboxamide transformylase by methotrexate and dihydrofolic acid polyglutamates.* Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(15): p. 4881-4885.

122. Cronstein, B.N., et al., *Methotrexate inhibits neutrophil function by stimulating adenosine release from connective tissue cells.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(6): p. 2441-2445.

123. Cronstein, B.N., D. Naime, and E. Ostad, *The antiinflammatory mechanism of methotrexate. Increased adenosine release at inflamed sites diminishes leukocyte accumulation in an in vivo model of inflammation.* The Journal of Clinical Investigation, 1993. **92**(6): p. 2675-2682.

124. Morgan, S.L., et al., *The effect of folic acid supplementation on the toxicity of low-dose methotrexate in patients with rheumatoid arthritis.* Arthritis & Rheumatism, 1990. **33**(1): p. 9-18.

125. Morgan, S.L., et al., *Supplementation with Folic Acid during Methotrexate Therapy for Rheumatoid Arthritis: A Double-Blind, Placebo-Controlled Trial.* Annals of Internal Medicine, 1994. **121**(11): p. 833-841.

126. McConkey B., A.R., Durham S., Forster PJG., Hubball S., Walsh L.,, *Sulphasalazine in rheumatoid arthritis.* Br Med J, 1980. **280**(6212): p. 442-444.

127. Houston JB., D.J., Walker J.,, *Azo reductino of sulphasalazine in healthy volunteers.* Br J Clin Pharmacol, 1982. **14**(3): p. 395-398.

128. Bird, H.A., *Sulphasalazine, Sulphapyridine or 5-Aminosalicylic Acid—which is the Active Moiety in Rheumatoid Arthritis?* Rheumatology, 1995. **XXXIV**(suppl 2): p. 16-19.

129. Yamasaki, Y., et al., *Pharmacogenetic Characterization of Sulfasalazine Disposition Based on NAT2 and ABCG2 (BCRP) Gene Polymorphisms in Humans.* Clin Pharmacol Ther, 2008. **84**(1): p. 95-103.

130. Baggott, J.E., Morgan, S.L., Ha, T., Vaughn, W.H., Hine, R.J., *Inhibition of folate-dependent enzymes by non-steroidal anti-inflammatory drugs.* Biochem J., 1992(282): p. 197-202.

131. Danis, V.A., et al., *Circulating cytokine levels in patients with rheumatoid arthritis: results of a double blind trial with sulphasalazine.* Annals of the Rheumatic Diseases, 1992. **51**(8): p. 946-950.

132. Wahl, C., et al., *Sulfasalazine: a potent and specific inhibitor of nuclear factor kappa B.* The Journal of Clinical Investigation, 1998. **101**(5): p. 1163-1174.

133. Hyrich, K.L., et al., *Predictors of response to anti-TNF-α therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register.* Rheumatology, 2006. **45**(12): p. 1558-1565.

134. Maini, R., et al., *Infliximab (chimeric anti-tumour necrosis factor [alpha] monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial.* The Lancet, 1999. **354**(9194): p. 1932-1939.

135. Nesbitt, A., et al., *Mechanism of action of certolizumab pegol (CDP870): In vitro comparison with other anti-tumor necrosis factor α agents.* Inflammatory Bowel Diseases, 2007. **13**(11): p. 1323-1332.

136. Van den Brande, J.M.H., et al., *Infliximab but not etanercept induces apoptosis in lamina propria T-lymphocytes from patients with Crohn's disease.* Gastroenterology, 2003. **124**(7): p. 1774-1785.

137. Gray, P.W., et al., *Cloning of human tumor necrosis factor (TNF) receptor cDNA and expression of recombinant soluble TNF-binding protein.* Proceedings of the National Academy of Sciences, 1990. **87**(19): p. 7380-7384.

138. Scallon, B., et al., *Binding and Functional Comparisons of Two Types of Tumor Necrosis Factor Antagonists.* Journal of Pharmacology and Experimental Therapeutics, 2002. **301**(2): p. 418-426.

139. Furst, D.E., et al., *Open-label, pilot protocol of patients with rheumatoid arthritis who switch to infliximab after an incomplete response to etanercept: the opposite study.* Annals of the Rheumatic Diseases, 2007. **66**(7): p. 893-899.

140. Roll, P., T. Dörner, and H.-P. Tony, *Anti-CD20 therapy in patients with rheumatoid arthritis: Predictors of response and B cell subset regeneration after repeated treatment.* Arthritis & Rheumatism, 2008. **58**(6): p. 1566-1575.

141. Dass, S., et al., *Highly sensitive B cell analysis predicts response to rituximab therapy in rheumatoid arthritis.* Arthritis & Rheumatism, 2008. **58**(10): p. 2993-2999.

142. Cambridge, G., et al., *Serologic changes following B lymphocyte depletion therapy for rheumatoid arthritis.* Arthritis & Rheumatism, 2003. **48**(8): p. 2146-2154.

143. Kormelink, T.G., et al., *Decrease in immunoglobulin free light chains in patients with rheumatoid arthritis upon rituximab (anti-CD20) treatment correlates with decrease in disease activity.* Annals of the Rheumatic Diseases, 2010. **69**(12): p. 2137-2144.

144. Nakou, M., et al., *Rituximab therapy reduces activated B cells in both the peripheral blood and bone marrow of patients with rheumatoid arthritis: depletion of memory B cells correlates with clinical response.* Arthritis Research & Therapy, 2009. **11**(4): p. R131.

145. van de Veerdonk, F.L., et al., *The anti-CD20 antibody rituximab reduces the Th17 cell response.* Arthritis & Rheumatism, 2011. **63**(6): p. 1507-1516.

146. Shan, D., J.A. Ledbetter, and O.W. Press, *Apoptosis of Malignant Human B Cells by Ligation of CD20 With Monoclonal Antibodies.* Blood, 1998. **91**(5): p. 1644-1652.

147. Szodoray, P., et al., *Apoptotic Effect of Rituximab on Peripheral Blood B Cells in Rheumatoid Arthritis.* Scandinavian Journal of Immunology, 2004. **60**(1-2): p. 209-218.

148. Bowles, J.A. and G.J. Weiner, *CD16 polymorphisms and NK activation induced by monoclonal antibody-coated target cells.* Journal of Immunological Methods, 2005. **304**(1–2): p. 88-99.

149. Van Der Kolk, L.E., et al., *Complement activation plays a key role in the side-effects of rituximab treatment.* British Journal of Haematology, 2001. **115**(4): p. 807-811.

150. Di Gaetano, N., et al., *Complement Activation Determines the Therapeutic Activity of Rituximab In Vivo.* The Journal of Immunology, 2003. **171**(3): p. 1581-1587.

151. Clayman, C., et al., *Toxicity of primaquine in Caucasians.* J. Am. Med. Assoc., 1952. **149**(17): p. 1563-1568.

152. Hughes, H.B., *On the metabolic fate of isoniazid.* Journal of Pharmacology and Experimental Therapeutics, 1953. **109**(4): p. 444-452.

153. Hughes, H., et al., *Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis.* Am. Rev. Tuberc., 1954. **70**(2): p. 266-273.

154. Lehmann, H. and E. Ryan, *The familial incidence of low pseudocholinesterase level.* Lancet, 1956. **271**(6934): p. 124.

155. Lockridge, O., *Genetic variants of human serum cholinesterase influence metabolism of the muscle relaxant succinylcholine.* Pharmacology & Therapeutics, 1990. **47**(1): p. 35-60.

156. Yang, J., et al., *GCTA: A Tool for Genome-wide Complex Trait Analysis.* The American Journal of Human Genetics, 2011. **88**(1): p. 76-82.

157. McGeachie, M.J., Stahl, E.A., Himes, B.E., Pendergrass, S.A., Lima, J.J., Irvin, C.G., Peters, S.P., Ritchie, M.D., Plenge, R.M., Tantisira, K.G., *Polygenic heritability estimates in pharmacogenetics: focus on asthma and related phenotypes.* Pharmacogenetics and Genomics, 2013. **23**(6): p. 324-328.

158. Samer, C.F., et al., *Genetic polymorphisms and drug interactions modulating CYP2D6 and CYP3A activities have a major effect on oxycodone analgesic efficacy and safety.* British Journal of Pharmacology, 2010. **160**(4): p. 919-930.

159. Takahashi, H., et al., *Metabolism of warfarin enantiomers in Japanese patients with heart disease having different CYP2C9 and CYP2C19 genotypes.* Clin Pharmacol Ther, 1998. **63**(5): p. 519-528.

