Surname: HARRISON	OSS ID No.: 2264169
First names (in full): PILLE	
Faculty /department: DIVISION OF MEDICAL SCIENCES DEPARTMENT OF ORTHOPAEDIC SURGERY	
Supervisor(s) name: PROF. B.P. WORDSWORTH; DR. J. POINTON	
Title of thesis (block capitals): GENETIC SUSCEPTIBILITY AND SEVERITY FACTORS IN RHEUMATOID ARTHRITIS.	

N.B. The title stated here must be precisely the same as that stated on the title page of the thesis submitted. **A candidate wishing to amend the title previously approved by the faculty/department must apply to the faculty/department for permission to do so.**

Declaration by the candidate as author of the thesis.

1. I understand that subject to the Statement of Policy on Intellectual Property promulgated from time to time by the University in its Statutes and Regulations, I am the owner of this thesis and the copyright will rest with me unless I specifically transfer it to another person.

2. I understand that the University requires that if I am granted leave to supplicate I shall deposit one copy of my thesis in the Bodleian or other appropriate University library where it shall be available for consultation and that reproductions of it may be made available to those who wish to consult it elsewhere. I understand that the library, before allowing the thesis to be consulted either in the original or in a reproduced form, will require each person wishing to consult it to sign a declaration that he or she recognises that the copyright of the thesis belongs to me and that no quotation from it and no information derived from it may be published without my prior written consent, except as permitted by law. **See notes overleaf.**

☑ I agree that my thesis shall be available for consultation from the date of deposit in accordance with paragraph 2 above.

☐ I also agree to deposit one digital version in the Oxford University Research Archive and understand that, although usage permissions will appear on the record for the thesis, people wishing to consult the digital copy of my thesis can do so freely without any requirement to sign a declaration as described above.

☐ I have made a formal request for Dispensation from Consultation on form GSO.3C. If the request is subsequently approved I agree that my thesis shall be available for consultation in accordance with paragraph 2 at the end of this or any additional approved period of dispensation. If the request is not approved I agree that my thesis shall be made available for consultation from the date of deposit.

Please tick all statements which apply

Signed _____ Date 11.05.2009

RSL
MS. D.Phil.
c. 21187

300416470V

Genetic Susceptibility and Severity Factors in Rheumatoid Arthritis



Pille Harrison
Green Templeton College
University of Oxford

D.Phil Thesis
Michaelmas Term 2008

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
01 JUN 2009
THESIS

Abstract

Genetic Susceptibility and Severity Factors in Rheumatoid Arthritis
Dr.Pille Harrison
Degree of Doctor of Philosophy
Green Templeton College
Michaelmas Term 2008

The aim of this D.Phil thesis was to enrich current knowledge of genetic susceptibility and severity factors in rheumatoid arthritis (RA), a common complex disease with an estimated heritability of approximately 60% in the Northern European population. A DNA resource from 760 clinically well-characterised British Caucasian RA patients was established for use in association studies to investigate candidate genes.

The previously reported association of *PTPN22* (1858C/T rs2476601) and rheumatoid factor positive disease was observed (p=0.004). Examination of *PTPN22* gene for differential allelic expression (DAE) in different peripheral blood mononuclear cell populations from RA patients was undertaken. A DAE of 20% or more in 8 RA patients ($p< 10^{-5}$) was observed, indicating possible existence of *cis*-acting genetic variation. *PTPN22* DAE in a number of cell types was demonstrated in 4 RA patients and there were also different allelic ratios between these cell populations ($p_c < 0.05$).

MHC class II expression is controlled at the transcription level by MHC class II transactivator (CIITA) and is mainly dependent on the transcription of its gene *MHC2TA*. The frequency of *MHC2TA* (-168A/G, rs3087456) polymorphism in RA cases did not differ from those in controls. Minor allele frequencies were higher in *HLA-DRB1* shared epitope negative and *DRB1*04* negative patients (p=0.01 and p=0.03 respectively). Further direct sequencing followed by single nucleotide polymorphism genotyping and haplotype analysis did not reveal significant skewing in the haplotype distribution.

Polymorphisms in the *IL-1A* and *IL-1B* promoter can influence gene expression and these were investigated for association. The *IL-1B* (-1464 C/G, rs1143623) G allele was found to be less common in the RA group (p=0.01). Meta-analysis revealed statistically significant association between *IL-1B* (-511 A/G, rs16944) and RA (p=0.02).

Although the importance of studied markers in predicting disease susceptibility is modest, together with future findings these results may help to provide better insight concering the RA disease process.

Attributions

The work undertaken in this thesis is my own original work. Parts of the thesis contain work undertaken in collaboration with others. All contribution and collaborations are listed below and in some cases within the text.

Chapter 2

DNA extraction from the whole blood for the first 500 RA patients was performed by Oxagen, Oxagen Ltd, Abingdon, Oxfordshire, UK.

Chapter 3

With patient recruitment I was assisted by research nurses Alice Harin, Veronica Ziel and Claire Farrar, Nuffield Orthopaedic Centre, Headington, Oxford. Radiographic scoring results analysis and severity percentile scores were calculated by statistical epidemiologist Andrew Roddam, Cancer Epidemiology Unit, University of Oxford, Richard Doll Building, Roosevelt Drive, Oxford, UK.

Chapters 6 and 7

PHASE analysis was performed by statistical geneticist Dr Kay Chapman, Institute of Musculoskeletal Sciences, The Botnar Research Centre, Oxford, UK.

Acknowledgements

Firstly I would like to thank all the patients who participated in this study and Oxagen who funded this research. I am grateful to Professor B. Paul Wordsworth, my supervisor, for allowing me the opportunity. I am equally grateful to Dr Jennifer J. Pointon my second supervisor for sharing with me her wealth of scientific expertise.

I want to thank members of the John Loughlin group, especially Lorraine Southam and Dr Kay Chapman for their expertise. I am grateful to Professor Rolf Rau from Rattingen for providing me with one-on-one instruction of his simple, but effective radiographic scoring method. I wish to say a big thank you to Dr Andrew Roddam, who helped me to analyse the radiographic data. I am also grateful to Dr Afsie Sabokbar for the use her equipment for cell sorting, Dr Guilliaume Mabilleau and Dr Racher Locklin assistance in understanding the cell work and FACS.

Finally I am grateful to my husband Alan, who stood by me through all this time.

Abbreviations

ACR- American College of Rheumatology

AGFAP- Antigen Genotype Frequencies among Patients

AIF-1- gene for allograft inflammatory factor 1

AITD – autoimmune thyroid disease

anti-CCP - anticyclic citrullinated peptide antibodies

anti-GPI - anti-glucose-6-phosphate isomerase antibodies

AP-2α- activator protein 2α

AP4B1 – gene for adapter related protein complex-4, beta 1 subunit

APCA- anti-citrullinated protein antibodies

APOE – gene for apolipoprotein E

ARMS - amplification refractory mutation system

Bat3 – gene for HLA-B-associated transcript 3

BLAST – Basic Local Alignment Search Tool

C5 – complement component 5

CARD - caspase-recruitment domain

CCP- cyclic citrullinated peptide

CCR5 - chemokine (C-C motif) receptor 5

CIITA - MHC class II transactivator

cM – centimorgan

CNV - copy number variations

CNVR - copy number variable regions

CREB - cyclic-AMP-responsive-element-binding protein

Csk - C-terminal Src kinase

CSNK – gene for casein kappa

CTH – C-terminal homology

CTLA-4 - cytotoxic T-lymphocyte antigen 4

CXCR3 – chemokine, CXC motif, receptor 3

CYP2D6 – Cytochrome P450 Debrisoquine 4-Hydroxylase

DAE – differential allelic expression

DCLRE1B – gene for DNA cross link repair 1B protein

ddNTP – Dideoxynucleotide Triphosphate

DMARD – disease-modifying anti-rheumatic drug

DNA - Deoxyribonucleic acid

dNTP – Deoxynucleotide Triphosphate
DZ- dizygotic

EDTA - Ethylenediamine Tetraacetic Acid

EIRA - Swedish Epidemiological Investigation of Rheumatoid Arthritis

EL buffer - Erythrocyte Lysis buffer

ENCODE - The Encyclopaedia of DNA Elements

ExoI – Exonuclease I

FACS - fluorescence-activated cell sorting

FAM - fluorophore

GM-CSF granulocyte/monocyte colony stimulating factor

GOLD - Graphical Overview of Linkage Disequilibrium

GWA - genome-wide association

GWAS - genome-wide association studies

GZMB – gene for granzyme B

HC – healthy controls

HIPK1-1 gene for homeodomain-interacting protein kinase 1

HLA- human leukocyte antigen

HRR- haplotype relative risk

HWE- Hardy-Weinberg equilibrium

IBD- identical by descent

ICC – intraclass correlation

IDDM- insulin dependent diabetes mellitus

IFNGR1 – gene for interferon-gamma receptor alpha chain precursor

IgG- immunoglobulin G

IL-1- interleukin 1 protein

IL-1RA interleukin -1 receptor antagonist protein

IL-1α - interleukin-1 alpha protein

IL-1β - interleukin -1 beta protein

IL1A – Gene for Interleukin 1 Alpha Protein

IL1B – Gene for Interleukin 1 Beta Protein

IL2RA – gene for interleukin 2 receptor alpha

IL2RB – gene for interleukin 2 receptor beta

IL-6 – interleukin 6

KAZALD1 – gene for kazal-type protease inhibitor domain-containing protein 1

KCl - Potassium Chloride

KIF5A - kinesin heavy chain isoform 5A

KLF12- kruppel-like transcriptional regulatory factor

LD- linkage disequilibrium

LOD- logarithm of the odds

LRRs - leucine-rich repeats

LYP- lymphoid tyrosine phosphatase

MAF – Minor Allele Frequency

MAGI-3 - membrane associated guanylate kinase related gene

MASC - Marker Association Segregation Chi-Square Method

MCP- metacarpophalangeal joint

M-CSF - mononuclear phagocyte colony-stimulating factor

MHC- major histocompatibility complex

MHC2TA – gene for CIITA

MMEL1 - membrane metallo-endopeptidase-like 1

MRI- magnetic resonance imaging

MTHFR – gene for 5,10-methylenetetrahydrofolate reductase

MTRR – gene for methionine synthase reductase

MS – gene for methionine synthase

MS_{beween} - variance between means

MS_{within} - variance within groups

MTP- metatarsophalangeal joint

MTX – methotrexate

MZ- monozygotic

NARAC- North American Rheumatoid Arthritis Consortium

NF-Y - nuclear transcription factor Y

NIMA- non-inherited maternal HLA antigens

NK – natural killer cells

OR – Odds Ratio

OLIG3- gene for oligodendrocyte lineage transcription factor 3

OPG –gene for osteoprotegerin

PADI4 – gene for peptidyl arginine deiminase, type IV

PBMC – peripheral blood mononuclear cells

PCR – polymerase chain reaction

PD-1- gene for programmed cell death-1

PHTF1 - putative homeodomain transcription factor 1

PIP- proximal interphalangeal joints

PODXL- gene for podocalyxin-like protein

PRKCQ – gene for protein kinase C

PTP – protein tyrosine phosphatase

PTPN22- gene for protein tyrosine phosphatase non-receptor type 22

RA- rheumatoid arthritis

RANK - receptor activator of nuclear factor kappa B

RANKL - receptor activator of nuclear factor kappa B ligand

rcf – Relative Centrifugal Force

RF- rheumatoid factor

RFC1 – gene for reduced folate carrier

RFLP- restriction fragment length polymorphism

rpm – Revolutions Per Minute

RFX – gene for heterotrimeric X-box-binding factor regulatory factor X

RNA - ribonucleic acid

RSBP1 – gene for round spermatid basic protein 1

RT – Room Temperature

SAP – Shrimp Alkaline Phosphatase

SDS - Sodium Dodecyl Sulfate

SE - shared epitope

SLC22A4 – gene for solute carrier family 22 (organic cation transporter), member 4

SLE – systemic lupus erythematosus

SNP – single nucleotide polymorphism

SQRD - square root of disease duration

STAT4 – gene for signal transducer and activator of transcription 4

T1D – Type I diabetes

TBE – Tris-Borate-EDTA

TCR- T-cell receptor

TDT - transmission disequilibrium test

TE - Tris-EDTA

TNF – tumor necrosis factor

TNFAIP3- gene for TNF- α –induced protein 3

TPMT - thiopurine methyl transferase

TRAF – gene for TNF associated factor

TS – gene for thymidylate synthase

WTCCC- Wellcome Trust Case Control Consortium

λ_R - familial recurrence risk ratio

χ^2 – Chi-Squared

Table of Contents

Abstract.....i

Attributions.....ii

Acknowledgements.....iii

Abbreviations.....iv

Table of contents.....ix

List of figures.....xiii

List of tables.....xiv

Chapter 1.....1
Introduction.....2
1.1 Rheumatoid arthritis.....2
1.1.1 Definition and clinical course.....2
1.1.2 Population prevalence.....7
1.1.3 Genetic epidemiology.....7
1.1.4 HLA role in rheumatoid arthritis.....9
1.1.4.1 HLA and susceptibility to rheumatoid arthritis.....9
1.1.4.2 HLA and severity of rheumatoid arthritis.....12
1.1.4.3 Other HLA genes.....14
1.1.4.4 Non-inherited maternal HLA antigens.....15
1.1.5 Pathogenesis of rheumatoid arthritis.....16
1.1.5.1 The role of T-cells.....16
1.1.5.2 Humoral immunity.....18
1.1.5.3 Cytokine network.....20
1.1.5.4 Synovial pathology.....22
1.1.5.5 Animal models.....25
1.1.6 Genetic studies in rheumatoid arthritis.....27
1.1.6.1 Linkage analysis.....27
1.1.6.2 Genome-wide linkage studies.....31
1.1.6.3 Association studies.....32
1.1.6.3.1 Epistasis.....34
1.1.6.3.2 Linkage disequilibrium.....35
1.1.6.3.3 Measures of linkage disequilibrium.....36
1.1.6.3.4 Patterns of linkage disequilibrium.....37
1.1.6.3.5 Association study design.....38
1.1.6.3.6 Statistical analysis of association.....40
1.1.6.3.7 Effect size, sample size, statistical significance and importance of association studies.....42
1.1.6.4 Candidate genes in rheumatoid arthritis: recent advances.....43

1.1.6.5 Genome wide association studies.....47
1.1.6.6 Genetic models of rheumatoid arthritis risk.....50
1.1.7 Conclusion.....53

Chapter 2.....54
Methods and Materials.....54
2.1 Laboratory methods.....54
2.1.1 DNA extraction and quantification.....54
2.1.2 RNA extraction from peripheral blood mononuclear cells (PBMC) and cDNA synthesis.....56
2.1.2.1 PBMC separation.....56
2.1.2.2 FACS analysis.....57
2.1.2.3 RNA extraction and quantification.....58
2.1.2.4 DNase treatment.....58
2.1.2.5 cDNA synthesis.....59
2.1.3 Polymerase chain reaction (PCR) design.....59
2.1.3.1 Primer design.....59
2.1.3.2 Restriction fragment length polymorphism (RFLP) design.....60
2.1.4 DNA amplification by PCR.....61
2.2 Genotyping.....62
2.2.1 SNP genotyping.....62
2.2.1.1 Amplification Refractory Mutation Systems (ARMS) PCR.....62
2.2.1.2 Restriction Fragment Length Polymorphism (RFLP).....63
2.2.1.3 SNP typing by KBiosciences using Taqman Assay.....64
2.2.1.4 Sequencing of the PCR products.....64
2.2.2 *HLA-DRB1* genotyping.....66
2.2.3 *HLA-DRB1*04* sub-typing.....67
2.2.4 Differential allelic expression (DAE) analysis by SNaPShot.....68
2.3 Statistical methods.....72
2.3.1 Power calculation.....72
2.3.2 Hardy-Weinberg Equilibrium (HWE).....72
2.3.3 GOLD.....73
2.3.4 Association analysis.....73
2.3.5 Haplotype construction and analysis by PHASE v. 2.0.2.....74
2.3.6 CLUMP for the comparison of haplotype counts.....75
2.4 Materials.....76
2.4.1 Reagents and suppliers.....76

Chapter 3.....78
Rheumatoid arthritis patients cohort- clinical, serological and *HLA-DRB1* characteristics.....78
3.1 Demographic and clinical characteristics.....78
3.2 Radiographic scoring and analysis of the extent of erosive disease.....80
3.3 *HLA-DRB1* associations.....87
3.4 Power of the RA patients collection.....89
3.5 Conclusions.....91

Chapter 4 92
Association analysis of the protein tyrosine phosphatase non-receptor type 22 (*PTPN22*) gene with rheumatoid arthritis and its clinical characteristics 92
4.1 Introduction 92
4.1.1 Lymphoid tyrosine phosphatase (Lyp; PTPN22) structure and function 92
4.1.2 Association of the *PTPN22* C1858T polymorphism with autoimmune diseases 95
4.2 Aims of the study 97
4.3 Subjects and methods 97
4.3.1 Study population 97
4.3.2 Methods and statistical analysis 97
4.4 Results 98
4.5 Discussion 102

Chapter 5 110
Cis-acting regulatory variation in the *PTPN22* gene of rheumatoid arthritis patients 110
5.1 Introduction 110
5.2 Aim of the study 112
5.3 Subjects and methods 112
5.3.1 Study population 112
5.3.2 *PTPN22* C1858T Genotyping 113
5.3.3 PBMC preparation and DAE analysis 113
5.3.4 Statistical analysis for DAE data 113
5.4 Results 114
5.5 Discussion 117

Chapter 6 123
Effects of *MHC2TA* genetic variation on susceptibility and severity of rheumatoid arthritis 123
6.1 Introduction 123
6.2 Aims of the study 126
6.3 Methods 127
6.3.1 Genotyping *MHC2TA* (-168A/G) (rs3087456) 127
6.3.2 Power calculation for *MHC2TA* (-168A/G) study 127
6.3.3 Analysis of the *MHC2TA* (-168A/G) genotyping data 128
6.3.4 Sequencing the *MHC2TA* gene 128
6.3.5 SNP selection and genotyping 128
6.3.6 Power calculation for the subsequent genotyping 129
6.3.7 Statistical analysis of the sequencing data 129
6.4 Results 130
6.4.1 *MHC2TA* (-168A/G) polymorphism 130
6.4.1 *MHC2TA* sequencing and subsequent genotyping 130
6.5 Discussion 136

Chapter 7 142
Interleukin-1 promoter region polymorphism role in rheumatoid arthritis and meta-analysis of *IL-1*(-511A/G) 142
7.1 Introduction 142
7.2 Aims of the study 143
7.3 Subjects and methods 144
7.3.1 Patients and study design for *IL-1A* and *IL-1B* promoter polymorphims 144
7.3.2 Genotyping 144
7.3.3 Statistical analysis 145
7.3.3.1 Case-control association analysis 145
7.3.3.2 Analysis of radiological progression 145
7.3.4 Meta-analysis 146
7.3.4.1 Search strategies and study selection 146
7.3.4.2 Evaluation of the statistical association 146
7.3.4.3 Power calculation for meta-analysis 147
7.4 Results 147
7.4.1 Case-control association study 147
7.4.2 Meta-analysis 152
7.5 Discussion 155

Chapter 8 159
General discussion and future work 159
8.1 Genome wide association studies past, present and future in rheumatoid arthritis genetic research 159
8.1.1 Common disease/common variant hypothesis 159
8.1.2 What has been achieved with genome wide association studies in searching rheumatoid arthritis genetic susceptibility factors 162
8.1.3 What the future will hold for rheumatoid arthritis genetic research? 164
8.2 Clinical importance of the genetic studies 167
8.3 Conclusion 172

References 173

Appendices 208
Appendix A 208
Appenidx B 215
Appendix C 217
Appendix D 218
Appendix E 219
Appendix F 224

Publications 228
Papers 228
Invited review 228
Abstracts 229

List of Figures

2.1 FACS analysis of a RA patient and matched healthy control.....57
2.2 Schematic overview of the differential allelic expression assay.....71
3.1 Examples for the different grades of the Rau-Ratingen score in MCP joints.....82
3.2 Example of the anatomical regions of the hand used for Rau-Ratingen method.....83
3.3 The relationship between radiographic score and disease duration with selected percentile lines.....85
3.4 Scatterplots of calculated radiographic severity scores and disease duration.....85
4.1 Overview of the PTPN22 structure.....94
4.2 Radiographic percentile severity scores according to the *PTPN22*genotypes.....101
5.1 Distribution of corrected allelic ratios of *PTPN22* in PBMC populations.....114
5.2 DAE of the *PTPN22* in different PBMC populations of 12 RA patients.....116
5.3 Comparison of corrected allelic ratios between different PBMC populations.....117
5.4 a The extensive LD block located in 1p13.....119
5.4 b Overview of genes in region 1p13.....119
6.1 Regulation of the transcription of MHC class II genes.....124
6.2 Transcription controls of the *MHC2TA*.....125
6.3 GOLD output for *MHC2TA* region.....135
7.1 GOLD output for *IL-1A-IL-1B* region 149
7.2 Scatterplots of Rau-Ratingen score according to the IL-1A –IL-1B genotype.....151
7.3 Forest plots for *IL-1B* (-511A/G) meta-analysis..... 154
A.1 Consent form 209
A.2 Questionnaire 211
C.1. Agarose gel view of RFLP genotyping of *PTPN22* C1858T.....217
C.2 An example of the electropherogram of *PTPN22* C1858T polymorphism.....217
D.1 Overview of 138 kb genomic region on 1p13.....218
E.1 Agarose gel view of ARMS genotyping of MHC2TA -168*A/G.....219
E.2 Electropherogram showing intronic SNP in *MHC2T*.....221

List of Tables

1.1 Genotype relative risk of *HLA-DRB1* for RA.....10
2.1 Limited *HLA-DRB1* typing primer sequences and PCR conditions.....67
2.2 Primers for high resolution *DRB1* typing.....67
2.3 *HLA- DRB1*04* sub-typing calculation 68
3.1 Relevant clinical characteristics in RA cohort.....79
3.2 Comparison of clinical characteristic in patients with mild and severe disease.....86
3.3 Different HLA genotype distribution in RA cohort..... 87
3.4 Power calculations.....90
4.1 Genotyping data and allele counts for *PTPN22*.....99
4.2 Comparison of clinical characteristics in RA patients according to *PTPN22*.....100
5.1 DAE analysis of *PTPN22* for 12 RA patients..... 115
6.1 Genotyping data and allele counts for *MHC2TA*.....132
6.2 Results of the direct sequencing of *MHC2TA* gene.....133
6.3 The list of the SNPs and genotyping methods used in the second phase for *MHC2TA*.....134
6.4 Ten most common *MHC2TA* 22 marker haplotypes for RA cases and HC.....136
6.5 Two and three marker haplotypes comparisons for *MHC2TA*.....140
7.1 Genotyping results for four SNPs in *IL-1A* and *IL-1B* promoter.....150
7.2 Extended *IL-1A-IL1-B* haplotype counts in RA cases and HC.....152
7.3 Genotype and allele counts for meta-analysis.....153
7.4 Results for *IL-1B* (-511A/G) meta-analysis.....153
B.1 *HLA-DRB1* allele counts in RA patients..... 215
B.2 The mean percentile severity score in RA patients according to *HLA-DRB1*.....215
B.3 Radiographic severity scores, mean and range according to *HLA-DRB1*04*.....216
B.4 Comparison of patients with mild or severe disease.....216
B.5 Comparison of seven hospital based RA cohorts 216
C.1 Primer sequences for *PTPN22* C1858T polymorphism genotyping.....217
E.1 Primer sequences for *MHC2TA* -168*A/G ARMS PCR 219
E.2 Primer sequences for *MHC2TA* sequencing.....220
E.3 Primer sequences and PCR conditions for additional *MHC2TA* genotyping.....222
F.1 Primer sequences and PCR conditions for *IL-1A* and *IL-1B* genotyping.....224
F.2 Results for IL-1A and IL-1B promoter region SNPs.....225
F.3 Results for extended *IL1A-IL1B* four marker haplotypes.....227

Chapter 1

Introduction

Human immunity, our natural defence against pestilence, which has underpinned two centuries of scientific research and clinical practice, is now perceived as a major force for both good and bad in medicine – "good" in the sense of its therapeutic and preventative potential, and "bad" in the sense that inflammatory responses underlie so many of the major chronic diseases pathologies of the twenty-first century. The biological mechanisms that link genetic variation and its phenotypic outcome stand as a central enigma in biology. Geneticists have approached this problem by trying to identify genetic variants that underlie the trait in question.

Rheumatoid arthritis is a common disease that has a complex aetiology, involving the action of multiple genes, as well as dynamic gene-environment interactions. Modern techniques of molecular genetics can be used to elucidate the mechanisms underlying disease susceptibility and progression, and to improve diagnosis and treatment. Causative DNA variants can be identified as the first step towards the eventual unravelling of the complex interplay between genes and environment.

1.1 Rheumatoid arthritis

1.1.1 Definition and clinical course

Rheumatoid arthritis (RA) [MIM 180300] is a chronic immune-meditated inflammatory disorder predominantly affecting synovial joints. RA is a clinically heterogeneous condition in terms of joint inflammation pattern, disease course, presence of extra-articular features and treatment response. It has been proposed that RA may be better classified as a heterogeneous syndrome rather than as a single disease entity.

Rheumatoid synovitis primarily affects the diarthrodial joints of the hands and feet, leading to progressive destruction of cartilage and bone in the affected joints, and consequently to severe loss of function. Subsequently the more proximal joints, including hip and knee may become involved leading to severe functional disability. Any synovial joint can be affected including the temporomandibular and cricoarytenoid joints. The peak age at onset is in the fifth decade, although disease onset can be at any time in adult life and it is a well-known fact that some forms of juvenile onset arthritis appear to be an early onset form of adult disease.

Overall the disease is three times more common in women, but this varies with the age of onset. While women are nearly ten times more frequently affected than men at the age of 30, this gender difference in incidence disappears by the age of 65. RA usually has a subacute (20%) or insidious (70%) onset with inflammatory symptoms of pain, swelling and stiffness, with the number of joints involved increasing over weeks to months. About 10% of patients have an acutely severe onset, and a few start with episodic symptoms that

progress into persistent disease. The course of RA is very variable; it may be associated with mild, reversible inflammation with destructive changes only occurring after many years; alternatively erosive disease may be present at the time of presentation. A small proportion of patients present with palindromic rheumatism characterised by short-lived episodes of synovitis, recurring at variable intervals. Positive rheumatoid factor (RF) and early involvement of the wrist and proximal interphalangeal (PIP) joints predict the subsequent development of RA. (Gonzalez-Lopez *et al.*1999*)*.

Age-related differences in RA clinical features and outcome have shown controversial results (Van Schaardenburg *et al.*1993; Peltomaa *et al.* 2000). Recently a study comparing hand radiographic scores between patient groups with disease onset before and after 55 years showed that the higher damage score seen in older patients was partly due to osteoarthritic joint damage. (Khanna *et al.* 2005).

Extra-articular manifestations are common and demonstrate the systemic nature of the disease. They can occur at any time in the disease (Turesson *et al.* 1999). In population based studies, 56% of male patients and 42% of female patients had evident rheumatoid organ disease (Weyand *et al.*1998). Smoking at the time of disease diagnosis and early disability seem to be associated with an increased risk of severe extra-articular manifestations. (Turesson *et al.* 2003)

RA can affect practically any organ and manifest as rheumatoid nodules, vasculitis, cardiopulmonary disease, haematologic abnormalities and neurovascular disease. These

extra-articular manifestations have a defining impact on disease outcome, including increased mortality in comparison to RA in general (Turesson *et al.* 2002; Gabriel *et al.* 2003). Some differences in extra-articular manifestations have been observed between men and women. Female patients describe symptoms related to secondary Sjogren's syndrome more frequently than men (Thomas *et al.* 1998). RA lung disease has been observed to occur more commonly in men who have: long standing-illness, positive RF and subcutaneous nodules. Pleural involvement, usually asymptomatic, is the most common manifestation of lung disease in RA and may occur concurrently with pulmonary nodulosis or interstitial disease (Anaya *et al.* 1995).

Cardiovascular manifestations in RA are classically presented as myocarditis or pericarditis. Recent population based studies have also revealed that the risk of myocardial infarction or congestive heart failure in RA patients is significantly higher than age and sex-matched control individuals (Gabriel *et al.* 1999; Nicola *et al.* 2005). The number of cardiovascular events in RA patients has been found to be 4 times higher than in individuals without RA of the same age and sex (Del Rincon *et al.* 2001). The reasons for increased cardiovascular morbidity and mortality are likely to be multifactorial. However, there is emerging evidence that systemic inflammation promotes endothelial dysfunction and induces premature atherosclerosis (Ridker *et al.* 2002). This is supported by results from recent studies showing that disease-modifying therapies with potent anti-inflammatory properties improve endothelial dysfunction and may reduce cardiovascular events in RA (Gonzalez-Juanatey C *et al.* 2004). Felty's syndrome has traditionally presented with a combination of chronic polyarthrits, neutropenia and

splenomegaly. Patients with Felty's syndrome often have high titres of RF and other antibodies as well as other extra-articular organ manifestations (Campion *et al.* 1990). Vasculitic manifestations in RA can range from isolated digital infarcts to systemic involvement with major organ damage. Peripheral neuropathy has been particularly associated with excessive mortality in RA (Turesson *et al.* 2002).

The diagnosis of RA is based on set of relatively non-specific diagnostic classification criteria for epidemiologic studies (Arnett *et al* 1988). There are 7 criteria; 5 are clinical, 1 laboratory and 1 radiographic parameter. The presence of four or more of these criteria is considered to define RA.

1. Morning stiffness for at least one hour and present for at least six weeks
2. Arthritis in 3 or more joint areas for at least six weeks (involvement of PIP, MCP, wrist, elbow, knee, ankle and MTP joints)
3. Arthritis of the hand joints (swelling of PIP, MCP, wrist joints)
4. Symmetric arthritis
5. Rheumatoid nodules
6. RF
7. Hand radiographic changes typical of RA, must include erosions or unequivocal bony decalcification localised or most marked adjacent to the involved joints.

The 1987 American College of Rheumatology (ACR) criteria for RA. Data from Arnett *et al* 1988.

RF has for a long time been the most recognised RA-associated antibody marker, recognising the Fc part of IgG molecules. Lately another class of autoantibodies has been recognised that appear more specific for RA. These antibodies are directed against proteins containing the non-standard amino acid citrulline (Schellekens *et al.* 1998; Girbal-Neuhauser *et al.* 1999).

Besides this high specificity (nearly 100% in RA patients), these anticyclic citrullinated peptide antibodies (anti-CCP) are present early in disease, even before clinical onset (Kurki *et al.* 1992; Rantapaa-Dahlqvist *et al.* 2003). These antibodies appear to be produced by local plasma cells in the inflamed synovium, since they constitute a higher proportion (7.5 fold) of IgG in synovial tissue than in paired serum samples (Masson-Bessiere *et al.* 2000). In addition, the existence of citrulline-reactive antibodies has been associated with a more active and severe disease (Bas *et al.* 2000; Paimela *et al.* 2001). A strong association with certain major histocompatibility complex (MHC) shared epitope (SE) haplotypes, has also been reported (van Gaalen *et al.* 2004).

Early RA can be difficult to differentiate from self-limiting or other forms of inflammatory arthritis. Frequently patients have no radiographic changes at baseline (van der Heijde *et al.* 1995). In addition, conventional radiographic methods do not provide adequate information about soft tissue structures and synovium. The role of ultrasonography and MRI as novel imaging techniques have shown promising results for detecting synovitis and tracking erosions with time (Lopez-Ben *et al.* 2004; Szkudlarek *et al.* 2004; Hoving *et al.* 2004).

1.1.2 Population prevalence

In the modern times the prevalence and incidence of RA have been relatively constant affecting around 0.5%-1% of the European and North American population making it one of the most common autoimmune disorders (Hazes *et al.* 1990; Gabriel *et al.* 2001) However, there are several variations observed in the prevalence of RA in different ethnic groups indicating the importance of environmental and genetic factors on disease aetiology. RA is exceptionally rare in most of sub-Saharan Africa and has also a relatively low prevalence in south-east Asia, China and Japan. Interestingly, significantly lower disease prevalence has been found in the Afro-Caribbean population living in Manchester, UK compared to the geographically matched white Caucasian population (MacGregor *et al.* 1994). In contrast RA is relatively common in some native American-Indian tribes (up to 7%). This might be explained by the high frequency of certain HLA (human leukocyte antigen) class II alleles among Chippewa and Yakima Indians that have been associated with RA (Harvey *et al.* 1981; Willkens *et al.* 1991).

1.1.3 Genetic epidemiology

RA is a classical example of a complex genetic disease characterised by incomplete penetrance, genetic heterogeneity and a role of multiple disease genes. Concordance rates in monozygotic (MZ) twins vary between studies and depend on the methodology of ascertainment and population being studied. A survey of twins in the UK and the Netherlands carried out by Lawrence and co-workers at the end of 1960s reported a concordance rate for MZ twins of 30% and 5% for dizygotic (DZ) for seropositive RA (Lawrence *et al.*1970). This was accepted as standard for many years, until in 1986 a

Finnish survey, based on nationwide twin registry that included less severe disease cases, reported much lower rates for MZ and DZ (12% and 4% respectively) (Aho *et al.*1986). Similar results to the Finnish study were obtained in the most recent UK study by Silman (15.4% MZ; 3.6% DZ) (Silman *et al.*1993). However, the concordance rates were highest in those twin pairs who carried *HLA-DR4* positive allele (MacGregor *et al.* 2000). Overall the excess rate for MZ and DZ twins as demonstrated in several studies confirms the importance of genetic factors in the aetiology of RA, with approximate heritability estimated around 60% (MacGregor *et al.* 2000). These results are in contrast with the most recent Danish nationwide study where the estimations were much lower. (Svendsen *et al.* 2002). The probandwise concordance rate was 0 (95% confidence interval 0 to 24.7) in monozygotic twins and 8.8 (1.9 to 23.7) in dizygotic twins. However these results are based on smaller numbers of twin pairs (13 MZ and 36 DZ twin pairs participated)

Sib recurrence risk of rheumatoid arthritis has been found to be variable depending on the clinical parameters used to define the illness. In the initial UK study by Lawrence the excess risk to the sib of individuals with mild seronegative disease was only just greater than the general population frequency (λ of 1.1) and increased sixfold for seropositive erosive disease(Lawrence *et al.*1970). Several larger systematic studies have shown that sib recurrence risk ratio tends to be greater in cases with younger age at onset in the proband, more severe disease in the proband, and as the number of affected relatives of the at-risk individual rises (Kwoh *et al.* 1996). On the other hand the sib recurrence risk ratio falls as the familial relationship becomes more distant. The siblings of those with RA who share both HLA haplotypes identical to the proband are at particular risk

(Deighton *et al.* 1993) since this group is more likely than others to exhibit the *HLA-DR* shared epitope (SE) compound heterozygote genotype, particularly strongly associated with RA (Wordsworth *et al.* 1992) . Overall in RA the estimations show an excess sibling recurrence risk ratio (λ_s) ~8 (Wordsworth *et al.* 1992) which is much lower when than other autoimmune diseases like ankylosing spondylitis and systemic lupus erythematosus (λ_s =50 and 20 respectively) (Hochberg *et al.* 1987; Vyse *et al.* 1996; Brown *et al.* 2002). It is important to emphasize that the observations indicate that genetic factors are particularly relevant to the susceptibility of seropositive erosive arthritis rather than milder non-erosive seronegative arthritis.

1.1.4 HLA role in rheumatoid arthritis

1.1.4.1 HLA and susceptibility to rheumatoid arthritis

Association of RA and *HLA-DRB1* was first noted over three decades ago. In 1976, Stastny demonstrated that, in allogenic mixed-leukocytic reactions, RA lymphocytes proliferate normally when stimulated by normal lymphocytes, but the response was diminished when the cells were stimulated by cells from other RA patients (Stastny *et al.* 1976). These experiments demonstrated similarities between RA patients and led to the discovery of the association with *HLA-DRB1* genes. The relationship between HLA and RA is by far the strongest genetic link in RA and has been confirmed by numerous association and genome-wide linkage studies (Gregersen *et al.* 1988; Wordsworth *et al.* 1989; Cornelis *et al.*1998).

MHC class II molecules are heterodimeric (ά and β chain) transmembrane glycoproteins displayed at the surface of specialised cells of the immune system. The α and β chain of each MHC II molecule are encoded by separate genes in the MHC class II regions. Three MHC class II isotypes exist in humans, designated HLA-DR, HLA-DQ and HLA-DP. All three isotypes serve similar functions and are critically involved in regulating the recognition of antigenic peptides by T cells. The contribution of HLA to RA is estimated to be around 30% (Deighton *et al.*1989). To date, the strongest association with RA has been reported with *HLA-DRB1* alleles that encode a conserved sequence of amino acids comprising residues 70-74 in the third hypervariable region of DRβ1 chain. The susceptibility epitope is glutamine-leucine-arginine- alanine-alanine (QKRAA) or glutamine-arginine-arginine-alanine-alanine (QRRAA). This sequence encoded by several *HLA-DRB1* genes, which are shown to be associated with RA, including *DRB1*0401, DRB1*0404, DRB1*0101* and *DRB1*1402.* There is a hierarchy of strength of the association of the different SE positive *HLA-DRB1* alleles (Table 1.1).

Table 1.1 Genotype relative risk of *HLA-DRB1* for RA

DRB1 genotype	Relative risk	P-value
DRX/DRX	1.0	
0101/DRX	2.3	10^{-3}
0401/DRX	4.7	10^{-12}
0404/DRX	5.0	10^{-9}
0101/0401	6.4	10^{-4}
0401/0401	18.9	10^{-19}
0401/0404	31.3	10^{-33}

*Adapted from Hall *et al.* 1996.

Studies in many populations have shown that there is a strong association between the susceptibility to RA and the presence of *HLA-DR1*04*. However, this association has been shown to be a subject of ethnic variability and not all alleles contribute equally to RA susceptibility. *HLA-DRB1*0401* and **0404* genotypes have been found most commonly in European Caucasian RA patients (Wordsworth *et al.*1989). In East Asian populations *DRB1*0401* is rare and *DRB1*0405* is the most frequent RA associated *DRB1*04* genotype (Kim *et al.*1995). *DRB1*0101* is common in the Northern European and North American population. Although many studies have confirmed the contribution of *DRB1*0101* to RA, its own relative risk is weak compared to *DRB1*0401* or **0404*. By itself some have proposed that it may also predispose to seronegative and mild RA (Polski *et al.*1994). The risk of developing RA increases when DRB1*0101 is combined with DRB1*04 alleles (Agrawal *et al.* 1995). *DRB1*1001* is rare in most European populations, but relatively common among Spanish (Sanchez *et al.* 1990). In addition to Spanish this allele has also been associated with RA in Greece, India and Singapore populations (Carthy *et al.* 1993; Taneja *et al.* 1996; Chan *et al.*1994). DRB1*1001 shares some similarities to *DRB1*0101*, as it also only weakly predisposes to RA by itself, but may increase the susceptibility to RA when combined with DR4 alleles. *HLA-DRB1*10* has a similar amino acid sequence (RRRAA) to other SE alleles. *HLA-DRB1*1402* is a common allele among Yakima American Indians and has been closely linked to RA among this Native American tribe (Willkens *et al.* 1991). It has the same SE sequence as *HLA-DRB1*0404*.

It is apparent that not all *HLA-DRB1*04* subtypes are equally associated with RA, and some have been shown to have a protective effect. *HLA-DRB1*0402* is most frequent among the Jews (Gao *et al.* 1991). Studies in Jewish and other populations have shown this allele not to be associated with RA and it is regarded as protective. It shares the similar amino acid motif IDEAA with other HLA-DRB1 alleles like *DRB1*0103*, *1102*, *1301*, *1302*, *1304* that are not associated with RA. Interestingly, when the transmission of DRB1 alleles was assessed between parents and RA patients, *DRB1*13* allele was shown to be less inherited by the RA patients than would have been expected by chance (Barrera *et al.* 2000).

The precise role of *HLA-DRB1* is probably more complex, as crystal structures of *HLA-DR* molecules show that the SE faces away from the antigen-binding groove of *HLA-DR* molecules that determines the presenting peptides to CD4+ T cells (Stern *et al.* 1994). In addition the SE model does not explain the additive effect of *DRB1*0403* and *DRB1*0901* alleles, and related haplotypes lacking the SE motif. On the other hand, the *DRB1*1406* allele harbours the SE, but has not been found to be associated with RA.

1.1.4.2 HLA and severity of rheumatoid arthritis

Initially proposed to explain mainly genetic susceptibility to RA the SE theory has undergone evaluation and further investigations have suggested that SE has a role in the development of more severe disease. Retrospective studies have shown that HLA genotyping provides information about the aggressiveness of rheumatoid synovitis. HLA *DRB1*0401* is found commonly in RF positive patients and appears to confer risk for the

formation of a more rapidly progressive synovitis. Patients combining *HLA-DRB1*04* and *DRB1*01* have been shown to require joint replacement earlier than patients who inherited only a single copy of the abovementioned alleles. In predicting progression, then *HLA-DRB1*04* positive individuals are more likely to develop early erosive disease (Goronzy *et al.* 2004). Recent meta-analysis by *Gorman et al.* confirmed the association between SE and erosive disease (OR 2.0, 95%CI 1.8-2.2), although most of studies, when taken individually failed to show the effects of SE on disease severity (Gorman *et al.* 2004). Heterogeneity between the studies was reduced when authors analysed the data according to geographical origin. Among Asians, Northern and Southern European Caucasians to *HLA-DRB1* SE seemed to confer risk for erosive disease with additional dose-dependent relationship evident in Asians and Southern European Caucasians. No association between SE and erosive disease was observed in the Greek RA population. Gender and RF seem to not add any additional risks. Relationship between SE and erosions remained robust even when *HLA-DRB1*0401* was excluded from the analysis.

The exact mechanisms of how HLA genes influence the inflammatory synovitis are unknown. Taking into an account that synovial lesions are not uniform, it is possible that HLA polymorphism could regulate the T cell selection in thymus, resulting in smaller T cell diversity and playing an important role in the formation of different lymphoid microstructures (Weyand *et al.* 1994). On the other hand, *HLA-DRB1*1301* and *DRB1*1302* have been reported to protect against severe erosions among patients who do not have SE positive alleles (Mattey *et al.* 1999).

A number of studies have indicated a role for *HLA-DR1*04* related genes in extra-articular RA (Turesson *et al.* 2004). Up to 95% of Felty's syndrome patients are *HLA-DRB1*04* positive, almost always carrying the *DRB1*0401* allele (Lanchbury *et al.*1991). Recently a study examining HLA impact on extra-articular manifestations in RA confirmed that patients with severe extra-articular RA were more likely to carry *HLA-DRB1*04* alleles. Genotypes featuring a double dose of *DRB1*04* alleles were associated with rheumatoid vasculitis, Felty's syndrome, and all extra-articular RA combined (Turesson *et al.* 2005). In addition severe extra-articular manifestations tend to cluster in individual patients with RA. This may in part be due to shared genetic factors such as *HLA-DRB1*04* alleles. Taken together, *DRB1*04* SE double gene dose is associated with disease severity in RA and certain alleles may contribute specifically to risk for severe extra-articular manifestations.

1.1.4.3 Other HLA genes

Considerable research has gone into trying to identify the further MHC genes that are likely to be involved in RA. The relatively large number of potentially relevant polymorphic genes and as well as the extent of linkage disequilibrium (LD) within the MHC, has complicated attempts to define precisely the disease causing genes. One of the most widely studied loci is TNF (tumor necrosis factor) that has been mapped 1000kb from *HLA-DRB1* in the MHC class III region. Results have been variable and very few studies have followed designs that control fully for LD with *HLA-DRB1*. Haplotype analysis conducted by *Newton et al* on *HLA-DRB1*04* positive cases and controls identified a single *LTA-TNF* haplotype that modified the *DRB1* association with RA

(overrepresented on case *DRB1*0404* haplotypes and under-represented on case **0401* haplotypes) (Newton *et al*. 2003).This haplotype spans 106 kb and encompasses at least 13 genes. Further analysis in this region identified single and two marker haplotypes involving *AIF-1*, *Bat3* and *CSNK* that were associated with RA (Harney *et al.* 2008 in press). A high degree of LD was observed between the markers making it difficult to determine which marker was associated with RA independent in this region.

Further evidence of the complexity of the MHC association with RA came from the work of Jawaheer et al who genotyped a total of 54 microsatellite markers across the MHC in multicase RA families (Jawaheer *et al.* 2002). There was a significant association with one microsatellite haplotype on the background of the ancestral A1-B8-DRB1*03 haplotype (this is a non-SE-encoding haplotype). There was also evidence of an additional susceptibility element on the background of *DRB1*0404* in the class I region of the MHC. These finding are consistent with other observations indicating different disease-modifying elements on the background of *DRB1*0401* and *DRB1*0404* (Newton *et al.* 2003).

1.1.4.4 Non-inherited maternal HLA antigens

Even in hospital based studies of RA, no more than 87% of patients carry one or two copies of *HLA-DRB1* SE alleles. It has been hypothesised that exposure to non-inherited maternal HLA antigens (NIMA) bearing SE could be important. This has been studied in auto-immune thyroid disease, scleroderma and transplant immunology (Nelson 2003). Initial studies in the Netherlands suggested that there is an excess of *HLA-DRB1*04*

NIMA in RA patients who lacked the SE alleles (van der Horst-Bruinsma *et al.*1998). This was replicated in the UK population where a weak excess of *DRB1*04* and SE positive NIMA in *HLA-DRB1*04* negative RA patients was observed (Harney *et al.* 2003). The mechanism underlying this effect remains to be elucidated. Microchimerism and effects on shaping the T cell repertoire have been suggested (Nelson 2003)

1.1.5 Pathogenesis of rheumatoid arthritis

1.1.5.1 The role of T-cells

The initiating event leading to the development of RA is unknown, but the disease initiation is probably T-cell driven and antigen specific. Substantial evidence exists to support the central role of T-cells in the pathogenesis of RA including the association with *HLA-DRB1* alleles, the presence of synovial T-cell infiltrates, the effects of T-cell derived cytokines, evidence of T-cell-dependent arthritis, successful disease transfer with T cells in experimental arthritis and the anti-arthritic effects of new T-cell targeted therapies. Until recently it was believed that following antigenic stimulation, naive $CD4^+$ T cells typically differentiate into 2 subsets, Th1 and Th2 cells, characterized by different cytokine production profiles and effector functions. Th1 cells produce large quantities of IFN-γ and mediate cellular immunity while Th2 cells, which are involved in humoral immunity, primarily produce IL-4, IL-5, and IL-13. RA was previously considered a classical "Th1" driven disease with a relative predominance of INF-γ and lack of Th2 cytokines leading to induction and maintenance of chronicity. This idea was challenged by the demonstration that a subset of IL-17-producing $CD4^+$ T cells which are distinct from those producing either IL-4 or IFN-γ were pathogenic in collagen-induced arthritis

(CIA) (Matthys *et al.* 1998). These IL-17–producing $CD4^+$ T cells were dubbed Th17 or $Th_{IL\text{-}17}$ cells. *In vitro* human studies followed and in support to the mouse model, revealed that IL-23 is implicated in IL-17 secretion (Hoeve *et al.* 2006). Many lines of evidence support the role of IL-17 in the pathogenesis of RA (Miossec 2003). Levels of IL-17 are elevated in the synovium of patients with RA (Kotake *et al.* 1999; Hwand *et al.* 2005). T-cells and lymphocytes isolated from synovial cultures from patients with RA spontaneously secrete IL-17 (Chabaud *et al.* 1999; 2001). IL-17 can induce several pathological mechanisms that lead to the progression of joint damage including upregulation of local inflammatory mediators and promoting extracellular matrix injury (Katz *et al.* 2001; Jovanovic *et al.* 2000).

Further T-cell abnormalities including a relative lymphopenia state characterised by a reduced number and function of naïve CD4 T cells repertoire contraction and hyporesponsivness to stimulation have been observed in RA. These data have been derived from animal models of arthritis where lymphopenia provoked compensatory homeostastic proliferation of autoreactive T-cell clones leading to a restricted T-cell repertoire and increased arthritis (Jang *et al.* 2006).

The clinical course of RA with its "flare ups" has led scientist to speculate the presence of "disequilibrium" in which immune response against self-components are dysregulated. The possibility arises that during these "active episodes of disease" the Th cell phenotype becomes skewed towards a proinflammatory profile (Th1 and Th17) and away from the anti-inflammatory regulatory phenotypes (T_{reg}) on the basis of the local cytokine environment and dendritic cell populations. The possible mechanisms of such a change

could either be a loss of dominance of T_{reg} phenotypes or a primary shift towards the proinflammatory pathways. The balance is regained during recovery from flares when the normal equilibrium re-establishes itself. However, the exact mechanisms and further more the central "catalyst" that tips the balance is yet to be discovered.

1.1.5.2 Humoral immunity

The first concepts of immune hyper-reactivity in the pathogenesis of RA were considered in mid-20th century. The clue to this was the identification of RF in the blood of affected patients. RF was originally observed by Waaler in 1939 and later on re-discovered by Rose in 1948. The seminal studies of *Kunkel et al.* ultimately characterised the unknown factor as antibodies that bind to the Fc portion of immunoglobulins (Kunkel *et al.*1964). Half a century ago it seemed reasonable to hypothesize that the patients' dilemma was indeed a scenario of "horror autotoxicus". RF became important as diagnostic tools, but why they were generated and how they participated in the pathogenesis of synovitis remained less well understood. The question of whether RFs are pathogenetic or instead an epiphenomenon of chronic autoimmune disease has been a subject of debate for many decades. Various hypotheses have been proposed to explain RA and its underlying synovial inflammation based on the known biologic functions of RF. It is thought that RFs can initiate immune complexes by binding Ig and subsequently activating complement.The subsequent release of chemotactic factors, such as C5a followed by the diversion of inflammatory cells into the inflamed tissues via the C5a receptor (CD88) could be highly relevant (Zvaifler *et al.*1973). There is indeed considerable data to support this hypothesis. Immune complex localization in the superificial layers of

cartilage have been described in ultrastructural studies. These fixed immune complexes provide a solid surface to facilitate neutrophil adherence and activation. The synovium itself is a rich source for the local production of complement proteins and RF (Neuman *et al.* 2002). However, many normal individuals and patients with other forms of chronic inflammatory disease produce RF, indicating that the mere presence of autoantibody is insufficient for a specific response. Because many immune-mediated and infective diseases are marked by the presence of immune complexes, this hypothesis does not provide an adequate explanation of the unique features of RA.

By the discovery of anti-cyclic citrullinated peptide, anti-RA33 and anti-GPI antibodies in the human and mouse, the impact of humoral autoimmunity in RA regained remarkable interest (Simon *et al.* 1993). These antibodies were first seen as binding reactivity to filaggrin, a protein that is not expressed in the synovium and is typically found during terminal differentiation of epithelial cells. It was assumed that filaggrin represented a cross-reacting antigen rather than the primary target of anti-citrullinated protein antibodies (APCA). However, other proteins, such as fibrin, that have undergone post-translational citrullination could serve as a target in the joint and thus represent the arthritogenic antigen recognized by pathogenic autoantibodies (Masson-Bessiere *et al.* 2001). Accumulation of fibrin on synovial and cartilage surfaces is a well-known feature of RA and therefore it is conceivable that deiminated fibrin deposited in the synovial membrane may form a preferred target structure for ACPA. Another target for ACPA was recently identified as the intermediate filament protein vimetin (De Rycke *et al.* 2005). Vimetin is a ubiquitously expressed cytoskeleton protein that occurs in various

citrullinated isoforms and therefore may form one of the primary targets of the ACPA response. It has been demonstrated that ACPAs are locally produced by synovial B-cells and therefore together with RF may contribute to the inflammatory and destructive processes in the rheumatoid joint (Masson-Bessiere *et al.* 2000). However, direct experimental evidence for pathgenetic involvement of ACPAs has not yet been obtained and very little is known about autoreactive T-cell driving the ACPA response. It has been definitely established that ASPAs are the most specific serological marker antibodies for RA with sensitivity about 68% and high specificity (98%) (Schellekens *et al.* 2000).

1.1.5.3 Cytokine network

TNF and interleukin 1 (IL-1) are the two main cytokines contributing to the joint inflammation in RA. TNF is a proinflammatory cytokine, produced by monocytes, T-lymphocytes, macrophages and fibroblasts. TNF stimulates prostaglandin E_2 and collagenase release from human synovial cells and potentiates osteoclastogenesis. IL-1 is typically synthesised and released simultaneously with TNF and the two cytokines share similar biologic properties. Elevated concentrations of proinflammatory cytokines have been found in peripheral blood and synovial fluid of RA patients and the concentrations correlate with disease activity and bone resorption (Eastgate *et al.* 1988; Neidel *et al.* 1995). Both, IL-1 and TNF, as well as granulocyte/monocyte colony stimulating factor (GM-CSF), stimulate fibroblast growth and pannus formation. These inflammatory cytokines also interact with each other and promote inflammatory process.

Normal bone remodelling depends on the coordinated activity of the osteoblasts and osteoclasts. Under normal physiological conditions osteoclasts are primarily responsible for bone resorption. These multi-nucleated cells, derived from the monocyte/macrophage line, also mediate focal bone resorption in RA. Osteoclast precursors have been identified in the bone resorption lacuna adjacent to the invading pannus (Bromley *et al.* 1984). The differentiation of osteoclast precursors into mature bone resorbing cells depends on cytokine network including TNF, IL-1 and IL-6. Osteoblasts and bone stromal cells produce a cytokine - receptor activator of nuclear factor κB ligand (RANKL), which in conjugation with macrophage-colony stimulating factor (M-CSF) is the essential factor for osteoclast differentiation. RANKL binds to its cell surface receptor RANK located on dentritic, T-lymphocyte, osteoclast precursors and osteoclasts. RANK mediates its signal transduction via TNF receptor-associated factor (TRAF) proteins leading to differentiation of precursors, fusion to form multi-nuclear osteoclasts and the activation of the osteoclasts to resorb bone. (Gravallese *et al.* 2001). Osteoblasts also secrete osteoprotegerin (OPG), a soluble decoy receptor for RANKL. By binding to RANKL with high affinity, OPG prevents RANKL from interacting with its cognate receptor (RANK). It is the relative amounts of RANKL and OPG that regulate osteoclast activity (Lancey *et al.*1998).

Evidence of a role for RANKL in bone resorption in inflammatory disease comes from the studies in the rat adjuvant-arthritis (Kong *et al.* 1999). Arthritic rats, treated with OPG at disease onset had minimal loss of cortical and trabecular bone, in contrast to the untreated control animals where bone loss was severe. OPG treatment also prevented

osteoclast accumulation. However, the treatment did not decrease joint inflammation, indicating that the effect was specific for osteoclast-mediated bone loss only. RANKL-knockout mice, in whom inflammatory arthritis was induced, demonstrated a reduced degree of bone erosions compared to wild-type control animals (Pettit AR 2001). In addition RA synovial tissue produces other cytokines like IL-1α, IL-1β and IlL-11 that enhance osteoclast formation, activity and survival (Deleuran *et al.* 1992).

1.1.5.4 Synovial pathology

The normal synovial membrane which is one to two cell layers thick is transformed in the rheumatoid joint into a proliferating cell mass, the pannus, which becomes invasive and destroys surrounding tissue and bone. The cellular composition of rheumatoid synovium is relatively simple; it consists of resident cells and inflammatory infiltrate. Synovial membrane includes resident fibroblast-like (type A) and macrophage-like (type B) synoviocytes, and endothelial cells. A complex cascade of cytokine-mediated events alters the phenotype of synoviocytes (especially type B cells) which show increased expression of oncogenes and are resistant to apoptosis (Yamanishi *et al.* 2005). The activated synoviocytes become highly proliferative, resulting in widespread joint inflammation and matrix degradation (Ritchlin *et al.* 2000). A broad spectrum of macrophage and fibroblast cytokines including IL-1, IL-6, IL-12, IL-15, IL17, IL-18, TNF, GM-CSF and many more are produced in the RA synovium that play a central role in maintaining synovial inflammation (Bucala *et al.* 1991).

The inflammatory infiltrate in RA synovitis consists of T cells, B cells, dentritic cells, macrophages, plasma cells, mast cells and granulocytes. Histology has shown that synovial infiltrates can form different architectural patterns. The most common form of rheumatoid synovitis is a diffuse infiltrate, where inflammatory cells are scattered among synovial cells without forming any higher organised structures. In about 40% of patients inflammatory cells are organised into follicular structures and can form germinal centres (Weyand *et al.* 2003). In a minority of patients, formation of granulomas, resembling rheumatoid nodules, have been observed. These different patterns are important, as they may influence therapeutic response.

T-lymphocytes have a central role in initiation and maintenance of chronic inflammatory process in RA. So, it was therefore a surprise that RA synovial fluid and tissue contain of relatively low concentrations of T cell cytokines such as interleukin-2 (IL-2) and interferon-γ (IFN- γ), in contrast to macrophage and fibroblast products that dominate (Firestein *et al.* 1990). Besides the accumulation of activated T cells in the RA synovium, the entire T cell pool phenotype and homeostasis is also abnormal. Significant telomere shortening found in CD4+ T cells infiltrating the rheumatoid synovium indicates that they have undergone intensive replication (Wagner *et al.* 1998). CD4+ T cells express the CCR5 and CXCR3 chemokine receptors. These chemo-attractants mark the subsets of cells with capacity for migration to inflammatory sites (Qin *et al.* 1998). In addition to the aberrations in the global and in the synovial T cell receptor repertoire, clonal proliferation of T cell population in RA patients has been observed. Proliferating CD4 cells are functionally different from classic T-helper cells and are uniformly present in

the circulation and infiltrate into synovial lesions (Goronzy *et al.*1994; Waase *et al.*1996). A set of clonally expanded CD4 T cells lack surface expression of the major co-stimulatory molecule CD28. This is caused by a transcriptional block of the CD28 gene, caused by absence of two gene specific nuclear proteins necessary for the initiation of CD28 transcription (Vallejo *et al.* 1998). The CD4+CD28- T cells are potentially cytotoxic, as they produce inflammatory cytokines (interferon-γ) and contain intracellular cytolytic substances (perforin and granzyme B) necessary to lyse target cells. In the absence of CD28 molecule these unusual cells use alternative co-stimulatory pathways and express killer cell immunoglobulin-like receptors (KIR2DS2, NKG2D) *de novo*. These cells may be related to NK cells as they express MHC class I-recognising receptors allowing them to participate in innate and adaptive immune processes. Furthermore the increased frequency of CD4+CD28- cells is a predictor of progressive erosive disease in early RA. (Goronzy *et al.* 2004)

There is no doubt that the pathogenic mechanisms involved in RA are complex and at least 6 different cell types, including synovial fibroblasts, macrophages, T-cells, B-cells, dendritic cells and mast cells, all hold critical positions in sustaining synovitis. Highly complex pathways of tissue damage and repair are involved in creating and maintaining the lesions that destroys cartilage and bone. The pannus could not grow unless supported by a network of blood vessels (Paleolog 2002). Rheumatoid synovitis is closely associated with the process of lymphoid neogenesis (Takemura *et al.* 2001) that provides a highly efficient micro-architecture for optimizing immune recognition events (Kang *et al.* 2002). The rheumatoid lesion itself produces growth and survival factors for B-cells

(Seyler *et al.* 2005), creating ideal conditions for their development and maturation. B-cells localised in the inflamed synovial membrane may release anti-CCP antibodies, thereby directly aggravating tissue injury. T cells recruited to the lesion encounter a micro-environment in which they receive stimulatory signals through legions of ligands expressed on nonlymphoid cells (Goronzy *et al.* 2004). All these processes are complex by themselves and in the inflamed joint they seem coexist.

1.1.5.5 Animal models

Experimental animal models of arthritis have greatly contributed to the understanding of the basic mechanisms of joint disease. The marked diversity that has been observed among numerous models may be explained by the different techniques used to induce the arthritis: immunisation with cartilage components, injection of non-specific immune stimuli, infection agents or they components and the manipulation of the genetic information in transgenic animals.

Of the many animal models collagen induced arthritis (CIA) is the best studied and has been induced in mice, rat and primates. In contrast proteglycan induced arthritis (PGIA) has only been successfully established in BALB/c mice. A hypothetical mechanism for the development of arthritis after immunisation with cartilage components is through recognition of specific epitopes presented in the context of MHC class II antigens (Wooley *et al.* 1981, Glant *et al.* 1987). The localisation of antibody and reactive cells in synovial joint causes immune-mediated cartilage damage that histologically resembles

RA. In general CIA tends to be more aggressive leading to joint ankylosis compared to PGIA.

The injection of non-specific agents with adjuvant activity can lead to the development of experimental arthritis in certain species (Billingham 1995), the classical example being adjuvant-induced arthritis (AA) in rats. The pathogenesis is still somewhat unclear but it is thought that heat shock proteins may be the autoimmune target in AA (van Eden *et al.* 1989). In susceptible strains of animals, phagocytosis of non-degradable oil components by macrophages, leads to the production of proinflammatory cytokines and subsequent activation of T and B-cells. Disease onset is generally rapid with features of arthritis appearing within 2 to 3 weeks.

The development of arthritis following infection is apparent in both natural and induced animal models. Many different organisms may induce joint disease in a broad range of animal species. The pathogenesis of the arthritis in these circumstances depends on the micro organism used. Chronic persistent infection of the joint can generate excessive local antigenic stimulation. Alternatively bacterial cell wall and mycoplasma membrane components may induce non-specific immune activation though mitogenic activity. Viral infection of synovial cells may elicit novel cell surface antigens or abnormal expression of activation antigens. Finally, infectious agent may disrupt normal immune regulation and induce the production of pro-inflammatory cytokines and promote immune activity against normal components of the joint.

Transgenic animals, defined as novel strains generated by genetic manipulations, have resulted in development of several models of joint disease. These animals have been

very useful for asking specific questions with regard to pathogenetic pathways in disease like over-expression of foreign proteins or inflammatory cytokines. Genetic modification can also alter susceptibility to the induction of experimental arthritis in these animals. For example, the mechanisms of antigen induced arthritis have been studied in *H2-c-fos* transgenic (*c-fos*+) mice which fail to generate specific IgG antibodies against immunizing antigens. These mice develop destructive arthritis very similar to the *c-fos*- animals, but the joints of *c-fos*+ mice contain very few lymphocytes, in contrast to the presence of copious mesenchymal cells that produce extensive amount of c-fos protein participating in generating joint destruction (Shiozawa *et al.* 1992).

So far, no single model has provided a definitive blueprint for of human disease, indicating that the data should be interpreted with caution. The key feature that is not reproduced by animal arthritis models is the ability to mount a high titre RF response.In contrast to the MRL/lpr mice readily produce these antibodies. Nevertheless, animal models provide valuable preclinical data for the development of therapies and insights into relevant mechanisms common to experimental arthritis and RA and may ultimately lead to the development of effective arthritic agents.

1.1.6 Genetic studies in rheumatoid arthritis

1.1.6.1 Linkage analysis

Genetic linkage analysis for locating disease genes in humans was originally used for single gene disorders. In this method, families with multiple affected relatives are investigated using a panel roughly evenly spaced genetic markers — typically short

tandem repeat markers ("minisattellites" and "microsatellites"), DNA sequences that show considerable variation among populations but often have no known functional consequences. Two genetic loci are said to be linked if they are transmitted together from parent to offspring more often than expected under independent inheritance. Two markers are in linkage disequilibrium (LD) if across the population as a whole, they are found together on the same haplotype more often than expected by chance alone. If one analyses a sufficient number informative meioses in a sufficiently large pedigree and uses enough markers — about 350 is typically sufficient for a genome-wide scan — at least one of these markers will lie sufficiently close to the disease gene that its inheritance pattern will closely match the inheritance pattern of the disease in the family. Moreover, in a collection of families with different mutations in this same disease gene, the correlation between marker and disease status will hold even though more than one disease mutation (and more than one variant at the marker locus) is represented in the group. When single families or collections of families are large enough that this correlation can be shown to be highly unlikely to have occurred by chance, the marker is said to be linked to the disease gene. Linkage, so defined, indicates only that the gene for the disease lies in the vicinity of the marker, but tight linkage may permit the precise localization of the disease gene using direct DNA sequencing and LD fine mapping.

Linkage is usually reported as the logarithm of the odds (LOD) score. This score reflects the odds that the observed pattern of inheritance could have occurred by chance, assuming that the marker is in or near the disease gene. It was first proposed by Morton in 1955 (Morton *et al.* 1955). It is a function of the recombination fraction θ between the marker and putative disease gene measured in centimorgans (cM), where 1 cM equates to

a genetic distance generating an average of 1 crossover between two loci per meiosis. The LOD score varies according to the value of θ being considered. Large positive scores are evidence for linkage and negative scores are evidence against it. The ultimate objective of the analysis is to estimate the recombination fraction between individual markers and the disease locus (two-point) or position of the disease locus relative to a fixed map of markers where the location of each marker is assumed to be known (multipoint). The best (maximum likelihood) estimate of θ or position is that which maximizes the lod score function: the maximum LOD score. The higher the LOD score the greater the evidence for linkage. Traditionally, a score of 3 was regarded as significant evidence of linkage for single gene disorders. However the common practice now is to maximise the LOD score over the recombination fraction. This increases the power to detect linkage and linkage is viewed as being excluded for all values of θ at which LOD score is less than -2.

Parametric or model-based linkage analysis is the analysis of co-segregation of genetic loci in pedigrees and is commonly applied to the analysis of monogenic traits. Loci that are close enough together on the same chromosome segregate together more often that do loci on different chromosomes. Loci on different chromosomes segregate together purely by chance. For any parametric linkage analysis the genetic model for the disease of interest must be specified. For multifactorial disease like RA, where several genes and environmental factors might contribute to disease risk, there is no clear mode of inheritance. Methods to investigate linkage have therefore been developed that to not require specification of the disease model; such methods are referred to as non-parametric or model-free. The rationale is that, excess sharing of disease haplotypes that are identical

by descent (IBD) will be observed, irrespective of the mode of inheritance. Various methods test whether IBD sharing at a locus is greater than expected under the null hypothesis of no linkage.

The power of linkage analysis to locate susceptibility loci for complex diseases is much greater in animal models than in humans, for a variety of reasons. First, inbred strains are used, which tends to reduce the genetic complexity and limit genetic effects to the loci that differentiate the original strains used in the breeding experiments. Second, by design, all matings are informative (e.g., all parents are heterozygous in an inter-cross). Third, all matings have the same phase (i.e., the genotypes at a presumed disease locus are known), and offspring from all matings can be combined into a single analysis. Thus, one strategy taken by human geneticists is first to localize a disease gene in an animal model, and then to examine the homologous region in human families. The assumption is that variation in the same gene may exist in the human population and influences susceptibility to the analogous disease.

This approach has a mixed track record. There are some striking examples of homology, such as the similar role of the MHC in human type 1 diabetes and in the non-obese diabetic mouse (Hattori *et al.* 1986). Swanberg *et al.* identified a *Mhc2ta* locus in advanced inbred rat strains that induced differences in MHC class II molecule expression after nerve injury (Swanberg *et al.* 2005). Subsequently polymorphism in the 5'flanking region of human *MHC2TA* -168A/G was associated with increased risk or RA, multiple sclerosis and myocardial infarction.On the other hand, in many cases it has proved impossible to demonstrate such homologies. This approach may fail either because the

gene of interest functions differently in humans and mice, or because functional variants comparable to those seen in the model system do not exist (or are uncommon) in the corresponding human gene.

One additional point relates to the issue of replication. When a genetic effect is large, most independent researchers can readily obtain similar results with strong levels of statistical significance. Most genes for Mendelian disorders have lived up to this expectation, and there have also been occasional successes in more complex cases, such as Alzheimer's disease. In this disorder, the risk to carriers of the *APOE-ε4* allele is substantially higher than to non-carriers, an effect that has been seen in virtually all studies published to date (Farrer *et al.* 1997). However, when genetic effects are weak and possibly context-dependent (e.g., they may vary by sex, ethnicity, or precision of diagnosis), replication may be particularly difficult, and very large samples may be required before confident conclusions can be drawn.

1.1.6.2 Genome-wide linkage studies

The first systematic genome-wide linkage study for RA involving French, Italian, Spanish and Belgian patient populations was carried out by *Cornelis et al* in 1998. This reported significant linkage only for HLA ($p < 2.5 \cdot 10^{-5}$) and nominal linkage ($p < 0.05$) for 19 markers in 14 other regions, among them IDDM6 and IDDM9. These two loci have been previously linked to insulin dependent diabetes mellitus (IDDM). Up to the present time, several other major genome-wide linkage studies for RA have been carried out in the UK, Japan and North America using microsatellite markers in affected sib-pair or multi-case RA families (Jawaheer *et al.* 2001; Jawaheer *et al.* 2003; John *et al.*2004;

Shiozawa *et al.* 1998; MacKay *et al.* 2002). These studies have demonstrated a number of regions of the genome likely to be linked with RA. However, apart from the well established HLA susceptibility locus, replication studies confirming the linkage of other reported regions, have essentially been unsuccessful. The limited success of linkage analysis for RA is at least partly due to the studies being too small to detect genes of modest effect. Simulation modelling shows that increasing the sample size will improve the likelihood of detecting the genes with weaker effect, but not indefinitely (Eaves *et al.* 1994, Risch *et al.*1995). Two meta-analyses of RA genome-wide linkage studies have been published (Fisher *et al.* 2003; Choi *et al.* 2006). Meta-analysis combines the linkage results from several studies potentially enhancing the evidence of linkage and increasing statistical power. Although, consistency was found in 4 regions: 6.2, 16.2, 6.4 and 8.1, there were also slight differences in the reported results. This may be explained to a degree by differences the studies included for analysis and also the genome search meta-analysis methods. Fisher *et al* analysed genome data obtained from linkage graphs, whereas Choi *et al* evaluated linkage scores with $P < 0.05$. Although by increasing the sample size meta-analysis may introduce heterogeneity arising from phenotypic differences between different populations and geographic differences, it is a powerful tool providing basis for further targeted linkage and candidate gene studies.

1.1.6.3 Association studies

Genetic association studies aim to detect association between one or more genetic polymorphisms and a trait, which might be some quantitative characteristic or a discrete attribute or disease status. Association differs from linkage in that the same allele (or

alleles) is associated with the trait in a similar manner across the whole population (due to LD), while linkage allows different alleles to be associated with the trait in different families. However, genetic associations arise only because human populations share common ancestry and it can be argued that association studies are really just a special form of linkage study in which the extended family is the wider population. In linkage analysis, data from distantly related individuals are more powerful for detecting small effects than data from closely related individuals, but this advantage is offset by the fact that, because of the increased risk of recombination events, linkage extends over shorter distances in distantly related individuals, making it necessary to use a greater density of markers.

Studies may be designed to test polymorphisms in genes considered *a priori* as putative causal variants (direct association). This type of study is the easiest to analyse and the most powerful, but the difficulty is the identification of candidate polymorphisms. A non-synonymous mutation is a candidate causal variant. However, it is likely that many causal variants responsible for heritability of common complex disorders will be non-coding and *vice versa*, not all coding polymorphisms lead to change in protein structure that has functional effect. Alternatively, studies may use polymorphisms as surrogates for the causal locus and this type of association allows researching for causal genes in indirect association studies. It is the presence of LD throughout the genome that can be exploited to map genetic susceptibility factors. However, indirect association are weaker than the direct associations they reflect, and it will usually be necessary to type several surrounding markers to have a high chance of picking up the indirect association. Indirect

association studies are more difficult to analyse and there is still debate as to the best methods. Finally, by contrast with direct studies, until the polymorphisms in a region have not been adequately mapped, there cannot be a definitive negative result, as it is impossible to exclude that the causal variant exists but is not picked up by the markers chosen.

1.1.6.3.1Epistasis

The fact that inheritance of some traits could only be explained by joint action of two unlinked loci was first demonstrated by William Bateson back in 1909, who called the effect epistasis (Bateson 1909). For example, variation of phenotype with genotype at one locus was only apparent in those with certain genotypes at the second locus. Thus epistasis was defined as one locus masking the effect of another. Fisher used a similar term, epistacy, to refer to a statistical interaction meaning deviation from the additive effects of two loci upon the trait mean (Fisher R 1918). The term epistacy soon evolved into epistasis and in modern genetics the two uses of the word coexist, often causing confusion.

Evidence consistent with epistatic interactions among susceptibility alleles has been reported in both animal models and human linkage studies of complex autoimmune diseases (Prins *et al.* 1993; Morel *et al.* 2000; Gray-McGuire *et al.* 2000) In relation to RA, it is thought that epistasis and natural selection are responsible for the high LD that can be seen in the MHC. LD within the MHC is an area of intense interest at present: it serves as a mapping tool for identification of disease susceptibility genes, gives an insight

into the demographic history of human populations and also the selective pressures acting therein. Although epistasis seems to be common, it is difficult to detect as it requires large samples due to the dimensionality of the evaluating combinations of multiple variables. When interactions among multiple polymorphisms are considered, there are many multilocus genotype combinations that have very few or no data points. This phenomena has been referred to as the curse of dimensionality (Bellman, 1961) and devalues the use of standard linear and logistic regression by increasing type I and II errors and decreases power. Alternative methods have been developed; for example the physiological epistasis model (Cheverud *et al.* 1995), exploratory data analysis like the combinatorial partioning method (Nelson *et al.* 2001), multifactor dimensionality reduction analysis (Ritchie *et al.* 2001) and many more methods that have successfully demonstrated evidence of epistasis where Fisherian analysis of variance approach has been unable to do so. As more powerful alternatives to traditional parametric statistical approaches are developed the detection of epistasis becomes easier.

1.1.6.3.2 Linkage disequilibrium

The identification of disease susceptibility causing polymorphisms in association studies by the indirect method relies on the existence of association between the causal variants and nearby markers. LD in which nearby genetic variant are correlated with each other more often than expected by chance, depends heavily on population history and on the genetic make-up of the founders of that population. If all examples of a particular region of DNA in a population derive a recent common ancestor, less recombination will have occurred and the variants will often be inherited together throughout the population.

However, in a different population established at an earlier time, the opportunities for recombination events will have been greater leading to a greater variety of haplotypes and less LD in the region. Furthermore, the particular arrangement of variants in the founders of a population will determine which variants are inherited together. Thus, it is possible that a polymorphism will be in LD with a nearby disease allele in one population but not in another, leading to variable results in association studies from different ethnic groups. To explore the possibility of LD, positive associations should be followed up by testing adjacent markers. If LD is present, the original marker associated with disease may not be the causal allele and further studies of the region are warranted to identify the causal variants accurately.

1.1.6.3.3 Measures of linkage disequilibrium

Various different measures of pairwise LD have been proposed, including Lewontin's D' (Lewontin 1964). Lewontin's D' is an important measure for identification of regions in which there has been little recombination and, therefore in which there is potential to map causal loci by indirect association studies. Lewotin's D was the earliest measure of LD that was proposed. It defined LD as the difference between the observed frequency of a two-locus haplotype and the frequency that would be expected to show if the tw0 alleles were segregating at random.

D=PAB – (PA x PB)

PA - frequency of allele at the first locus, PB- frequency of allele at second locus.

D' is determined by dividing D by its maximum possible value given the allele frequencies. D' therefore equals 1 (complete LD) if two alleles have not been separated by recombination. However, this measure does not directly determine the power of indirect association studies. Formally, the power of test for indirect association depends largely on the index r^2 that is the correlation of the alleles at two sites. r^2 could be calculated by dividing D^2 with the product of the four allele frequencies at the two loci, where PA and PB are the common variant allele frequencies and Pa and Pb are the rare variant allele frequencies

$r^2 = D^2/(PA \times PB \times Pa \times Pb)$ (Hill *et al.* 1994)

Even when loci are in complete disequilibrium (D'=1), the pairwise r^2 values can vary widely because they are related to the allele frequencies and to the position of the corresponding polymorphisms in the genealogy. An indication of the increase in sample size needed to have the same power to detect association using an indirect marker is given by $1/r^2$. This demonstrates the need for high values of r^2 to avoid loss of information and power.

1.1.6.3.4 Patterns of linkage disequilibrium

LD is also relevant to the presence of "haplotype blocks" (Daly *et al.* 2001). It has been suggested that genetic loci across large areas of the genome divide into blocks characterised by little LD between blocks with limited haplotype diversity within blocks. These two aspects of blocks, physical extent and haplotype diversity are, in a sense,

reflected by the measures D' and r^2, respectively. However, they are not necessarily linked since they are determined by the different random processes of recombination and mutation. Further haplotypic diversity almost inevitably increases as more polymorphisms are discovered.

There has been some discussion as to whether blocks have clear boundaries, coincident with so called recombination hot spots or whether they arise as a result of purely random forces (Ardlie *et al.* 2002, Wall *et al.* 2003). The debate is important in relation to the stability of block structures across populations and to the sharpness of block boundaries. In general, haplotype blocks are shorter in populations of African ancestry (average 11 kb) than in European or Asian groups (average 22 kb) (Gabriel *et al.* 2002) consistent with the theory that the Euro-Asian population arose from a limited number of founders of African origin approximately 50,000 years ago.

1.1.6.3.5 Association study designs

Association can be tested in case-control studies or using a variety of intrafamilial (case-parent triad design, case-parent-grandparent design or analysis of general pedigrees) tests. Most of the association studies have a case-control study design, in which allele or genotype frequencies in unrelated patients are compared with frequencies in an unaffected control population. This study design is subject to population stratification due to ethnic admixture, which occurs when the cases and controls are unintentionally drawn from two or more ethnic groups. If one of these subgroups has higher disease prevalence than the others, stratification occurs, because that subgroup will be over-represented in

the cases and under-represented in the controls. Intrafamilial tests (including transmission disequilibrium test (TDT) and haplotype relative risk (HRR) are less prone to population stratification (despite not being matched for age and gender). Disadvantages of the TDT are that family based samples are relatively difficult obtain and that 50% more genotyping is required than in case-control studies to achieve similar power. For these reasons, the case-control approach is generally preferred, although, family based approaches provide a useful complementary strategy because of their robustness to population stratification and because they allow estimates of effects due to direct maternal genotype or maternal-fetal interactions and parent of origin (imprinting) effects (Weinberg *et al.* 1998, Sinsheimer *et al.* 2003, Cordell *et al.* 2004).