160. Schroth, W., et al., *Association Between CYP2D6 Polymorphisms and Outcomes Among Women With Early Stage Breast Cancer Treated With Tamoxifen.* JAMA: The Journal of the American Medical Association, 2009. **302**(13): p. 1429-1436.

161. Pasanen, M.K., et al., *SLCO1B1 polymorphism markedly affects the pharmacokinetics of simvastatin acid.* Pharmacogenetics and Genomics, 2006. **16**(12): p. 873-879.

162. Hampras. S. S., et al., *Genetic polymorphisms of ATP-binding cassette (ABC) proteins, overall survival and drug toxicity in patients with Acute Myeloid Leukemia.* Int J Mol Epidemiol Genet, 2010. **1**(3): p. 201-207.

163. Ross, J.S., et al., *Biomarker-Based Prediction of Response to Therapy for Colorectal Cancer.* American Journal of Clinical Pathology, 2010. **134**(3): p. 478-490.

164. Ross, J.S., et al., *Commercialized Multigene Predictors of Clinical Outcome for Breast Cancer.* Oncologist, 2008. **13**(5): p. 477-493.

165. D'Andrea, G., et al., *A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin.* Blood, 2005. **105**(2): p. 645-649.

166. Aithal, G.P., et al., *Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications.* The Lancet, 1999. **353**(9154): p. 717-719.

167. Lettre, G., C. Lange, and J.N. Hirschhorn, *Genetic model testing and statistical power in population-based association studies of quantitative traits.* Genetic Epidemiology, 2007. **31**(4): p. 358-362.

168. Ge, D., et al., *Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance.* Nature, 2009. **461**(7262): p. 399-401.

169. Conaghan, P.G., et al., *Persistently moderate DAS-28 is not benign: loss of function occurs in early RA despite step-up DMARD therapy.* Rheumatology, 2010. **49**(10): p. 1894-1899.

170. Ulrich, C.M., et al., *Pharmacogenetics of methotrexate: toxicity among marrow transplantation patients varies with the methylenetetrahydrofolate reductase C677T polymorphism.* Blood, 2001. **98**(1): p. 231-234.

171. Frosst, P., et al., *A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase.* Nat Genet, 1995. **10**(1): p. 111-113.

172. van der Put, N.M.J., et al., *A Second Common Mutation in the Methylenetetrahydrofolate Reductase Gene: An Additional Risk Factor for Neural-Tube Defects?* The American Journal of Human Genetics, 1998. **62**(5): p. 1044-1051.

173. Berkun, Y., et al., *A1298C polymorphism of the methylenetetrahydrofolate reductase gene is associated with methotrexate related side effects in patients with rheumatoid arthritis.* Arthritis and Rheumatism, 2003. **48**(9): p. 224.

174. Berkun, Y., et al., *Methotrexate related adverse effects in patients with rheumatoid arthritis are associated with the A1298C polymorphism of the MTHFR gene.* Annals of the Rheumatic Diseases, 2004. **63**(10): p. 1227-1231.

175. Bohanec Grabar, P., et al., *Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism.* European Journal of Clinical Pharmacology, 2008. **64**(11): p. 1057-1068.

176. Dervieux, T., et al., *Gene-gene interactions in folate and adenosine biosynthesis pathways affect methotrexate efficacy and tolerability in rheumatoid arthritis.* Pharmacogenetics and Genomics, 2009. **19**(12): p. 935-944

177. Taniguchi, A., et al., *Validation of the associations between single nucleotide polymorphisms or haplotypes and responses to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a proposal for prospective*

pharmacogenomic study in clinical practice. Pharmacogenetics and Genomics, 2007. **17**(6): p. 383-390.

178. Urano, W., et al., *Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses.* Pharmacogenetics, 2002. **12**(3): p. 183-190.
179. Weisman, M.H., et al., *Risk genotypes in folate-dependent enzymes and their association with methotrexate-related side effects in rheumatoid arthritis.* Arthritis & Rheumatism, 2006. **54**(2): p. 607-612.
180. Wessels, J.A.M., et al., *Efficacy and toxicity of methotrexate in early rheumatoid arthritis are associated with single-nucleotide polymorphisms in genes coding for folate pathway enzymes.* Arthritis and Rheumatism, 2006. **54**(4): p. 1087-1095.
181. Aggarwal, P., et al., *Correlation between methotrexate efficacy & toxicity with C677T polymorphism of the methylenetetrahydrofolate gene in rheumatoid arthritis patients on folate supplementation.* Indian Journal of Medical Research, 2006. **124**(5): p. 521-526.
182. Ghodke, Y., et al., *Are Thymidylate synthase and Methylene tetrahydrofolate reductase genes linked with methotrexate response (efficacy, toxicity) in Indian (Asian) rheumatoid arthritis patients?* Clinical Rheumatology, 2008. **27**(6): p. 787-789.
183. Lee, Y.H. and G.G. Song, *Associations between the C677T and A1298C Polymorphisms of MTHFR and the Efficacy and Toxicity of Methotrexate in Rheumatoid Arthritis A Meta-Analysis.* Clinical Drug Investigation, 2010. **30**(2): p. 101-108.
184. Ranganathan, P., et al., *Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis.* Journal of Rheumatology, 2008. **35**(4): p. 572-579.
185. Stamp, L.K., et al., *Polymorphisms within the folate pathway predict folate concentrations but are not associated with disease activity in rheumatoid arthritis patients on methotrexate.* Pharmacogenetics and Genomics, 2010. **20**(6): p. 367-376.
186. Zeng, Q.Y., et al., *Pharmacogenetic study of 5,10-methylenetetrahydrofolate reductase C677T and thymidylate synthase 3R/2R gene polymorphisms and*

methotrexate-related toxicity in Chinese Han patients with inflammatory arthritis. Annals of the Rheumatic Diseases, 2008. **67**(8): p. 1193-1194.

187. Lee, Y.C., et al., *Investigation of candidate polymorphisms and disease activity in rheumatoid arthritis patients on methotrexate.* Rheumatology, 2009. **48**(6): p. 613-617.
188. Sharma, S., et al., *Purine biosynthetic pathway genes and methotrexate response in rheumatoid arthritis patients among north Indians.* Pharmacogenetics and Genomics, 2009. **19**(10): p. 823-828.
189. Owen, S.A., et al., *Genetic polymorphisms in key methotrexate pathway genes are associated with response to treatment in rheumatoid arthritis patients.* Pharmacogenomics J, 2013. **13**(3): p. 227-234.
190. Dervieux, T., et al., *Polyglutamation of methotrexate with common polymorphisms in reduced folate carrier, aminoimidazole carboxamide ribonucleotide transformylase, and thymidylate synthase are associated with methotrexate effects in rheumatoid arthritis.* Arthritis and Rheumatism, 2004. **50**(9): p. 2766-2774.
191. Takatori, R., et al., *ABCB1 C3435T polymorphism influences the methotrexate sensitivity in rheumatoid arthritis patients.* Arthritis and Rheumatism, 2005. **52**(9): p. 570.
192. Drozdzik, M., et al., *The effects of C3435T MDR-1 gene polymorphism of methotrexate (MTX) treatment outcome in patients with rheumatoid arthritis.* Acta Pharmacologica Sinica, 2006. **27**: p. 228-228.
193. Sharma, S., et al., *Interaction of genes from influx-metabolism-efflux pathway and their influence on methotrexate efficacy in rheumatoid arthritis patients among Indians.* Pharmacogenetics and Genomics, 2008. **18**(12): p. 1041-1049.
194. Takatori, R., et al., *ABCB1 C3435T polymorphism influences methotrexate sensitivity in rheumatoid arthritis patients.* Clinical and Experimental Rheumatology, 2006. **24**(5): p. 546-554.
195. Hider, S.L., et al., *Polymorphisms within the adenosine receptor 2a gene are associated with adverse events in RA patients treated with MTX.* Rheumatology, 2008. **47**(8): p. 1156-1159.
196. Dervieux, T., N. Greenstein, and J. Kremer, *Pharmacogenomic and metabolic biomarkers in the folate pathway and their association with methotrexate effects during dosage escalation in rheumatoid arthritis.* Arthritis and Rheumatism, 2006. **54**(10): p. 3095-3103.

197. Wessels, J.A.M., et al., *A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis.* Arthritis and Rheumatism, 2007. **56**(6): p. 1765-1775.

198. Owen, S.A., et al., *Testing pharmacogenetic indices to predict efficacy and toxicity of methotrexate monotherapy in a rheumatoid arthritis patient cohort.* Arthritis & Rheumatism, 2010. **62**(12): p. 3827-3829.