Some study designs try to reduce the genotyping effort. Thus, in some studies only individuals at the extremes of the phenotype distributions are typed. DNA pooling studies reduce the amount of genotyping by typing DNA pooled from a group of individuals as opposed to genotyping each person separately (Sham *et al.* 2002). With the haplotype tagging approach, genotypes from an initial sample are used to select loci to genotype in the larger sample (Johnson *et al.* 2001). This strategy initially exploits the indirect association approach to gene mapping. Many different methods have been proposed for the choice of such tagging single nucleotide polymorphisms (SNPs) (Johnson *et al.* 2001, Stram *et al.* 2003, Chapman *et al.* 2003; Byng *et al.* 2003; Carlson *et al.* 2004). Consideration of the power to detect indirect association via haplotype tagging SNPs suggests that the important criterion is the coefficient of determination, a generalisation of r^2 to multiple regression models and usually denoted by R^2. This quantity measures the

ability of a set of tag SNPs to predict another dimorphism (Chapman *et al.* 2003; Clayton *et al.* 2004).

1.1.6.3.6 Statistical analysis of association

The methods for analysing data in association studies depend on study designs. In the simplest case, familial methods such as χ^2 for association, odd ratios and logistic regression may be suitable. The simplest and most powerful statistical analyses arise in the direct association studies in which causal hypothesis and hence analyses are specific to a single typed polymorphism. In indirect studies, which exploit LD between typed markers and causal variants, analysis of individual marker loci one sequentially might not be ideal – the r^2 between single marker and the causal locus could be much lower that the R^2 for prediction of the causal locus from a group of markers. For such studies, the multilocus approach to analysis is more desirable.

Multilocus approaches are generally assumed to involve consideration of haplotypes. Analysis at the haplotype level can reveal an effect marked by an ancestral haplotype but has two main drawbacks: since the number of haplotypes could be large, the potential gain might be offset by an excessive increase in the degrees of freedom in the test. In addition, the haplotype phase will often be uncertain.

There are several ways in which to tackle these problems. The simplest and most common is to pool rare haplotypes, which will certainly sacrifice some information. Instead some have proposed grouping strategies based on cladistic considerations

(Seltman *et al.* 2001). However, for markers in regions in which LD is strong, it is questionable whether the use of any haplotype information justifies the increase in degrees of freedom. In such circumstances, a simple multiple regression equation with one parameter per marker can achieve prediction of untyped loci with R^2 only slightly worse than haplotype-based predictions (Chapman *et al.* 2003; Clayton *et al.* 2004). When LD is less strong, haplotype analyses remain important, especially for fine mapping (Cordell *et al.* 2002). The resolution of phase can present a serious practical problem. Various computational algorithms address the phase estimation problem, both in unrelated individuals (Stephens *et al.* 2001, Fallin *et al*. 2000; Clayton *et al.* 2005) and in families (O'Connell 2000). In a two stage procedure (whereby haplotype scoring based on a first haplotype analyis stage are used in a second stage test for association), these algorithms can be satisfactory in population-based studies since significance can be assessed by permutation arguments. In family studies, oversimplified two-stage approaches break down since information about assoicationn is given by transmission patterns, which are relevant to haplotype phase resolution.

Alternatively, as in population-based studies, phase uncertainty can be accounted for explicitly in the association analysis (Clayton 1999). The problem of uncertain phase can be avoided altogether by the use of molecular methods for haplotyping (Michalatos-Beloin *et al.* 1996; Eitan *et al.* 2002) but such methods usually have lower throughput and are more expensive that those that yield only diplotypes.

1.1.6.3.7Effect size, sample size, statistical significance and importance of association studies

Evidence from studies in *Drosophila melanogaster* and from animal models of human disease suggests that there are few loci with large effects and larger number of loci with small effects (Barton *et al.* 2002, Mackay *et al.* 1992) Disease susceptibility variants in humans that have been identified to date have allelic odd ratios in the range of 1.1-3.0 and therefore make small overall contribution to disease risk (Risch *et al.* 1996)

Associations can be real but nonetheless difficult to replicate if the underlying genetic effect is weak. Inadequate sample size is one of the most common criticisms of association studies performed to date and reflects difficulties in sample ascertainment and the pressure to limit genotyping costs. If the subsequent studies are small in size, they will be underpowered to reliably detect weak effects and may therefore fail to achieve real effects. This difficulty is heightened by the "jackpot" effect, in which the first group to publish a significant association involving a weak locus is more likely to have overestimated than underestimated the true effect size of the polymorphism. Subsequent studies thus may therefore need to include much large numbers of patients to achieve statistical significance. For markers with minor allele frequencies of less than 10% and with allelic odd ratios of less than 1.3, unrealistically large sample (greater than 10,000 case-control pairs) are required to provide 80% power at a genome-wide significance level of $p < 10^{-6}$. Sample sizes of a few hundred case-control pairs with allele frequencies greater than 10% provide adequate power to detect allelic odd rations of 2.0 or more only. In addition, none of these estimates take into account high levels of complexity

including gene-gene or gene environment interactions that would further modify the power estimates.

The standards of statistical proof that have become acceptable in the general biomedical literature are not appropriate for genetic association studies. Approaches such as the Bonferroni correction are probably not appropriate because it is not the number of tests in any one investigation that is important. Rather it is that the vast majority of loci tested will not be associated, so that even a small false positive probability will mean that most positive results will turn out to be false. Thus, it is the a-priori probability of association that needs to be accounted for, rather than the number of tests. Therefore, it has been suggested that Bayesian methods are more appropriate (Wacholder *et al.* 2004, Thomas *et al.* 2004) when the prior probability of association is known, since they allow calculation of the posterior probability that an association is genuine. However, this approach requires knowledge not only of the prior probability of association but also of the distribution of the size of effects that will be encountered.

1.1.6.4 Candidate gene studies in rheumatoid arthritis: recent advances

The identification of disease susceptibility polymorphisms in association studies relies on capturing all relevant genetic variation between cases and controls. Genes can be considered as candidates in RA if the following are fulfilled: 1. Genes in question are known to play a role in inflammation/infection. 2. They fall within a linkage interval defined by family studies. 3. They have been identified in previous association studies. 4.

They have been identified using animal models. Taking these four points into an account, it has become apparent that a high proportion of the 35,000-40,000 genes thought to comprise the human genome could, in one way or another, be considered as candidates.

Many candidate genes studies have been performed in RA, but only few have been replicated. Furthermore, it has become clear that some candidate gene polymorphisms are variable between different ethnic groups. Two large and comprehensive association studies on Japanese RA patients have identified two genes to be significantly associated with RA: *PADI4* (on 1p36) and *SLC22A4* (on 5q31) (Suzuki *et al.* 2003; Tokuhiro *et al.* 2003). *PADI4* is an excellent candidate gene as it is responsible for the conversion of positively charged arginine to polar but uncharged citrulline by deamination. The presence of anti-citrulline antibodies is highly specific for RA. Although polymorphism in *PADI4* has been replicated in a Korean cohort of RA patient, association studies in Caucasian population have been week and inconsistent. This may be partly explained by ethnic differences in the frequency of allele variants. Meta-analysis of *PADI4_94* variant (rs2240340) was recently reported in a large collection of North American Rheumatoid Arthritis Consortium (NARAC) and Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) cohorts (n=2,370 RA patients and 1,757 controls) and was found to be associated with the disease in the NARAC group (p=0.001) (Plenge *et al* 2005). Meta-analysis of previously published reports combined with NARAC and EIRA data revealed a small significant result (OR 1.12, 95%CI 1.03-1.23).

Recently a missense SNP in the gene encoding a protein tyrosine phosphatase, *PTPN22* (on 1p13), has been associated with multiple autoimmune diseases. The *PTPN22* gene encodes the protein Lyp, an important negative regulator of T cell activation. The presence of the *PTPN22* 1858 T/T genotype increases the risk for RA more than twofold. This discovery demonstrates effectively that some autoimmune diseases have a common pathway. (Please refer to chapters 4 and 5).

CTLA-4 (on 2q34) is a co-stimulatory factor that acts as a negative regulator of T cell activation. It counteracts the activating T cell signals derived from CD80 activation of CD28 and inducible stimulator (ICOS). *CTLA-4* is related to CD28, but in contrast to CD28, it is not present on resting T-cells, but is expressed several days after T cell activation. Polymorphisms in *CTLA-4* gene have been associated with multiple autoimmune diseases (Nistico *et al.*1996; Donner *et al.* 1997; Heward *et al.*1999). Studies in RA have been controversial, with positive association found in some populations, but not replicated in others (Milicic *et al.* 2001; Barton *et al.* 2004; Lei *et al.* 2005; . Interestingly, in the combined analysis of the US and UK linkage data, 2q34 locus was identified in the early onset RA subgroup. In more recent the Wellcome Trust Case Control Consortium (WTCCC) genome-wide association studies (GWAS) two polymorphisms in the *CTLA-4* (rs3087243 and rs11571300) demonstrated nominally significant association (p=0.085 and p=0.034). *CTLA-4* CT60 variant (rs3087243) was also studied in a collection of 4,127 US NARAC and Swedish EIRA cohorts (Plenge *et al* 2005). Authors reported association of this polymorphism with anti-CCP positive cases in NARAC (p=0.003) but not in EIRA, however after pooling, the results remained

significant (p=0.0006). Meta-analysis of all published studies also confirmed modest effect (OR 1.08 95%CI 1.01-1.17) indicating genetic association.

The Programmed Death (PD-1) molecule is another co-stimulatory factor belonging to the immunoglobulin receptor superfamily and is a negative regulator of T cells (Nishimura *et al.* 2001). PD-1 -/- mice on the C57BL/6 background show spontaneous development of lupus-like glomerulonephritis and destructive arthritis (Nishimura *et al.* 1999). The human gene for PD-1, Programmed Death Cell Domain 1 (*PDCD1*) is localised on 2q37.3, telomeric to CD28/CTLA4/ICOS. Initially 4 SNPs in PDCD1 have been associated with lupus. Recent studies have found independent SNPs in association with RA (James *et al.* 2005; Kong *et al.* 2004).

Association of *STAT4* and RA was recently reported following initial investigations of 13 genes in a 52MB region under a linkage peak on chromosome 2q in US RA families (rs7574865, p=0.002) (Remmers *et al.* 2007). This result was subsequently confirmed by fine mapping of the 209kb *STAT1-STAT4* region and followed by further replication in a large sample of US, Swedish, UK and Korean populations (Lee *et al.* 2007, Barton *et al.* 2008). Association with rs7574865 was also found in a meta-analysis of SLE patients ($p=1.78 \times 10^{-9}$), suggesting that like *PTPN22* this locus may harbour a risk variation for multiple autoimmune diseases (Remmers *et al.* 2007). *STAT4* encodes a transcription factor that transmits signals induced by several key cytokines, including interleukin-12 (IL-12), interleukin-23 (IL-23) and type 1 interferons. There may be multiple mechanisms by which genetic variations in *STAT-4* can influence immune response. IL-

12 (signalling through *STAT4*) has been known to allow T cell lineage commitment towards a Th1 phenotype producing IFN-γ as its signature cytokine (Szabo *et al.* 2000). On the other hand, IL-23 induces IL-17 expression in human T-cells (Hoeve *et al.* 2006).The latter has been shown to activate and enhance all mechanism of tissue injury that has been previously described in RA.

1.1.6.5 Genome wide association studies

The genetic studies of human diseases have enjoyed a surge of recent success due to the advent of GWAS that have enabled to combine two fundamental advantages of previous approaches: candidate gene studies and linkage scans. The timing reflects recent progress in three key areas. First, development of the International HapMap resource with the aim of cataloguing common human genetic variants. Second, advances in technology, leading to the development of dense genotyping chips that have allowed simultaneous genotyping of hundreds of thousand of markers and that has made GWAS technically and financially feasible. Finally collections of appropriately large well-characterised clinical samples have been assembled for many common diseases.

To date four GWAS in RA have been performed. The first, ground-breaking GWAS on British population was undertaken by the WTCCC. In this unprecedented study, analysis of 500,000 SNPs in 2000 samples from seven common complex diseases, including RA and 3000 shared controls was made. This approach enabled not only identification of strong candidate regions (in total 24 independent regions achieving stringent significance thresholds $p < 5x10^{-7}$ across all diseases) but also the identification of common

susceptibility regions between different diseases. Two regions, *HLA* and *PTPN22*, previously well established susceptibility loci for RA, emerged strongly from GWAS ($p < 10^{-7}$). Interestingly RA was the sole disease for which the sex-differentiated analysis revealed different genetic effects of a marker in chromosome 7 near *PODXL* (rs11761231) in males and females. This genotype did not have any effect on disease status in males but had a strong apparently additive effect in females ($p=0.68$ and $p < 10^{-7}$ respectively). Nine furhter loci showed evidence of association with *P* values in the range of 10^{-5}-10^{-7}. None of these nine mapped to loci that have been previously associated with RA. Polymorphisms mapping close to both the alpha and beta chains of the IL2 receptor (rs2104286 for *IL2RA* and rs743777 for *IL2RB*) raised particular interest as good candidates. The IL2 receptor is known to mediate T lymphocyte stimulation and is thought to have an important role in preventing autoimmunity. Further 49 SNPs with nominal significance in the range 10^{-4} to 10^{-5} were found that map to genes with plausible biological relevance (*TNFAIP3* in TNF pathway, *GZMB* and *PRKCQ* in T-cell function regulation and *KAZALD1* in bone regeneration).

Initial GWAS by WTCCC was quickly followed by publication of two more GWAS conducted North-American and Swedish RA cohorts (Plenge *et al.* 2007 a,b). These studies identified and replicated *TRAF-C5* (Chr.9, rs3761847, $p < 10^{-7}$) and TNF induced protein 3 (*TNFAIP3*) (Chr.6, rs10499194 $p < 10^{-7}$) as new genetic loci strongly associated with anti-CCP positive RA.

A subsequent replication study to validate the association of the nine polymorphisms that showed strong evidence for association with disease in the initial WTCCC was conducted in 5,063 UK RA patients and 3,849 controls (Thomoson *et al.* 2007). One polymorphism (rs6920220) in the intergenic region of 6q23 was unequivocally replicated ($p < 10^{-7}$) in anti-CCP and RF positive RA patients. This SNP maps between oligodendrocyte lineage transcription factor 3 *OLIG3* and *TNFAIP3*, the region that was also independently identified in North-American and Swedish GWAS. This block itself contains no known genes or transcripts but encompasses part of the *PTPN11* pseudogene. It has been suggested that up to 20% of pseudogenes may be transcribed, but no functional effects so far has been demonstrated (Zeng *et al.* 2007). It is possible the polymorphism regulates function of *TNFAIP3.* The latter is more attractive candidate compared to *OLIG3* as it has been shown to negatively regulate the transcription factor NF-kB in response to TNF and toll-like receptor activation (Boone *et al*. 2004, Wertz *et al.* 2004). Imputed data analysis of four SNPs in the *TRAF-C5* region from the WTCCC did not reveal any significant association (Barton *et al.* 2008). Further independent genotyping in the UK replication cohort (> 3,000 RA cases and > 3,000 controls) and performing combined analysis with WTCCC confirmed the association (rs10760130, OR 1.08 95%CI1.03-1.14). Although the association appeared stronger in the anti-CCP positive group, the direct comparison with anti-CCP negative group revealed no statistically significant difference.

More recently, a fourth GWAS was carried out on Spanish RA patients. This identified and replicated polymorphisms in the member of the Kruppel-like transcriptional regulatory factor (*KLF12*) (rs1887346 and rs9318228) as an additional new locus

strongly associated with RA ($p < 10^{-5}$) (Julia *et al.* 2008). These polymorphisms were modestly associated with previously reported variations in WTCCC (rs1887346, p=0.005) and BRASS (rs9318225, p=0.002) RA cohorts. *KLF12* is known to suppress the transcription of activator protein 2α (*AP-2α*) (Schierer *et al.* 2001). *AP-2α* is known to regulate expression of *TNF* genes involved in malignant transformation and stress responses (Kannan *et al.* 1994, Wajapeyee *et al.* 2003). An independent replication study confirming this result in a bigger sample is desirable. No association with *TRAF1-C5* was confirmed but the samples were not stratified according to CCP status. Further epistasis analysis in Spanish cohort revealed several SNP pairs to be significantly associated. However, no previously known strong candidates like HLA class II variants seem to be the main contributors. Further studies are needed to clarify the complex gene-gene interactions.

1.1.6.6 Genetic models of rheumatoid arthritis risk

The SE hypothesis put forward to explain disparate *HLA-DR* association with RA (Gregersen *et al.* 1987) implies the presentation of a limited range of potentially arthritogenic peptides by antigen presenting cells within the joint to a subset of T cells capable of initiating a specific inflammatory response. This would produce a dominant model of susceptibility but several studies have shown that the opposite is the case.

Rigby *et al.* applied the Antigen Gentotype Frequencies among Patients (AGFAP) method in 111 multiplex families to investigate the *HLA-DRB1* component of RA and found that for *HLA-DRB1*04* and *HLA-DRB1*01* the AGFAP data were close to a

recessive model. A dominant mode of inheritance where the majority of probands are heterozygotes was rejected (Rigby *et al.* 1991). Similar results were obtained by an independent method applied by the same authors and based on calculations of *HLA-DR* transmission from affected parent to the affected children. The deviation from random expectation was observed but data were insufficient to discriminate between recessive and dominant modes. The effect of the HLA-linked locus on familiality accounted for only a 1.61-fold increased risk to sibs over the population prevalence, compared to an overall sibling recurrence risk of 3.90 (Deighton *et al.* 1989). The authors concluded that at least one other non-HLA locus must exists that predisposes to RA. Similar conclusions have been reported by several other authors using a variety of statistical methods; the exception is a study by Wallin *et al* (Deighton *et al.* 1993, Evans *et al.* 1995, Lynn *et al.* 1995, Wallin *et al.* 1991). In the latter paper using the same AGFAP method the authors demonstrated dominant f inheritance for HLA-linked RA. These differences were explained by the possibility of disease heterogeneity. It would be argued that none of these studies were adequately powered or designed to analyse modes of inheritance definitively.

In the subsequent study Rigby *et al*. investigated the risk of RA based on the *HLA-DRB1* genotypes of 85 affected sibships (Rigby *et al.* 1998). The highest genotype specific risks were seen for sibs sharing two haplotypes with either a *HLA-DRB1*0401/0404* (12.5%, 95%CI 6.9-15.2) or a *HLA-DRB1*0401/*0408* (11.1%, 95%CI 4.5-15.1) proband. The *HLA-DRB*0404* and **0408* alleles differ by an amino acid substitution (Val→Gly) at position 86 (Weyand *et al.* 1992) and this could lead to different disease risk.

Another model that has been applied to investigate the effects of HLA on disease susceptibility is the Marker Association Segregation Chi-Square Method (MASC) (Clerget-Darpoux *et al.* 1988). This method is designed for modelling and testing the role of a candidate region and has advantages from AGFAP. It uses two kinds of information on markers in this region: allelic association between the marker and the disease susceptibility locus through genotype distribution among unrelated index cases and linkage between marker and disease through haplotype sharing between index case and one affected sib. The data are categorized into appropriate classes and the expected distribution in each category under a given genetic model computed. The observed distribution is compared to this expected distribution using the χ^2 test. Using the MACS method Genin *et al* investigated the role of *HLA-DRB1* and *TNF-LT* loci in 49 multiplex families. When only *HLA-DRB1* data was used, the MACS result demonstrated that the SE did influence risk for RA and under a recessive model, similar to AGFAP. When the *TNF* data was added to the model, both SE and recessive modes of inheritance were rejected. The authors noted that the result obtained by analysing the *HLA-DRB1* data was very dependent on the frequencies of the alleles at this locus. They also conclude that by using addition *TNF* data provided further information that allowed more precise discrimination between genetic models of RA than previously possible with only *HLA-DRB* data.

For complex human disease like RA, a simple mode of genetic inheritance is not operant, and indeed multiple contributing genetic loci are likely to be involved. In this setting, whole genome models of disease risk have been developed. In early 1990s, Risch

demonstrated that the familial recurrence risk ratio, lambda R (λ_R), (which is defined as the risk ratio for a relative of a proband compared to the population prevalence) can be used to evaluate patterns of inheritance in genetically complex diseases. By using Risch's theory Rigby *et al.* examined several multiple locus models of inheritance in RA. Low sibling recurrence risk of RA compared to the general population, generally "late" onset of the disease, sparse published risk estimates in affected kinships in RA are the main obstacles to analysis and should be taken into an account when interpreting the results of Risch theory in RA. Given the paucity of data, the theory supported at least two-locus model of inheritance, assuming that one locus in HLA and suggested that the true risk ratio in siblings is more likely to appear at the lower end of the spectrum.

1.1.7 Conclusion

Rheumatoid arthritis is clinically heterogeneous disease with multifactorial aetiology, intricate pathogenesis and complex pattern of inheritance. With the advent of WGAS and improved technologies further progress will be made in revealing and clarifying pathogenesis of the disease mechanisms, leading to the development of better diagnostic, therapeutic and prognostic methods.

This D.Phil project was designed as an association study and aimed to investigate several candidate gene regions with potential functional importance, in relation of RA susceptibility and severity in a well characterised, large UK Caucasian patient sample.

Chapter 2

Methods and Materials

2.1 Laboratory methods

Generally applicable laboratory methods used in this thesis are described in this chapter. More specific analytical methods are described in the individual experimental sections were appropriate.

2.1.1 DNA extraction and quantification

DNA extraction from the whole blood of the first 500 RA patients was performed by Oxagen using "Puregene" kit (Gentra Systems, Minnesota, USA). The rest of the patient DNA was extracted in the Botnar Research Centre using standard salting out method.

Peripheral venous blood (20ml) was collected into BD EDTA Vacutainer® tubes (BD, USA) and stored at -20°C prior to proceeding to the DNA extraction. 10ml of defrosted blood was mixed with 25ml of cold Erythrocyte Lysis buffer (EL buffer) (see section 2.4) in the 50ml Falcon tubes, held in room temperature for 10min, occasionally mixing by inversion. The mix was centrifuged at 2500 x g for 10min. The supernatant was removed and the remaining pellet was re-suspended in 20ml of EL buffer and vortexed until the pellet was completely dispersed, prior to a further centrifugation for 10min at 2500 x g. The pellet was then re-suspended in 500µl of 5 mM sodium perchlorate. 2ml of solution

B (section 2.4) was added to the solution and incubated at 65°C for minimum 25min. The tubes were then cooled to room temperature; 2ml of chloroform ($CHCl_3$) was added and mixed well by inversion and centrifuged at 3000 x rpm for 10min. The upper phase was collected into the clean 50ml Falcon tube, using Pasteur pipette and 2.5ml cold 100% ethanol was added. The tubes were mixed well by inversion in order to induce DNA precipitation. The DNA was then spooled out on the end of a Pasteur pipette and allowed to air dry for a moment or two and then re-suspended in 500µl of 1 x Tris-EDTA (TE buffer) (section 2.5).

The extracted DNA was quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA). 1.5µl of DNA was used to ensure that the liquid sample column formed and the light path was completely covered by the sample. The sample concentration is calculated based on absorbance at 260nm and selected analysis constant according to Beer's law:

$$A = E*b*c$$

Where A is the absorbance presented in absorbance units, E is the wavelength-dependent molar absorptivitycoefficient with unit of litre/mol-cm , b is the path length in cm and c is the analyte concentration in moles/litre or molarity (M).

The calculation is done automatically by software and the sample concentration is presented in ng/µl.

2.1.2 RNA extraction from peripheral blood mononuclear cells (PBMC) and cDNA synthesis

2.1.2.1 PBMC separation

PBMC of 12 RA patients were isolated from fresh heparinized blood by Ficoll-Hypaque density gradient centrifugation. Briefly, 18ml of diluted cell suspension was laid over Ficoll-Paque (Ficoll-Paque^{TM}Plus, Amersham, Biosciences, UK) and centrifuged at 2300 x g for 25min at 4°C. Cells were collected from the interface, washed twice in MACS running buffer (Miltenyi Biotec, UK) and finally suspended in 1ml MACS running buffer. The cell number was determined after lysis of red cells with a 5% v/v acetic acid solution and using tryptophan blue. The following formula was used to calculate the cell number:

No of cells= cell number in 5 squares in the row x 5 x 10 x 10^4 x 2

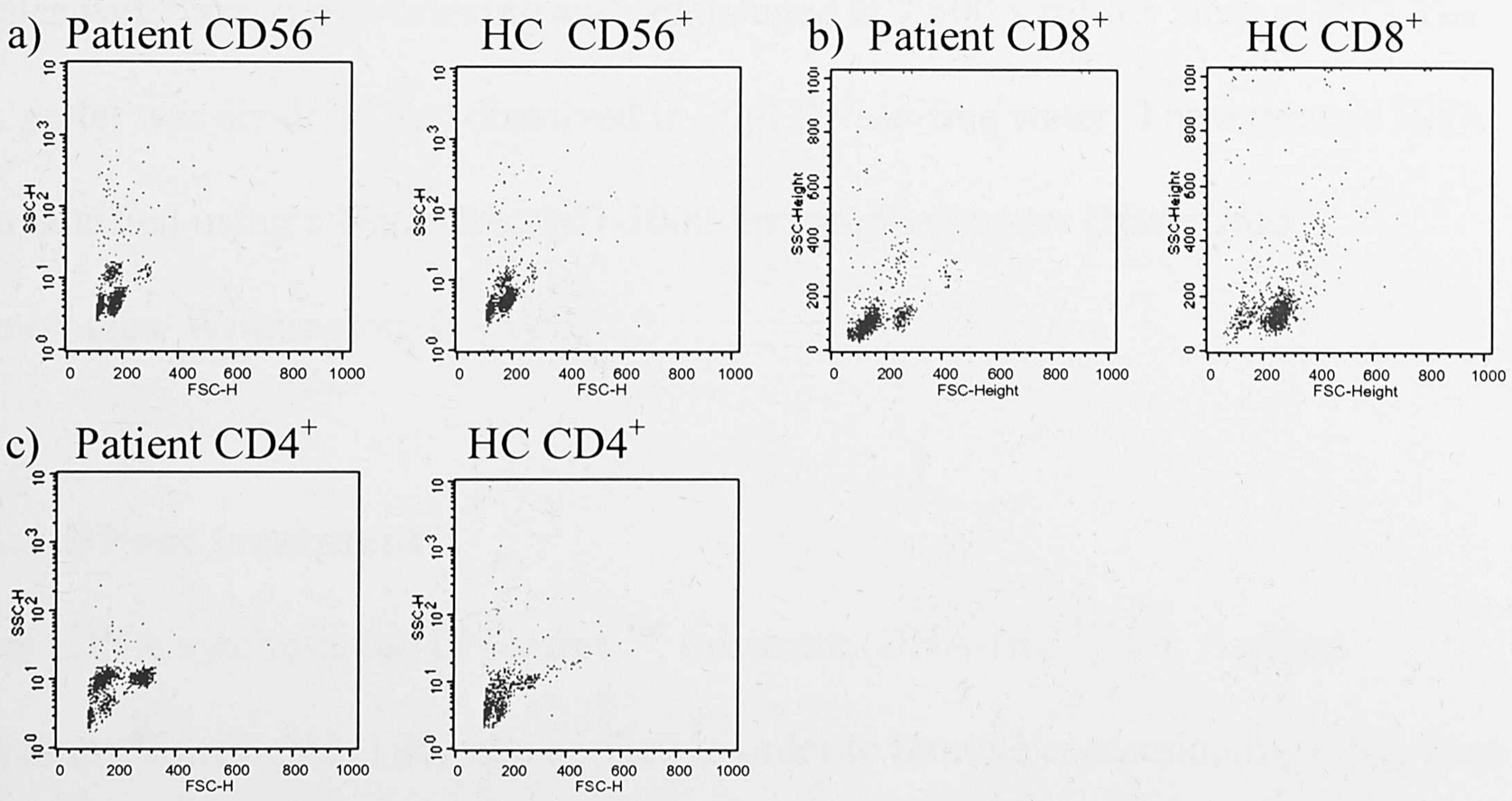
Further cell separation was then performed according to the manufactures guidelines. Briefly, cell suspension was centrifuged at 1500 x g at 4°C for 10min. After removing the supernatant completely, the cell pellet was re-suspended in 80µl of MACS running buffer per 10^7 total cells and 20µl of microbeads per 10^7 cells was added (Miltenyi Biotec, UK). The bead-cell mix was incubated for 15min at 4-8°C. Cells were washed by adding 2ml of running buffer per 10^7 cells and centrifuged at 1500 x rpm for 10min. The supernatant was removed and cells were suspended in 500µl of running buffer and the magnetic separation was performed using AutoMax separator (Miltenyi Biotech, UK). Cells were sorted by positive selection and in the following order: $CD56^+$, $CD8^+$ and $CD4^+$. The

number of cells was counted after each separation. 20µl of cell suspension was put aside for flow cytometry, the rest was used for RNA extraction.

2.1.2.1 FACS analysis

Each aliquot of 20µl sorted cells were re-suspended in 80µl MACS running buffer and labelled with specific flow cytometry antibody (CD56-PE, CD8-FITC, CD4-PE) or with its negative control (Mouse IgG1-PE or Mouse Ig2A-FITC isotype) (Miltenyi Biotech, UK). The tubes were incubated in the dark for 30min at 4°C, washed once in MACS running buffer and re-suspended in 300µl PBS. The level of fluorescence was measured using a FACSCalibur flow cytometer (Becton Dickinson). Example of the FACS analysis and a comparison with a healthy volunteer is presented in Figure 2.1.

Figure 2.1 FACS analysis of a RA patient and matched healthy control. The results show overall similarity in separated cell counts in a RA patient and control sample. $CD56^+$(a), $CD8^+$(b) and $CD4^+$(c) cells.



2.1.2.3 RNA extraction and quantification

Total RNA was isolated from freshly purified $CD56^+$, $CD8^+$, $CD4^+$ and the remaining PBMC according to the protocol for Trizol Reagent (Invitrogen, Paisley, UK). Cells were pelleted by centrifugation at 4,900 x rcf for 5min. Cells were then lysed in 800µl of Trizol reagent by repetitive pipetting The homogenized sample was incubated at 30°C to permit the complete dissociation of nucleoprotein complexes. 160µl of chloroform was added. The tubes were capped and shaken vigorously for 15sec and incubated at room temperature for 2 to 3 minutes. The samples were centrifuged at 12.000 x rcf for 15min at 8°C. Following centrifugation, the mixture separated into a lower red, phenol-chloroform phase, an interphase, and a colourless upper aqueous phase. RNA is in the aqueous phase. This aqueous phase was transferred to a fresh tube and the RNA was precipitated by mixing with 400µl isopropyl alcohol. The samples were incubated at room temperature for 10min and centrifuged at no more than 12,000 x rcf for 10min at 8°C. The supernatant was removed and the RNA pellet was washed in 500µl of 75% ethanol. Samples were mixed by vortexing and centrifuged at 7,500 x rcf for 5min at 8°C. The RNA pellet was air-dried and dissolved in 20µl RNase-free water. The extracted RNA was quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA).

2.1.2.4 DNase treatment

Before cDNA synthesis the DNA-free ™ treatment (DNA-free ™ kit, Applied Biosystems Warrington, UK) was applied in order to remove contaminating DNA from the RNA preparations. In brief, 0.1 volume of 10x /DNase I buffer and 1µl rDNase I was

added to 8µl of RNA and mixed gently. The tubes were incubated at 37°C for 20-30min. 2µl of re-suspended DNase Inactivation reagent was added and mixed well. The tubes were then incubated at room temperature for 2min and mixed occasionally. Finally the tubes were centrifuged and 10.000 x rcf for 1.5min and the RNA was transferred to the fresh tube.

2.1.2.5 cDNA synthesis

For the cDNA synthesis, SuperScript ™ kit was used according to the manufactures guidelines. (Invitrogen, Paisley,UK). Approximately 800ng of RNA was incubated with 1µl of oligo$(dT)_{20}$ and 1µl 10mM dNTP mix at 65°C for 5min. To each reaction, 4µl 5x First-Strand-Buffer, 1µl 0.1M DTT, 1µl RNaseOUT ™ was added. Two reverse transcription (RT) reactions were performed for each cell population of each patient: with reverse transcriptase (+RT) (SuperScript ™ III RT) and without reverse transcriptase (-RT). The tubes were incubated at 55°C for 60min and 70°C for 15min. Finally 2units of *E.coli* RNase H was added to each tube and samples incubated at 37°C for 20min.

2.1.3 Polymerase chain reaction (PCR) design

2.1.3.1 Primer design

Specific oligonucleotide primers were designed using Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) web-based search program and checked for specificity by comparison with the human genome sequence using the National Centre for Biotechnology Information nucleotide-nucleotide Basic Local Alignment Search Tool (BLAST) facility for short, nearly exact matches

(http://www.ncbi.nlm.nih.gov/BLAST/). Primers were generally between 17 and 25 base pairs (bp) in length and were chosen if GC content did not exceed ~ 50-60% and primer melting temperatures were similar (difference ≤ 5°C).

Primers for RFLP investigation of SNPs were designed to give a 200 to 300bp product, with the nearer one end, allowing a cleaved uneven product (± 80bp) to be visualised by agarose gel electrophoresis (see section 2.1.4). Sequencing primers were generally designed to give a product < 600bp. Occasionally, a product exceeding 1000bp was generated. In this case, internal primers were designed, to ensure full sequence coverage. Primers for exonic sequencing were positioned ~ 100bp from the intron-exon boundaries to ensure full coverage of exonic and intron-exon boundaries.

2.1.3.2 Restriction Fragment Length Polymorphism (RFLP) design

Sequence containing the polymorphism of interest (~ 200bp either side) were obtained from genomic browsers, mainly from Ensembl database (http://www.ensembl.org/Homo_sapiens/). Appropriate restriction enconucleases were identified by NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/index.php). The program finds natural restriction sites and checks for restriction endonucleases that cleave more than once in designated sequence. If no natural restriction site existed for the SNP of interest, the sequence variant was entered into dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html) a program which designs mismatched primers in order to "force" a restriction site (Neff *et al.* 2002).

2.1.4 DNA amplification by PCR

PCR formed the basis of all subsequent polymorphism detection and genotyping and was performed using standard materials and methods, as described by *Sambrook et al.* (Sambrook *et al.* 2001).

Primers were synthesised commercially by MWG Biotech (Ebersberg, Germany) and Applied Biosystems (Warrington, UK). Primers were diluted in dH_2O according to the manufacturer's guidelines. Primers were stored at -20°C at a stock concentration of 100µM and a working concentration of 10µM. Optimal PCR conditions for each primer pair were established by amplifying over a range of annealing temperatures and different magnesium chloride concentrations under standard PCR conditions. A variety of different *Taq* polymerases were used and this reflected the requirement for hot-start assay as well as cost.

In general, 50ng of DNA was used for PCR reactions. The PCRs were performed in a 10µl or 15µl reaction volume for the SNP genotyping and 50µl for sequencing reactions. The PCRs were performed in 96 well ThermoQuick PCR plates (Greiner Bio, Stonehouse, Glosterchire, UK) overlaid with mineral oil (oil was not used for sequencing PCR reactions, instead heated leads were applied), in Hybaid Multibloc Thermal Cyclers

.

The success of each reaction was assessed by electorphoresis of 10-15µl of product through an agarose gel .Two per cent agarose gels were most commonly used and prepared by dissolving either 3g agarose (Sigma, Poole, UK) in 150ml 1 x tris-borate-

EDTA (TBE, section) or 10g of agarose in 500 ml 1 x TBE and boiling until dissolved in a microwave. Ethidium bromide solution (0.01% ethidium bromide (Sigma, Poole, UK) was added. Equal volumes of loading buffer (25% blue dextran 50 mg/ml (Applied Biosystems, Warrington, UK), 25% glycerol (Sigma, Poole, UK) and 50% dH_2O) and PCR product were combined prior to loading. 100bp ladder (New England Biolabs, Hitchen, UK) was used as a size standard. Gels were electrophoresed in 1 x TBE at 200 V for 30-60 minutes and then visualised by UV light transillumination, on an AlphaImager (Flowgen, Staffordshire, UK).

Images of the gels were printed using a videographic printer.

2.2 Genotyping

2.2.1 SNP genotyping

2.2.1.1 Amplification Refractory Mutation Systems (ARMS) PCR

ARMS PCR is a simple, rapid and reliable method for detecting any mutation involving single base changes or small deletions. ARMS PCR is based on the use of sequence-specific primers that allow amplification of test DNA only when the target allele is contained within the sample and will not amplify the non-target allele. Following an ARMS reaction the presence or absence of a PCR product is diagnostic for the presence or absence of the target allele. Additional deliberate mismatches were commonly introduced at the penultimate base of the ARMS primer to increase the specificity of the ARMS reaction. In order to distinguish the true absence of target allele from PCR failures, control primers were included in each reaction (forward: 5'-

TGCCAACTCCAGCACCCAA-3'; reverse: 5'-GCATCTTGCTCTGTGCAGAT-3')

which amplify a 796bp region of the *HLA-DRB1* gene.

The PCR reaction took place in 96 well microtitre plates under standard PCR conditions described in section 2. Each reaction was done in duplicates and all 96 well plates included negative controls (water) to identify DNA contamination. PCR products were run on 2 to 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light. All genotypes were read twice and results cross-checked.

2.2.1.2 Restriction Fragment Length Polymorphism (RFLP)

RFLP is another reliable, rapid and cost effective method of SNP typing. The design of the RFLP has been discussed already in section 2. The reactions were carried out in 96 well microtitre plates under standard PCR conditions. The PCR products were tested initially by running an aliquot of a few samples on a 2% agarose gel. The remaining PCR products were digested overnight at temperature according to manufacturers (New England Biolabs, Hitchen, UK) guidelines that were appropriate for the enzyme used. Products were separated by electrophoresis on 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light. Each plate included two negative controls (water) in order to identify PCR contamination. All genotypes were read twice and results cross-checked.

2.2.1.3 SNP typing by KBiosciences using Taqman® Assay

An additional 441 unaffected parents from ankylosing spondylitis families were included and analysed as healthy controls for the association study on the *MHC2TA* -168 A/G polymorphism (rs3087456).This work is discussed in Chapter 6. The additional samples were genotyped by KBiosiences (www.kbioscience.co.uk; Hoddesdon UK) using TaqMan® assay.

TaqMan ® assay is a high throughput method of SNP typing that uses 7900HT Fast Real-Time PCR System (Applied Biosystems, Warrington, UK). The probes that are used to detect specific sequences are n nucleotides long and their 5' ends are labelled with reporter dyes that are either VIC or FAM. At the 3' end the probes contain fluorescent quenchers. When the probe is intact, the reporter dye remains close to the quencher, resulting in suppression of the reporter fluorescence. During the extension phase of the PCR, the probe hybridizes to its complementary sequence, making it possible for the Taq DNA polymerase to cleave the reporter dye from the 5' end thus allowing it to fluoresce. During the entire amplification process the light emission increases exponentially and the final level is measured by fluorimetry after termination of the PCR. The increase of the fluorescence intensity of the reporter dye is only achieved when the allele specific probe has hybridised with its target sequence and amplification has occurred.

2.2.1.4 Sequencing of the PCR products

Following the initial DNA amplification described in section 2.1.4, the PCR products were purified in order to remove all unincorporated dNTPs and primers using a vacuum

filtration system on a membrane plate (Millipore Manu 3050 Filter Plate, Stonehouse, UK). The clean PCR products were then re-suspended in 50µl dH_2O and taken forward for the sequencing reaction.

To the 1-9µl of purified PCR template (the amount of PCR needed to take forward depended on the intensity and size of the PCR product) 2µl of BigDye Terminator read mix (Applied Biosystems, Warrington, UK), 2µl 5x BigDye Sequencing Buffer (Applied Biosystems, Warrington, UK) and 4µl of forward or reverse primer was used; the mix was made up to 20µl by adding the dH_2O. The sequencing reactions was performed in Hybaid Multibloc Thermal Cyclers using the following cycling parameters: 96°C for 5min for the denaturation and enzyme activation this was followed by 25 cycles of 96°C for 10sec (denaturation phase), 50°C for 5sec (annealing phase) and followed by 60°C 4min (extension phase).

The sequencing reactions were subsequently cleaned from unincorporated BigDye Terminator mix using Sephadex. Sephadex G-50 Superfine (Amersham Biosciences, Little Chalfont, UK) is a bead gel that had been prepared by cross-linking dextran with epichloroldyrin. The Sephadex solution (0.5g/ml) was previously prepared and 300-400µl pipetted on each well of membrane plates (Millipore Manu 45 Filter Plate, Stonehouse, UK). The plate was centrifuged at 4000 rpm (ALC International Rome, Italy) for 4sec in order to remove the excess dH_2O. The total volume (20µl) of sequencing reaction product was loaded to the Sephadex-filled Manu 45 plate and the plate was centrifuged at 4000 x rpm for 10min (ALC International Rome, Italy), the purified product was collected in the

96 well non-skirted microtitre plates (ABgene, Epsom, UK) and 10µl of Hi-Di Formamide (Applied Biosystems, Warrington, UK) was added. The samples were then loaded on an ABI 3700 semi-automatic sequencer (Applied Biosystems, Warrington, UK).

The sequencing results were compared with the template sequence that was derived from the online genome browser (http://www.ensembl.org/Homo sapiens/) and analysed using SeqScape® V3.5 software (Applied Biosystems, Warrington, UK).

2.2.2 *HLA-DRB1* genotyping

Limited *HLA-DRB1* typing (*DRB1*01-16*) was performed on all RA cases using a sequence based method that involved eight group-specific PCR amplification reactions, as previously described (Jaakkola *et al.* 2006). Primer sequences and PCR conditions are presented in Table 2.1.

PCR products were separated on a 3% agarose gel stained with ethidium bromide and visualised under ultraviolet light. Representative samples were subsequently sequenced using M13-21 Big Dye primer sequencing kit (Applied Biosystems, Warrington, UK). The details of the sequencing have been described in section 2.2.1.4. All samples were analysed in duplicates and positive and negative controls were included.

Table 2.1 Limited *HLA-DRB1* typing primer sequences and PCR conditions.

HLA-DRB1	Primer	Sequence	$MgCl_2$	Annealing Temp (°C)
HLA-DRB1*01	F	5'CAGTGTCTTCTCAGGTGGCT	1.5	60
HLA-DRB1*15/16	F	5'GGCCGCCTTGTGACCGGATG	2.5	55
HLA-DRB1 *03/08/11/13/14	F	5'GCCTCAGGAAGACAGAGGAG	1.5	60
HLA-DRB1*04	F	5'CTTGGGATCAGAGGTAGATTTT	2.5	55
HLA-DRB1*07	F	5'CGGCGTCGCTGTCAGTGTT	2.5	60
HLA-DRB1*09	F	5'CAGTTAAGGTTCCAGTGCCA	2.5	60
HLA-DRB1*10	F	5'CCCACAGCGTTCTTGGAGG	2.5	55
HLA-DRB1*12	F	5'AGTGTCTTCTCAGGACGCCA	2.5	60
REVERSE M-13-21 TAG	R	5'TGTAAAACGACGGCCAGTGCCG CTGCACTGTGAAGCTCTC		

2.2.3 *HLA-DRB1*04* sub-typing

Further sub-typing was performed in *HLA-DRB1*04* positive patients. Primer sequences and PCR conditions are presented in Table 2.2 and 2.3.

Table 2.2 Primers for high resolution *DRB1* typing.

Primer Code		Primer sequence	PCR fragment (bp)
491	R	5'-CTCCGTCACCGCCCGGT	139
2512	R	5'-CACCGCGGCCCGCCTCT	208
2510	R	5'-GGCTGTTCCAGTACTCGGCATC	173
501	R	5'-CTGCAGTAGGTGTCCACCG	222
498	R	5'-GTCCACCGCGGCCCGCT	211
37	R	5'-CTGCACTGTGAAGCTCTCAC	259
38	R	5'-CTGCACTGTGAAGCTCTCCA	259
500	R	5'-TGCAGTAGGTGTCCACCT	221
57	R	5'-TGTTCCAGTACTCGGCGCT	170
2516	R	5'-TCCACCGCGGCCCGCC	210
47	F	5'-GTTTCTTGGAGCAGGTTAAACA	

Table 2.3 *HLA- DRB1*04* sub-typing calculation table.

	0401	0402	04031	04032	0404	0405	0406	040701	040702	04073	0408
491	x	x	x	x	x	x	x	x	x	x	x
2512	x		x	x	x	x	x	x	x	x	**x**
2510	x	x	x		x		x	x	x	x	x
498	x	x									
501	x	x			x	x					x
37	x					x		x	x	x	x
500			x	x			x	x	x	x	
38		x	x	x	x		x				
57						x					
2516					x	x	x	x		x	x

2.2.4 Differential allelic expression (DAE) analysis by SNaPShot

cDNA was synthesised from the PBMC populations of patients who were heterozygous at *PTPN22* C1858T polymorphism (see section 2.1.2). From each +RT reaction, 12 individual PCR amplifications were carried out using forward primer 5'-CAACTGCTCCAAGGATAGATGATG and reverse primer 5'-CTTGCTTGGTCTAAGTATCACAGAG. These primers were spanned over exon 13 and 14 producing a PCR product of 215bp for cDNA. The –RT controls did not yield any detectable PCR products. The primer, which is located immediately adjacent to *PTPN22* C1858T polymorphism, was used for the single base extension assay (SBE). To obtain an accurate estimate of the expression peak for an assumed 1:1 ratio between alleles, I performed five individual genomic DNA (gDNA) measurements. The same PCR primers and SBE were used on the gDNA of each of the 12 patients to give a total of 56 gDNA measurements (4 PCR reactions failed for technical reasons).

PCR amplification was carried out in a total reaction volume of 15µl, containing 1x PCR buffer, 0.1mM dNTPs, 0.20mM primers and 0.1 IU amplitaq "Taq Gold" polymerase (Qiagen, Crawley, UK). Thermocycling conditions consisted of an initial denaturation step of 95˚C for 10 minutes, followed by 35 cycles of 95˚C for 30sec, annealing of 58˚C for 30sec and 72˚C for 30sec, with a final extension step of 72˚C for 10min.

Amplified samples were incubated with 1.5U shrimp alkaline phosphatase (Amersham. Bucks UK) and 2U exonuclease I (Amersham) for 60min at 37˚C followed by 15min at 75˚C prior to primer extension reactions. These products then served as the template used to test for DAE by single base extension (SBE) assay using the SNaPshot Multiplex Kit (PE Applied Biosystems, Warrington, UK). SBE involved the extension by a single base of a primer located immediately downstream of the SNP of interest (Figure 2.2A).

Reactions were performed in a total volume of 10µl, containing 1.2µl treated PCR product, 1.5µl SNaPshot kit and 0.1µM extension primer. Primer extension termocycling conditions were as following: 25 cycles of 95˚C for 5sec, 43˚C for 5sec and 60˚C for 5sec. Primer extension reaction products were treated with 1.0 shrimp alkaline phosphatase (Amersham) for 60min at 37˚C and then for 15min at 75˚C. Samples were further diluted and 1µl of 1:100 diluted Snapshot reaction products were combined with 9µl Hi-Di formamide. As a size standard, Liz 125 was used. The fluorescently labelled extension products were separated by capillary electrophoresis using the 3100 DNA Genetic Analyser (Applied Biosystems). The data was analysed using GeneMapper 3.5 software (Applied Biosystems) that produced electropherograms that listed peak heights

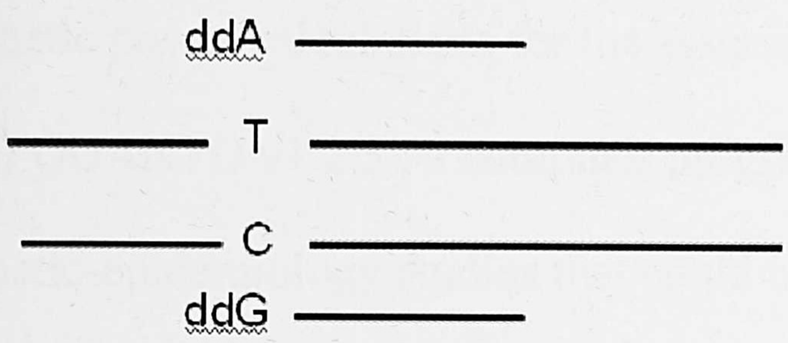
to the two alleles, which are proportional to the amount of each allele present (Figure 2.2B)

Each gDNA sample was assayed alongside the corresponding heterozygous cDNA. Using the same analytic conditions for the cDNA and gDNA measurements enabled me to use the average of the 56 genomic DNA allelic ratio measurements (representing the assumed 1:1 ratio between alleles) to correct the allelic ratios obtained form the cDNA measurements and thus to account for differences in fluorescent yield and terminator dye incorporation specific to the assay. The corrected allelic ratio was calculated using the following equation:

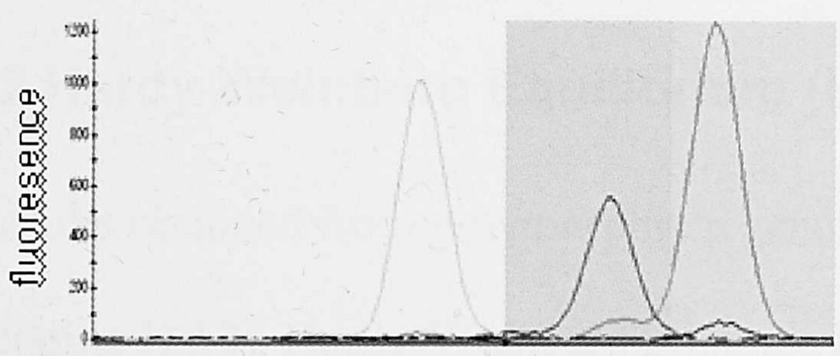
corrected allelic ratio=cDNA peak height ratio (or genomic DNA peak height ratio)/average genomic DNA peak height ratio.

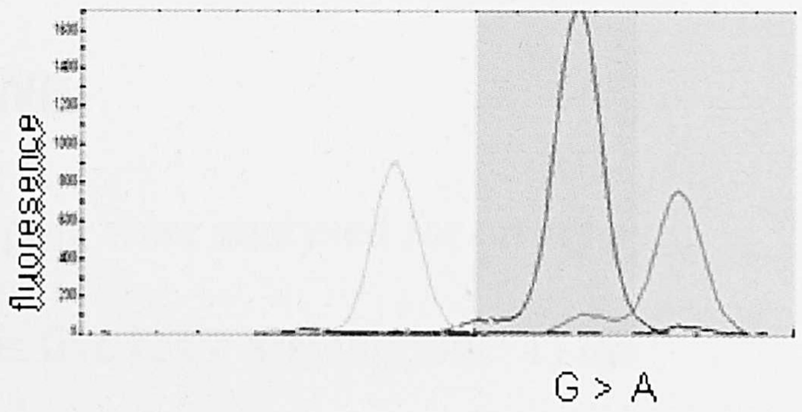
Figure 2.2 Schematic overview of the differential allelic expression assay. (A) For the single base extension reaction, a primer is annealed just downstreatm of the transcribed SNP and extended into the polymorphic site with fluorescently labelled dideoxy nucleotides (ddNTPs). (B) These extension products are separated by gel electrophoresis and analyzed to produce electropherograms giving the peak heights for the two alleles. Scenarios are depicted by more message from the T allele relative to C allele (G<A), more message from the C allele relative to T allele (G>A) and equal amounts of message from both alleles (G=A)

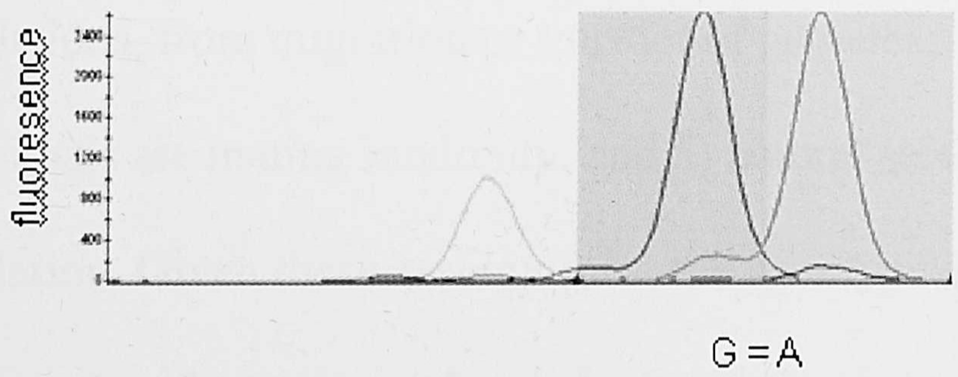
A



B







2.3 Statistical methods

A variety of statistical methods and statistical packages were applied in this thesis. The reasoning behind the application of each method for each project is outlined in the relevant chapter.

2.3.1 Power calculation

Genetic power calculations for the association studies were performed by using the online tool QUANTO v1.2.3 , a computer program for power and sample size calculations for genetic-epidemiology studies that could be found at http://hydra.usc.edu/gxe (Gauderman *et al.* 2006). User defined variables (sample size, allele frequencies) were adjusted accordingly for each particular study.

2.3.2 Hardy-Weinberg Equilibrium (HWE)

All results obtained from polymorphism genotyping were analysed for errors by calculating HWE. Hardy-Weinberg principle has five basic assumptions: 1) the population is large (i.e., there is no genetic drift); 2) there is no gene flow between populations, from migration or transfer of gametes; 3) mutations are negligible; 4) individuals are mating randomly; and 5) natural selection is not operating on the population. Given these assumptions, the genotype and allele frequencies in a given population will remain unchanged over successive generations.

For example the equation for allele frequencies when the common allele is p and the rare allele is q, where

$$p + q = 1$$

and the equation for genotype frequencies

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

Differences between the observed and expected frequencies are calculated by χ^2 test.

2.3.3 GOLD

Graphical Overview of Linkage Disequilibrium (GOLD) program is designed to give a graphical summary of pair-wise measures of LD. This program could be found at http://sph.umich.edu/csg/abecasis/GOLD/ (Abecasis *et al.* 2000). The program uses an input in the form of founder haplotype estimates and calculates commonly used genetic measures of LD (such as D, D' and Δ^2) Finally, it provides graphical overview of disequilibrium patterns in a form of colour plot. The horizontal and vertical axes are scaled according to physical (or genetic) distances between the studied markers. For each marker, the pair-wise disequilibrium statistics are colour coded (bright red and dark blue represent opposite ends of the scale).

2.3.4 Association analysis

Alleles and genotypes were compared between cases and controls and further analysis incorporating only RA cases stratified by clinical and laboratory parameters were by χ^2 or Fisher's exact test implemented in SPSS version 15.5 software (SPSS Inc. Illinois, USA).

P values were calculated from the χ^2 distribution and values $p < 0.05$ were considered statistically significant.

Meta-analysis for the IL-1B -511 A/C polymorphism in chapter 7 was done using Mantel-Heinzel test implemented by StatsDirect v. 2.6.5 software for Windows (StatsDirect Ltd. Cheshire, UK)

In all association tests, p values are reported as uncorrected. Where necessary multiple testing was corrected by using a false discovery rate approach, as Bonferroni correction would have been overly conservative.

2.3.5 Haplotype construction and analysis by PHASE v. 2.0.2

PHASE version 2.0.2 was used in assigning haplotypes from unphased genotype data for the sample of unrelated patients and controls (Stephens *et al.* 2003). The software could be accessed at http://stephenslab.uchicago.edu/software.html. The theory behind PHASE is based on maximum likelihood method, implemented via the expectation-maximisation (EM) algorithm (Excoffier *et al.* 1995), but with additional more sophisticated assumptions about the conditional distribution and is not just dependent on previous genotypes. PHASE uses Bayesian statistical method and reconstructs unknown haplotypes by combining prior information and the likelihood, to estimate posterior distribution, from where the actual haplotypes themselves can then be estimated. The PHASE method has been shown to provide improvement over other methods for haplotype reconstruction with the mean error rate being about half that of straightforward

EM algorithm. However, it is important to bear in mind that for a given dataset different prior assumptions will generally lead to different posterior distributions and different haplotype estimates. Another complicated issue with applying Bayesian approach is that for a real population data set the posterior distribution cannot be calculated exactly. Instead, computational methods like Markov chain Monte Carlo are used to approximate it. Therefore different computational methods or different numbers of iterations will change the quality of approximations produced by Bayesian methods.

2.3.6 CLUMP for the comparison of haplotype counts

The CLUMP software program (http://linkage.rockefeller.edu/soft/clump.html) is designed to assess the significance of the departure of observed values in the contingency table from the expected values, conditional on the marginal totals. The significance is assessed using a Monte Carlo approach, by performing repeated simulations to generate tables having the same marginal totals as the one under consideration and counting the number of times that a chi-squared value associated with real table is achieved by the randomly simulated data (Sham *et al.* 1995). The original feature of CLUMP is the novel chi-squared value which is produced by "clumping" columns together into a new two-by-two table in a way that is designed to maximise the chi-squared value. The advantage of this method is that it allows one to gain a robust assessment of the post-hoc hypothesis that certain columns tend to have higher values in one row than the rest. Each column is rotated around to see if this increases the chi-square value. This process is repeated until no further moves can be found that increase the chi-squared value for the table. This chi-square value is inflated and would be expected to be higher than an ordinary chi-squared

statistic with one degree of freedom, and its significance is assessed using Monte-Carlo simulations.

2.4 Materials

2.4.1 Reagents and suppliers

All chemicals used for the assays were supplied by Sigma (Poole, UK), Invitrogen (Paisley UK) or Amersham Biosciences (Little Chalfont, UK). Primers were mainly synthesised by MWG Biotech (Ebersberg, Germany). Restriction endonucleases were supplied by New England Biolabs (Hitchen UK). For the cell separation Miltenyi Biotec (Miltenyi Biotec Ltd, UK) reagents, magnetic beads and antibodies for FACS analysis were used.

The compositions of solutions made *ex tempore* in house are outlined below:

Lysis buffer:

Sucrose	109.4g
Tris-HCl (ph 7.8)	1.0ml
4.9 M $MgCl_2$	1.2ml
Triton X-100	10.0ml

Add ~ 500ml distilled water and mix by swirling

Solution B

Tris HCl	48.44g
EDTA	22.33g
NaCl	8.766g

Adjust the pH to 8 using 40% NaOH

10% SDS	100ml

Make up to 1 litre with dH_2O

1 x TE

10 mM Tris-HCl

1 mM EDTA

Adjust pH to 8 using 40% NaOH

1 x TBE

Tris Base 108g

Boric Acid 55g

0.5M EDTA 20ml

Make up to 10 liters with dH_2O

Chapter 3

Rheumatoid arthritis patients cohort- clinical, serological and *HLA-DRB1* characteristics

3.1 Demographic and clinical characteristics

A total of 756 British Caucasian RA patients attending the Nuffield Orthopaedic Centre, Oxford, UK were recruited for the study. All patients satisfied the 1987 ACR criteria for RA. Written informed consent was obtained from all subjects and the study was approved by the Oxford Research Ethics Committee (OxREC No. C02.032) (Appendix A). All patients completed a structured questionnaire on demographic characteristics and clinical features of the disease including age at onset of continuous joint symptoms, history of autoimmune diseases, surgical history and current treatment (see appendix A for an example of the questionnaire). All demographic and clinical information was coded and entered into a FileMarkerPro database. The clinical features of the patient cohort are summarised in Table 3.1.

Medical records were examined to determine the RF status of each patient. Patients who tested positive by the particle agglutination test at two different time points were considered to have RF positive disease. Where data was not available, RF was measured using nephelometric technique and a titre over 40 IU/ml was considered positive. Based on collected data 586 (77.5%) of RA patients were RF positive. Positive RF was observed more commonly in male than female patients and this difference was

Table 3.1 Relevant clinical characteristics in RA cohort.

	All RA patients	Females	Males
Total number of patients	756	542 (71.7%)	214 (28.3%)
RF positive	586 (77.5%)	409 (75.5%)	177 (82.7%)
Disease onset (yrs, range)	48(16-86)	48 (16-86)	49 (16-81)
Nodules	201 (26.6%)	137 (25.3%)	64 (29.9%)
Joint replacement	165 (21.8%)	128 (23.6%)	37 (17.3%)
Other autoimmune conditions*	73 (9.7%)	63 (11.6%)	10 (0.5%)

* Presence of autoimmune thyroid disease and Type I diabetes.

statistically significant (p = 0.03, OR 1.55 95%CI 1.02-2.40). The mean age of RA onset was 48 years and similar in both gender groups. The range of disease onset varied widely, the minimum age at onset was 16 and maximum 86, confirming the well known fact that RA can affect subjects at any age.

Severity of RA was determined by the presence of rheumatoid nodules or the requirement of major (hip, knee or shoulder) replacement as established from medical records. Presence of nodules during the disease course was documented in 26.6% of patients. This is similar to previously reported prevalence of ~20%, although higher frequencies of nodules among hospital based RA cohorts have also been described (Vlak *et al.* 2003; Meyer *et al.*1999).

Familial clustering and co-occurrence of more than one autoimmune disease have long been recognised. In order to characterise the co-existence of other autoimmune diseases in our RA cohort I collected data on the presence of other autoimmune conditions

affecting different organ systems. Autoimmune disease diagnoses were confirmed on the basis of a review of medical records. Co-occurrence of a second autoimmune disease was observed more commonly in female patients (p=0.004, OR 2.68 95%CI 1.33-5.97). Autoimmune thyroid disease (AITD) was the most commonly co-existing pathology, with female patients being more frequently affected than male patients (57 and 5 cases respectively). This was followed by Type I diabetes (T1D) that was recorded in 11 cases (5 cases in male and 6 in female RA patients). Increased frequency of thyroid antibodies and co-occurrence of AITD in RA patients have been well described (Pongratz *et al.* 2000; Criswell *et al.* 2005). The association of RA with T1D is more controversial. A large study involving 1,460 RA patients admitted to two Finnish hospitals conducted by Hakala *et al* demonstrated the prevalence of T1D in RA patients of 0.6%; this did not exceed the prevalence of T1D in the middle aged population of Finland in general (0.5-0.6%) (Hakala *et al.* 1992). In our cohort the frequency of T1D in RA was 1.4%; this is much higher than in the UK general population (0.40% in men and 0. 28% in females) (Harvey JN *et al.* 2002). However, our RA cohort is too small to make confident conclusions and a bigger sample is needed to confirm the result.

A group of 630 unrelated healthy British Caucasian blood donors were used as controls.

3.2 Radiographic scoring and analysis of the extent of erosive disease

Hand radiographs were scored using Rau-Ratingen method (Rau *et al.* 1998). This method restricts scoring of an individual joint to define changes with regard to erosions and joint destruction. The extension of the erosion into the bone is not

considered. Grading of an individual joint is based solely upon the amount of joint surface destruction, no matter if the surface is destroyed by one large or by several small erosions. Each joint is divided into 5 equal areas, representing 20% of total joint surface area. The amount of joint surface destruction is defined by the length of the clear visible interruption of the cortical plate in relation to the total joint surface. Grading is performed according to the following rules: Grade 1 represents one or several definite erosions summing up destruction of no more than 20% of the total surface; grade 2 represents a joint surface destruction between 21% and 40%, grade 3 between 41% and 60%, grade 4 between 61% and 80%, grade 5 greater than 80% (Figure 3.1).

Scoring is performed in the following joints or areas: 10 proximal interphalangeal (PIP), 10 metacarpophalangeal (MCP), 4 sites in each wrist (scaphoid, lunate, radius, ulna) (Figure 3.2). The original method also includes the metatarsophalangeal joints (MTP2-5) and IP of the big toe on both sides. Feet radiographs were not included in this data set.

Hand radiographs were available for 206 (150 female and 56 male) patients with disease duration over 3 years. Mean onset of the disease was 46 years (range 19-85). The majority of patients (n=154, 74.8%) were RF positive but 51 (24.8%) were RF negative (RF was unavailable for 1 patient).

Figure 3.1 Examples for the different grades of the Rau-Ratingen score in MCP joints. From left to right: a - presents a healthy joint with a score 0, b- destruction of the joint surface < 20% (grade 1); c – destruction between 21 and 40% (grade 2), d- destruction between 41-60% (grade 3), e – destruction between 61 and 80% (grade 4) and f- destruction > 80% joint surface area (grade 5).

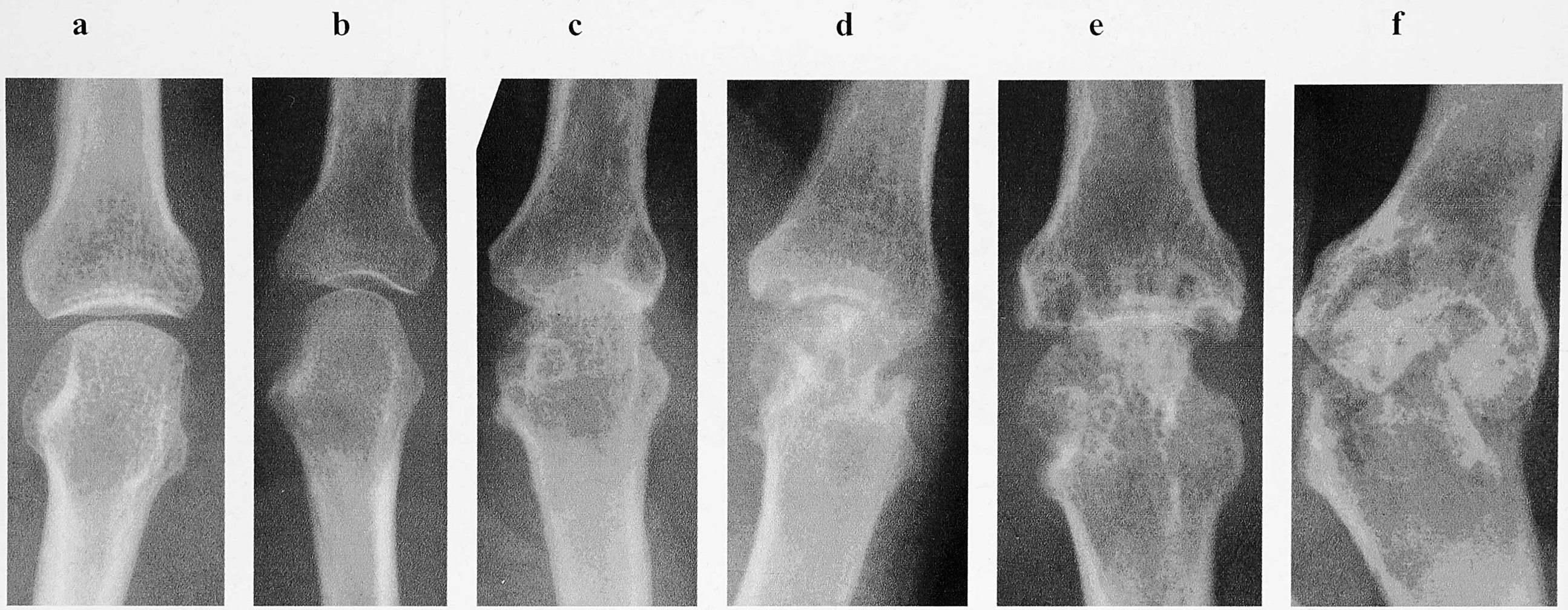
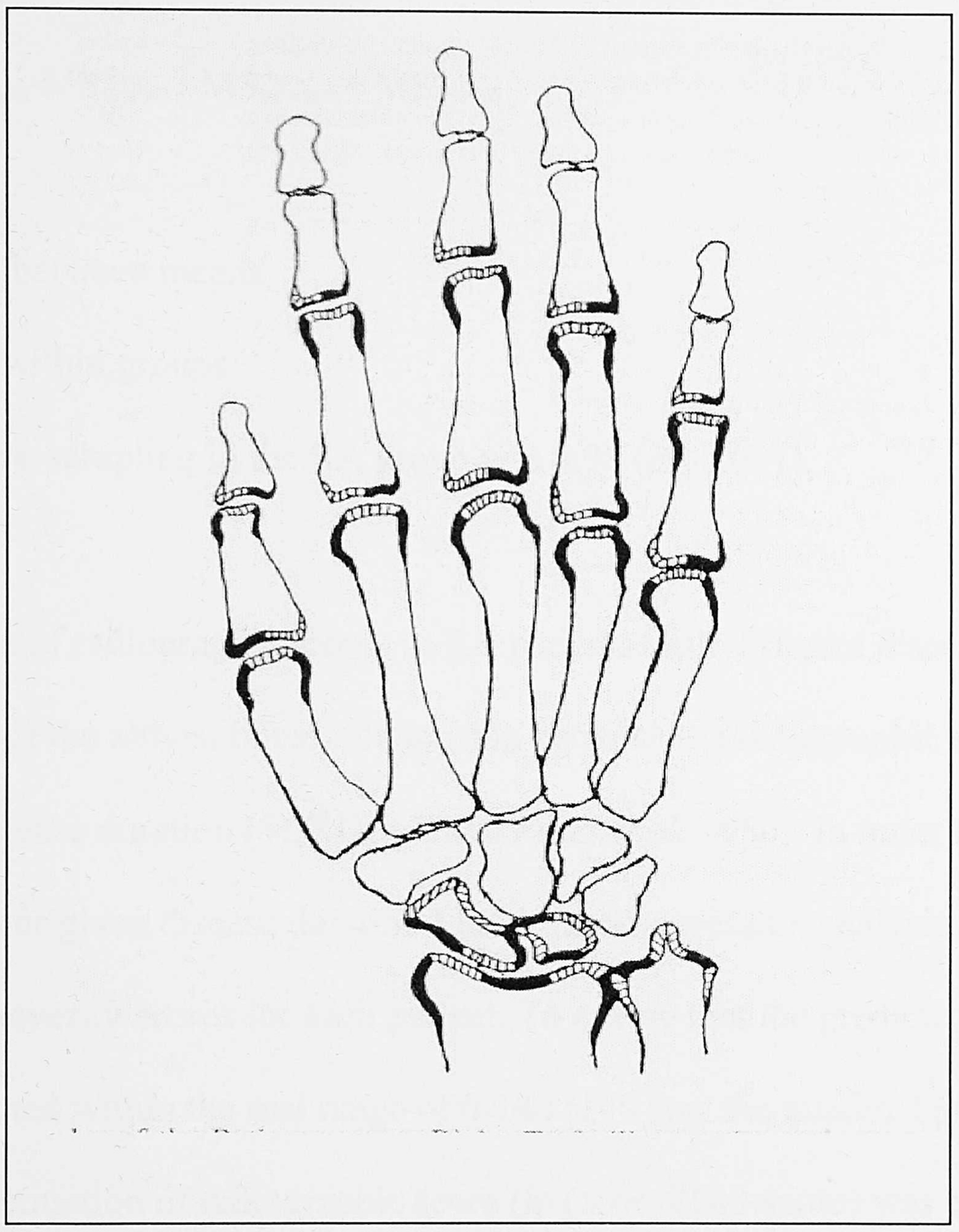


Figure 3.2 Example of the anatomical regions of the hand used for scoring according to Rau-Ratingen method. Each joint scored is divided in 5 equal areas (illustrated as solid and hatched), representing 20% of the total joint surface.



Radiographs were scored by one reader and the majority of radiographs were evaluated twice. Intraclass correlation (ICC) for unordered pairs was calculated using the web-based program "VassarStats" (http://faculty.vassar.edu/lowry/VassarStats.html) . This program uses the following formula for the calculation:

$$ICC = MS_{beween} - MS_{within} / MS_{beween} + MS_{within}$$

MS_{beween} = Variance between means

MS_{within} = Variance within groups

The ICC after 1000 re-sampling in the RA group was 0.98 ($p < 0.001$).

To allow comparison of radiographic scores in RA patients with different disease duration I made use of the almost linear relationship between the radiographic score and the square root of disease duration (SQRD) (de Carvalho *et al.* 1980). In order to predict radiographic scores for given disease duration I used linear regression analysis and assigned percentile severity scores for each patient. To ensure that the predicted scores on all percentiles remained within the real range of 0-140 (140 was the maximal possible score), a logit transformation of radiographic score (ln (score/(140-score) was entered as the dependent variable with SQRD as the independent variable. The final regression equation was Logit score= (0.66 x SQRD) – 4.10. The relationship between the radiographic score and disease duration by this equation, with selected percentile scores is shown in figure 3.3. A graph displaying the relationship between assigned radiographic percentile severity scores and disease duration in shown in figure 3.4.

Figure 3.3 The relationship between radiographic score and disease duration as described by the regression equation with selected percentile lines.

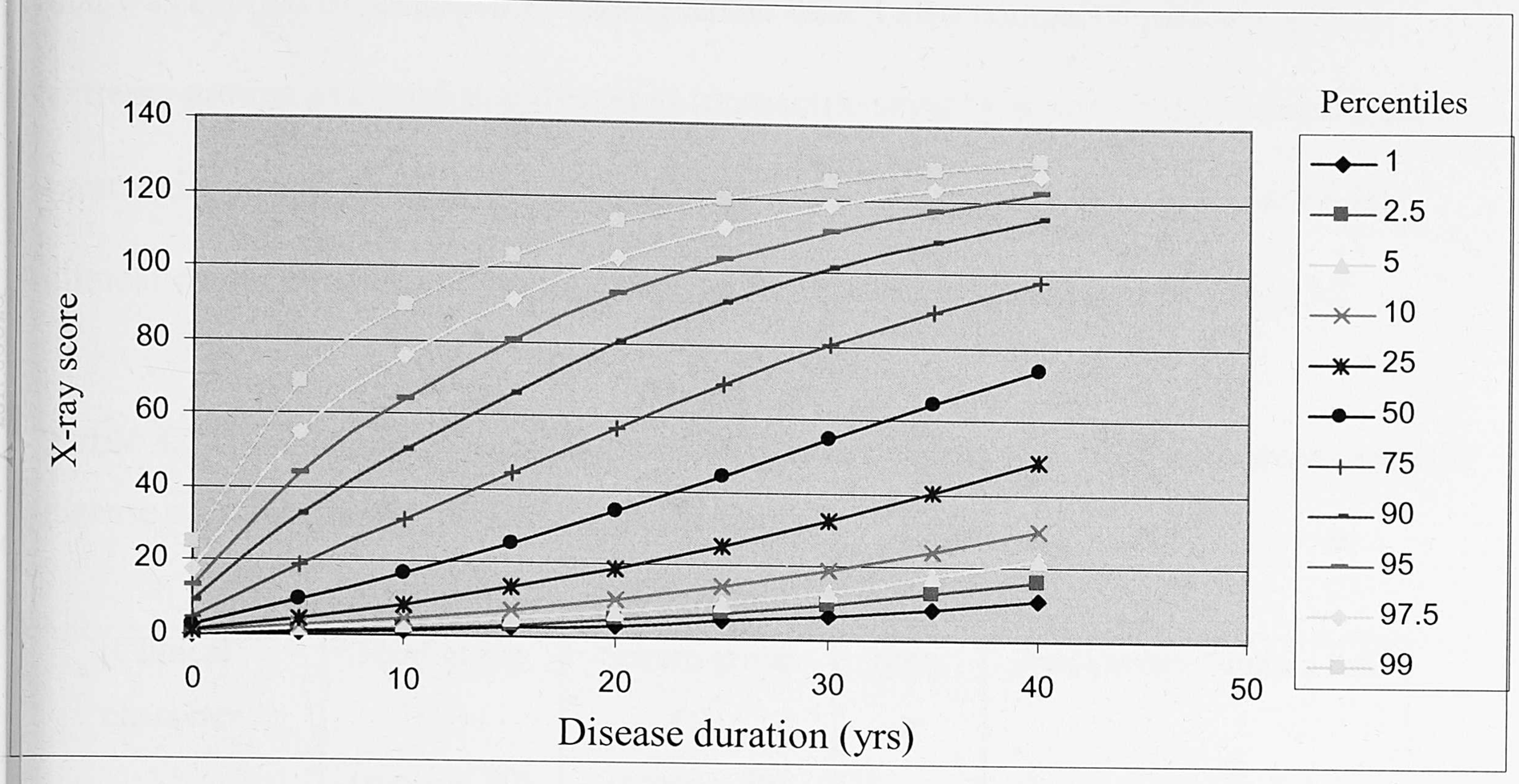
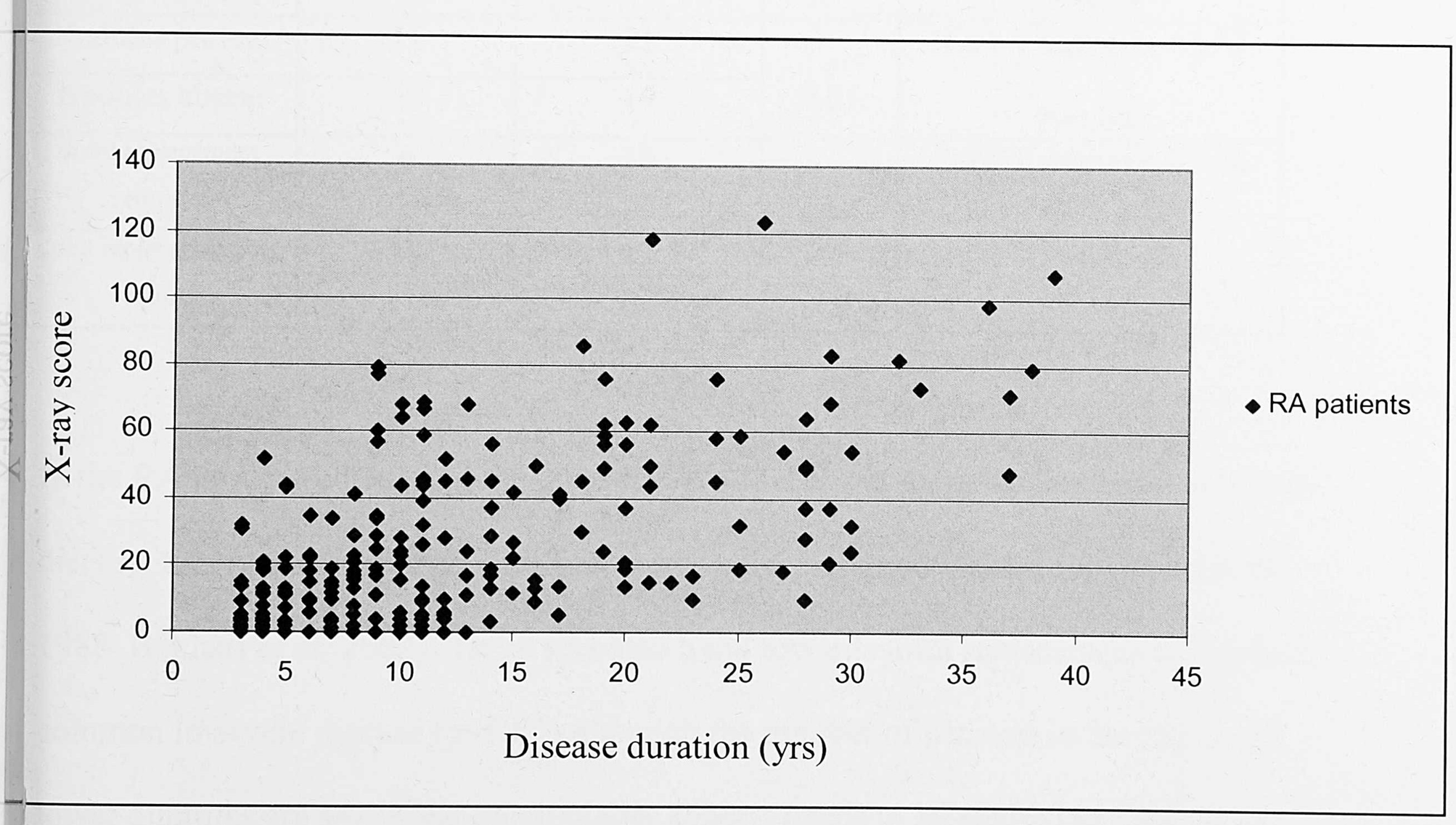


Figure 3.4 Scatterplots of calculated radiographic severity scores and disease duration in RA group.