199. Jorgensen, A.L. and P.R. Williamson, *Methodological quality of pharmacogenetic studies: Issues of concern.* Statistics in Medicine, 2008. **27**(30): p. 6547-6569.

200. Bohanec Grabar, P., et al., *Genetic determinants of methotrexate treatment in rheumatoid arthritis patients: a study of polymorphisms in the adenosine pathway.* Annals of the Rheumatic Diseases, 2010. **69**(5): p. 931-2.

201. Wessels, J.A.M., et al., *Relationship between genetic variants in the adenosine pathway and outcome of methotrexate treatment in patients with recent-onset rheumatoid arthritis.* Arthritis and Rheumatism, 2006. **54**(9): p. 2830-2839.

202. Drozdzik, M., et al., *The effect of 3435C > T MDR1 gene polymorphism on rheumatoid arthritis treatment with disease-modifying antirheumatic drugs.* European Journal of Clinical Pharmacology, 2006. **62**(11): p. 933-937.

203. Drozdzik, M., et al., *Reduced folate carrier-1 80G > A polymorphism affects methotrexate treatment outcome in rheumatoid arthritis.* Pharmacogenomics Journal, 2007. **7**: p. 404-407.

204. Pullar, T., J.A. Hunter, and H.A. Capell, *Effect of acetylator phenotype on efficacy and toxicity of sulphasalazine in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 1985. **44**(12): p. 831-837.

205. Bax, D.E., M.S. Greaves, and R.S. Amos, *Sulphasalazine for Rheumatoid Arthritis: Relationship Between Dose, Acetylator Phenotype and Response to Treatment.* Rheumatology, 1986. **25**(3): p. 282-284.

206. Tanaka, E., et al., *Adverse effects of sulfasalazine in patients with rheumatoid arthritis are associated with diplotype configuration at the N-acetyltransferase 2 gene.* The Journal of Rheumatology, 2002. **29**(12): p. 2492-2499.

207. Kumagai, S., et al., *N-Acetyltransferase 2 Genotype-Related Efficacy of Sulfasalazine in Patients with Rheumatoid Arthritis.* Pharmaceutical Research, 2004. **21**(2): p. 324-329.

208. Soejima, M., et al., *Prospective study of the association between NAT2 gene haplotypes and severe adverse events with sulfasalazine therapy in patients with rheumatoid arthritis.* The Journal of Rheumatology, 2008. **35**(4): p. 724.

209. Urquhart, B.L., et al., *Breast cancer resistance protein (ABCG2) and drug disposition: intestinal expression, polymorphisms and sulfasalazine as an in vivo probe.* Pharmacogenetics and Genomics, 2008. **18**(5): p. 439-448

210. James, H.M., et al., *Common polymorphisms in the folate pathway predict efficacy of combination regimens containing methotrexate and sulfasalazine in early rheumatoid arthritis.* Journal of Rheumatology, 2008. **35**(4): p. 562-571.

211. James, H.M., et al., *Common polymorphisms in the folate pathway predict efficacy of combination regimens containing methotrexate and sulfasalazine in early rheumatoid arthritis.* The Journal of Rheumatology, 2008. **35**(4): p. 562-571.

212. Pawlik, A., et al., *The effect of 677C>T and 1298A>C MTHFR polymorphisms on sulfasalazine treatment outcome in rheumatoid arthritis.* Brazilian Journal of Medical and Biological Research, 2009. **42**: p. 660-664.

213. Kuuliala, K., et al., *Polymorphism at position +896 of the toll-like receptor 4 gene interferes with rapid response to treatment in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2006. **65**(9): p. 1241-1243.

214. Laivoranta-Nyman, S., Mottonen, T., Hannonen, P., Korpela, M., Kautiainen, H., Leirisalo-Rep, M., Julkunen, H., Luukkainen, R., Hakala, M., Vuori, K., Laine, A.P., Toivanen, A., Ilonen, J., FIN-RACo Trial Group.,, *Association of tumour necrosis factor a, b and c microsatellite polymorphisms with clinical disease activity and induction of remission in early rheumatoid arthritis.* Clin Exp Rheum, 2006. **24**(6): p. 636-642.

215. Wilson, A.G., et al., *Effects of a polymorphism in the human tumor necrosis factor α promoter on transcriptional activation.* Proceedings of the National Academy of Sciences, 1997. **94**(7): p. 3195-3199.

216. Seitz, M., et al., *The −308 tumour necrosis factor-α gene polymorphism predicts therapeutic response to TNFα-blockers in rheumatoid arthritis and spondyloarthritis patients.* Rheumatology, 2007. **46**(1): p. 93-96.

217. Lee, Y., et al., *Association of TNF-alpha –308 G/A polymorphism with responsiveness to TNF-α-blockers in rheumatoid arthritis: a meta-analysis.* Rheumatology International, 2006. **27**(2): p. 157-161.

218. Lee, Y.H., et al., *Associations Between Tumor Necrosis Factor-α (TNF-α) −308 and −238 G/A Polymorphisms and Shared Epitope Status and Responsiveness to TNF-α Blockers in Rheumatoid Arthritis: A Metaanalysis Update.* The Journal of Rheumatology, 2010. **37**(4): p. 740-746.

219. Pavy, S., et al., *Tumour necrosis factor α −308G→A polymorphism is not associated with response to TNFα blockers in Caucasian patients with rheumatoid arthritis: systematic review and meta-analysis.* Annals of the Rheumatic Diseases, 2010. **69**(6): p. 1022-1028.

220. Maxwell, J.R., et al., *Association of the tumour necrosis factor-308 variant with differential response to anti-TNF agents in the treatment of rheumatoid arthritis.* Human Molecular Genetics, 2008. **17**(22): p. 3532-3538.

221. O'Rielly, D.D., et al., *TNF-α -308 G/A polymorphism and responsiveness to TNF-α blockade therapy in moderate to severe rheumatoid arthritis: a systematic review and meta-analysis.* Pharmacogenomics J, 2009. **9**(3): p. 161-167.

222. Coulthard, L.R., et al., *Genetic variants within the MAP kinase signalling network and anti-TNF treatment response in rheumatoid arthritis patients.* Annals of the Rheumatic Diseases, 2011. **70**(1): p. 98-103.

223. Potter, C., et al., *Polymorphisms spanning the TNFR2 and TACE genes do not contribute towards variable anti-TNF treatment response.* Pharmacogenetics and Genomics, 2010. **20**(5): p. 338-341

224. Toonen, E.J.M., et al., *The tumour necrosis factor receptor superfamily member 1b 676T>G polymorphism in relation to response to infliximab and adalimumab treatment and disease severity in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2008. **67**(8): p. 1174-1177.

225. Mathews, R.J., et al., *Evidence of NLRP3-inflammasome activation in rheumatoid arthritis (RA); genetic variants within the NLRP3-inflammasome complex in relation to susceptibility to RA and response to anti-TNF treatment.* Annals of the Rheumatic Diseases, 2013: p. [Epub ahead of print].

226. Cui, J., et al., *Rheumatoid arthritis risk allele PTPRC is also associated with response to anti–tumor necrosis factor α therapy.* Arthritis & Rheumatism, 2010. **62**(7): p. 1849-1861.

227. Plant, D., et al., *Replication of association of the PTPRC gene with response to anti–tumor necrosis factor therapy in a large UK cohort.* Arthritis & Rheumatism, 2012. **64**(3): p. 665-670.

228. Potter, C., et al., *Association of rheumatoid factor and anti-cyclic citrullinated peptide positivity, but not carriage of shared epitope or PTPN22 susceptibility variants, with anti-tumour necrosis factor response in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2009. **68**(1): p. 69-74.

229. Potter, C., et al., *Association between anti-tumour necrosis factor treatment response and genetic variants within the TLR and NFκB signalling pathways.* Annals of the Rheumatic Diseases, 2010. **69**(7): p. 1315-1320.

230. Coenen, M.J.H., et al., *Genetic Variants in Toll-Like Receptors Are Not Associated with Rheumatoid Arthritis Susceptibility or Anti-Tumour Necrosis Factor Treatment Outcome.* PLoS ONE, 2010. **5**(12): p. e14326.

231. Liu, C., et al., *Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis.* Mol Med, 2008. **14**(9-10): p. 575-581.

232. Umićević Mirkov, M., et al., *Genome-wide association analysis of anti-TNF drug response in patients with rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2013. **72**(8): p. 1375-1381.

233. Cui, J., et al., *Genome-Wide Association Study and Gene Expression Analysis Identifies CD84 as a Predictor of Response to Etanercept Therapy in Rheumatoid Arthritis.* PLoS Genet, 2013. **9**(3): p. e1003394.

234. Fabris, M., et al., *The TTTT B lymphocyte stimulator promoter haplotype is associated with good response to rituximab therapy in seropositive rheumatoid arthritis resistant to tumor necrosis factor blockers.* Arthritis & Rheumatism, 2013. **65**(1): p. 88-97.

235. Ruyssen-Witrand, A., et al., *Association between -871C>T promoter polymorphism in the B-cell activating factor gene and the response to rituximab in rheumatoid arthritis patients.* Rheumatology, 2013. **52**(4): p. 636-641.

236. Fabris, M., et al., *The CC homozygosis of the - 174G>C IL-6 polymorphism predicts a lower efficacy of rituximab therapy in rheumatoid arthritis.* Autoimmunity Reviews, 2012. **11**(5): p. 315-320.

237. Ruyssen-Witrand, A., et al., *Fcγ receptor type IIIA polymorphism influences treatment outcomes in patients with rheumatoid arthritis treated with rituximab.* Annals of the Rheumatic Diseases, 2012. **71**(6): p. 875-877.

238. Quartuccio, L., et al., *The 158VV Fcgamma receptor 3A genotype is associated with response to rituximab in rheumatoid arthritis: results of an Italian multicentre study.* Annals of the Rheumatic Diseases, 2013: p. [Epub ahead of print].

239. Robledo, G., Marquez, A., Davila-Fajardo, C.L., Ortego-Centeno, N., Rubio, J.L., Garrido Ede, R., Sanchez-Roman, J., Garcia-Hernandez, F.J., Rios-Fernandez, R., Gonzalez-Escribano, M.F., Garcia, M.T., Palma, M.J., Avala Mdel, M., Martin, J.,, *Association of the FCGR3A-158F/V gene polymorphism wih the response to rituximab treatment in Spanish systemic autoimmune disease patients.* DNA Cell Biol, 2012. **31**(12): p. 1671-1677.

240. Morgan, A.W., et al., *FcγRIIIA-158V and rheumatoid arthritis: a confirmation study.* Rheumatology, 2003. **42**(4): p. 528-533.

241. Robinson, J.I., et al., *Dissection of the FCGR3A association with RA: increased association in men and with autoantibody positive disease.* Annals of the Rheumatic Diseases, 2010. **69**(6): p. 1054-1057.

242. Cartron, G., et al., *Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor Fcgamma RIIIa gene.* Blood, 2002. **99**(3): p. 754-758.

243. Kastbom, A., et al., *Influence of FCGR3A genotype on the therapeutic response to rituximab in rheumatoid arthritis: an observational cohort study.* BMJ Open, 2012. **2**(5): p. e001524.

244. Robledo, G., Davila-Fajardo, C.L., Marquez, A., Ortego-Centeno, N., Rubio, J.L.C., de Ramon Garrido de, E., Sanchez-Roman, J., Garcia-Hernandez, F.J., Rios-Fernandez, R., Gonzalez-Escribano, M.F., Garcia, M.T.C., Palma, M.J.C., del Mar Ayala, M., Martin, J.,, *Association between -174 Interleukin-6 gene polymorphism and biological response to rituximab in several systemic autoimmune diseases.* DNA Cell Biol, 2012. **31**(9): p. 1486-1491.

245. Felson, D.T., et al., *American college of rheumatology preliminary definition of improvement in rheumatoid arthritis.* Arthritis & Rheumatism, 1995. **38**(6): p. 727-735.

246. van der Heijde, D.M., et al., *Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score.* Annals of the Rheumatic Diseases, 1990. **49**(11): p. 916-920.

247. van der Heijde, D.M., et al., *Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 1992. **51**(2): p. 177-181.

248. Spiegel, T.M., et al., *Measuring disease activity: comparison of joint tenderness, swelling, and ultrasonography in rheumatoid arthritis.* Arthritis & Rheumatism, 1987. **30**(11): p. 1283-1288.

249. Smolen, J.S., et al., *A simplified disease activity index for rheumatoid arthritis for use in clinical practice.* Rheumatology, 2003. **42**(2): p. 244-257.

250. van Gestel, A.M., et al., *Development and validation of the european league against rheumatism response criteria for rheumatoid arthritis: Comparison with the preliminary american college of rheumatology and the world health organization/international league against rheumatism criteria.* Arthritis & Rheumatism, 1996. **39**(1): p. 34-40.

251. Larsen, A., *How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies.* J Rheumatol, 1995. **22**(10): p. 1974-1975.

252. Sharp, J.T., et al., *Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities.* Arthritis & Rheumatism, 1971. **14**(6): p. 706-720.

253. Sharp, J.T., et al., *How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis?* Arthritis & Rheumatism, 1985. **28**(12): p. 1326-1335.

254. Kaye, J.J., Callahan, L.F., Nance, E.P. Jr., Brooks, R., Pincus, T., *Bony akylosis in rheumatoid arthritis. Associations with longer duration and greater severity of disease.* Invest Radiol, 1987. **22**(4): p. 303-309.

255. Van Der Heijde, D., et al., *Effects of Hydroxychloroquine and Sulphasalazine on Progression of Joint Damage in Rheumatoid Arthritis.* The Lancet, 1989. **333**(8646): p. 1036-1038.

256. Larsen, A., Dale, K., Eek, M., *Radiographic evalutation of rheumatoid arthritis and related conditions by standard reference films.* Acta Radiol Diagn (Stockh), 1977. **18**(4): p. 481-491.

257. Bruynesteyn, K., et al., *The Sharp/van der Heijde method out-performed the Larsen/Scott method on the individual patient level in assessing radiographs in early rheumatoid arthritis.* Journal of Clinical Epidemiology, 2004. **57**(5): p. 502-512.

258. Morgan, A., et al., *Re-evaluation of the interaction between HLA-DRB1 SE alleles, PTPN22 and smoking in determining the susceptibility to autoantibody positive and negative rheumatoid arthritis in a large UK Caucasian population.* Arthritis Rheum, 2009. **60**: p. 2565 - 2576.

259. Fransen, J. and P.L.C.M. Van Riel, *The Disease Activity Score and the EULAR response criteria.* Clin Exp Rheum, 2005. **23**(Suppl 39): p. S93-S99.

260. Morgan, A., et al., *Evaluation of the rheumatoid arthritis susceptibility loci HLA-DRB1, PTPN22, OLIG3/TNFAIP3, STAT4 and TRAF1/C5 in an inception cohort.* Arthritis Research & Therapy, 2010. **12**(2): p. R57.

261. Young, A., *Short-Term Outcomes in Recent-Onset Rheumatoid Arthritis.* Rheumatology, 1995. **XXXIV**(suppl 2): p. 79-86.

262. Jones, G., et al., *Comparison of tocilizumab monotherapy versus methotrexate monotherapy in patients with moderate to severe rheumatoid arthritis: the AMBITION study.* Annals of the Rheumatic Diseases, 2010. **69**(01): p. 88-96.

263. Tak, P.P., et al., *Inhibition of joint damage and improved clinical outcomes with rituximab plus methotrexate in early active rheumatoid arthritis: the IMAGE trial.* Annals of the Rheumatic Diseases, 2011. **70**(1): p. 39-46.

264. Vuong, Q.H., *Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses.* Econometrica, 1989. **57**(2): p. 307-333.

265. Zou, G., *A Modified Poisson Regression Approach to Prospective Studies with Binary Data.* American Journal of Epidemiology, 2004. **159**(7): p. 702-706.

266. Purcell, S., et al., *PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.* The American Journal of Human Genetics, 2007. **81**(3): p. 559-575.

267. Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results.* Bioinformatics, 2010. **26**(18): p. 2336-2337.

268. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-265.

269. Howie, B.N., P. Donnelly, and J. Marchini, *A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies.* PLoS Genet, 2009. **5**(6): p. e1000529.

270. van der Linden, M.P.M., et al., *Association of a single-nucleotide polymorphism in CD40 with the rate of joint destruction in rheumatoid arthritis.* Arthritis & Rheumatism, 2009. **60**(8): p. 2242-2247.

271. Orozco, G., et al., *Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus.* Annals of the Rheumatic Diseases, 2011. **70**(3): p. 463-468.

272. WTCCC, *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-678.

273. Hardy, G.H., *Mendelian Proportions in a Mixed Population.* Science, 1908. **28**(706): p. 49-50.

274. Higgins, J.P.T. and S.G. Thompson, *Quantifying heterogeneity in a meta-analysis.* Statistics in Medicine, 2002. **21**(11): p. 1539-1558.