Associations between particular characteristic (clinical or genetic) and radiographic severity score were assessed by using Jonckheere-Terpstra test for ordered alternatives that was applied to all available radiographic data. I also compared patients in two extreme groups assigned as either mild (percentile severity score <20) or severe disease (percentile severity score >80) using Fisher's exact probability test. Association of clinical characteristics and radiographic severity score are presented in Table 3.2.

Table 3.2 Comparison of clinical characteristic in patients with mild and severe articular disease on radiography. N= number of patients

Clinical character	Mild group (N) (score < 20)	Severe group (N) (score > 80)	Sign. p	Jonckheere-Terpstra test
RF present	30	33	0.22	Std.J-T statistics = 1.0 P=0.18
RF absent	14	8		
Nodules present	11	21	0.01	Std.J-T statistics = 2.68 P=0.004
Nodules absent	33	19		
Joint replacement required	6	13	0.06	Std.J-T statistics = 1.66 P=0.05
Joint replacement not required	38	27		

In the RA group studied only the presence of nodules was associated with more severe disease, this is in accordance with previously published reports (van der Heijde *et al.* 1988, Bukhari *et al.* 2002). There was also trend towards joint replacement being more common in severe disease (p=0.05) although the number of patients in the upper and lower quintiles are small and much bigger cohorts needs to be studied to confirm the results.

3.3 *HLA-DRB1* associations

Limited *HLA-DRB1* typing and further sub-typing in *HLA-DRB1*04* positive patients was performed as described in chapter 2. Results of the genotyping and frequencies of *HLA-DRB1* alleles in male and female patients are presented in Appendix B, Table B.1. The frequency of *HLA-DRB1*04* positive alleles was slightly lower than expected in the hospital based RA population (overall 0.42; 0.41 for female and 0.44 in male patients in the group studied) and closer to the values reported in the newly diagnosed RA patients recruited in the community (Thomson *et al*. 1993). Distribution of genotypes is presented in Table 3.3 and revealed that 47% of RA patients carried one and 30% carried two copies of *HLA-DRB1* SE alleles and this is similar to previously published reports (Hinks *et al*. 2008). When the data was stratified according to the presence of *HLA-DRB1*04* status, then 47% had one and 16% had two copies of the higher risk allele. The frequencies of *HLA-DRB1* type in our RA cohort are similar to previously published data from UK and USA (Mattey *et al.* 1999, Gorman *et al.* 2004).

Table 3.3 Different HLA genotype distribution in RA cohort.

	All RA patients n=756 (%)	Females n=542 (%)	Males N=214 (%)
DRx/DRx	176 (23.3)	135 (24.9)	41 (19.2)
DRSE/DRx	355 (47.0)	248 (45.8)	107 (50.0)
DRSE/DRSE	255 (29.8)	159 (29.3)	66 (30.8)
DR04/DRx	358 (47.3)	256 (47.2)	102 (47.7)
DR04/DR04	123 (16.3)	85 (15.7)	38 (17.8)

SE- shared epitope positive allele: *01, *0401, *0404, *0405, *0408, *10
DR04- *0401, *0404, *0405, *0408

Association of *HLA-DRB1* SE genotypes and clinical characteristics of RA have been studied extensively and generally demonstrate positive correlation of the SE with erosive disease. Studies investigating associations of the SE with other clinical severity features have revealed less consistent results. Logistic regression analysis of my data showed association of SE with RF and the presence of rheumatoid nodules (p=0.03, 95%CI 1.0-2.4 and p=0.04 95%CI 1.0-2.5 respectively). The positive association with nodules is different from the recently published meta-analysis that incorporated data from 3,272 RA patients (Gorman *et al.* 2004). In their analysis, the presence of the *HLA-DRB1* SE did not increase the risk of rheumatoid nodules among Caucasian patients with RA, although significant heterogeneity was present. However, sensitivity analysis after removing ambiguous studies revealed significant association between the SE and development of nodular disease (OR1.4 95%CI 1.1-1.8). Although the authors did not make any comments about these differences and overall the study concluded no link between the *HLA-DRB1* and nodules, the discrepancy in results after removing ambiguous studies might indicate that quality of the data is as important as quantity.

There was no association with SE alleles and the requirement for joint replacement in our cohort (p = 0.53, 95%CI 0.74-1.78); this is in accordance with a previously published report (Gossec *et al.* 2004).

In order to optimise the chance of detecting true associations between *HLA- DRB1* and radiographic progression in my dataset, I compared the percentile severity scores between the groups with and without each particular *HLA-DRB1* genotype. I also assessed percentile severity scores in *DRB1*04* homozygotes, *DRB1*04* heterozygotes and non-

*DRB1*04*. Finally I also compared the frequencies of *HLA-DRB1* alleles in subjects with mild or severe articular disease. The results are presented in Appendix B Tables B.2-B.4. Using a variety of statistical approaches I was unable to demonstrate a significant relationship between *HLA-DRB1* and percentile severity score. Possible explanation for this could be that patients with more recent onset of RA (< 15 years disease duration) have progressively benefited from the recent improvements in treatment that have been demonstrated to reduce joint destruction. This has reduced the power of studies based on radiographic analysis (Sokka *et al.* 1999). A comparatively novel scoring method was used, and that could have accounted for decreased reliability. However, this scoring method has been assessed and in comparison with other well known established methods like the Sharp and Larsen scores, has performed equally well (Guillemin *et al.* 2005).

3.4 Power of the RA patients collection

This collection of RA cases was initiated with the aim of providing a resource for the investigation of genetic factors influencing susceptibility and severity of RA. Sample size is one of the several important factors that have to be taken into an account when designing genetic studies. Table 3.4 presents the power of 756 patients and 630 controls to detect significant association in the presence of different genotypic odd ratios with the assumptions: a log-additive model, disease prevalence of 1% and assuming 80% power with significance level of 0.05. Calculations were performed using program "Quanto v1.2.3" (http://hydra.usc.edu/gxe).

Table 3.4 Power of 756 patients and 630 controls to detect association assuming a log-additive model and disease prevalence 1%.

Genotype relative risk	Allele frequency	Power (%) (α =0.05)
2.0	0.10	100
	0.20	100
	0.50	100
1.5	0.10	93
	0.20	99
	0.50	100
1.25	0.10	45
	0.20	67
	0.35	81
	0.50	83

The calculations demonstrate that for loci with a moderate effect size (genotype relative risk $\geq$ 1.5) with the minor allele frequency of 10%, this RA cohort provides greater than 90% power at significance level of 0.05. Power remains sufficient for testing polymorphisms with minor allele frequency $\geq$ 35% and genotype relative risk of 1.25. In practice this implies that this collection is adequately powered to investigate previously unexplored candidate genes for a primary test of association. It is also well positioned to replicate previously identified associations and further explore the effect of different RA phenotypes. It has also proven valuable in combination with other resources, especially in the WTCCC replication study for RA.

3.5 Conclusions

The aim of establishing this collection of RA cases was to provide a well characterised resource to facilitate genetic association studies in RA. In this chapter the demographic characteristics and clinical features, as well as distribution of *HLA-DRB1* genotypes in the whole collection have been described. Overall the collection is typical of other hospital-based RA cohorts with established disease. Comparison of seven hospital based RA cohorts that were used in GWAS and replication studies is presented in Appendix B, Table B.5. Although the presence of SE alleles was associated with seropositivity for RF and nodular disease, I did not observe any significant relationship between *HLA-DRB1* genotypes and radiographic progression.This suggests that *HLA-DRB1* typing alone is unlikely to of great benefit in predicting outcomes of RA in the clinical setting. This collection of RA patients, supported by the power calculations, provides a useful resource for the primary genetic association and replication studies in RA.

Chapter 4

Association analysis of the protein tyrosine phosphatase non-receptor type 22 (*PTPN22*) gene with rheumatoid arthritis and its clinical characteristics

4.1 Introduction

4.1.1 Lymphoid tyrosine phosphatase (Lyp; PTPN22) structure and function

In eukaryotic cells, protein tyrosine phosphatases (PTPs) are important regulators of a broad set of physiological processes, including cell signalling, cell growth and differentiation. By counterbalancing the action of protein tyrosine kinases, PTPs regulate the reversible phosphorylation of proteins that is necessary for cell function. A growing number of PTPs have been implicated in various pathologies, such as developmental defects, neoplasias and immunodeficiency.

The human genome contains a total of 107 genes encoding either experimentally verified PTPs or proteins with a domain homologous to the catalytic domain of these PTPs (Alonso *et al.* 2004). On the basis of the primary structure of the catalytic domain, PTPs can be divided into 4 evolutionarily distinct classes. Class I PTPs constitute the largest group with 99 members, including 38 classical PTPs and 61 VH-1 like (dual specific)

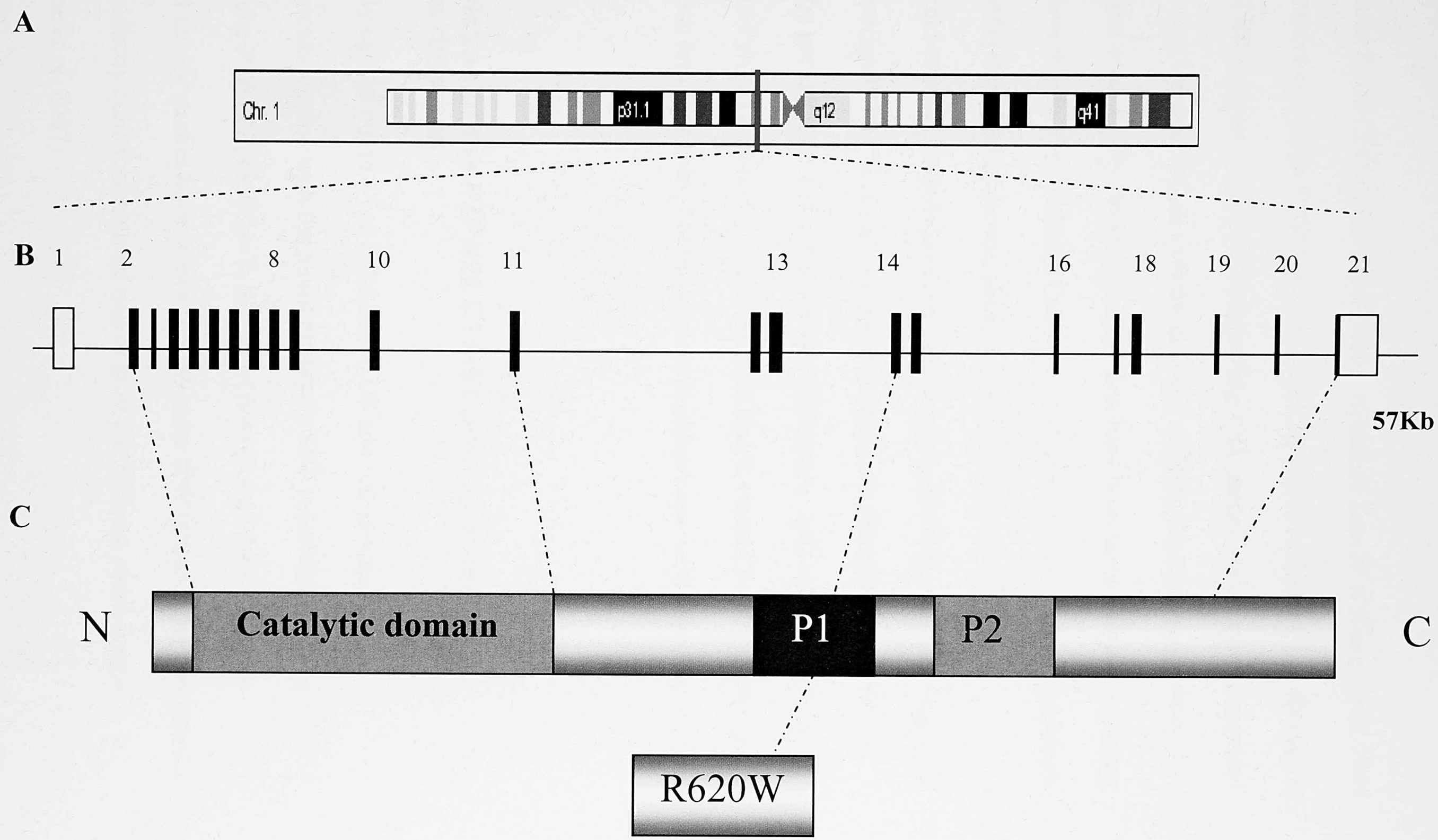
phosphatases. Most cells express 30% to 60% of the entire complement of PTPs, among them neuronal and hematopoietic cells have a tendency to express a relatively high number of PTPs.

Lymphoid-tyrosine phosphatase (LYP) belongs to the PEST group of non-receptor classical class I PTPs (Cohen *et al.* 1999). The protein is encoded by human protein tyrosine phosphatase non-receptor type 22 gene (*PTPN22*) that maps on chromosome 1p13.

LYP contains an N-terminal PTP domain, a central region of unknown function and a C-terminal portion of approximately 200 amino acids containing four proline-rich motifs referred P1-P4 (Figure 4.1). P4 is a part of the so called C-terminal homology domains (CTH), which is characteristic in all members of PEST group PTPs (Spencer *et al*. 1997, Cote *et al*. 2002). LYP can also exist as an alternatively spliced variant called LYP2 that has a shorter C terminus, resulting in the absence of P2, P3 and P4 (Cohen *et al.* 1999). Much of the current knowledge regarding LYP is derived from studies of its mouse ortholog PEP (Matthews *et al.* 1992)

Both PEP and LYP inhibit T cell receptor (TCR) signalling by acting immediately downstream of the TCR. More precisely, PEP has been implicated in the dephosphorylation of the positive regulatory tyrosine residue in the activation loop of the

Figure 4.1 Overview of the PTPN22 structure. **A** chromosomal location, **B** genomic organisation and **C** primary structure of LYP protein with position of R620W polymorphism. P- proline rich motif



Src family kinases FynT (Y417) and Lck (Y394), as well as Zap-70 (Cloutier *et al.* 1999, Gjorloff-Wingren *et al.* 1999). LYP is able to bind Csk and this complex was thought to promote LYP/PEP function in early TCR signalling. Csk phosphorylates the C-terminal negative regulatory tyrosine in Lck (Y505) and FynT (Y528) (Cloutier *et al.* 1999, Gjorloff-Wingren *et al.* 1999). However, recent data have been somewhat contradictory. It has been shown that PEP cannot bind Csk when the latter interacts with the membrane protein Cbp/PAG (Csk binding protein/phosphoprotein associated with glycosphingolipid-enriched membrane domains) which exclusively partitions into lipid rafts. More importantly PEP is in fact not found in lipid rafts (Davidson *et al.* 2003). Therefore, at the present time it is unknown how PEP targets lipid raft-resident proteins such as Lck and FynT. There is now some evidence that in human T-cells a small fraction of LYP partitions into lipid rafts, but this process is independent of Csk binding (Vang T *et al.* 2007).

4.1.2 Association of the *PTPN22* C1858T polymorphism with autoimmune diseases

The initial study carried out in Type I diabetes in US and Italian populations demonstrated an association with the minor allele of a SNP (rs2476601) in a coding region of the gene *PTPN22* (1858C→T, R620W) (Bottini *et al*. 2004). Further association and family studies have shown that the same SNP is one of the first genetic variants that evidently confer substantial risk of multiple different autoimmune phenotypes (Lee *et al*. 2007).

Association between a functional SNP in the *PTPN22* coding region and RA in Caucasians was first described by *Begovich et al* and has been replicated by several other groups worldwide (Begovich *et al*. 2004, Lee *et al.* 2005, Steer *et al.* 2005, Hinks *et al*. 2005, Orozco *et al.* 2005, Zhernakova *et al.* 2005) . The presence of the *PTPN22* 1858 T/T genotype increases the risk for RA more than twofold. However, little is known about the effects on severity and co-association with other autoimmune diseases.

Association with 1858 T allele has also been demonstrated in juvenile onset arthritis (Viken *et al.* 2005; Hinks *et al.* 2005), systemic lupus erythematosus (Kyogoku *et al.* 2004; Criswell *et al.* 2005), Grave's disease (Smyth *et al.* 2004; Velaga *et al.* 2004), myasthenia gravis (Vandiedonck *et al.* 2006), generalised vitiligo (Canton *et al.* 2005) and Wegener's granulomatosis (Jagiello *et al.* 2005). However, other autoimmune diseases have not been shown to be associated with the 1858T allele, including multiple sclerosis (Criswell *et al.* 2005; Hinks *et al.* 2005; Begovich *et al.* 2005), Crohn's disease (van Oene *et al.* 2005; Wagenleiter *et al.* 2005; Martin *et al.* 2005), ulcerative colitis (Martin *et al.* 2005; Prescott *et al.* 2005), celiac disease (Viken *et al.* 2005; Rueda *et al.* 2005), primary sclerosing cholangitis (Viken *et al.* 2005), primary biliary cirrhosis (Milkiewicz *et al.* 2006), ankylosing spondylitis (Orozco *et al.* 2006) psoriasis and psoriatic arthritis ((Hinks *et al.* 2005). Association with Hashimoto's thyroiditis, primary Sjogren's syndrome and systemic sclerosis has been variable and more studies are needed.

4.2 Aims of the study

The aim was to study the distribution of the *PTPN22* C1858T polymorphism in a large sample of British Caucasian RA patients for four reasons: to confirm the strength of the association with RA; to evaluate epistatic interactions with *HLA-DRB1* alleles; to evaluate the effects of the polymorphism on clinical characteristic and to explore the influence of *PTPN22* on the co-occurrence of other autoimmune diseases in those with RA.

4.3 Subjects and methods

4.3.1 Study population

The study population has been described in detail in Chapter 3. Briefly 756 British Caucasian RA patients and 566 British Caucasian healthy blood donors were genotyped for the *PTPN22* C1858T polymorphism. Further analysis was done in RA population according to the presence of RF, *HLA- DRB1* SE alleles and clinical severity characteristics.

4.3.2 Methods and statistical analysis

Cases and controls were genotyped for C1858T polymorphism by RFLP analysis of PCR products. The C→T transition at codon 620 creates in the T allele a restriction site for *Xcm*I (New England Biolabs). The primer sequences and PCR conditions are described in Appendix C (Table C.1, Figure C.1). To confirm the genotypes obtained by the RFLP direct sequencing for *PTPN22* C1858T variants on representative samples from each

genotype was performed using an ABI 3100 capillary sequencer and GeneMapper software (Applied Biosystem, Foster City CA) (Appendix C, Figure C.2). Limited *HLA-DRB1* typing of RA cases was performed as previously described (Chapter 2, section 2.2.2 and 2.2.3)

Statistical analysis for association and Hardy-Weinberg equilibrium were evaluated using chi-square test. P values less than 0.05 were considered significant. The results were also expressed as odds ratios (± 95% confidence intervals, CI). For analyzing continuous variables software SPSS 15.0 for Windows was used.

4.4 Results

The *PTPN22* genotypes were in Hardy-Weinberg equilibrium in cases and controls. The genotyping data are shown in Table 4.1. The risk allele (1858T) was more prevalent in RA patients (14%) than in the healthy controls (10%) (p=0.004 OR 1.4 95% CI 1.1-1.8). Interestingly, while the T allele was significantly increased in the RF positive cases (15%), its prevalence in the RF negative cases (12%) did not differ significantly from the controls (10%).

The different *PTPN22* genotypes were equally distributed among female and male patients. The median age of RA onset for the entire group was 48 yrs and did not vary significantly between the different genotype subgroups (p = 0.19)

Table 4.1 Genotyping data and allele counts of 756 RA patients and 566 healthy controls.

	Allele counts			**Genotyping counts**			
	C No (%)	**T No (%)**	**OR, 95% CI**	**CC No (%)**	**CT No (%)**	**TT No (%)**	**OR CT/TT vs CC**
RA cases n=756	1292 (85)	212 (14)	1.42 (1.11- 1.82) p=0.0038	551 (73)	190 (25)	11 (2)	1.52 (1.17-2.01) p=0.0015
Controls n=566	1015 (90)	117 (10)		457 (81)	101 (18)	8 (1)	
*RF neg n=158	279 (88)	37 (12)	1.15 (0.78-1.70) *p=0.48	122 (77)	35 (22)	1 (1)	1.23 (0.78-1.92) *p=0.33
*RF pos n=586	993 (85)	171 (15)	1.50 (1.15-1.94) *p=0.0016	420 (72)	153 (26)	9 (2)	1.61 (1.21-2.15) *p=0.0006
RF pos *v* RF neg			1.29 (0.88-1.95) p=0.18				1.31 (0.85-2.03) p=0.20

* versus healthy controls

I extended the study by stratifying RA cases by clinical features and *HLA-DRB1* type. No association between the *PTPN22* and the presence of *HLA-DRB1* SE alleles was detected

There was a weak (insignificant) trend towards major joint replacement (including hip, knee, shoulder and elbow joint) being more common in the *PTPN22* 1858T allele group (p=0.07, OR 1.36 95% CI 0.96 -1.91) (Table 4.2). No significant impact on radiographic progression was observed (Figure 4.2).

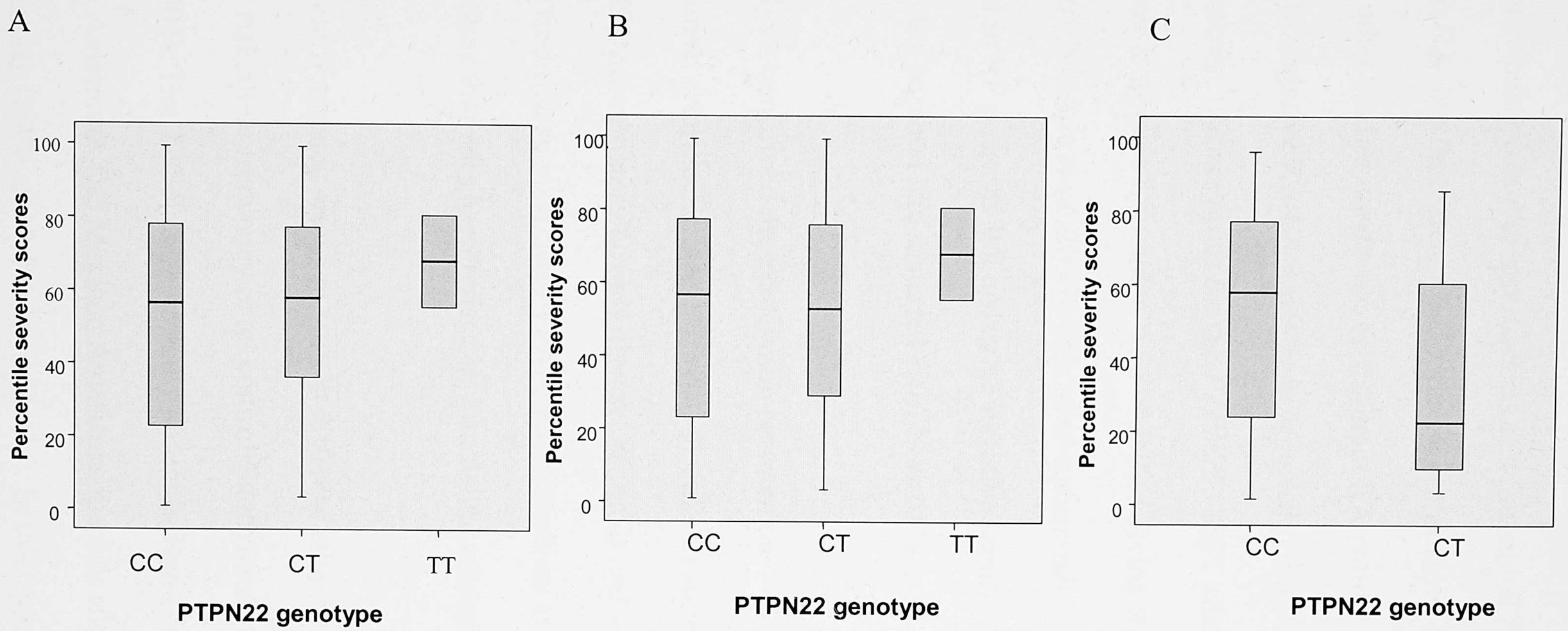
Finally, the effect of the *PTPN22* polymorphism on the co-occurrence of RA and either Type I diabetes or autoimmune thyroid disease was made; only a weak, non significant trend was noted with Type I diabetes in the *PTPN22* C1858T group (p = 0.07, OR 2.30 95% CI 0.91-5.80) and no significant association was observed with autoimmune thyroid disease.

Table 4.2 Comparison of clinical characteristics in RA patients according to *PTPN22* C1858T allele distribution and genotypes.

	Allele counts			Genotyping counts			
	C No (%)	**T No (%)**	**OR**	**CC No (%)**	**CT No (%)**	**TT No (%)**	**OR CT/TT vs CC**
Joint replacement			1.36 95% CI (0.96-1.91) p=0.07				1.39 95% CI (0.93-2.06) p=0.09
Not required n=567	982 (87)	152 (13)		422 (74)	138 (24)	7 (1)	
Required n=164	271 (83)	57 (17)		111 (67)	49 (30)	4 (3)	
Type I diabetes			p*=0.07				p*=0.16
Absent n=725	1247 (86)	203 (14)		532 (73)	183 (25)	10 (2)	
Present n=11	16 (73)	6 (27)		6 (56)	4 (36)	1 (10)	
Thyroid pathology			1.35 95% CI (0.82-2.18) p=0.23				1.44 95% CI (0.83-2.49) p=0.20
Absent n=674	1167 (87)	187 (13)		497 (73)	167 (25)	10 (2)	
Present n=62	102 (82)	22 (18)		41 (66)	20 (32)	1 (2)	

* Fisher's exact test was used to compare the groups.

Figure 4.2 Boxplots for the radiographic percentile severity scores according to the *PTPN22* C1858T genotypes. Jonckheere-Terpstra test for ordered alternatives did not show any statistically significant difference between the groups in overall RA patients (A), RF positive (B) and RF negative patients (C) (one tailed p=0.45, p=0.17 and p=0.05 respectively).



4.5 Discussion

The data confirm previously reported associations of a functional *PTPN22* 1858T polymorphism with RA, but this effect was limited to RF positive RA cases. This result is similar to those of other studies conducted on different RA patient groups although associations with RF negative RA have also been reported (Hinks *et al.* 2005, Simkins *et al.* 2005). More recently, linkage proof for the *PTPN22* 1858T allele was confirmed in 465 French and West European Caucasian "trio" families giving the linkage estimate T=63% ($p<0.00070$) (Michou *et al.* 2006). This over-transmission of the *PTPN22* 1858T allele was observed in RF positive, but not in RF negative families. There was no difference between the parental and maternal T (67% vs 61%, $p=0.4$).

A recent meta-analysis confirmed the association between the *PTPN22* 1858T allele and T1D, RA juvenile onset arthritis, SLE and Graves' disease; in contrast there was no association with MS, inflammatory bowels disease, psoriasis and Addison's disease (Lee *et al*. 2007). These findings suggest that *PTPN22* polymorphism is coupled to the set of autoimmune conditions that are characterised by circulating autoantibodies.

PTPN22 1858T allele frequency seems to vary in different ethnic populations. *Begovich et al* found that the risk allele frequency in all control Caucasians studied was 8.7% (Begovich *et al.* 2004). The T allele was virtually absent in Han Chinese and Africans, whereas in African Americans and Mexican Americans the risk allele was detected in intermediate frequencies (2.4% and 3.5% respectively). In the control population that I

studied, the frequency of the 1858T risk allele was 10.3%, similar to another UK study, indicating that it is relatively common in Caucasians (Hinks *et al*. 2005). This study did not reveal additional effects arising from homozygosity in the *PTPN22* 1858T risk allele. However, the ability to detect subtle differences was limited by the relatively low frequency of the T allele and the small number of TT homozygotes. There also appears to be a geographical gradient among European Caucasian populations as the 1858T allele seems to be relatively rare among southern European cohorts (2% in Italy, 6% in Spain) perhaps with increasing frequency towards the north (12% in Sweden and 15.5% in Finland). It is premature to assert that the association with 1858T allele and various autoimmune diseases is population independent as not enough studies have been conducted.

The highly robust association of the *PTPN22* C1858T polymorphism with multiple autoimmune diseases in Caucasians, and the lack of this polymorphism in the Asian and African populations naturally raises the possibility that additional risk alleles may exist in these populations, and may also be present in Caucasians as well. *Carlton et al.* performed the first comprehensive analysis of the *PTPN22* gene by sequencing the coding regions in 48 Caucasian RA patients (Carlton *et al.* 2005). They identified 15 novel SNPs, including 2 coding SNPs in the catalytic domain. By combining the results of the initial sequencing data with the data from public database they genotyped two large case-control cohorts.

Analysis of the results predicted 10 distinct common haplotypes (frequency >1%). The haplotype data supported the hypothesis that the *PTPN22* 1858T allele is indeed causative. However, the presence of the *PTPN22* C1858T SNP did not fully explain the association between the gene and RA in the two sample sets of *Carlton et al*, since significant differences remained between the cases and controls after the haplotype data were stratified by C1858T. Three polymorphisms that were consistently associated with RA in two independent ways were identified. Two SNPs, in nearly complete LD (r^2>0.98), one located in the 3'UTR region (rs3811021) and the second in the predicted transcriptional factor binding-site (rs3789604) on a single common haplotype were indeed associated with RA independent of the original non-synonymous SNP. Subsequent studies carried out independently on Dutch, Mexican, UK and Norwegian RA patient cohorts have not confirmed these initial findings (Wesloy *et al.* 2006, Gregersen *et al.* 2006, Hinks *et al.* 2007; Viken *et al.* 2007). The third SNP identified (rs1310182) that located in the predicted transcriptional factor binding-site was independently associated with RA in the first sample set, but only a marginal effect was demonstrated in the second sample. When all three polymorphisms were jointly modelled no independent effects were obtained. Haplotype regression analysis was unable to demonstrate a separate role of rs3789604 and rs1310182 on RA susceptibility.

More recently, *Kawasaki et al.* carried out another comprehensive study on *PTPN22* in a Japanese population by re-sequencing the gene (Kawasaki E *et al.* 2006). They identified 5 novel SNPs, two of which were common; the first was in the promoter region (-1123

G>C) (rs2488457) and the second in the 3'UTR region (+2740 C>T) (rs1217412). The authors suggested an association between the promoter SNP allele G and subgroup of T1D. The authors subsequently tested the -1123 G allele in a T1D UK Caucasian sample using the transmission disequilibrium test (TDT) that demonstrated a significant over-transmission of the -1123G allele compared to the +1858T allele, indicating that -1123G was more likely to be a causative variant.

The association of -1123G>C SNP with T1D was apparent in heterozygotes and their analysis was not corrected for multiple testing. Among the Caucasian population, a strong LD has been demonstrated between these SNPs (Choi *et al.* 2006). The over-transmission hypothesis of the -1123G allele was subsequently tested in French Caucasian trio families with RA. No significant departure from Mendel's first law was observed, indicating that the *PTPN22* -1123G allele is not independently associated with RA genetic susceptibility (Dieudé *et al.* 2008). A recent study conducted on Norwegian RA populations also demonstrated significant association of the *PTPN22*-1123G>C SNP with the disease (Viken *et al.* 2007). However, the association with the *PTPN22* 1858T allele was statistically indistinguishable from the -1123C allele.

It is still possible that there are other polymorphisms in *PTPN22* yet to be discovered that may confer risk to various autoimmune conditions. Although the C1858T polymorphism has not been found to be associated with psoriasis, a more recent study has indicated that a susceptibility locus for the disease could exist somewhere else in the *PTPN22* or in its

vicinity (Hüffmeier *et al.* 2006). The authors reported association with two haplotypes spanning the two genes adjacent to *PTPN22*: round spermatid basic protein 1 (*RSBN1*) and the neighbouring gene putative homeodomain transcription factor 1 (*PHTF1*).

The C1858T polymorphism in *PTPN22* changes amino acid 620 in LYP from an arginine to a tryptophan (Bottini *et al.* 2004). The residue 620 is located within the P1 region, which interacts with the SH3 domain of Csk. The latter contains two peptide recognition surfaces: the usual polyproline helix-binding surface of SH3 domains and a hydrophobic patch that interacts with hydrophobic residue C-terminal to the polyproline helix of the PEP P1 region. The model that LYP R620W has an effect on LYP-Csk complex formation was confirmed by series of experiments (Vang *et al.* 2005, Begovitch *et al.* 2004). Based on previous reports describing the effect of PEP-Csk complex as down modulating TCR signalling in a cooperative manner, it was assumed that the inability of LYP*W620 to bind Csk would result in less efficient inhibition of TCR signalling. However, experimental data so far have demonstrated completely opposite effects, indicating that LYP*W620 is a gain of function variant (Vang *et al.* 2005).

It is not clear how variation in *PTPN22* can affect T cells and induce the development of autoimmunity. So far two common theories exist and these are not mutually exclusive. First, the reduced signal transduction ability downstream of the TCR due variations in *PTPN22* may result in a skewing of the TCR repertoire toward generation of more autoreactive T cells in the thymus. Failure to delete these autoreactive thymocytes by

negative selection gives rise to a peripheral T cell population with enhanced ability for self-recognition. Alternatively, the presence of the *PTPN22* polymorphism in Treg cells may inhibit signalling in theses cells and reduce the ability to control the activity of autoreactive T cells in peripheral blood.

Studies published so far have not revealed firm evidence of genetic interactions between *PTPN22* and *MHC* genes in RA. However, two reports have demonstrated a strong connection between the *PTPN22* minor allele T and the occurrence of anti-cyclic citrullinated peptide antibodies in RA (Plenge *et al.* 2005, Johansson *et al.* 2006). Analysis showed that, when combined together, the presence of the 1858T allele and anti-cyclic citrullinated peptide antibodies, led to an up to 350-fold increased risk of developing RA. The concurrent presence of these two factors also predicted the later development of RA with very high accuracy (Johansson *et al.* 2006). Similar findings have been demonstrated between the 1858T allele and increased risk of development of insulin antibodies and further progression from insulitis to clinical T1D (Hermann *et al.* 2006).

Little is known about the effects of the *PTPN22* risk allele on the clinical course of RA. Orozco *et al* studied 826 Spanish RA patients and suggested differences in extra-articular disease in patients carrying CC genotype, although these were not robust after correction for multiple comparisons (Orozco *et al.* 2005). In my study, *PTPN22* status did not generally associate with specific RA clinical characteristics, although a weak trend for

the requirement for major joint replacement was observed in those with the T allele. This has been supported by Snyder *et al* who also observed an increased rate of joint replacement in the RA patients who were positive for *PTPN22* 1858T allele. However, the study cohort was small and the results were not controlled for multiple testing (Snyder *et al.* 2007). Stratifying the *PTPN22* C1858T genotyping data according to the radiographic scoring results did not reveal any association between the percentile severity score and *PTPN22* risk genotype. This is in accordance with another study from the Netherlands by Wesoly *et al* who did not demonstrate association with *PTPN22* genotypes and radiographic outcome in their patient cohort (Wesoly *et al.* 2005). Most recently another group from the Netherlands observed increased progression rate for radiographic damage in patients who tested positive for the *PTPN22* 1858T allele (Lie *et al.* 2007). This finding indicates that *PTPN22* may also have an effect on disease severity. Clearly more work is required to clarify the effects of 1858T allele on RA clinical course and severity.

An excess of various other autoimmune diseases is well recognized in RA patients and their relatives. For example, thyroid dysfunction and T1D have been found in higher frequencies among RA patients and their family members (Grennan *et al.* 1986 Silman *et al.* 1989). Studies have shown that co-association of autoimmune illnesses might be slightly higher in patients carrying the *PTPN22* risk allele (Criswell *et al.* 2005, Wu *et al.* 2005). The prevalence of thyroid pathology and T1D was slightly higher in the RA population compared to the general population in the current study (8.1 vs 1% and 1.4 vs

0.4%, respectively). Only a weak non-significant association with T1D was demonstrated in the *PTPN22* 1858T group (p = 0.07, OR 2.30 95% CI 0.91-5.80) and no significant association was observed with autoimmune thyroid disease. Since associations have already been described between *PTPN22* 1858T and T1D and Graves disease independently of RA, the trend is likely to be relevant. However but this study was underpowered to confirm it reliably.

In summary, this study confirmed the *PTPN22* 1858T association with exclusively RF positive RA patients and this was independent from the *HLA-DRB1* type. There was no association with specific clinical characteristics although weak insignificant trend was observed between carriage of the T allele and requirement for joint replacement.

Chapter 5

Cis-acting regulatory variation in the *PTPN22* gene of rheumatoid arthritis patients

5.1 Introduction

The known association of the *PTPN22* W620 allele with RA and the strong LD across the gene (Carlton *et al.* 2005) have had a confounding effect on the ability to determine which if any of the other polymorphisms in *PTPN22* are independently associated with susceptibility to RA. Although several strong candidate SNPs have been identified in addition to *PTPN22* C1858T (they are described in more detail in Chapter 4), subsequent robust replication has not been successful. Therefore the "genetic enigma" behind *PTPN22* and increased susceptibility to various autoimmune diseases still requires clarification.

Research into the mechanisms by which LYP*W620 exerts its stronger inhibitory capacity on TCR signaling has been inconclusive. Although the residue 620 is more than 300 amino acid residues C-terminal to the PTP domain, LYP*W620 is approximately 1.5 fold more active PTP compared to LYP*R620 (Vang T *et al*. 2005). The question remains, how a polymorphism so far away from the PTP domain can affect its activity?

There may be other ways by which LYP function could be influenced including through polymorphism in *cis*-acting elements that regulate the expression of *PTPN22*.

Polymorphisms in the coding regions of human genes have generally been well studied

but functional variants that affect gene regulation remain much harder to determine. Interestingly, these regulatory polymorphisms have increasingly been linked to phenotypic diversity observed in natural populations. Regulatory polymorphisms represent DNA elements that modify the expression level of a transcript or isoform. The concentration of a given mRNA allele is controlled both by *cis*-acting factors (such as DNA polymorphisms and methylation) in the flanking region of the gene and *trans*-acting modulators (such as transcription factors) that are themselves regulated by other genetic and environmental factors in the cell. Therefore, c*is*-acting variants affect transcript synthesis in an allele-specific manner, whereas *trans*-acting variants can affect both alleles of a gene.

Since the full components of all regulatory sequences for any given gene are generally unknown, comprehensive direct analyses of such elements are currently difficult to conduct. Furthermore, direct evidence that gene expression is commonly under the influence of *cis*-acting elements is scarce, as the majority of studies of gene expression have failed to distinguish between *cis*- and *trans*-acting influences.

Whereas direct studies are difficult, the presence of allele-specific expression of transcripts can be detected indirectly. The central assumption of this type of investigation is that in the absence of *cis*-regulatory polymorphism both copies of an autosomal gene will contribute equally to the mRNA pool. The detection of differential allelic expression (DAE) would provide evidence of heterozygosity at a *cis*-regulatory site for the gene being tested (Bray *et al.* 2003). This means that one can use the *PTPN22* C1858T SNP as

an indirect marker for assessing the DAE. This is under the presumption that if a difference in allelic expression is observed that would indicate the presence of regulatory polymorphism that is in high LD with the *PTPN22* C1858T SNP.

PTPN22 is expressed primarily in haematopoietic tissues. However, there appears to be a hierarchy; NK cells and neutrophils express the most *PTPN22* message while B cells and CD4+ T cells express the least (Begovich *et al.* 2004). Previous studies have reported that, in some genes that demonstrated DAE, the allelic imbalances were tissue and cell specific (Lo *et al.* 2003; Koch *et al.* 2006) and generally of modest in magnitude (typically 1.5 to 2 fold) (Yan H *et al.* 2002; Bray *et al.* 2003; Pastinen *et al.* 2004).

5.2 Aim of the study

This study aimed to ascertain whether *cis*-regulatory polymorphisms exist in the *PTPN22* gene by measuring the allelic expression of *PTPN22* gene in various PBMC populations from patients with RA.

5.3 Subjects and methods

5.3.1 Study population

UK Caucasians with RF positive RA were recruited following signed consent. Please refer to Chapter 3 for detailed characteristics of RA patients. The average age of onset of disease among patients recruited for this study was 38 (range 22 to 52).

5.3.2 *PTPN22* C1858T Genotyping

Genomic DNA was prepared from whole peripheral blood of the 12 patients recruited for DAE analysis. *PTPN22* genotyping for the C1858T polymorphism was performed by the RFLP method described in chapters 2 and 4, to confirm the presence of heterozygosity at this particular locus.

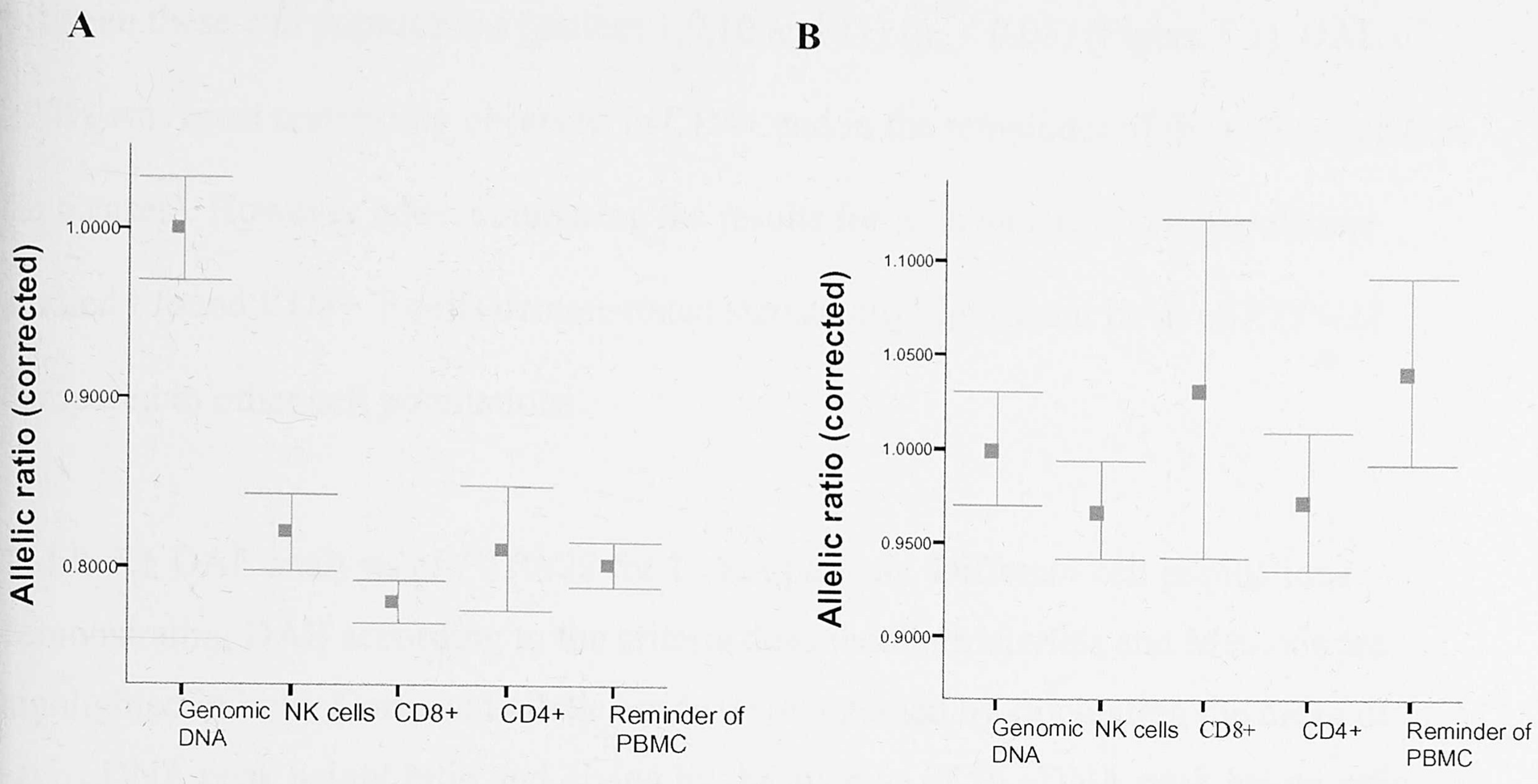
5.3.3 PBMC preparation and DAE analysis

PBMC separation of 12 RA patients, FACS analysis, followed by nucleic acid extraction and quantification and DAE analysis are all described in detail in Chapter 2 sections 2.1.2 and 2.2.4.

5.3.4 Statistical analysis for DAE data

To determine if the there was a significant difference in allelic expression for each cell type of individual patients, I compared the cDNA allelic ratios for that patient to the pooled genomic allelic ratios. A particular sample was considered to demonstrate DAE only if all the following stringent criteria were met: i) after correction by the average genomic DNA peak height ratio, the fold difference between the alleles was ≥20%; ii) the distribution of the 12 corrected cDNA replicates was significantly different than the distribution of the 56 corrected DNA replicates at $p < 0.05$ (two tailed Mann-Whitney test); (iii) the 95%CI for the mean of the 12 corrected cDNA replicates did not overlap the 95%CI for the mean of the 56 corrected genomic DNA replicates (Figure 5.1).(Bray *et al.* 2003).

Figure 5.1 Distribution of corrected allelic ratios of *PTPN22* in different PBMC populations is demonstrated in 2 RA patients. Patient A demonstrates significant DAE in different cell types and was used in further analysis; no significant DAE was observed in patient B PBMC. Error bars show 95%CI of mean.



A 20% fold difference was set as the threshold, as it has been demonstrated in previous studies that the quantitative allelic analysis method used in our study can detect differences in allelic output at or greater than this threshold (Yan *et al.* 2002, Bray *et al.* 2003). Statistical analysis was performed using SPSS 15.0 software for Windows.

5.4 Results

From the twelve individual RA patients studied, eight revealed relative DAE of 20% or more in at least one of their cell populations. The corrected allelic ratios for all the RA patients are summarized in Table 5.1 and Figure 5.2.

I observed that patients with DAE of *PTPN22* in one cell population did not necessarily demonstrate DAE in other cell types; only two patients demonstrated DAE in all cell groups. There were 7 RA patients who demonstrated DAE of *PTPN22* in a number of their cell types and 4 patients, who also had statistically significant different allelic ratios between these cell populations (patient 1,9,10 and 11) ($p_c < 0.05$) (Figure 5.3). DAE of ≥20% was most commonly observed in CD4+ and in the remainder of the cell population (in 6 cases). However when comparing the results for each patient's cell population studied I found CD8+ T cells demonstrated statistically significant DAE of *PTPN22* compared to other cell populations.

Table 5.1 DAE analysis of *PTPN22* for 12 RA patients. Different cell populations demonstrating DAE according to the criteria described in Materials and Methods are highlighted in bold. Corrected allelic ratios were obtained by calculating the mean of each cDNA peak height ratio and diving by the average of 56 gDNA peak height ratio.

	Sex	Age onset	Allelic ratio (corrected) ± 1 SD			
			NK	CD8+	CD4+	Remainder
1	F	22	0.82±0.04	**0.78±0.02**	0.81±0.06	**0.80±0.02**
2	M	48	0.97±0.04	1.03±0.14	0.97±0.06	1.04±0.07
3	F	28	1.11±0.02	1.08±0.08	1.16±0.06	1.14±0.25
4	M	31	0.84±0.04	0.82±0.06	0.86±0.05	N/A
5	F	49	1.06±0.05	1.07±0.02	1.06±0.04	1.04±0.10
6	M	52	**0.80±0.08**	**0.75±0.06**	**0.76±0.05**	**0.79±0.06**
7	M	42	**1.32±0.38**	1.12±0.04	**1.28±0.33**	0.96±0.07
8	F	33	0.84±0.03	0.87±0.15	**0.80±0.02**	0.84±0.06
9	F	31	0.88±0.11	**0.80±0.03**	**0.76±0.03**	**0.78±0.02**
10	M	38	0.82±0.02	**0.79±0.05**	0.86±0.42	**0.74±0.02**
11	F	29	**0.80±0.07**	**0.73±0.05**	**0.80±0.07**	**0.79±0.01**
12	F	49	0.86±0.01	0.86±0.03	**0.80±0.12**	**0.76±0.03**

Figure 5.2 DAE of the *PTPN22* in different PBMC populations of 12 RA patients. The corrected allelic ratios between the two *PTPN22* alleles and the error bar for 1SD are shown for the cell populations. The grey bar demarcates a relative difference in expression of 20% between two alleles. GDNA is genomic DNA. **$p>10^{-10}$, *$p>10^{-5}$. PT-patient.

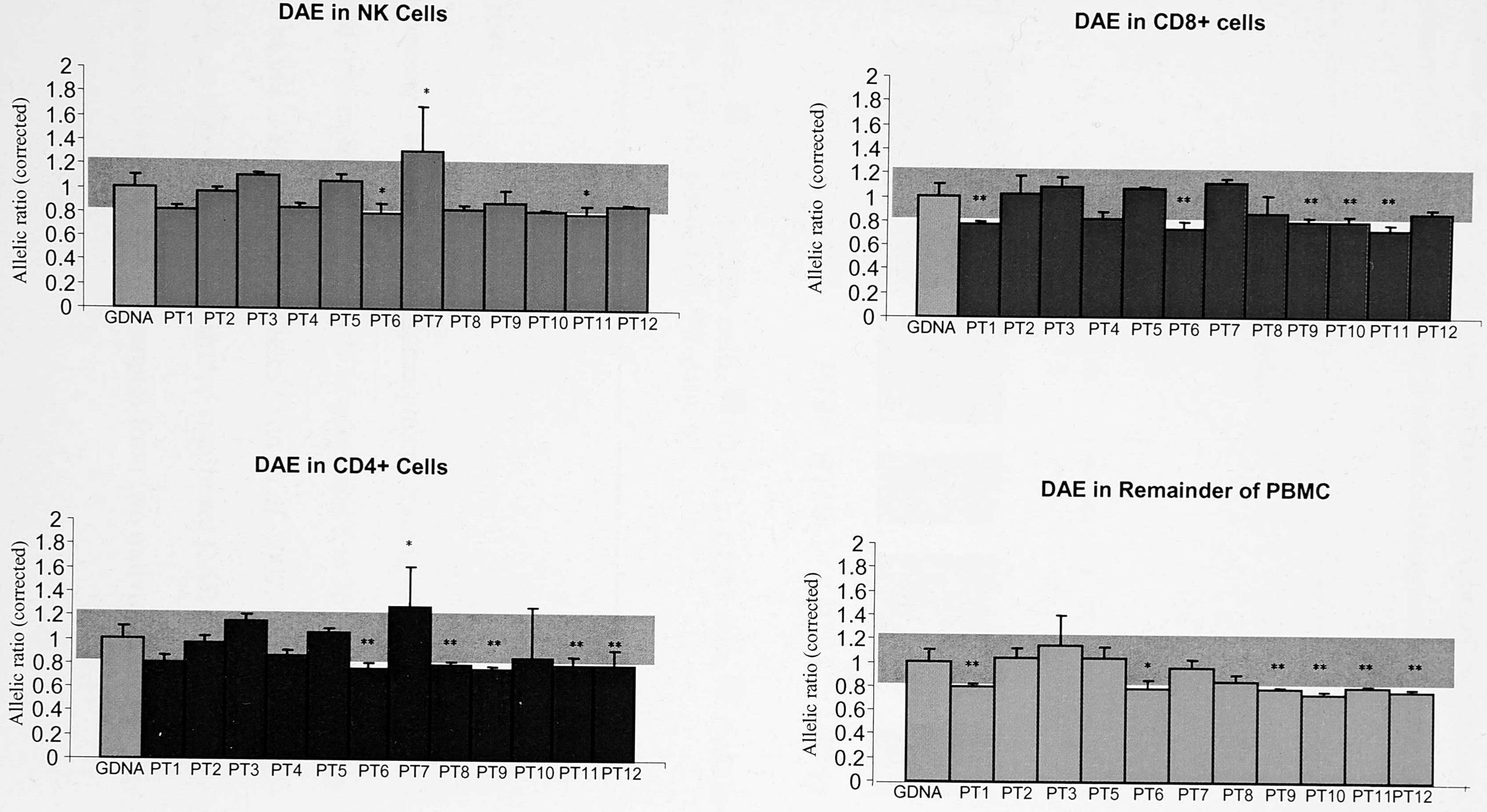
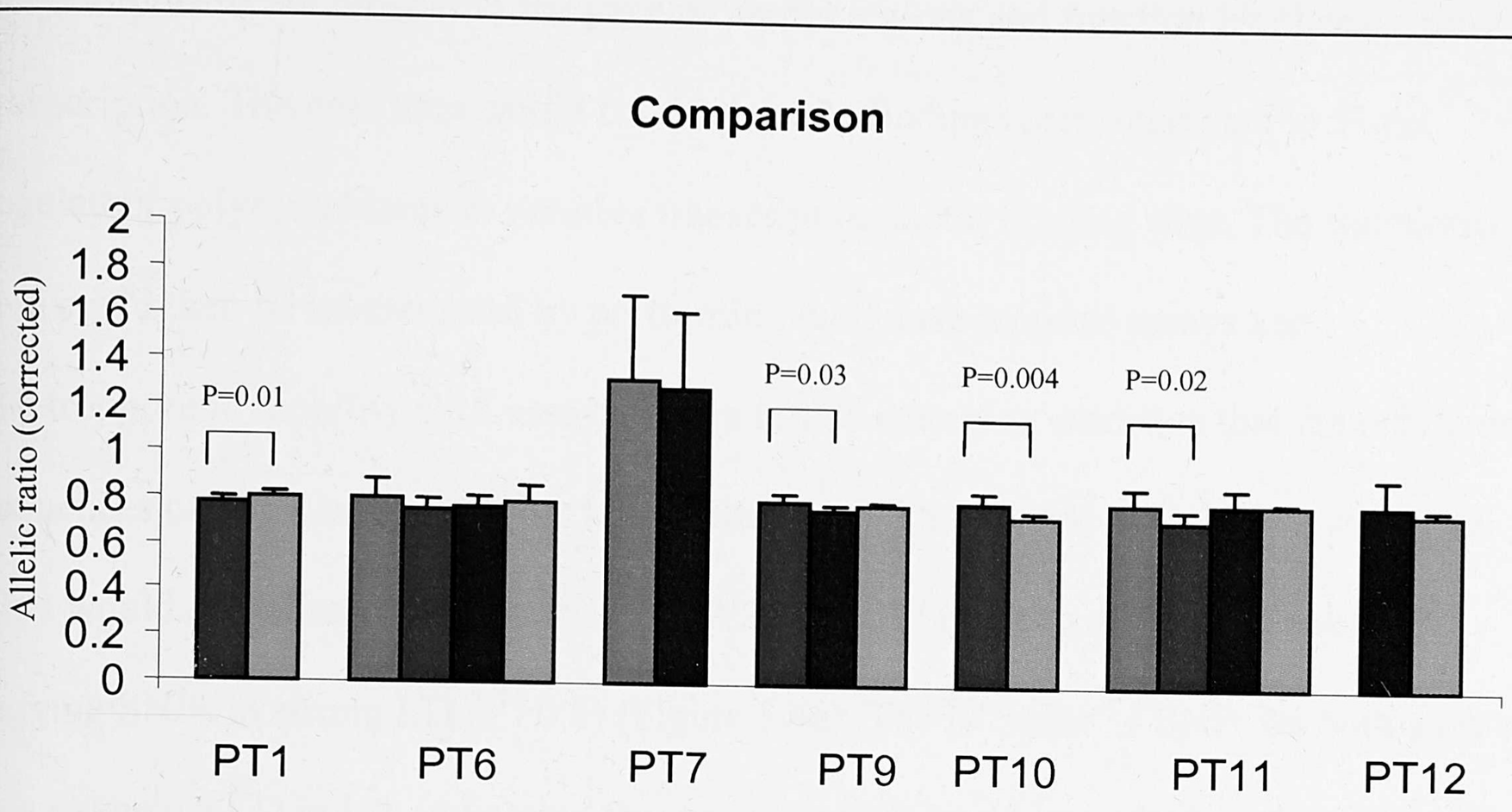


Figure 5.3 Comparison of corrected allelic ratios between different PBMC populations within 7 individual patients showing DAE. The corrected allelic ratios (+1SD) between the two *PTPN22* alleles are shown. Comparison between different cell populations was made using the two-tailed Mann-Whitney test; p values were corrected for multiple comparisons.



DAE in NK cells, DAE in CD8+ cells, DAE in CD4+ cells, DAE in the remainder of the PBMC population, PT-patient

5.5 Discussion

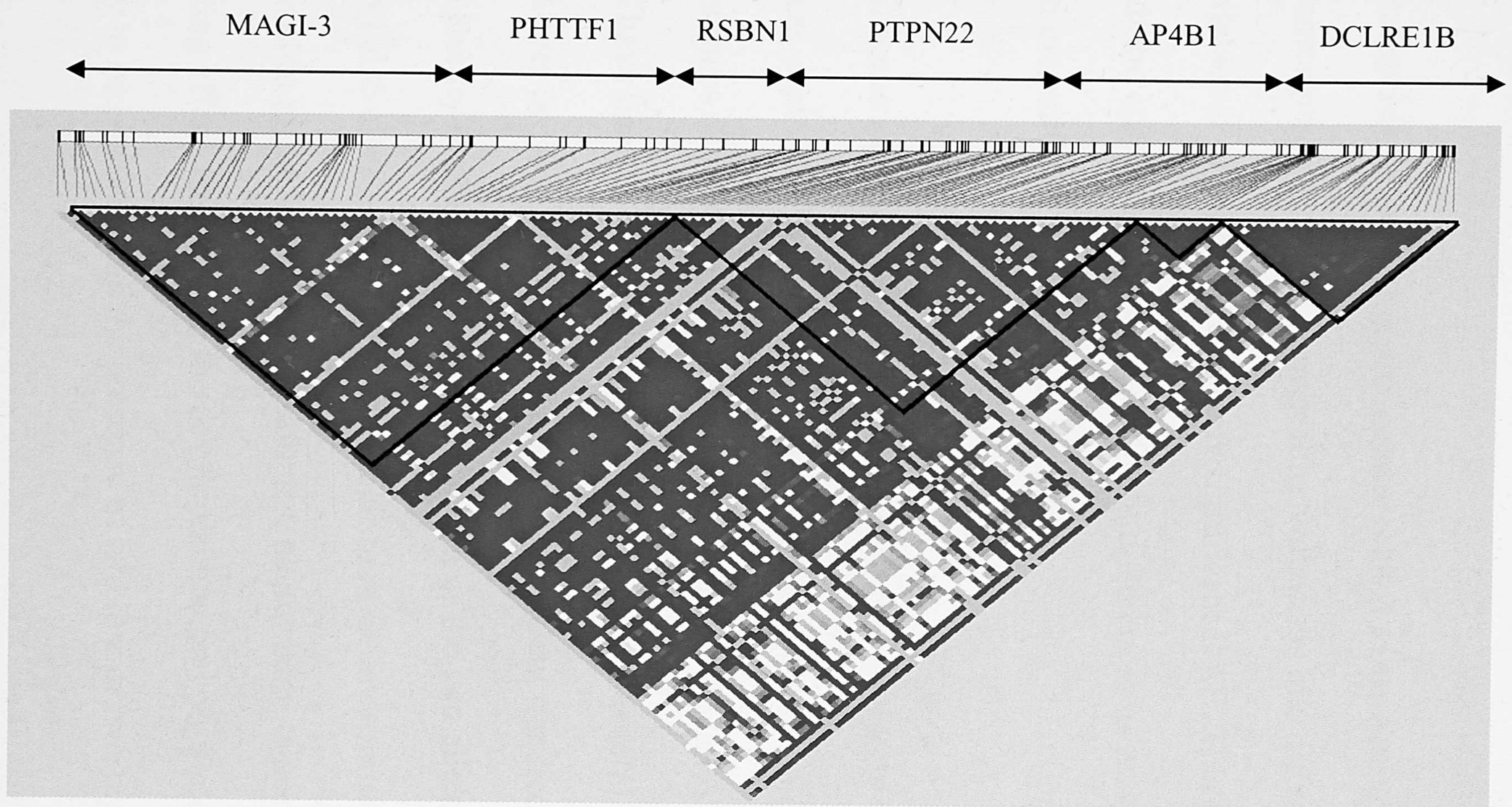
It has become increasingly apparent that polymorphisms leading to changes in gene expression levels are common, with 25 to 50% of genes and 5 to 25% of heterozygotes showing evidence of unequally expressed alleles (Yan *et al.* 2002; Lo *et al.* 2003; Pastinen *et al.* 2004). In this study, I demonstrated significant DAE of *PTPN22* in various PBMC populations from patient with RA. Samples from two individuals in this study

showed a decrease in ratio in all the cell populations examined, four showed significant decrease in two cell populations with non-significant decrease in others and one patient demonstrated increase of DAE in two cell populations. These results would indicate the possible existence of *cis*-acting genetic variation. Most of the known regulatory polymorphisms are located in the gene promoter regions and function by altering gene transcription. The next step would be to relate the findings presented here to 5'-*cis* regulatory polymorphisms in putative transcription factor binding sites. The functional role could then be investigated by performing luciferase reporter assays and electrophoretic mobility shift assays. There is also emerging evidence that the regulatory sequences can be distributed over long distances and be located outside promoter area. This would complicate the matter as *PTPN22* R620W maps to a 293 kb block of 27 tagging SNPs in strong LD (r^2>0.8) (Figure 5.4a). The D' value of SNPs on both ends of this extensive LD is 1.0, indicative that genetic effect could be contributed to other polymorphisms located anywhere in the long segment.

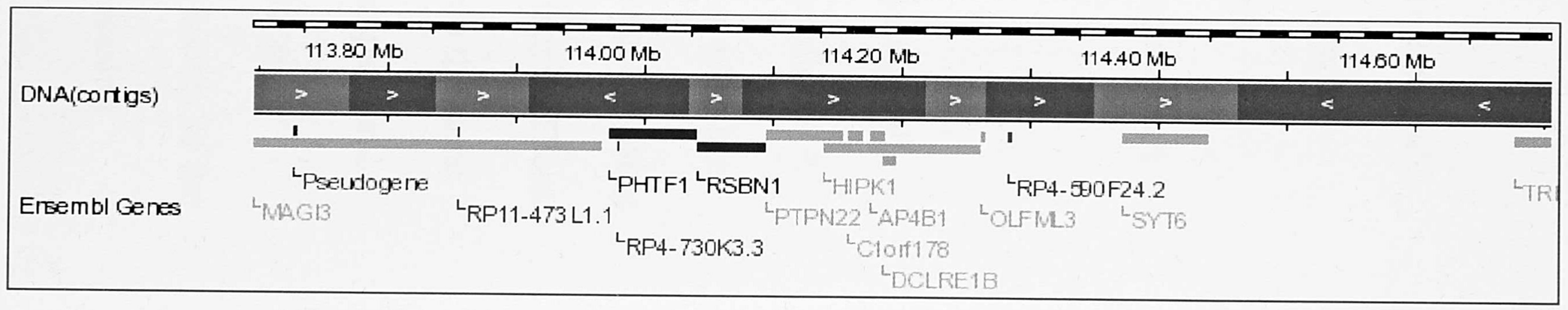
Six other known genes (Figure 5.4b) map to this LD block: the 3'part of the membrane associated guanylate kinase related gene (*MAGI*-3), putative homeodomain transcription factor 1 (*PHTF1*), the round spermatid basic protein 1 (*RSBN1*, also known as FLJ11220), homeodomain-interacting protein kinase 1 (*HIPK1-1*). *HIPK1-1* has nine transcripts, one of which, HIPK1-203, overlaps with adapter related protein complex-4, beta 1 subunit (*AP4B1*) and the 5' part of DNA cross link repair 1B protein (*DCLRE1B*) (Appendix D, Figure D.1).

Figure 5.4 a The extensive LD block located in 1p13. The genotypic data are delivered from the International HapMap Project. The haplotype map was generated by Haploview v. 2.03 software based on genotyping results from HapMap Genome Browser Phase 1 & 2 B36. **b** overview of genes in region 1p13 (Ensembl release 50).

a



b



Altogether the 293 kb block contains 219 marker SNPs according to the HapMap Genome Browser Phase 1 & 2 B36, of which sixteen are non-synonymous (ns) and twenty (including 8 nsSNPs) demonstrate high human-mouse conservation and therefore could have a potential regulatory role.

A previous investigation of *cis*-acting and gender related *trans*-acting factors that could influence the transcription of *PTPN22* in healthy volunteers and different stages of pregnancy by Nielsen *et al.* did not reveal any significant DAE. This apparent discrepancy with the current study could be due to differences in study population and methods used to measure DAE (Nielsen *et al.* 2007).

The difference in gene expression between individuals that are likely to be biologically relevant is unclear, and is likely to be gene dependent. Current techniques can detect expression differences as low as 1.2 fold between samples and alleles. Both *in vitro* and *in vivo* methods have been used to detect the DAE. There are several advantages in using *in vivo* methods over *in vitro* ones (mainly involving transient transfection assays). These include the opportunity to observe relative expression of the two alleles within the same tissue sample and also alleles are expressed in their normal environment. Comparison of alleles is made within rather than between samples, maximizing the sensitivity of detecting *cis*-acting effect. In addition, the developmental and physiological history of the tissue is unlikely to be perturbed by the presence of two low or two high expressing alleles.

The variation of the DAE in different tissues and cells has been observed before and is linked to allele-specific binding of nuclear factors in different cell types. Koch *et al.* reported that one allele of a *cis*-regulatory polymorphic site in the interferon-gamma receptor alpha chain precursor (*IFNGR1*) gene showed 4-fold reduction in gene expression in endothelial cells, a 30% increased in gene expression in B-lymphocytes, and no significant effect on expression in T-lymphocytes (Koch *et al.* 2006). In my sample DAE was observed in all cell groups, most commonly in $CD4^+$ and the reminder of the PBMC. However, comparison of DAE for each patient's cell group revealed that $CD8^+$ cells in particular were significantly different (Figure 5.3).

A number of recent studies have highlighted the previously unappreciated prevalence of structural genetic variations including insertions, deletions and duplications which have been collectively defined as copy number variations (CNV) (Redon *et al.* 2006, Feuk *et al*. 2006). Copy number variable regions (CNVR), which can encompass genes, functional regulatory elements and conserved non-coding sequences, have been reported to influence gene expression and have impact on complex phenotypes. Such variants can lead to an imbalance in the level of expressed RNA at the allelic level and may therefore generate a DAE phenotype related to a particular gene.

In summary, I have demonstrated a significant difference in expression of *PTPN22* alleles (averaging 20%) in RA patients that may indicate the existence of relevant *cis*-acting genetic variation. The differences in *PTPN22* allelic expression could also be cell type specific. Overall, the effects of these regulatory polymorphisms are likely to be subtle, but nonetheless important in explaining some of the complexity of multi-factorial disease like RA.

Chapter 6

Effects of *MHC2TA* genetic variation on susceptibility and severity of rheumatoid arthritis

6.1 Introduction

Genes in the HLA region remain the most powerful disease risk genes in RA. MHC class II expression is controlled mainly at the transcription level, by a highly conserved regulatory module that is situated approximately 150-300 base pairs upstream of the transcription-initiation site in all MHC class II isotypes (HLA-DP, HLA-DQ and HLA-DR), in the first intron of the gene encoding invariant chain (Ii) and in the non-classical MHC class II molecules HLA-DM and HLA-DO (Ting *et al.* 2002, Boss *et al.* 2003). This MHC-class-II-specific regulatory module known as the SXY module consists of four sequences – S, X, X2 and Y boxes that are present in a tightly constrained order, orientation and spacing. Four key *trans*-acting factors bind cooperatively with SXY module and regulate the transcription (Accolla *et al.* 1983, Accolla *et al.* 1986). This multiprotein complex, which is known as the MHC class II enhanceosome, acts as a "landing pad "for the MHC class II transactivator (CIITA) (Masternak *et al.* 2000; Ting *et al.* 2002; Boss *et al.* 2003). CIITA is a non-DNA-binding co-activator that acts as a platform for the assembly of several transcription factors at MHC promoters (Figure 6.1).

CIITA expression is mainly dependent on the transcription of its gene *MHC2TA*, although it can be further modulated by changes in mRNA and protein stability. Transcription of the *MHC2TA* is further controlled by four functional promoters which are known as pI, pII, pIII and pIV (Muhlethaler-Mottet et al. 1997). Each promoter is activated in a tissue-specific manner and leads to the generation of different isoforms

Figure 6.1 Regulation of the transcription of MHC class II genes. The SXY module that is present in all classical MHC class II genes, as well as in the genes encoding invariant chain, HLA-DM, HLA-DO and MHC class I molecules, is bound cooperatively by four factors: the heterotrimeric X-box-binding factor regulatory factor X (RFX); the X2-box-binding factor cyclic-AMP-responsive-element-binding protein (CREB); the Y-box-binding factor nuclear transcription factor Y (NF-Y); and as yet un-identified S-box-binding factor. These four factors form a multiprotein complex –MHC class II enhanceosome. CIITA binds with several components of the enhanceosome and coordinates the recruitment of additional factors that are involved in chromatin modifications and remodeling (Adapted from *Friese et al*. 2005). (Figure 6.2).

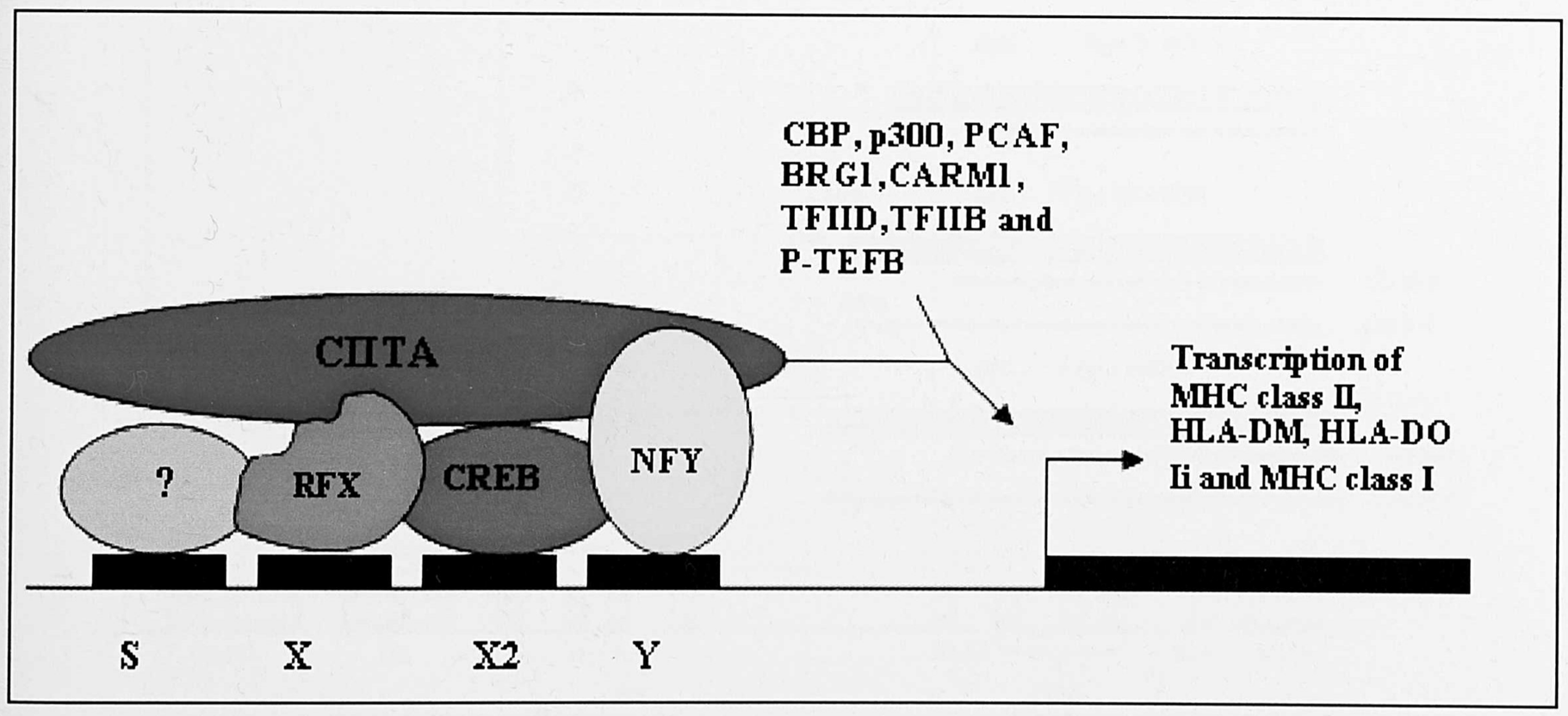
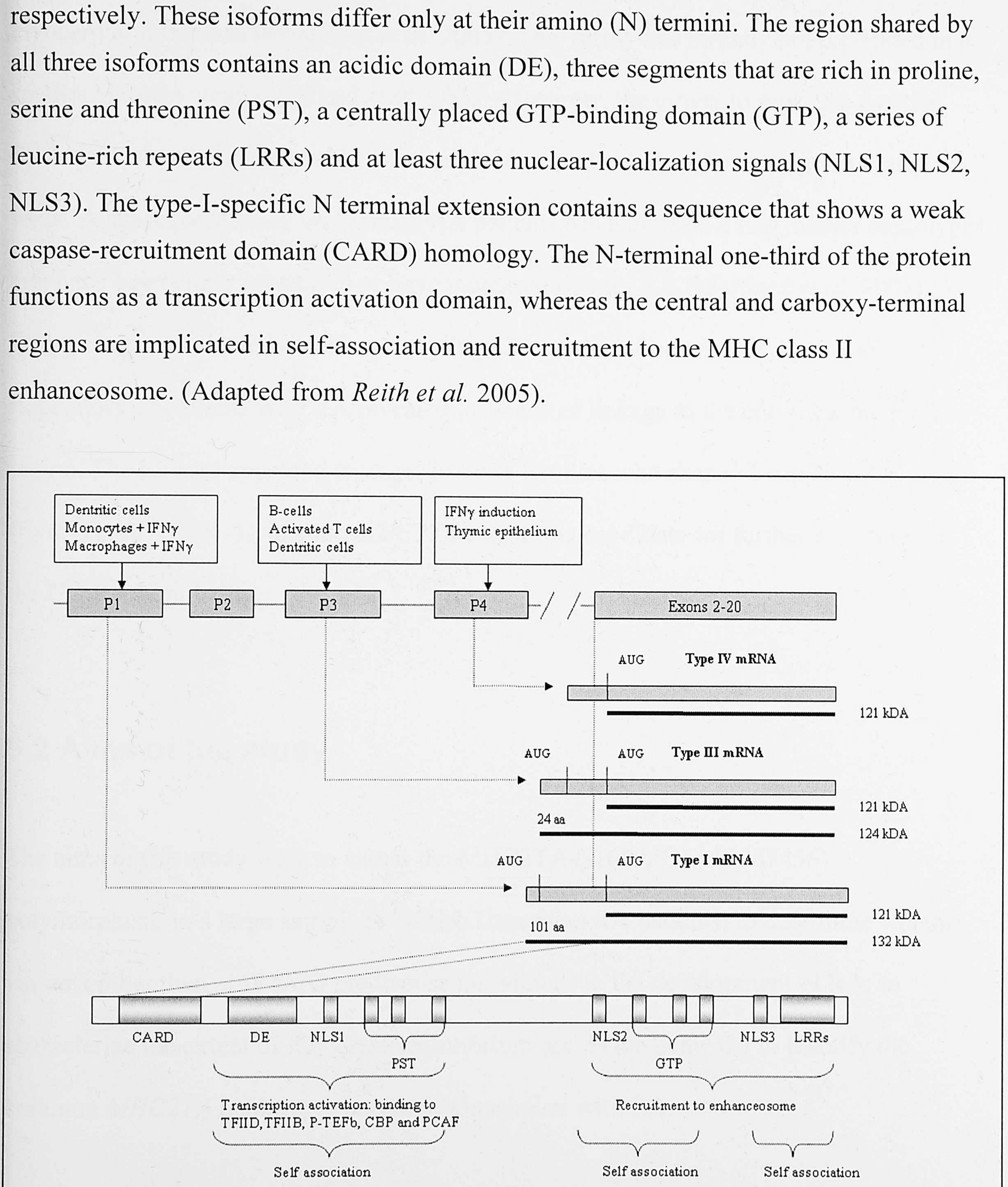


Figure 6.2 Transcription of the *MHC2TA* is controlled by 3 alternative promoters that exhibit cell type specific activity. These promoters precede alternative first exons that are spliced to the shared downstream exons. A fourth promoter (pII) is present in the human gene but is not conserved in mice. The 3 types of mRNA encoding CIITA (type I, III and IV) are derived from pI, pIII and pIV and encode their isoforms of 132, 124 and 121 kDA respectively. These isoforms differ only at their amino (N) termini. The region shared by all three isoforms contains an acidic domain (DE), three segments that are rich in proline, serine and threonine (PST), a centrally placed GTP-binding domain (GTP), a series of leucine-rich repeats (LRRs) and at least three nuclear-localization signals (NLS1, NLS2, NLS3). The type-I-specific N terminal extension contains a sequence that shows a weak caspase-recruitment domain (CARD) homology. The N-terminal one-third of the protein functions as a transcription activation domain, whereas the central and carboxy-terminal regions are implicated in self-association and recruitment to the MHC class II enhanceosome. (Adapted from *Reith et al.* 2005).