275. Teare, M.D., et al., *Allele dose association of the C5orf30 rs26232 variant with joint damage in rheumatoid arthritis.* Arthritis & Rheumatism, 2013. **65**(10): p. 2555-61.

276. Grundberg, E., et al., *Mapping cis- and trans-regulatory effects across multiple tissues in twins.* Nat Genet, 2012. **44**(10): p. 1084-1089.

277. Fairfax, B.P., et al., *Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles.* Nat Genet, 2012. **44**(5): p. 502-510.

278. *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

279. *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

280. Rosenbloom, K.R., et al., *ENCODE Data in the UCSC Genome Browser: year 5 update.* Nucleic Acids Research, 2013. **41**(D1): p. D56-D63.

281. Schaub, M.A., et al., *Linking disease associations with regulatory information in the human genome.* Genome Research, 2012. **22**(9): p. 1748-1759.

282. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Research, 2012. **22**(9): p. 1790-1797.

283. de Vries-Bouwstra, J.K., et al., *Progression of joint damage in early rheumatoid arthritis: Association with HLA–DRB1, rheumatoid factor, and anti–citrullinated protein antibodies in relation to different treatment strategies.* Arthritis & Rheumatism, 2008. **58**(5): p. 1293-1298.

284. Suwannalai, P., et al., *Low-avidity anticitrullinated protein antibodies (ACPA) are associated with a higher rate of joint destruction in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2013.

285. Salaffi, F., et al., *Relationship between time-integrated disease activity estimated by DAS28-CRP and radiographic progression of anatomical damage in patients with early rheumatoid arthritis.* BMC Musculoskelet Disord, 2011. **12**: p. 120.

286. Plant, M.J., et al., *Relationship between time-integrated C-reactive protein levels and radiologic progression in patients with rheumatoid arthritis.* Arthritis & Rheumatism, 2000. **43**(7): p. 1473-1477.

287. Knevel, R., et al., *Evaluation of the contribution of cumulative levels of inflammation to the variance in joint destruction in rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2013. **72**(2): p. 307-308.

288. Ge, B., et al., *Global patterns of cis variation in human cells revealed by high-density allelic expression analysis.* Nat Genet, 2009. **41**(11): p. 1216-1222.

289. Kasowski, M., et al., *Variation in Transcription Factor Binding Among Humans.* Science, 2010. **328**(5975): p. 232-235.

290. Stranger, B.E., et al., *Population genomics of human gene expression.* Nat Genet, 2007. **39**(10): p. 1217-1224.

291. Zeller, T., et al., *Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility.* PLoS ONE, 2010. **5**(5): p. e10693.

292. Prevoo, M.L.L., et al., *Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis.* Arthritis & Rheumatism, 1995. **38**(1): p. 44-48.

293. Suzuki, T., et al., *PADI4 and HLA-DRB1 Are Genetic Risks for Radiographic Progression in RA Patients, Independent of ACPA Status: Results from the IORRA Cohort Study.* PLoS ONE, 2013. **8**(4): p. e61045.

294. Viatte, S., et al., *Investigation of Rheumatoid Arthritis Genetic Susceptibility Markers in the Early Rheumatoid Arthritis Study Further Replicates the TRAF1 Association with Radiological Damage.* The Journal of Rheumatology, 2013. **40**(2): p. 144-156.

295. Martinez, A., et al., *Association of the major histocompatibility complex with response to infliximab therapy in rheumatoid arthritis patients.* Arthritis & Rheumatism, 2004. **50**(4): p. 1077-1082.

296. O'Dell, J.R., et al., *HLA-DRB1 typing in rheumatoid arthritis: predicting response to specific treatments.* Annals of the Rheumatic Diseases, 1998. **57**(4): p. 209-213.

297. Criswell, L.A., et al., *The influence of genetic variation in the HLA–DRB1 and LTA–TNF regions on the response to treatment of early rheumatoid arthritis with methotrexate or etanercept.* Arthritis & Rheumatism, 2004. **50**(9): p. 2750-2756.

298. Knevel, R., et al., *Genetic predisposition of the severity of joint destruction in rheumatoid arthritis: a population-based study.* Annals of the Rheumatic Diseases, 2012. **71**(5): p. 707-709.

299. Nica, A.C., et al., *The Architecture of Gene Regulatory Variation across Multiple Human Tissues: The MuTHER Study.* PLoS Genet, 2011. **7**(2): p. e1002003.

300. Stranger, B.E., et al., *Patterns of cis Regulatory Variation in Diverse Human Populations.* PLoS Genet, 2012. **8**(4): p. e1002639.

301. Simpfendorfer, K.R., et al., *The autoimmunity-associated BLK haplotype exhibits cis-regulatory effects on mRNA and protein expression that are prominently observed in B cells early in development.* Human Molecular Genetics, 2012. **21**(17): p. 3918-3925.

302. Knevel, R., et al., *Genetic variants in IL15 associate with progression of joint destruction in rheumatoid arthritis: a multicohort study.* Annals of the Rheumatic Diseases, 2012. **71**(10): p. 1651-1657.

303. Krajinovic, M., et al., *Role of polymorphisms in MTHFR and MTHFD1 genes in the outcome of childhood acute lymphoblastic leukemia.* Pharmacogenomics J, 2003. **4**(1): p. 66-72.

304. Fisher, M.C. and B.N. Cronstein, *Metaanalysis of Methylenetetrahydrofolate Reductase (MTHFR) Polymorphisms Affecting Methotrexate Toxicity.* The Journal of Rheumatology, 2009. **36**(3): p. 539-545.

305. Owen, S.A., et al., *MTHFR gene polymorphisms and outcome of methotrexate treatment in patients with rheumatoid arthritis: analysis of key polymorphisms and meta-analysis of C677T and A1298C polymorphisms.* Pharmacogenomics J, 2011.

306. Day, I.N.M. and S.E. Humphries, *Electrophoresis for Genotyping: Microtiter Array Diagonal Gel Electrophoresis on Horizontal Polyacrylamide Gels, Hydrolink, or Agarose.* Analytical Biochemistry, 1994. **222**(2): p. 389-395.

307. Xiao, H., et al., *Associations between the genetic polymorphisms of MTHFR and outcomes of methotrexate treatment in rheumatoid arthritis.* Clinical and Experimental Rheumatology, 2010. **28**(5): p. 728-733.

308. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials.* Controlled Clinical Trials, 1986. **7**(3): p. 177-188.

309. Mikkelsen T.S., T.C.F., Yang J.J., Ulrich C.M., French D., Zaza G., Dunnenberger H.M., Marsh S., McLeod H., Giacomini K., Becker M.L., Gaedigk R., Leeder J.S., Kager .L., Relling M.V., Evans W., Klein T.E., Altman R.B.,, *PharmGKB summary: methotrexate pathway.* Pharmacogenetics and Genomics, 2011. **21**(10): p. 679-686.

310. Giusti, B., et al., *Role of C677T and A1298C MTHFR, A2756G MTR and -786 C/T eNOS Gene Polymorphisms in Atrial Fibrillation Susceptibility.* PLoS ONE, 2007. **2**(6): p. e495.

311. Toffoli, G., et al., *Effect of methylenetetrahydrofolate reductase 677C→T polymorphism on toxicity and homocysteine plasma level after chronic methotrexate treatment of ovarian cancer patients.* International Journal of Cancer, 2003. **103**(3): p. 294-299.

312. Aplenc, R., et al., *Methylenetetrahydrofolate Reductase Polymorphisms and Therapy Response in Pediatric Acute Lymphoblastic Leukemia.* Cancer Research, 2005. **65**(6): p. 2482-2487.

313. Grarup, N., et al., *Genetic Architecture of Vitamin B_{12} and Folate Levels Uncovered Applying Deeply Sequenced Large Datasets.* PLoS Genet, 2013. **9**(6): p. e1003530.

314. Lee, J., et al., *Plasma folate, methylenetetrahydrofolate reductase (MTHFR), and colorectal cancer risk in three large nested case–control studies.* Cancer Causes & Control, 2012. **23**(4): p. 537-545.

315. Levine, A.J., et al., *Genetic Variability in the MTHFR Gene and Colorectal Cancer Risk Using the Colorectal Cancer Family Registry.* Cancer Epidemiology Biomarkers & Prevention, 2010. **19**(1): p. 89-100.

316. Stamp, L.K., et al., *Expression of Methotrexate Transporters and Metabolizing Enzymes in Rheumatoid Synovial Tissue.* The Journal of Rheumatology, 2013. **40**(9): p. 1519-1522.