MHC2TA was originally identified by the cloning of a gene with a null mutation in hereditary MHC class II deficiency, also known as bare lymphocyte syndrome (Steimle *et al.* 1993). Recently a Swedish group described a SNP (-168A/G, rs3087456) in the type III promoter of *MHC2TA* associated with susceptibility to RA, multiple sclerosis and myocardial infarction (Swanberg *et al.* 2005). This result was initially not replicated in a smaller German sample (Akkad *et al.* 2006). However, the power to replicate positive findings is heavily influenced by the sample size, particularly where the genetic effect is weak. A subsequent study on Spanish RA patients has identified 2 two marker haplotypes that have been either protective or have increased risk for RA (Martinez *et al*. 2007). A linkage study conducted on affected sibling pair families from the UK National Repository of patients with RA revealed evidence of linkage to the chromosome 16p13 region. The linkage appeared stronger in these families who shared 2 copies of SE alleles (Eyre S *et al*. 2004). Hence the *MHC2TA* is a strong candidate for further evaluation in the UK RA population.

6.2 Aims of the study

The aims of this study were to assess the MHC2TA (-168A/G) (rs3087456) polymorphism in a large sample of British Caucasian RA patients; to determine whether variants other than -168*A/G predispose individuals to the development of RA; to characterize the extent of linkage disequilibrium across the gene and to identify the common *MHC2TA* haplotypes and their association with RA.

6.3 Methods

6.3.1 Genotyping *MHC2TA* (-168A/G) (rs3087456)

All RA cases (n = 756) were genotyped for *MHC2TA* (-168A/G) polymorphism using ARMS PCR with additional mismatch (see appendix E Table E.1 for the primer sequences and PCR conditions; see Figure E1 for the agarose gel view; see Chapter 2 for the more detailed methods description). All samples were genotyped in duplicate and positive and negative controls were included. Healthy controls for this study were obtained from 3 sources; I genotyped 613 healthy blood donors that have been characterized in Chapter 3 using the same ARMS PCR method that was used to genotype the RA cohort; 441 unaffected parents from ankylosing spondylitis families were genotyped by KBiosiences. SNP rs3087456 was included in Welcome Trust Case Control Consortium study and genotypes for 1916 subjects from 1958 British birth cohort were available in a public database (http://www.b58cgene.sgul.ac.uk/search.php). All together genotypes and allele frequencies for 3,015 healthy controls were available for association study. Further stratification analysis was undertaken in RA patients, investigating whether association was present in specific patient subsets based on RF, SE alleles and clinical severity characteristics.

6.3.2 Power calculation for *MHC2TA* (-168A/G) study

Sample sizes were calculated based on published minor allele frequency of 0.22, so that association study of the functional RA causal SNP had 88% power to detect a gene

conferring odd ratio of 1.3 as reported in the original Swedish study at the 5% significance level assuming a dominant model for inheritance.

6.3.3 Analysis of the *MHC2TA* (-168A/G) genotyping data

Statistical analysis for association and Hardy-Weinberg equilibrium were evaluated using χ^2 test. P values less than 0.05 were considered significant. For continuous variables the statistical software package SPSS version 15.0 for Windows was used.

6.3.4 Sequencing the *MHC2TA* gene

In order to understand the full genetic diversity within *MHC2TA* in the context of RA, I sequenced the gene in a set of RA patients and HC. Sequencing DNA from 24 patients and 24 healthy controls (i.e. 48 chromosomes) provided 95% power to detect alleles with ≥ 5% frequency in the population (Kruglyak *et al.* 2001). I performed bidirectional sequencing on all the exons and exon-intron boundaries. Primer frequencies and PCR conditions are available in the appendix E Table E.2. Sequencing was performed using the 3100 DNA Genetic Analyser (Applied Biosystems). The GeneMapper 3.5 software was used to align and compare chromatographs. (Applied Biosystems). An example of the electrophenogram can be viewed in the appendix E (Figure E.1).

6.3.5 SNP selection and genotyping

Following the initial sequencing, a set of polymorphisms of 17 previously identified SNPs (MAF over 5%) and 5 tagging SNPs from International HapMap project were

subsequently genotyped in 250 RA patients and 250 healthy controls. Primers and PCR conditions can be seen in appendix E (Table E.3).

6.3.6 Power calculation for the subsequent genotyping

With a sample size of 250 RA cases and healthy controls, using a dominant model for inheritance, with 10% allele frequency, I had there was 80% power to detect a relative risk of 1.8.

6.3.7 Statistical analysis of the sequencing data

For single marker analysis P values were calculated using 2 x 2 contingency tables for allele and 2 x 3 tables for genotype counts. I did not apply corrections for multiple testing as the genotypic data was not solely independent and therefore Bonferroni correction would most likely have been overly conservative.

Linkage disequilibrium measuring *D*' and r^2 was calculated between every marker pair using LDMAX program in the GOLD package (Abecasis *et al*. 2000).

Haplotypes were constructed using Bayesian method of Phase v.2.1.1. (Stephens M *et al.* 2001). To assess the significance difference of haplotype frequencies between the groups I used CLUMP program (Sham *et al.* 1995).

6.4 Results

6.4.1 *MHC2TA* (-168A/G) polymorphism

All the genotypes were in Hardy-Weinberg equilibrium and there was no association of the *MHC2TA* (-168A/G) with susceptibility to RA (Table 6.1). When the RA population was analysed according to the RF status, the minor allele was less common in RF positive compared to RF negative patients (minor allele frequencies 0.25 and 0.30, respectively), but this did not reach statistical significance.

Subsequently I analysed RA patients population in two groups according to the presence of *HLA* SE and *HLA- DRB1*04*. Minor allele frequencies were higher in *HLA* SE negative compared to *HLA* SE positive patients (0.32 and 0.25 respectively) (p=0.01). A similar trend was observed in the *DRB1*04* negative and positive patients cohort (MAF 0.30 and 0.24, respectively) (p=0.03). The presence of G allele and A/G +G/G genotype seemed to provide protective effect against *HLA* SE and *DRB1*04* positive RA (Table 1).

Further subgroup analysis according to the RA severity characteristic did not reveal any associations with *MHC2TA* -168*A/G and the presence of nodules, requirement for joint replacement or radiographic progression.

6.4.1 *MHC2TA* sequencing and subsequent genotyping

The sequencing data revealed 21 previously described SNPs and 3 novel variations that are presented in Table 6.2. A collection of polymorphisms, including 17 SNPs identified from the initial sequencing project (MAF over 5%) and 5 tagging SNPs from the

International HapMap project were subsequently genotyped in 250 RA patients and 250 healthy controls. The results are presented in Table 6.3. Except for SNP15 (rs7500869), all other variations were in Hardy-Weinberg equilibrium. Analysis of LD data of 22 SNPs (including also the promoter SNP rs3087456) revealed that there was a small separate disequilibrium block involving the promoter area and first exon, separate from the rest of the gene (Figure 6.3). Another disequilibrium block was observed stretching from intron 10-11 area (rs45481191) to exon 18 (rs34538398). Comparing these findings with LD block obtained from the HapMap project revealed similar results.

Single marker analysis revealed that only SNP10 in exon 11 was modestly associated with RA ($p=0.04$). Incorporating all 22 SNPs studied and carrying out a PHASE analysis, produced 156 different haplotypes with $p=0.9$, suggesting no statistically significant skewing in the haplotype distribution. The haplotype-specific results (by χ^2 test) for the ten most common 22 marker haplotypes are shown in Table 6.4.

Further two- and three-marker haplotype analysis was carried out, with SNPs rs3087456/rs4774 and rs3087456/rs4774/rs2229320. There were no globally significant differences in the frequencies of two- and three-marker haplotypes between RA and HC (with empirical $p=0.11$ and $p=0.09$ respectively). Consequent analysis of RA cohort according to the presence of *HLA-DRB1* SE also did not reveal any significant difference in two- or three-marker haplotype frequencies ($p=0.22$ and $p=0.36$ respectively).

Table 6.1 Genotyping data and allele counts for *MHC2TA* (-168A/G) polymorphism in RA cases and HC.

	Allele counts			Genotyping counts				Genotyping counts G-dominant model		
	A (%)	G (%)	OR, 95% CI	AA (%)	AG (%)	GG (%)	P value	AA (%)	AG+GG (%)	OR, 95%CI
RA cases n=734	1082 (74)	386 (26)	1.04 (0.92-1.19) p=0.51	403 (55)	276 (37)	55 (8)	0.66	403 (55)	331 (45)	1.03 (0.88-1.21) p= 0.71
Controls n=3015	4495 (75)	1535 (25)		1678 (56)	1139 (38)	198 (6)		1678 (56)	1337 (44)	
*RF neg n=156	218 (70)	94 (30)	1.26 (0.99-1.62) *p=0.07	76 (49)	66 (42)	14 (9)	0.18	76 (49)	80 (51)	1.32 (0.96-1.82) p=0.09
*RF pos n=567	846 (75)	288 (25)	1.00 (0.86-1.15) *p= 0.97	320 (56)	206 (37)	41 (7)	0.73	320 (56)	244 (44)	0.96 (0.80-1.15) p=0.63
RF pos *v* RF neg			1.27 (0.96-1.67) p= 0.09				0.23			1.38 (0.97-1.97) p=0.08
SE present n=567	854 (75)	280 (25)	0.71 (0.54-0.92) p=0.01	324 (57)	206 (36)	37 (7)	0.04	324 (57)	243 (43)	0.67 (0.48-0.95) p=0.03
SE absent n=167	228 (68)	106 (32)		79 (47)	70 (42)	18 (11)		79 (47)	88 (53)	
DRB1*04 present n=474	716 (76)	232 (24)	0.77 (0.61-0.98) p=0.03	274 (58)	168 (35)	32 (7)	0.10	274 (58)	200 (42)	0.71 (0.53-0.97) p=0.03
DRB1*04 absent n=260	366 (70)	154 (30)		129 (50)	108 (42)	23 (8)		129 (50)	131 (50)	

*versus healthy controls.

Table 6.2 Results of the direct sequencing of *MHC2TA* gene, *minor allele frequency ** Exon 11 present in CIITA-001, CIITA-002 and CIITA-004 transcripts. ***Exon 18 in CIITA-003 transcripts, exon 20 in CIITA-001 and CIITA-002 transcripts. † - positions are according to Ensembl database

SNP	Rs number	Position†	Function	SNP	MAF* RA	MAF* HC
SNP1	rs45456897	29824	Intron 2-3	C→T	0.04	0.02
SNP2	rs649821	33169	Intron 4-5	G→T	0.16	0.16
SNP3	rs4781016	36853	Intron 6-7	C→A	0.25	0.4
SNP4	rs45545642	39311	Intron 9-10	C→T	0.06	0.02
SNP5	rs4774	41302	Exon 11**	G→C	0.22	0.31
SNP6	novel	41876	Exon 11**	C→A	0.04	0.0
SNP7	rs13330686	42145	Exon 11**	C→T	0.06	0.04
SNP8	rs2229320	42197	Exon 11**	G→A	0.1	0.14
SNP9	rs2228238	43358	Exon 10	G→A	0.2	0.38
SNP10	rs45481191	43611	Intron 10-11	C→T	0.13	0.06
SNP11	rs4781019	44604	Intron 11-12	C→T	0.33	0.58
SNP12	rs7500869	50401	Intron 12-13	C→A	0.17	0.17
SNP13	rs6498130	50604	Intron 12-13	G→T	0.56	0.44
SNP14	rs7404786	53004	Intron 14-15	G→C	0.54	0.38
SNP15	rs7201430	56499	Exon 15	C→T	0.08	0.06
SNP16	rs4781024	57512	Intron 16-17	G→A	0.38	0.36
SNP17	novel	58181	Exon 18***	C →T	0.0	0.02
SNP18	rs4780336	58182	Exon 18***	G→C	0.02	0.04
SNP19	rs11074940	58856	Exon 18***	C→T	0.08	0.16
SNP20	rs34538398	58901	Exon 18***	G→A	0.08	0.08
SNP21	rs1139564	59076	Exon 18***	T→C	0.21	0.13
SNP22	rs45466393	59077	Exon 18***	G→A	0.04	0.02
SNP23	Rs45538933	59093	Exon 18***	G→A	0.02	0.02
SNP24	novel	59319	3'downstream	G→A	0.0	0.02

Table 6.3 The list of the SNPs and genotyping methods in the second phase. * - tagging SNPs added from the International HapMap Project.

SNP	dbSNP ID	Variation	Genotyping freq. in RA			HW in RA	Genotyping freq. in HC			HW in HC	P value
			11	12	22		11	12	22		
1	rs3087456	A/G	139	89	13	0.80	139	104	12	0.18	0.65
2	rs12928665*	A/G	144	85	13	0.92	135	104	11	0.10	0.41
3	rs8043545*	G/C	139	88	14	0.98	119	110	13	0.05	0.13
4	rs6498119*	T/C	214	30	0	0.31	226	29	1	0.95	0.59
5	rs6498121	T/G	158	78	8	0.67	155	78	10	0.96	0.72
6	rs4781016	C/A	117	97	29	0.21	127	106	18	0.52	0.22
7	rs45545642	C/T	212	28	1	0.94	209	37	2	0.80	0.22
8	rs4774	G/C	122	92	26	0.18	125	108	15	0.18	0.13
9	rs13330686	C/T	210	30	0	0.30	208	38	2	0.86	0.19
10	rs2229320	G/A	168	67	8	0.71	156	86	14	0.95	0.04
11	rs2228238	G/A	121	89	29	0.05	127	106	20	0.75	0.51
12	rs45481191	C/T	197	42	2	0.88	194	51	2	0.50	0.41
13	rs4781019	C/T	82	108	45	0.38	76	121	53	0.71	0.30
14	rs4781020*	G/A	132	95	15	0.70	122	101	24	0.64	0.13
15	rs7500869	C/A	154	77	10	0.92	161	80	7	0.05	0.43
16	rs6498130	G/T	89	105	44	0.19	89	118	41	0.86	0.94
17	rs9788916*	G/A	182	56	3	0.57	198	56	1	0.15	0.47
18	rs7201430	C/T	194	45	2	0.73	197	51	4	0.74	0.44
19	rs4781024	G/A	90	112	37	0.83	83	115	36	0.71	0.74
20	rs11074940	C/T	209	20	0	0.49	214	34	0	0.25	0.10
21	rs34538398	G/A	211	24	0	0.41	226	25	0	0.41	0.93
22	rs1139564	T/C	160	65	12	0.12	164	72	9	0.75	0.87

Figure 6.3 Pairwise linkage disequilibrium was calculated based on the genotyping results of 250 RA patients using the program GOLD. Orange areas represent a D' of 1 which means maximum LD. The D' decreases as the colours go from orange through green to blue. Dark blue areas represent D' = 0, meaning absence of LD. The X and Y axes are mirror images. The SNPs are represented along the X and Y axes as small vertical and horizontal lines respectively. The SNPs have been assigned numbers reflecting their position along chromosome 16 from telomere to centromere.

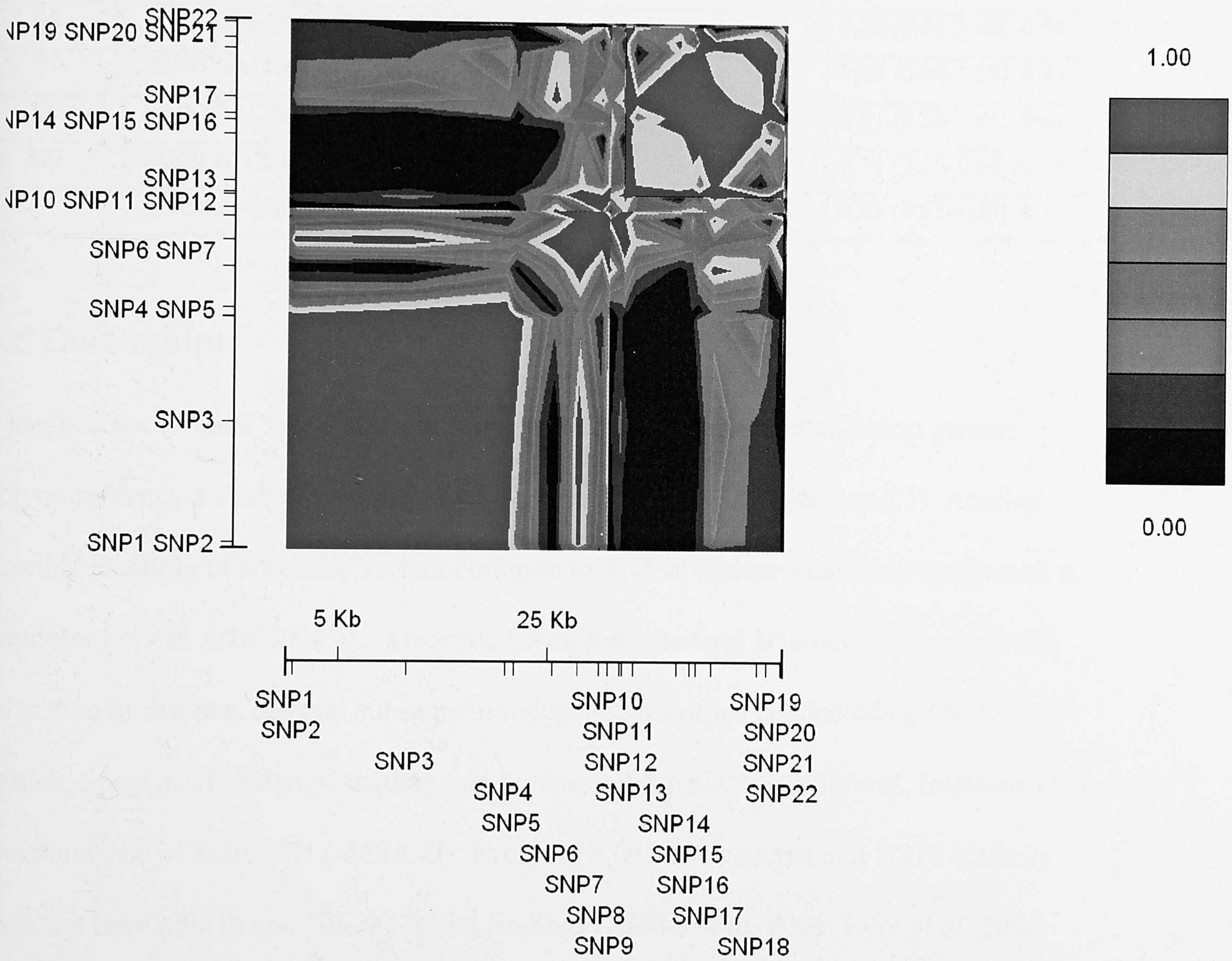


Table 6.4 Ten most common *MHC2TA* 22 marker haplotypes for RA cases and HC.

Haplotype No	Haplotype	No. of RA	No. of HC	OR 95%CI, P
2	AAGTGCCGCGACTGCTGCGCGC	66	55	1.29 (0.87-1.94) 0.18
9	AAGTGCCGCGACTGAGGCGCGC	17	18	1.00 (0.50-1.95) 0.98
11	AAGTGCCGCGACTGAGGCACGC	15	12	1.32 (0.61-2.85) 0.47
24	AAGTGCCGCGATTGCTGCGCGT	19	13	1.57 (0.76-3.22) 0.21
33	AAGTGCCGCAACCACGGCACGC	43	44	1.03 (0.66-1.60) 0.91
49	AAGTGCTCTGGCTGCTGTGCGC	17	28	0.62 (0.33-1.15) 0.13
64	AAGTGACCCGGCCGCGACGCGC	25	22	1.20 (0.67-2.16) 0.54
65	AAGTGACCCGGCCGCGACGCGT	24	20	1.27 (0.69-2.33) 0.43
70	AAGTGACCCGGCCACGGCACGC	15	13	1.21 (0.53-2.81) 0.61
122	GGCTTCCGCAACCACGGCACGC	11	16	0.71 (0.29-1.66) 0.39
Remaining haplotypes		236	271	0.83 (0.65-1.06) 0.15

6.5 Discussion

It has become evident that distinct autoimmune diseases can share common genetic polymorphisms, a classic example of which is in *PTPN22* (Chapter 4 and 5). Another possible example of a genetic variant common to several diseases has been suggested: a promoter SNP in *MHC2TA* was associated with RA, multiple sclerosis and myocardial infarction in Swedes. Several subsequent independent studies conducted on German, British, Spanish, Austrian, Canadian and Japanese RA patient populations, followed by a meta-analysis of *MHC2TA* (-168A/G) involving 6,861 RA patients and 9,270 controls have not been able to confirm the initial findings (Akkad *et al.* 2006, Eyre *et al.* 2006, Yazdani-Biuki *et al.* 2006, Newman *et al.* 2006, Iikuni *et al.* 2007, Bronson PG *et al.* 2008). The likely explanation for this apparent discrepancy may lie in the differences in

the minor allele frequencies found between different control populations. It is striking that the minor allele frequencies are very similar in the RA patients from each population (24.4% in Swedes, 26% in German, 26% in British, 30% in Canadian, 28% in Austrian) but in contrast, it is only 20.5% in the Swedish controls compared with 26% in the Germans, 25% in the British, 30% in Canadian, 31% in Austrian. Such a difference in the control samples could have occurred by chance, generating a spurious association with RA in the Swedes.

In their analysis of the Swedish the population authors indicated a significant effect for the GG genotype in a co-dominant model of susceptibility (Swanberg *et al.* 2005). This equates with an excess of 21 patients with the GG genotype in the RA group compared to predictions (n= 62) based on the genotype frequency in controls (4.8%). This would be a tiny patient subgroup (constituting only 1.6% of their study population) compared with other known RA associations, such as *HLA-DRB1*.

Different results were reported by Japanese investigators who genotyped *MHC2TA* (-168A/G) in 1,121 RA patients but only 450 healthy controls. They found significant association between the *MHC2TA* (-168A/G) SNP and susceptibility to RA ($p = 0.0026$ OR 1.37 95%CI 1.11-1.70) (Iikuni *et al.* 2007). The risk is associated with GG genotype, however the frequency of G allele in Japanese RA and HC population was much higher (0.86 and 0.82, respectively) compared to Europeans (on average 0.25). The dbSNP database indicates that the G allele is ancestral. Studies in the Caucasian population have demonstrated the variant allele A is more common, although this was not the case in

Japanese cohort. As no further replication study of the *MHC2TA* (-168A/G) polymorphism in an Asian population has been conducted it is yet unknown whether the Japanese findings were unique.

A replication study carried out by Orozco *et al.* involved a small cohort of Swedish, Spanish and Argentinean RA patients. Interestingly they reported higher A allele frequencies in Spanish RA patients and this difference was statistically significant (p=0.01, OR 1.23 95%CI 1.04-1.47) (Orozco *et al.* 2006). No such association was observed in the Swedish or Argentinean cohorts. The G allele frequencies in the Swedish healthy controls were similar to other published reports but different from the initial Swanberg *et al.* study.

Ethnic differences can account for different allelic frequencies, and it is not uncommon to see significant differences in allele frequencies between ethnic groups. However, this does not explain why *MHC2TA* (-168A/G) SNP is associated with RA in these particular, ethnically very different populations. Genetic heterogeneity due to differences in the haplotype structure of *MHC2TA* locus across populations could be one explanation. It is also possible that the association with the A allele and RA found in the Spanish population could be due to type I error, a hypothesis supported by the authors themselves.

In the original Swedish study a surprisingly low proportion (62.5%) of the RA group were RF positive compared to our sample. *MHC2TA* may therefore be more relevant to the aetiology of seronegative RA and I did find a small (non-significant) difference

between our seropositive and negative patients. Furthermore there was a modest association of the G allele and A/G + G/G genotypes with *HLA* SE negative and *HLA-DRB1*04* negative RA. It is however unclear how the polymorphisms in the *MHC2TA* could be associated with certain *HLA-DRB1* alleles. One possibility is that certain *MHC2TA* haplotypes can increase the chances of certain MHC class II isotypes being differentially produced.

The investigation if whether different *MHC2TA* haplotypes can have an effect on RA susceptibility was initially carried out on 350 Italian patients. The authors demonstrated that the two marker rs3087456/ rs4774 (SNP located in exon 11) A/C haplotype conferred protection (p = 0.024; OR 0.72 (0.54-0.97) and while the G/C haplotype marginally increased the risk to RA (p=0.05; OR=1.77 (0.95-3.3) (Martinez *et al.* 2007). A similar protective trend with the AC haplotype (p=0.08) was also observed in a Norwegian RA population (O'Doherty *et al*. 2007). I could not confirm the presence of this apparently protective haplotype in our cohort. This may be due to differences in the haplotype frequencies. In the Italian study cohort AC hapolytpe was observed in 15.3% of RA patients and 20.5% HC, in the Norwegian population the the AC haplotype occurred in 23% of RA and 27% of HC; in contrast in my study group the haplotype frequencies were 27% and 26%, respectively. Investigations into the effect of the GC haplotype revealed that this haplotype did increase risk of RA in the British population which was similar to the Italian population (p=0.03) (Table 6.5). However, the GG haplotype occurred less frequently in British RA and HC in comparison with the Italian

Table 6.5 Two-(rs3087456/ rs4774) and three-marker (rs3087456/rs4774/rs2229320) haplotypes association with RA cases and controls and *HLA* SE positive and negative RA cases.

Haplotype	No of RA	No of HC	OR 95%CI p	RA HLA SE +	RA HLA SE-	OR 95%CI, p
AG	243	250	1.04 (0.80-1.33) 0.76	192	51	1.14 (0.72-1.80) 0.54
AC	130	134	1.02 (0.77-1.36) 0.87	104	26	1.18 (0.71-2.04) 0.50
GG	101	123	0.82 (0.61-1.12) 0.21	76	25	0.83 (0.50-1.45) 0.50
GC	14	5	*0.03	8	6	*0.08
AGG	193	188	1.13 (0.87-1.47) 0.36	154	39	1.20 (0.75-1.93) 0.41
AGA	48	61	0.81(0.53-1.23) 0.29	36	12	0.83 (0.40-1.83) 0.61
ACG	132	136	1.03 (0.77-1.37) 0.86	106	26	1.22 (0.74-2.00) 0.43
GGG	66	70	0.99 (0.68-1.44) 0.95	51	15	0.96 (0.51-1.79) 0.90
GGA	35	54	0.65 (0.41-1.04) 0.06	25	10	0.69 (0.32-1.48) 0.34
GCA	14	5	* 0.03	8	6	*0.09

* Fisher's exact tests was used to calculate the p value

and Norwegian studies. Overall, in all three populations studied the GC haplotype was generally relatively rare, with a frequency of less than 10%.

Study of the three-marker haplotype (rs3087456/rs4774/rs2229320) did not add additional information to the two-marker analysis. I observed likewise that the presence of the G/C/G haplotype increased the risk of RA (p=0.03). However, once again this haplotype was rare with frequency less than 10%.

Association of the G/G haplotype with increased risk to SE positive RA was reported in the Italian RA cohort (Martinez *et al.* 2007). However, in our British RA population I could not confirm this finding. There was a trend towards the G/C haplotype being more common in SE negative patients (p=0.08), however but the frequency of the G/C haplotype in both groups was too small for this to have a major impact.

In conclusion, although I could not confirm the association of the *MHC2TA* (-168A/G) polymorphism with RA, a modest association found with SE negative patients and replication of the effects of GC haplotype could indicate a subtle influence towards increased susceptibility to RA. *MHC2TA* and its regulation are attractive candidates for further studies in assessing the mechanisms of autoimmune diseases. Given its pivotal role in regulating the expression of HLA class I and class II proteins and many other factors involved in antigen presentation and processing, its marginal dysregulation could have an effect on the immune response and increase the susceptibility to various autoimmune disease

Chapter 7

Interleukin-1 promoter region polymorphism role in rheumatoid arthritis and meta-analysis of *IL-1*(-511A/G)

7.1 Introduction

Cytokines are small regulatory proteins that are produced by cells of the immune system and act as intercellular mediators in the generation and control of immune and inflammatory responses. In RA the two main cytokines at the apex of the cytokine network are TNF and IL-1. They encompass a central role in joint inflammation and destruction. IL-1 is typically synthesized and released simultaneously with TNF and the two cytokines share many similar biological properties.

The most studied of the IL-1 proteins are interleukin-1 alpha (IL-1α) and interleukin -1 beta (IL-1β). IL-1α and IL-1β act as agonist molecules that activate target cells by binding to the IL-1 receptor. The production and activity of IL- α and IL-1 β are regulated by transcriptional and post-translational mechanisms, as well as from a natural competitive antagonist (IL-1RA).

The *IL-1* gene cluster (comprising ten related genes) maps to a 430kb region on chromosome 2q12. Both *IL-1A* (MIM 147760) and *IL-1B* (MIM 147720) are thought to

have arisen through gene duplications, and hence encode similar protein products (Nicklin *et al.* 2002). Strong linkage disequilibrium has been demonstrated across this gene cluster in Caucasian populations (Nicklin *et al.*1994).

Polymorphisms of the *IL-1* region have been related to stable inter-individual differences in IL-1 production and association with susceptibility to common inflammatory diseases (Hall *et al.*2004; Timms *et al.* 2004; diGiovine *et al.* 2000; Kornman *et al.*1997). Genetic linkage study in European RA families provided suggestive evidence of linkage of RA to the *IL-1* locus (Barrera *et al.* 2001). The results indicating the association of certain *IL-1* gene cluster polymorphisms with susceptibility and severity, especially in the more erosive form of RA, have been variable and not widely replicated in different populations (Cox *et al*. 1999; Cantagrel *et al.* 1999; Jouvenne *et al.* 1999; Genevay *et al.* 2002). These studies have generally been conducted on relatively small RA patient samples, apart from one recent exception (Johnsen *et al*. 2008). In the latter case the authors performed a comprehensive analysis of *IL-1A* and *IL-1B* in a large sample of North American and Spanish patients with RA, but identified no major contribution to RA from these genes.

7.2 Aims of the study

To examine four common SNPs and haplotype structures in the *IL-1A* and *IL-1B* promoter region and their association with susceptibility to RA in a large group of well characterised British Caucasian patients. In addition, to evaluate the effects of four

polymorphisms on the clinical characteristics and radiological outcome of RA. Finally, based on previously published studies and incorporating genotyping data on *IL-1B* (-511A/G) to perform meta-analysis of published series to evaluate the effect of this polymorphism on susceptibility to RA among Caucasians.

7.3 Subjects and methods

7.3.1 Patients and study design for *IL-1A* and *IL-1B* promoter polymorphims

The study sample and the clinical characteristics have been described in chapter 3. A total of 756 British Caucasian RA patients and 625 healthy blood donors were genotyped for *IL-1A* and *IL-1B* polymorphisms.

7.3.2 Genotyping

Genomic DNA was extracted from whole blood. Cases and controls were genotyped for *IL-1A* (-889 C/A, rs17561), *IL-1B* (-511 A/G, rs16944), *IL-1B* (-1464 C/G, rs1143623), and *IL-1B* (3737 G/A, rs4848306) SNPs using ARMS and RFLP assays. (See appendix F for primer sequences and PCR conditions, Table F1). For technical reasons *IL-1A* rs1800587 G/A was genotyped instead of *IL-1A* (-889 C/A, rs17561) with which it is in complete LD (r^2= 1.0). Limited *HLA-DRB1* typing and further *DRB1*04* subtyping was performed as previously described (Chapter 2). The *HLA-DRB1* SE was defined by the

presence of one or two copies of the *HLA DRB1*01,*0401, *0404, *0405, *0408*, and **10* alleles.

7.3.3 Statistical analysis

7.3.3.1 Case-control association analysis

Statistical analysis for association and Hardy-Weinberg equilibrium were evaluated using χ^2. P values less than 0.05 were considered significant. Corrections for multiple testing were applied were necessary (Sabatti *et al.* 2003). For continuous variables, statistical software packages SPSS version 15.0 for Windows was used. Haplotype estimations and frequency comparison were performed using PHASE v. 2 and CLUMP software.

7.3.3.2 Analysis of radiological progression

Analysis of the radiological status in 206 RA patients has been described in detail in Chapter 3, section 3.2. Associations between particular *IL-1* gene cluster polymorphisms and radiographic severity score were assessed by using the Jonckeheere test for ordered alternatives. I also compared the frequencies of *IL-1* SNPs in subjects with mild (percentile severity score <20) or severe disease (percentile severity score >80). Finally I examined how Rau-Ratingen score evolved as a function of disease duration in subgroups of patients according to the polymorphism status. This analysis was done using the "Cystat" program and locally weighted scatter plot smoothing regression (lowess) was applied (Cleveland *et al.* 1979). This method does not impose a functional form to the

relationship between the variables. It is very closely related to a moving average: the mean Rau-Ratingen score is plotted as a function of time elapsed since diagnosis

7.3.4 Meta-analysis

7.3.4.1 Search strategy and study selection

Meta-analysis was conducted for *IL-1B* (-511A/G) SNPs based on the PubMed database using the search terms "rheumatoid arthritis" and "interleukin-1". Publications in English language and studies conducted on Caucasian populations were included. Study selection criteria were: 1) RA patients fulfilled 1987 ACR classification criteria for RA (excluding juvenile onset RA); 2) controls were derived from a population in the same geographic area and with the same ethnic background as RA patients; 3) authors provided original genotype frequencies; 4) RA patient and control groups did not overlap between studies. There were six studies that met the selection criteria (Cantagrel *et al.* 1999; *Crilly et al.*; 2000. Buchs *et al.*2001; Genevay *et al.* 2002; Tolusso *et al.* 2006; Johnsen *et al.* 2008). All six publications were original articles and had case-control study design. Including our dataset, the meta-analyses incorporated 3,712 RA patients and 2,331 HC.

7.3.4.2 Evaluation of the statistical association

Though previous studies describe *IL-1B* (-511) SNP as a G→A change, the dbSNP database indicates that the A allele is ancestral, thus the variant allele is G. However, the G allele occurs more commonly (allele frequencies among HC are 0.67). In the meta-analysis, I compared the variant G allele with ancestral A as well as recessive (G/G *vs*

G/A + A/A) and dominant (G/G + G/A *vs* A/A) models for the G allele. Genotype specific risks such as G/G *vs* A/A and G/A *vs* A/A were also analysed. Statistical analyses were performed using StatsDirect software. Single odd ratios (OR) and 95% confidence intervals (95%CI) were estimated for each study. Two tailed *P*- values were calculated using the χ^2 test of independence. Between-study heterogeneity was assessed with Cochrane's Q test-statistics. The pooled OR and 95% CI for the variant allele in RA patients against HC were calculated from fixed effects model, as no significant heterogeneity was observed. Forest plots were generated in StatsDirect to display the OR and 95%CI.

7.3.4.3 Power Calculations for meta-analysis

Calculations to determine the genetic power of this meta-analysis were done using software program Quanto v1.2.3 assuming a disease prevalence of 1% and G allele frequency of 0.67. Under log-additive model I calculated the effect of 0.05 with OR of 1.2 (weighted average OR of all combined studies).

7.4 Results

7.4.1 Case-control association study

All polymorphisms tested were in Hardy-Weinberg equilibrium in both RA cases and controls. The control allele frequencies of the four SNPs tested were consistent with previous published reports for the Caucasian populations. There was a high level of LD between the SNPs in the *IL-1B* promoter region (Figure 7.1). The genotyping results and

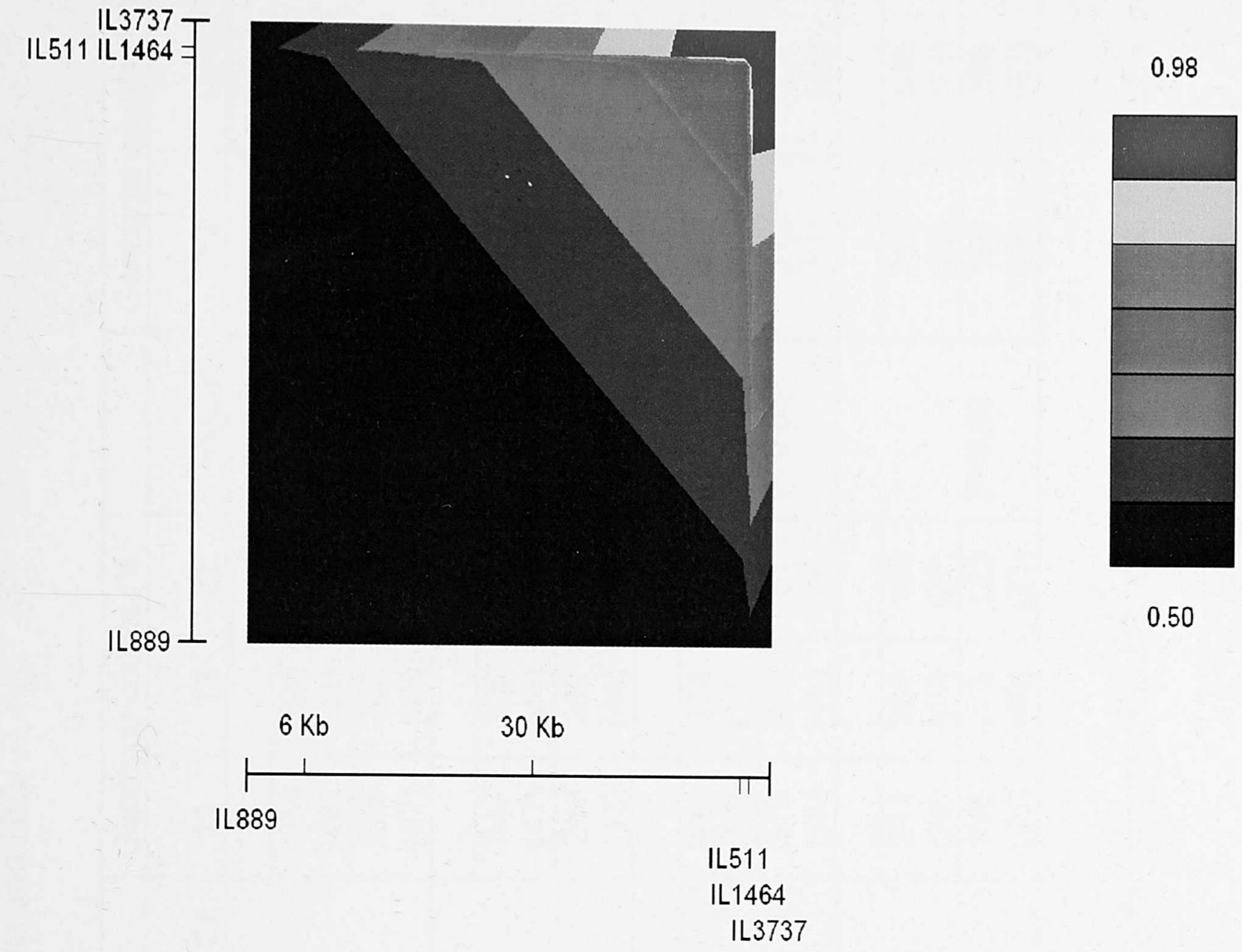
allele counts in RA patients and HC are presented in Table 7.1. *IL-1B* (-1464 C/G) G allele was less common among RA patients (p= 0.01, OR 1.24 95%CI 1.04-1.48). The minor allele frequencies for cases and controls were 24% and 29% respectively. The association remained significant after applying corrections for false discovery rate.