317. Hirata, S., et al., *A multi-biomarker score measures rheumatoid arthritis disease activity in the BeSt study.* Rheumatology, 2013. **52**(7): p. 1202-1207.

318. Curtis, J.R., et al., *Validation of a novel multibiomarker test to assess rheumatoid arthritis disease activity.* Arthritis Care & Research, 2012. **64**(12): p. 1794-1803.

319. Centola, M., et al., *Development of a Multi-Biomarker Disease Activity Test for Rheumatoid Arthritis.* PLoS ONE, 2013. **8**(4): p. e60635.

320. Bruhns, P., et al., *Specificity and affinity of human Fcγ receptors and their polymorphic variants for human IgG subclasses.* Blood, 2009. **113**(16): p. 3716-3725.

321. Niederer, H.A., et al., *FcγRIIB, FcγRIIIB, and systemic lupus erythematosus.* Annals of the New York Academy of Sciences, 2010. **1183**(1): p. 69-88.

322. Debets, J.M., et al., *Cross-linking of both Fc gamma RI and Fc gamma RII induces secretion of tumor necrosis factor by human monocytes, requiring high affinity Fc-Fc gamma R interactions. Functional activation of Fc gamma RII by treatment with proteases or neuraminidase.* The Journal of Immunology, 1990. **144**(4): p. 1304-10.

323. Anderson, C.L., et al., *Monoclonal antibodies to Fc receptors for IgG on human mononuclear phagocytes. Antibody characterization and induction of superoxide production in a monocyte cell line.* Journal of Biological Chemistry, 1986. **261**(27): p. 12856-12864.

324. Anderson, C.L., et al., *Phagocytosis mediated by three distinct Fc gamma receptor classes on human leukocytes.* The Journal of Experimental Medicine, 1990. **171**(4): p. 1333-1345.

325. Karpovsky, B., et al., *Production of target-specific effector cells using hetero-cross-linked aggregates containing anti-target cell and anti-Fc gamma receptor antibodies.* The Journal of Experimental Medicine, 1984. **160**(6): p. 1686-1701.

326. Scallon, B.J., et al., *Chimeric anti-TNF-α monoclonal antibody cA2 binds recombinant transmembrane TNF-α and activates immune effector functions.* Cytokine, 1995. **7**(3): p. 251-259.

327. Lügering, A., et al., *Infliximab induces apoptosis in monocytes from patients with chronic active Crohn's disease by using a caspase-dependent pathway.* Gastroenterology, 2001. **121**(5): p. 1145-1157.

328. Smeets, T.J.M., et al., *Tumor necrosis factor α blockade reduces the synovial cell infiltrate early after initiation of treatment, but apparently not by induction of apoptosis in synovial tissue.* Arthritis & Rheumatism, 2003. **48**(8): p. 2155-2162.

329. Cooper, D.L., et al., *FcγRIIIa Expression on Monocytes in Rheumatoid Arthritis: Role in Immune-Complex Stimulated TNF Production and Non-Response to Methotrexate Therapy.* PLoS ONE, 2012. **7**(1): p. e28918.

330. Willcocks, L.C., et al., *Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake.* The Journal of Experimental Medicine, 2008. **205**(7): p. 1573-1582.

331. Lee VW., Coan MH., and Galloway CJ., *Conjugated anti-TNF antibodies enhance the binding and degradation of TNF.* Biotechnol Ther, 1992. **3**(1-2): p. 1-62.

332. Sbarra, A.J. and M.L. Karnovsky, *The Biochemical Basis of Phagocytosis: I. Metabolic Changes During the Ingestion of Particles by Polymorphonuclear Leukocytes.* Journal of Biological Chemistry, 1959. **234**(6): p. 1355-1362.

333. Robinson, J.J., et al., *Stimulation of neutrophils by insoluble immunoglobulin aggregates from synovial fluid of patients with rheumatoid arthritis.* European Journal of Clinical Investigation, 1992. **22**(5): p. 314-318.

334. Shah, D., A. Wanchu, and A. Bhatnagar, *Interaction between oxidative stress and chemokines: Possible pathogenic role in systemic lupus erythematosus and rheumatoid arthritis.* Immunobiology, 2011. **216**(9): p. 1010-1017.

335. Clarkson, S., et al., *Blockade of clearance of immune complexes by an anti-F(cγ) receptor monoclonal antibody.* The Journal of Experimental Medicine, 1986. **164**(2): p. 474-489.

336. Fossati, G., et al., *Differential role of neutrophil Fcγ receptor IIIB (CD16) in phagocytosis, bacterial killing, and responses to immune complexes.* Arthritis & Rheumatism, 2002. **46**(5): p. 1351-1361.

337. Warmerdam, P.A., et al., *The human low affinity immunoglobulin G Fc receptor IIC gene is a result of an unequal crossover event.* Journal of Biological Chemistry, 1993. **268**(10): p. 7346-7349.

338. Bailey, J.A., et al., *Recent Segmental Duplications in the Human Genome.* Science, 2002. **297**(5583): p. 1003-1007.

339. Sharp, A.J., et al., *Segmental Duplications and Copy-Number Variation in the Human Genome.* The American Journal of Human Genetics, 2005. **77**(1): p. 78-88.

340. Fredman, D., et al., *Complex SNP-related sequence variation in segmental genome duplications.* Nat Genet, 2004. **36**(8): p. 861-866.

341. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-454.

342. Aitman, T.J., et al., *Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans.* Nature, 2006. **439**(7078): p. 851-855.

343. Fromont, P., et al., *Frequency of the polymorphonuclear neutrophil Fc gamma receptor III deficiency in the French population and its involvement in the development of neonatal alloimmune neutropenia.* Blood, 1992. **79**(8): p. 2131-2134.

344. Huizinga, T., et al., *Maternal genomic neutrophil FcRIII deficiency leading to neonatal isoimmune neutropenia.* Blood, 1990. **76**(10): p. 1927-1932.

345. Breunis, W.B., et al., *Copy number variation at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B.* Human Mutation, 2009. **30**(5): p. E640-E650.

346. Reilly, A., et al., *Variation in human FCGR2C gene copy number.* Immunogenetics, 1994. **40**(6): p. 456-456.

347. Breunis, W.B., et al., *Copy number variation of the activating FCGR2C gene predisposes to idiopathic thrombocytopenic purpura.* Blood, 2008. **111**(3): p. 1029-1038.

348. Niederer, H.A., et al., *Copy number, linkage disequilibrium and disease association in the FCGR locus.* Human Molecular Genetics, 2010. **19**(16): p. 3282-3294.

349. Robinson, J.I., et al., *Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis.* Human Mutation, 2012. **33**(4): p. 741-749.

350. McKinney, C., et al., *Association of variation in Fcγ receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples.* Annals of the Rheumatic Diseases, 2010. **69**(9): p. 1711-1716.

351. Machado, Lee R., et al., *Evolutionary History of Copy-Number-Variable Locus for the Low-Affinity Fcγ Receptor: Mutation Rate, Autoimmune Disease, and the Legacy of Helminth Infection.* The American Journal of Human Genetics, 2012. **90**(6): p. 973-985.

352. Mueller, M., et al., *Genomic Pathology of SLE-Associated Copy-Number Variation at the FCGR2C/FCGR3B/FCGR2B Locus.* The American Journal of Human Genetics, 2013. **92**(1): p. 28-40.

353. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-712.

354. WTCCC, *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.* Nature, 2010. **464**(7289): p. 713-720.

355. Armour, J.A.L., et al., *Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats.* Nucleic Acids Research, 2007. **35**(3): p. e19.

356. Ramsland, P.A., et al., *Structural Basis for FcγRIIa Recognition of Human IgG and Formation of Inflammatory Signaling Complexes.* The Journal of Immunology, 2011. **187**(6): p. 3208-3217.

357. Warmerdam, P.A., et al., *A single amino acid in the second Ig-like domain of the human Fc gamma receptor II is critical for human IgG2 binding.* The Journal of Immunology, 1991. **147**(4): p. 1338-43.

358. Bredius, R.G., Fijen, C.A., De Haas, M., Kuijper, E.J., Weening, R.S., Van de Winkel, J.G., Out, T.A.,, *Role of neutrophil Fc gammaRIIa (CD32) and Fc gamma RIIIb (CD16) polymorphic forms in phagocytosis of human IgG1- and IgG3-opsonized bacteria and erythrocytes.* Immunology, 1994. **83**(4): p. 624-630.

359. Sanders, L.A., et al., *Human immunoglobulin G (IgG) Fc receptor IIA (CD32) polymorphism and IgG2-mediated bacterial phagocytosis by neutrophils.* Infection and Immunity, 1995. **63**(1): p. 73-81.

360. Stein, M.-P., et al., *C-reactive protein binding to FcγRIIa on human monocytes and neutrophils is allele-specific.* The Journal of Clinical Investigation, 2000. **105**(3): p. 369-376.

361. Warmerdam, P.A., et al., *Molecular basis for a polymorphism of human Fc gamma receptor II (CD32).* The Journal of Experimental Medicine, 1990. **172**(1): p. 19-25.

362. Li, X., et al., *A novel polymorphism in the Fcγ receptor IIB (CD32B) transmembrane region alters receptor signaling.* Arthritis & Rheumatism, 2003. **48**(11): p. 3242-3252.

363. Kyogoku, C., et al., *Fcγ receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: Contribution of FCGR2B to genetic susceptibility.* Arthritis & Rheumatism, 2002. **46**(5): p. 1242-1254.

364. Ernst, L., et al., *Allelic polymorphisms in the FcγRIIC gene can influence its function on normal human natural killer cells.* Journal of Molecular Medicine, 2002. **80**(4): p. 248-257.

365. Metes, D., et al., *Expression of Functional CD32 Molecules on Human NK Cells Is Determined by an Allelic Polymorphism of the FcγRIIC Gene.* Blood, 1998. **91**(7): p. 2369-2380.

366. Wu, J., et al., *A novel polymorphism of FcgammaRIIIa (CD16) alters receptor function and predisposes to autoimmune disease.* The Journal of Clinical Investigation, 1997. **100**(5): p. 1059-1070.

367. Li, L.H., Yuan, H., Pan, H.F., Li, W.X., Li, X.P., Ye, D.Q.,, *Role of the Fcγ receptor IIIA-V/F158 polymorphism in susceptibility to systemic lupus erythematosus and lupus nephritis: a meta-analysis.* Scandinavian Journal of Rheumatology, 2010. **39**(2): p. 148-154.

368. Morgan, A.W., et al., *Fcγ receptor type IIIA is associated with rheumatoid arthritis in two distinct ethnic groups.* Arthritis & Rheumatism, 2000. **43**(10): p. 2328-2334.

369. Kastbom, A., et al., *The 158V polymorphism of Fc gamma receptor type IIIA in early rheumatoid arthritis: increased susceptibility and severity in male patients (the Swedish TIRA project).* Rheumatology, 2005. **44**(10): p. 1294-1298.

370. Salmon, J.E., J.C. Edberg, and R.P. Kimberly, *Fc gamma receptor III on human neutrophils. Allelic variants have functionally distinct capacities.* The Journal of Clinical Investigation, 1990. **85**(4): p. 1287-1295.

371. Huizinga, T., et al., *Biallelic neutrophil Na-antigen system is associated with a polymorphism on the phospho-inositol-linked Fc gamma receptor III (CD16).* Blood, 1990. **75**(1): p. 213-217.

372. Bux, J., et al., *Characterization of a New Alloantigen (SH) on the Human Neutrophil Fcγreceptor IIIb.* Blood, 1997. **89**(3): p. 1027-1034.

373. Carr, I.M., et al., *Inferring relative proportions of DNA variants from sequencing electropherograms.* Bioinformatics, 2009. **25**(24): p. 3244-3250.

374. Clayton, D.G., et al., *Population structure, differential bias and genomic control in a large-scale, case-control association study.* Nat Genet, 2005. **37**(11): p. 1243-1246.

375. Carpenter, D., et al., *Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders.* BMC Genomics, 2011. **12**(1): p. 418.

376. Kibbe, W.A., *OligoCalc: an online oligonucleotide properties calculator.* Nucleic Acids Research, 2007. **35**(suppl 2): p. W43-W46.

377. Korch, C. and H. Drabkin, *Manganese citrate improves base-calling accuracy in DNA sequencing reactions using rhodamine-based fluorescent dye-terminators.* Nucleic Acids Research, 1999. **27**(5): p. 1405-1407.

378. Barnes, C., et al., *A robust statistical method for case-control association testing with copy number variation.* Nat Genet, 2008. **40**(10): p. 1245-1252.

379. Lee, J., *A note on the comparison of group means based on repeated measures of the same subject.* JChronDis, 1980. **33**: p. 673-675.

380. Perry, G.H., et al., *Copy number variation and evolution in humans and chimpanzees.* Genome Research, 2008. **18**(11): p. 1698-1710.

381. Pinkel, D., T. Straume, and J.W. Gray, *Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization.* Proceedings of the National Academy of Sciences, 1986. **83**(9): p. 2934-2938.

382. van der Heijden, J., et al., *Phenotypic Variation in IgG Receptors by Nonclassical FCGR2C Alleles.* The Journal of Immunology, 2012. **188**(3): p. 1318-1324.

383. Stewart-Akers, A.M., et al., *FcγR expression on NK cells influences disease severity in rheumatoid arthritis.* Genes Immun, 2004. **5**(7): p. 521-529.

384. Bonagura, V.R., et al., *Mapping studies reveal unique epitopes on IgG recognized by rheumatoid arthritis-derived monoclonal rheumatoid factors.* The Journal of Immunology, 1993. **151**(7): p. 3840-52.

385. Ramirez, J., et al., *FCGR2A/CD32A and FCGR3A/CD16A Variants and EULAR Response to Tumor Necrosis Factor-α Blockers in Psoriatic Arthritis: A Longitudinal Study with 6 Months of Followup.* The Journal of Rheumatology, 2012. **39**(5): p. 1035-1041.

386. Iles, M.M., *What Can Genome-Wide Association Studies Tell Us about the Genetics of Common Disease?* PLoS Genet, 2008. **4**(2): p. e33.

387. Gibson, G., *Rare and common variants: twenty arguments.* Nat Rev Genet, 2012. **13**(2): p. 135-145.

388. Westra, H.-J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations.* Nat Genet, 2013. **45**(10): p. 1238-1243.

389. Plant, D., et al., *Genome-wide association study of genetic predictors of anti–tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci.* Arthritis & Rheumatism, 2011. **63**(3): p. 645-653.

390. Mishima, Y., et al., *High reproducible ADCC analysis revealed a competitive relation between ADCC and CDC and differences between FcγRIIIa polymorphism.* International Immunology, 2012. **24**(8): p. 477-483.

391. Hart, S.P., et al., *Characterization of the Effects of Cross-Linking of Macrophage CD44 Associated with Increased Phagocytosis of Apoptotic PMN.* PLoS ONE, 2012. **7**(3): p. e33142.

Appendix I – Buffers, reagents and primers

50X Tris-acetate EDTA (TAE)

242g Tris (base)

57.1ml acetic acid

100ml 0.5M EDTA (pH 8.0)

Make up to 1 litre with dH_2O

10X Mg^{2+} polymerase chain reaction buffer

5ml 1M Tri HCl (pH 9.0)

25ml 1M KCl

500µl 1.5M $MgCl_2$

500µl Triton X-100

19ml dH_2O

1X 100bp DNA marker

100µl New England Biolabs Inc. 100bp ladder

83µl 6X New England Biolabs Inc. loading dye

Make up to 500µl with dH_2O

Appendix Table 1. PCR primer sequences				
Primer label	Sequence 5’-3’	Tm	GC%	Nt
High resolution *HLA-DRB1* typing primers				
DRB95001 forward	TTGTGGCAGCTTAAGTTTGAAT	52.1	36	22
DRB95002 forward	TTCCTGTGGCAGCCTAAGAGG	56.7	57	21
DRB95003 forward	GTTTCTTGGAGTACTCTACGTC	51.1	45	22
DRB95004 forward	GTTTCTTGGAGCAGGTTAAAC	49.5	43	21
DRB95005 forward	AGTACTCTACGGGTGAGTGTT	51.9	48	21
DRB95006 forward	CCTGTGGCAGGGTAAGTATA	50.4	50	20
DRB95031 forward	GTTTCTTGAAGCAGGATAAGTT	50.1	36	22
DRB1*10 forward	GCGGTTGCTGGAAAGACGCG	65	59.7	20
DRB95008b reverse	CTCGCCGCTGCACTGTGAAG	65	60.3	20
DRB96001GT reverse	CTGCACTGTGAAGCTCTCAC	54.3	55	20
DRB96002TG reverse	CTGCACTGTGAAGCTCTCCA	55.4	55	20
FCGR locus primers				
TriPSV forward	AGTTCAGCTGGGAGCCAGGGA	58.8	62	21
TriPSV reverse	GCCTCAGTCTTACAGCCCCTA	55.6	57	21
2C ORF QSV 2 forward	TCAGTTGCAGAACCATTCTGGG	55.2	50	22
2C ORF QSV 2 reverse	GCTGTCATTGTTGTTGGCCTTG	55.1	50	22
IIIA forward	TCACATATTTACAGAATGGCAATGG	53.5	36	25
IIIA reverse	CAGGAAACAGCTATGACCCTTGAGTGATGGTGATGTTCA	67.2	46	39
rs9427399F	CCTGTGCATCTGACTGTGCTT	55.8	52	21
rs9427399R	CTCCTGCCTTAGAGCCTTTAG	54.7	52	21
2A PCR1	GGAGAAACCATCATGCTGAG	51.2	50	20
2A R2	TCAATACTTAGCCAGGCT	47.4	44	18

Appendix II – Experimental protocols

Qiagen QIAmp gDNA extraction – whole blood

Sample blood	200µl
Proteinase K	20µl
Buffer AL	200µl
96-100% EtOH	200µl
Buffer AW1	500µl
Buffer AW2	500µl
Buffer AE	400µl

Method:

1. Pipette 20µl of proteinase K into a 1.5ml (or2ml) microcentrifuge tube. (For a 400µl sample add 40µl, for 300µl add 30µl, etc.)
2. Add 200µl sample to the microcentrifuge tube. Using 200µl whole blood, plasma, serum, etc. (or $5x10^6$ lymphocytes in 200µl PBS).
3. Add 200µl buffer AL to sample and mix by pulse-vortexing for 15. (If using 400µl increase buffer amount to match 1:1).
4. Incubate mixture at 56°C, checking every 10mins. The tube contents should turn brown/green after ~30-40mins of incubation.
5. Briefly spin the tube down to remove any droplets from the lid.
6. Add 200µl (96-100%) ethanol and mix by pulse-vortexing for 15s. Briefly centrifuge to remove any droplets from inside the lid. (Adjust ethanol volume in 1:1 with original sample where necessary).
7. Carefully add the total solution to the top of the spin column (~800µl for each column), being careful not to wet the rim of the tube. Close the cap properly and centrifuge at 8000rpm (6000g) for 1min.

8. Place the column in a new clean collection tube (2ml) and discard the filtrate and old tube.
9. Open the tube and carefully add 500µl buffer AW1 without wetting the rim. Close the cap and spin at 8000rpm (6000g) for 1min. Discard the filtrate and place the column in a new clean 2ml collection tube.
10. Open the spin column and add 500µl buffer AW2 without wetting the rim. Close the cap and spin at 14000rpm for 3mins.
11. Place the spin column in a new clean collection tube and spin at 14000rpm for 1mins.
12. Place the column in a new clean microcentrifuge tube and discard the old collection tube with the filtrate. Carefully add 200µl AE buffer and incubate at room temperature (15-25°C) for 5mins. Spin at 8000rpm for 1mins. (Do not discard the filtrate at the end of this step).
13. Repeat step 12, but reserve all the filtrate.

Invitrogen Gene Catcher DNA extraction

Magnetic beads	45µl
Lysis buffer L13	14ml
Sample blood	3ml
Protease buffer	3ml
Protease	40µl
100% isopropanol	5ml
50% isopropanol	3ml
Wash buffer W12	500µl
Elution buffer E5	1ml

Method:

Binding DNA

1. Set water bath to 65°C.
2. Add lysis buffer and beads to a falcon tube, volume dependent on the amount of sample blood used (e.g. 3ml = 45µl beads and 9ml lysis buffer L13)
3. Add required volume of blood to each tube (e.g. 3ml) and gently invert the capped tube to mix with the beads.
4. Incubate at room temperature for 5mins, allowing the beads to bind the DNA. Gently invert the capped tube to mix occasionally.
5. Place the tubes onto the magnetic separator rack for 5mins.
6. Using a Pasteur pipette remove the supernatant without disturbing the pellet, carefully!
7. Remove the tube from the rack and add 5ml lysis buffer L13, regardless of the volume, and mix carefully. The pellet does not need to be resuspended.
8. Incubate for 30s then place back on the magnetic rack for 20s.

9. Remove the supernatant without disturbing the pellet using a Pasteur pipette.
10. Repeat steps 7-9 once only.

Protease digestion

11. Add 3ml protease buffer (depending on blood volume) and 40µl protease to the tube containing the pellet and vortex for ~30s to mix.
12. Incubate at 65°C water bath for 10mins then remove and allow to cool to room temperature (10-20mins).
13. Gently invert the tubes to resuspend the pellet.

Washing and eluting DNA

14. Add an equal volume of 100% isopropanol (~5ml).
15. Twist and slowly invert the tube repeatedly until an aggregate has formed and there are no remaining clumps of pellet.
16. Place the tube on the magnetic rack and carefully remove all the supernatant with a Pasteur pipette.
17. Add 3ml of 50% v/v isopropanol (regardless of volume). Gently mix and invert 5 times.
18. Place the tube back on the rack and leave for 1mins. Remove all of the supernatant. Leave the pipette in the bottom of the tube to remove any residual liquid.
19. Add 250µl of Wash buffer W12 on the opposite side of the tube to the pellet, being careful not to dislodge it. Incubate for 1mins at room temperature.
20. Remove the liquid and repeat step 19 once only.
21. Remove the tubes from the rack and add 1ml of elution buffer E5 directly onto the pellet. Gently swirl the tube to remove all pellet from the side.
22. Incubate at 65°C in a water bath for 60-90mins. Swirl occasionally (~30mins).

23. Gently swirl the tube to disperse the pellet.
24. Place the tube on the magnetic rack until the solution becomes clear. This can be left overnight if necessary.
25. Transfer the supernatant containing the DNA with a 1ml pipette to a labelled clean tube; without removing the tube from the magnetic rack. Discard the magnetic beads and tubes.

Life Technologies ChargeSwitch™ PCR purification

Magnetic beads	1.5µl
Purification buffer (N5)	equal volume µl
dH_2O	300µl
Elution buffer (E5)	10µl

Magnetic rack

Method:

1. Add 1.5µl of ChargeSwitch magnetic beads to each PCR reaction with an equal volume of purification buffer (), i.e. if there is 40µl of PCR reaction volume add a further 40µl of purification buffer
2. Carefully pipette to properly mix the beads and incubate at room temperature for 1mins.
3. Place the tubes on a suitable magnetic rack and allow the beads to form a tight pellet.
4. Remove the solution with a clean pipette tip, being careful not to disturb the magnetic bead pellet
5. Remove the tube(s) from the magnetic rack and add 150µl of dH_2O to each, then carefully pipette mix to resuspend the magnetic beads
6. Place the tube(s) back on the magnetic rack and allow the beads to form a tight pellet.
7. Remove the supernatant carefully using a clean pipette tip and discard.
8. Repeat steps 5-7
9. Add 10µl of elution buffer () and pipette mix thoroughly to resuspend the magnetic beads. Incubate at room temperature for at least 1mins.
10. Place the tube(s) on the magnetic rack and allow the beads to form a pellet. Remove the supernatant containing the purified PCR amplicons into a fresh tube. Discard the magnetic beads

Sanger sequencing ethanol precipitation

125mM EDTA	5µl
100% Ethanol	60µl
70% (v/v) Ethanol	150µl
Formamide Hi-Di	10µl

Method:

1. Add 5µl 125mM EDTA to each tube to halt the sequencing reactions
2. Add 60µl 100% ethanol to each well and carefully pipette mix an equal number of times (i.e. all wells must be mixed the same number of times)
3. Seal the reaction tubes and incubate at room temperature for 10mins
4. In a balanced centrifuge spin the plates at 3900rpm for 30mins at 20°C
5. Carefully tip off the supernatant onto a clean paper towel and invert the plate in the centrifuge
6. Pulse spin up to 180rpm and immediately stop the centrifuge
7. Add 150µl of 70% (v/v) ethanol to each well and incubate at room temperature for 5mins. Do not pipette mix
8. Carefully tip off the ethanol by inverting the plate onto a clean paper towel. Pulse spin the plate over a clean paper towel up to 180rpm and immediately stop the centrifuge
9. Air-dry the plate in a darkened place at room temperature (I.e. a closed box) for a minimum of 30mins.
10. Prior to loading the sequencing reactions onto a suitable capillary gel electrophoresis machine add 10µl formamide Hi-Di to each well and pipette mix.