Haplotype analysis revealed a global difference in haplotype frequencies between the RA cases and controls (p=0.01). Further analysis revealed that this effect was entirely due to the *IL-1B* (-1464 C/G) and there was no single over-represented haplotype. Frequencies of the extended *IL1A-IL1B* four-marker haplotypes can be viewed in Table 7.2. Out of thirteen extended four-marker haplotypes generated, four of them represented 87% of the chromosomes in both RA patients and HC. The haplotype-specific results (by χ^2 test) for these four most common extended *IL1A-IL1B* haplotypes are shown in Appendix F, Table F.3.

Each of the polymorphisms was investigated in RA cohort according to RF and *HLA-DRB1* SE status and presence of high risk *HLA-DRB1**04 (*DRB1*0401, 0404, 0405* or *0408*) or a low risk allele (*DRB1*01, 0406, 0407* and *10*). No significant difference was observed between the groups. The data can bee seen in appendix F, Table F.2.

Further analysis according to various features of severity of RA did not reveal any independent association with the presence of nodules, requirement for major joint replacement or radiographic progression (Figure 7.2.) (see also Appendix F Table F.2).

Figure 7.1 Pairwise linkage disequilibrium of the *IL-1A-IL-1B* region according to the program GOLD. The four polymorphisms tested on 756 RA patients are represented along the X and Y axes. The SNPs have been assigned numbers reflecting their position along chromosome 2 from telomere to centromere.



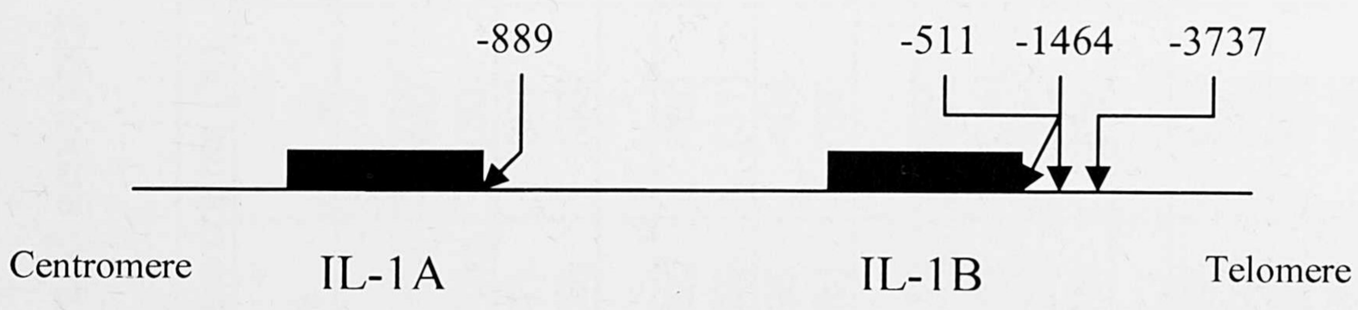


Table 7.1 Distribution of *IL-1A* (rs1800587), *IL-1B* (-511 A/G, rs16944), *IL-1B* (-1464 C/G, rs1143623) and *IL-1B* (-3737 G/A, rs4848306) alleles and genotypes among RA patients and HC.

	Allele counts (%)			Genotype counts (%)				Genotype counts (%)		
	1	**2**	**P value, OR 95%CI**	**11**	**12**	**22**	**P value**	**11**	**12+22**	**P value, OR 95%CI**
IL-1A										
RA	1031 (70)	447 (30)	P=0.93 OR=1.01 95%CI 0.85-1.19	355 (48)	321 (43)	63 (9)	P=0.91	355 (48)	384 (52)	P=0.80 OR=1.03 95%CI 0.84-1.28
HC	841 (70)	367 (30)		286 (47)	269 (45)	49 (8)		286 (47)	318 (53)	
IL-1B-511*										
RA	1022 (69)	460 (31)	P=0.14 OR=1.13 95%CI 0.96-1.33	347 (47)	328 (44)	66 (9)	P=0.20	347 (47)	394 (53)	P=0.39 OR=1.10 95%CI 0.88-1.37
HC	796 (66)	404 (34)		267 (45)	262 (44)	71 (11)		267 (45)	333 (55)	
IL-1B-1464										
RA	1116 (76)	362 (24)	P=0.01 OR=1.24 95%CI 1.04-1.48	421 (57)	274 (37)	44 (6)	P=0.05	421 (57)	318 (43)	P=0.03 OR=1.28 95CI 1.02-1.59
HC	871 (71)	352 (29)		312 (51)	250 (41)	51 (8)		312 (51)	301 (49)	
IL-1B-3737										
RA	778 (54)	676 (46)	P=0.18 OR=0.90 95%CI 0.77-1.05	211 (29)	356 (49)	160 (22)	P=0.26	211 (29)	516 (71)	P=0.10 OR=0.82 95%CI 0.65-1.05
HC	669 (56)	523 (44)		198 (33)	273 (46)	125 (21)		198 (33)	398 (67)	

* Alleles were named according to their frequency in the general population i.e. allele 1 being the frequent or more common allele, allele 2 the rare form

Figure 7.2 Scatterplots of Rau-Ratingen score (severity of erosions on radiographs) by duration of disease (interval between onset of RA and radiograph) in patients positive (one or two copies) (x) and negative (o) for the *IL-1A* (-889) rare allele [A], *IL-1B* (-511) rare allele [B], *IL-1B* (-1464) rare allele [C] and *IL-1B* (-3737) rare allele [D]. The lines represent nonparametric regression lines.

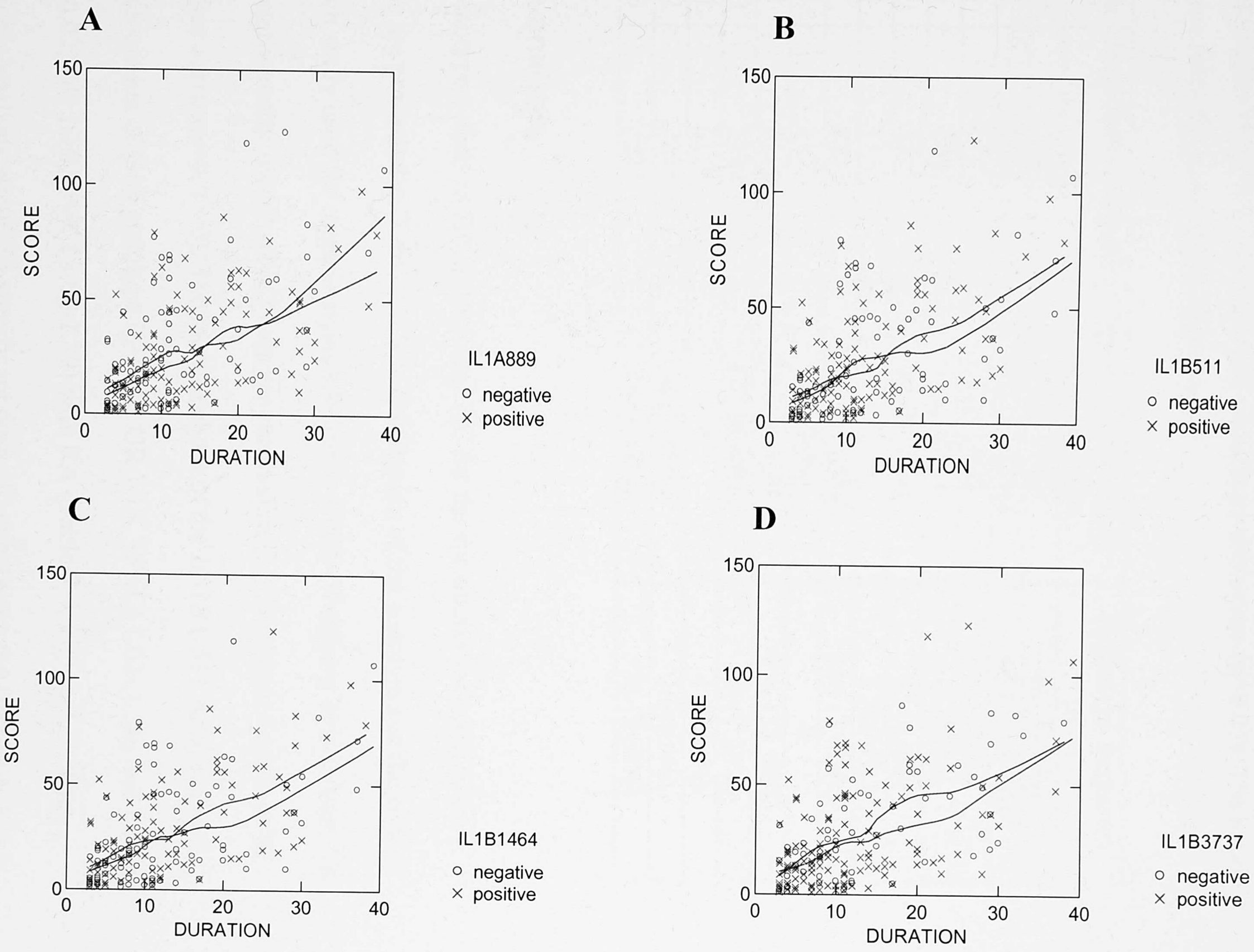


Table 7.2 Extended *IL-1A-IL1-B* haplotype counts in RA cases and controls. Haplotypes that all together represented 87% of the chromosomes are marked in bold. When I combined the rare haplotypes (No.4,6,8,11) and compared them with the rest of the haplotypes there was no statistically significant difference (p=0.07). The main effect came from the group of rare haplotypes.

No	Haplotype structure	Haplotype frequency in RA patients	Haplotype frequency in HC
1	1111	58 (3.84%)	46 (3.68%)
2	**1112**	572 (37.83%)	443 (35.44%)
3	1121	4 (0.26%)	5 (0.4%)
4	1122	1 (0.07%)	5 (0.4%)
5	1211	81 (5.36%)	55 (4.4%)
6	1212	0	3 (0.24%)
7	**1221**	341 (22.55%)	316 (25.28%)
8	1222	2 (0.13%)	0
9	**2111**	274 (18.12%)	226 (18.08%)
10	**2112**	127 (8.40%)	99 (7.92%)
11	2121	0	6 (0.48%)
12	2211	28 (1.85%)	20 (1.6%)
13	2221	24 (1.59%)	26 (2.08%)

7.4.2 Meta-analysis

Allele and genotype counts in RA patients and HC for the six studies included in the meta-analysis are presented in Table 7.3. The distribution of the genotypes in the control groups of each study used for meta-analysis adhered to Hardy-Weinberg equilibrium. No significant heterogeneity was observed between the studies (Moment-based estimate of between studies variance=0, I^2=0). The meta-analysis of the *IL-1B* (-511 A/G) showed significant association of G/G genotype (p= 0.02 OR 1.13, 95%CI 1.02- 1.26) and variant allele G (p= 0.02, OR 1.10 95%CI 1.02-1.19) with RA (Table 7.4, Figure 7.3). The genotype specific analysis showed a significant dose effect between the G/G *vs* A/A (p= 0.04, OR 1.20 95%CI 1.01-1.44), but not between G/G *vs* G/A and G/A vs A/A (p= 0.06 and p= 0.45, respectively).

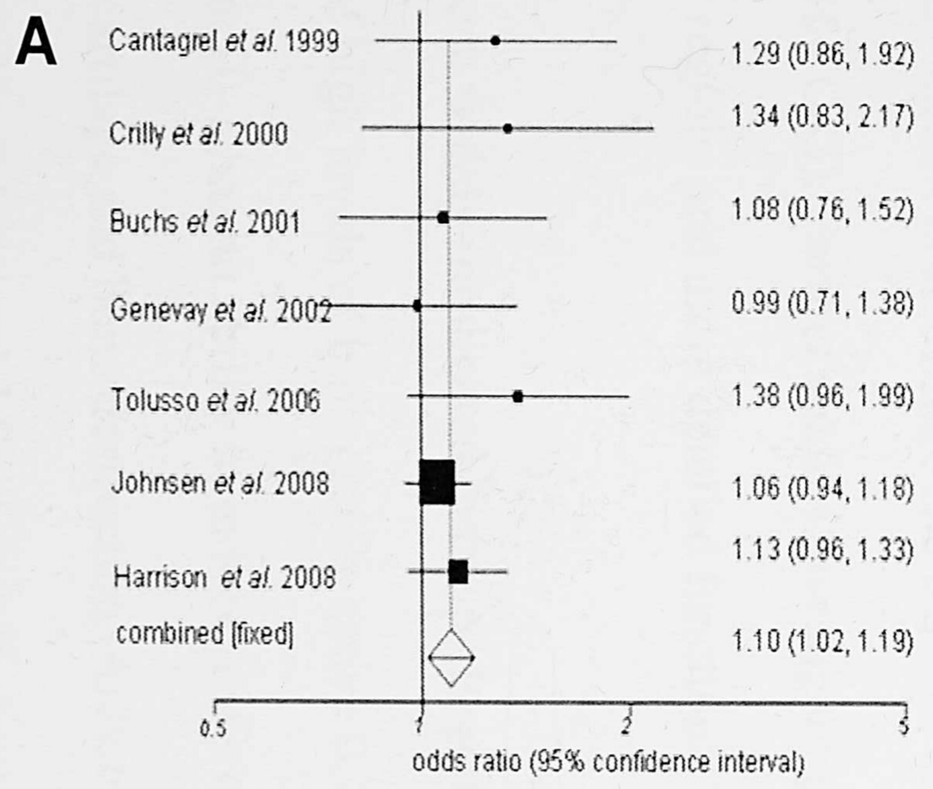
Table 7.3 Genotype and allele counts in RA patients and healthy controls for six studies included in a meta-analysis of the IL-1 (-511 A/G, rs16944) polymorphism.

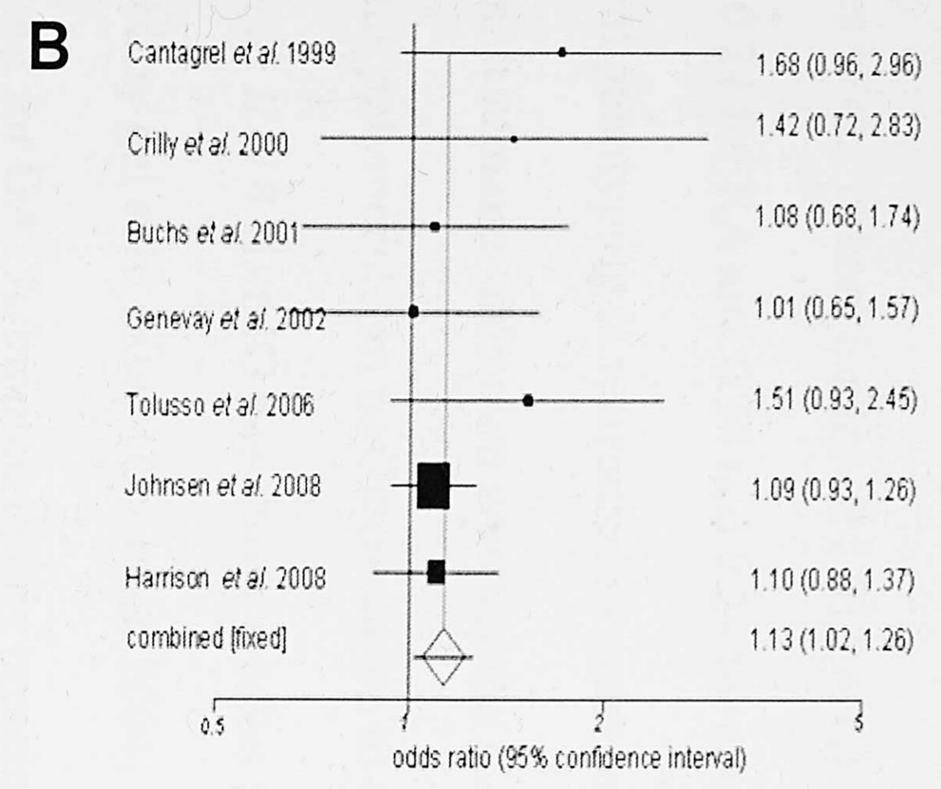
No	Reference	Country	Study population	Allele counts		P value	Genotype counts			P value	Power (%) (α = 0.05, OR=1.2)
				1	2		11	12	22		
1	Cantagrel,1999	France									
			RA n=108	139	73	P=0.19	50	39	17	P=0.1	15%
			HC n=128	148	100		43	62	19		
2	Crilly, 2000	Scotland									
			RA n=100	132	66	P=0.21	46	40	13	P=0.48	12%
			HC n=66	79	53		25	29	12		
3	Buchs, 2001	France									
			RA n=378	368	176	P=0.64	119	130	23	P=0.87	20%
			HC n=108	145	75		46	53	11		
4	Genevay, 2002	Switzerland									
			RA n=233	311	151	P=0.96	106	99	26	P=0.98	21%
			HC n=148	189	91		64	61	15		
5	Tolusso, 2006	Italy									
			RA n=126	181	71	P=0.07	66	49	11	P=0.19	18%
			HC n=178	231	125		75	81	22		
6	Johnsen, 2008	USA									
			RA=2026	2595	1319	P=0.33	868	859	230	P=0.55	89%
			HC=1433	1433	769		466	501	134		

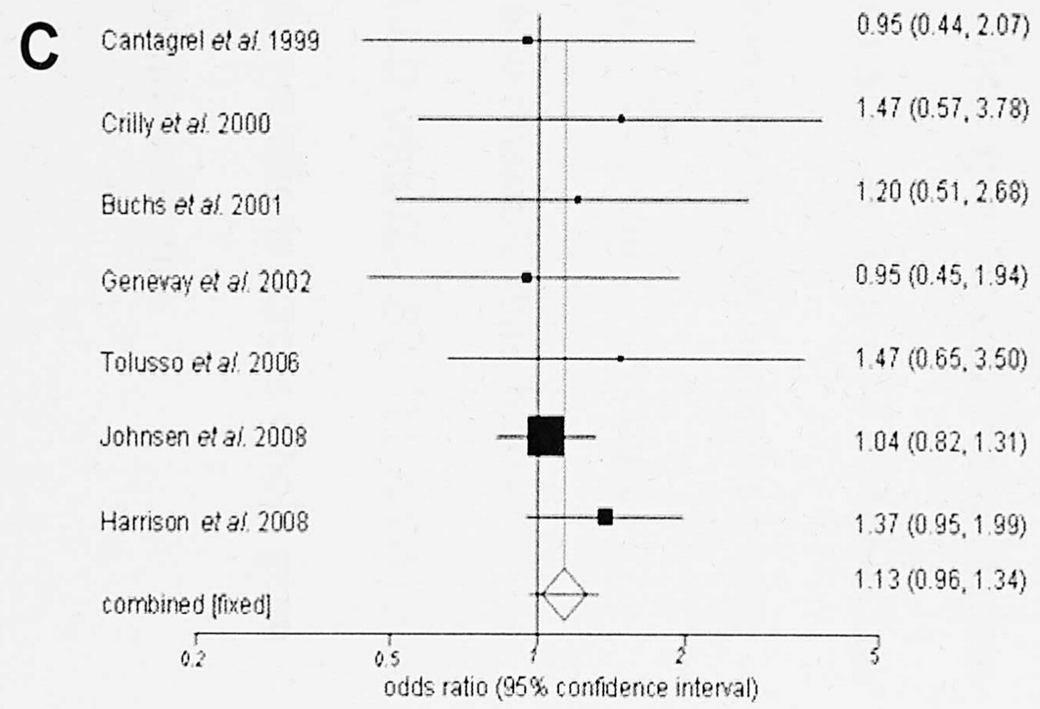
Table 7.4 Pooled OR and 95%CI for the G risk allele and the G risk genotype in RA patients and HC for six studies included in a meta-analysis of the *IL-1B* (-511) polymorphism.

G versus A	OR = 1.10 95%CI 1.01-1.19	P=0.02
G/G versus G/A + A/A	OR = 1.13 95%CI 1.02-1.26	P=0.034
G/G + G/A versus A/A	OR = 1.13 95%CI 0.96-1.33	P=0.15

Figure 7.3 OR and 95%CI from studies examining the association between (A) *IL-1B* (-511) G allele versus A allele, (B) *IL-1B* (-511) G/G versus G/A + A/A (recessive contrast), (C) *IL-1B* (-511) G/G + G/A versus AA (dominant contrast) (OR and 95%CI for each study is shown by a black square on a horizontal line and the summary OR and 95%CI shown by a dotted vertical line extending from a white diamond).







7.5 Discussion

In this study I have demonstrated a novel association of the *IL-1B* (-1464 C/G) SNP with RA in a Caucasian population. In our dataset *IL-1B* (-1464 C/G) is in near complete LD with *IL-1B* (-511 A/G) (D'>0. 90, r^2>0.80). However, the minor allele frequencies in our cohort for *IL-1B* (-511 A/G) and *IL-1B* (-1464 C/G) are dissimilar (0.31 and 0.24 for RA and 0.34 and 0.29 for controls, respectively), indicating that independent genotyping is warranted in order to investigate the *IL-1B* (-1464 C/G) association. It remains unknown whether the *IL-1B* (-1464 C/G) SNP is the "causal" variant. This polymorphism has been shown to exhibit allele specific difference in gene expression as *IL-1B* -1464G has shown weaker promoter activity than *IL-1B* -1464C on electromobility gel shift assay (Lee *et al.* 2004) This promoter region harbours a core binding motif for GATA-binding family proteins and *IL-1B* (-1464 C/G) could therefore affect protein binding (Lee *et al.* 2004). It is also possible that the real "protective" variant is elsewhere and exhibits different LD with *IL-1B* (-511 A/G) and *IL-1B* (-1464 C/G). Further studies investigating genetic variations across whole gene regulatory regions and more detailed functional analysis are essential.

IL-1 has been widely implicated in the pathogenesis of RA. This is based on the presence of high levels of biologically active IL-1 in RA synovial fluid and synovial cell culture. IL-1 has an ability to induce many of the pathological changes seen in RA, including cartilage and bone destruction. Although both IL-1α and IL-1β mRNA have been detected in RA synovial fluid, it is IL-1α that has been found to be strongly

expressed at the cartilage-pannus junction, especially by those cells invading cartilage (Deleuran *et al.* 1992).

Several studies have investigated the effect of *IL-1* polymorphisms on severity of RA with variable results. In a study of affected sib pair RA families *Cox et al* found four *IL-1* gene cluster markers associated with erosive RA, particularly in sib pair families with 0 probability of sharing two *HLA-DRB1* alleles identical by descent (Cox *et al.* 1999). Evidence of less severe disease progression after 20 years in Swiss RA patients carrying one or two copies of the *IL-1B* (-511) minor allele has also been reported (Genevay *et al.* 2002). Using several different analysis strategies I was unable to demonstrate any significant relationship between *IL-1* polymorphisms and radiographic progression (Figure 3). Similar conclusions were reached in the recent study of *Johnsen et al* where no association was observed with erosive disease in a large sample of the Brigham Rheumatoid Arthritis Sequential Study cohort.

Requirement for joint surgery before 15 years of disease duration has been examined before and a much higher frequency of *IL-1* (-511) rare allele was found in patients who required surgery (Crilly *et al*. 2000). I had joint replacement data for 735 patients, ~20% of them requiring joint replacement before 15 years of disease duration. The allele frequencies and genotyping data for all the polymorphisms studied did not differ significantly between the groups. For the *IL-1B* (-511) polymorphism our patients who required joint replacement had a much lower frequency of the rare allele compared to

the Crilly *et al* group (34% versus 40% respectively) which would account for differences in final results.

Although replication studies should ideally be carried out on large, ethnically homogenous cohorts, this is not always possible. By conducting a meta-analysis one can maximize the sample size and provide more robust findings. This meta-analysis presented here is the largest to have investigated the effects of *IL-1* (-511) in RA. I demonstrated a significant association of GG genotype and G allele with RA. The power of this meta-analysis was over 90% to detect modest effect (OR 1.2 at a significance level of 0.05). However, the magnitude of the genetic risk associated with this polymorphism is relatively small and further studies on a much larger RA sample are desirable to confirm the true association.

Although *IL-1B* (-511) is near a binding site for a glucocorticoid response element, so far no specific change in gene regulation associated with this marker has been described. *IL-1* (-511) is essentially in complete linkage disequilibrium with another promoter base change in *IL-1* (-31) that changes the TATA box and is associated with differential binding of transcription factors (Guasch *et al.* 1996). *IL-1* (-31) has also been found to enhance susceptibility to Helicobacter pylori infection and gastric cancer (El-Omar EM *et al.* 2000). It is thought that the influence of *IL-1* (-511) on transcription may be due to subtle changes in the vicinity of the marker that may affect protein binding to regulatory elements or be

due to indirect interactions with *trans*-activating proteins that bind to regulatory elements in this region (Chen *et al.* 2006).

In summary, the role of the genetic factors in RA is complex and the magnitude of the effects is known to vary greatly. I demonstrated an association of *IL-1B* (-511) with RA patients of Caucasian origin by conducting a meta-analysis. The protective influence of *IL-1B* (-1464) found in this study will require further conformation in other populations.

Chapter 8

General discussion and future work

This D.Phil project focused on the identification of genetic susceptibility and severity factors to rheumatoid arthritis using a case-control approach. The selection of the investigated candidate genes was made largely on the basis of the functional evidence. The strengths and weaknesses of the conducted work involving each candidate gene have been discussed in the context of previously published works within each of the relevant chapters. In this final chapter, the wider implications of these findings with regard to future genetic studies in RA are discussed. Special emphasis has been placed on the role of GWAS. Without doubt, GWAS have major impact on widening current knowledge of complex autoimmune disease. Furthermore GWAS form a base for future functional studies and hopefully help to achieve the ultimate goal which involves taking the knowledge "from bench to bedside". Genetic factors can have a wide impact on clinical care and this is discussed in the final part of this chapter.

8.1 Genome wide association studies past, present and future in rheumatoid arthritis genetic research

8.1.1 Common disease/common variant hypothesis

Investigation of genetic determinants and their regulatory mechanisms that underlie phenotypic diversity is one of the central challenges in today's biomedical research.

Although genetic impact of the MHC on RA susceptibility and severity has been known for several decades, over the last five years great progress has also been made in identifying and robustly replicating RA susceptibility genes outside *HLA*. The discovery of the *PTPN22* C1858T polymorphism initially in T1D, followed by RA, SLE and autoimmune thyroid disease in Caucasian populations is one of the most remarkable success stories and strengthens the idea that common pathogenetic mechanisms underlie different autoimmune diseases.

The exciting work concerning the genetic aspects of RA has continued and the year 2007 saw the publication of the results of three GWAS conducted in British, Northern American and Swedish Caucasian RA populations. Although the timing reflects recent technological improvements in high throughput genotyping together with the knowledge drawn from the International HapMap Project and the collection of large patient samples, the idea behind these studies can be traced back more than a decade ago to two key publications (Risch *et al.* 1996; Lander *et al.* 1996). The authors in these papers proposed that common variants may underlie many common diseases and that these would be more easily found using population-based association studies rather than family based linkage analysis even if this required testing every gene in the genome.

The common disease/common variant hypothesis assumes that much of the genetic variation of common complex diseases is due to relatively few common variants. If multiple rare genetic variants were the primary cause, association studies would have very little power to detect them. There are studies that support the common

disease/common variant hypothesis, the first and earliest evidence was drawn from research into *APOE* gene which demonstrated increased risk for Alzheimer's and heart disease (Corder *et al.* 1993). Subsequently many other genes that confirm increased risk for multiple diseases, including multiple autoimmune diseases (*PTPN22* and *STAT4*) have been convincingly replicated in Caucasian RA and SLE populations. However, there are also been examples of rare variants influencing common diseases (Cohen *et al.* 2004; Romeo *et al.* 2007). This would indicate that both rare and common variants may influence common diseases, although it is remains unknown which is more important. The great success of GWAS at uncovering many common alleles associated with common complex diseases suggests that the hypothesis is true to an extent. Furthermore, by providing useful information about the underlying genetic basis of human disease, GWAS will help to design future studies to identify those genetic variations that have a direct impact.

In essence, GWA represents a method for capturing a new class of disease-associated genetic variants and offers the advantage of susceptibility locus identification as no prior assumption regarding gene function is needed. GWA relies on population-based samples and so detects common variants of more modest effects that are unlikely to be found using traditional linkage based approaches. Therefore, association and linkage study approaches have different benefits and shortcomings. However the ideal method would be a synthesis of both approaches that allows the association analysis of both rare and common variants at high throughput and at a practical cost.

8.1.2 What has been achieved with genome wide association studies in searching rheumatoid arthritis genetic susceptibility factors

Although the two strongest associations that emerged from the initial UK WTCCC were known beforehand (*HLA-DRB1* and *PTPN22*), the most important outcome of GWA study is the discovery of new biological association with genes and/or regions previously unrecognized to have role in RA. Some of these novel markers are located in genes that encode proteins with previously known function. For example, two polymorphisms (rs2104286 and rs743777) in the gene encoding the IL2 receptor alpha and beta chains, respectively are associated with RA. The IL2 receptor mediates IL2 stimulation of T-lymphocytes and is thought to have an important role in preventing autoimmunity. A rare 4-base-pair deletion of *IL2RA* has been associated with development of severe autoimmune disease (Sharfe *et al.* 1997). However, several associated regions either contain genes of unknown function (*ANAPC4* rs3816587, chromosome 4p15; CRYL1 rs9550642, chromosome13q12) or do not contain annotated genes (rs11162922 chromosome 1p31 and rs2837960, chromosome 21q22) or map to an intragenic region (*OLIG3/TNFAIP3* rs6920220, chromosome 6q23).

Genetic effects common to T1D and RA have long been suspected, given the increased family incidence of these conditions together. Indeed the excess of RA, T1D, SLE and autoimmune thyroid disease in families suggests common immunopathology in several autoimmune diseases. The best known and the most replicated of these in the Caucasian populations, is the association with *PTPN22*. As a result of WTCCC new interactions

were discovered between RA and T1D locating in *PTPN2* (chromosome 18p11), *IL2RA* (chromosome 10p15) and the chromosome 12q24 region.

Although these recent GWA findings in RA have given the field considerable new momentum several important challenges remain. Even though the sample size tested in GWA studies is generally in the thousands, it might be expected that such studies are well powered. Several reports have demonstrated that, given the strict genome-wide significance criteria ($p < 5 \times 10^{-7}$) as in WTCCC, power is much less that might be imagined (Wang *et al.* 2005; Jorgenson E *et al.* 2006). Many of the findings from GWA studies that have not quite reached the genome-wide significance may be genuine and could be uncovered by combining the results from several studies, including meta-analysis of published GWA. Replication studies on a larger independent sample are therefore essential.

In the UK, six different centres (Manchester, Aberdeen, Leeds, Sheffield, London and Oxford) have joined forces to create validation samples, comprising 5,063 RA cases and 3,849 independent healthy volunteers, have been assembled. Five previously identified genetic loci have been successfully replicated in this validation cohort: *HLA-DRB1*, *PTPN22*, *OLIG3/TNFAIP3*, *STAT4* and *TRAF1/ C5*. Further validation studies of nine polymorphism with nominal significance of $p = 10^{-5}$-10^{-7} and 49 SNPs with nominal significance of $p = 10^{-4}$-10^{-5} have been carried out incorporating an even bigger RA cohort (combining WTCCC and validation cohorts). Recently published results report successful replication of a further six loci that were initially identified from WTCCC GWAS (Tier

2: rs10910099, chromosome 1p36 LD with rs6684865 *MMEL1*; rs2104286, chromosome 10p15 *IL2RA*; rs743777 chromosome 22q13 *IL2RB*; Tier 3: rs4750316 chromosome 10p15; rs3218253 chromosome 12q13 KIF5A; rs3218253 chromosome 22q13 IL2RB) (Barton *et al.* 2008). This brings the total number of replicated loci to 11.

The results from the GWA studies so far indicate that individual polymorphisms have very modest effects in the population: except for *HLA-DRB1,* associated SNPs rarely show odd ratios >2.0 and more typically fall into < 1.5 range. This would mean that the remaining undiscovered common variants are most likely to have similar or smaller effects.

8.1.3 What the future will hold for rheumatoid arthritis genetic research?

Despite substantial progress, many of the genes contributing to the heritability of RA remain unidentified. To some extent, the magnitude of the associations discovered is currently underestimated because the full spectrum of causal variants at each locus has yet to be defined. It is now necessary to obtain a full picture of genotype-phenotype correlation at these loci and to find the remaining loci. Rarely will the polymorphism used to discover each locus prove causal. Imputation methods allow association analysis to be extended to a large set of variants (HapMap SNPs for which reliable imputation is possible), but this still comprises a minority of all common SNPs and representation of rare variants is poorer still (Zeggini *et al.* 2005). Further targeted re-sequencing of each region of interest is likely to be necessary to discover all causal variants and fully define genotype-phenotype correlations. It is very likely that multiple independent common

variants and also rare variants will be found at the same locus. In addition, sequencing of exons in each associated region may identify coding polymorphisms of stronger effect, which may be easier to study functionally and in individual subjects.

It is well recognized that some susceptibility loci may be missed by well-powered GWA studies, simply because none of the causal variants are in LD with SNPs on the genotyping arrays. Some of these may be found by genome-wide measurement of copy number variation, a project currently initiated by the Welcome Trust.

The studies have found a substantial fraction of associations outside transcription units. This is not surprising, as coding sequences make up less than half of the evolutionary conserved DNA in the human genome. Investigations of these functional non-coding associations are very important in order to unravel the molecular and cellular roles of non-coding DNA. Recent work has highlighted the prevalence of heritable *cis* or *trans* acting gene expression changes and their likely impact in disease susceptibility (Morley *et al.* 2004; Vawter *et al.* 2006). In a microarray study of 8,500 genes in immortalized human B cells, 3,554 showed significant inter-individual differences in expression (Morley *et al.* 2004). Analysis of genotypes of 2756 SNP markers was then combined with linkage analysis and the results revealed that 984 of the differentially expressed genes showed significant evidence of linkage to at least one region of the genome, indicating a genetic contribution to gene expression.

Chapter 8 - General Discussion

Once all the statistical evidence has been extracted and putative causal variants identified, substantial challenges will still remain. One of the most prominent of these challenges is the need to obtain functional confirmation that the variants implicated are truly causal and to reconstruct the molecular and physiological mechanisms whereby they have an impact on the particular phenotype. Because some of the susceptibility variants so far identified in RA (*TRAF1/C5* and *OLIG3/TNFAIP3*) map to sequence of unknown function that is some distance from the nearest coding sequence, the design of appropriate functional assays is not straightforward. Work towards developing new tools, frameworks and analytical methods for the functional annotation of the genome has already started. The Encyclopaedia of DNA Elements (ENCODE) Project aims to provide a more biologically informative representation of the human genome by using high-throughput methods to identify, and catalogue the functional elements it encodes (ENCODE Project consortium 2007). In its pilot phase, roughly 30 Mb, ~1% of the human genome was subjected to pilot testing of multiple experimental and computational methods. The results have greatly advanced our knowledge of transcription of the human genome and yielded new understanding of transcription start sites and their relationship to specific regulatory sequences and features of chromatin accessibility and histone modification. A more sophisticated view has emerged about chromatin structure and its relationship with DNA replication and transcriptional regulation. The comparative analysis of inter- and intraspecies sequences has lead to the discovery of novel mechanisms and enriched the evolutionary insights about the functional landscape of the human genome. Although more work will be required to enhance existing assays, invent new techniques, and develop new data-analysis methods, the generation of genome-wide experimental

datasets, the results from the ENCODE pilot phase have provided an impressive platform for future genome exploration efforts.

8.2 Clinical importance of the genetic studies

Ultimately, the value of all the genetic research lies in the translation of the findings into advances in clinical patient care. The mechanistic insights generated by gene discovery might identify new therapeutic targets and lead to novel pharmaceutical and preventative approaches.

In classic Mendelian disease, the predictable nature of disease inheritance and its comparatively straightforward relationship with the genotype at disease susceptibility loci has led to the development of genetic testing for diagnostic and prognostic purposes, and to informed lifestyle and reproductive choices. In contrast, genetic testing in complex disease like RA is not straightforward and currently not applied in everyday clinical practice. Even the well-established association of *HLA-DRB1* alleles with RA, has not resulted in routine testing being offered to help with diagnosis, prediction of the outcome or in the guidance of therapy. This is explained by the relatively low penetrance of high risk *HLA-DRB1* among Caucasian population (the frequency of *HLA-DRB1*04* is estimated to be around 30%). However, RA affects less than 1% of population.

Furthermore, a significant proportion of RA patients do not test positive for *HLA-DRB1* high-risk alleles. Therefore on their own, high risk *HLA-DRB1* alleles are not to be sufficient to lead to the development of RA. However most recent data combining the

presence of anti-CCP antibodies, *HLA-DRB1* SE alleles and testing for *PTPN22* C1858T polymorphism , showed high sensitivity and specificity for RA (Orozco *et al.* 2008). Further longitudinal studies are necessary to assess and confirm the utility of combinations of these laboratory and genetic markers in predicting RA susceptibility.

Another important issue lies in the genetic variation and its impact on drug pharmacocinetics, pharmacodynamics and development of side effects. Despite recent developments, very few tests are universally accepted and used in everyday practice. Pharmacogenetic testing of the thiopurine metabolizing enzyme, thiopurine methyl transferase (TPMT), prior to prescription of azathioprine is regularly used in clinical context. Although helping to avoid some adverse drug effects, testing for TPMT polymorphisms has limited value, as the majority of adverse thiopurine reactions have not been found to be associated with TPMT polymorphism (van Aken *et al.* 2004). The variants of cytochrome P450 enzyme CYP2D6 are known to be important in the metabolism of a large number of medicines, but these are not tested routinely in clinical practice (Wolf *et al.* 1999). The adverse reactions are generally not serious and more than 70 known alleles make testing and interpretation very complex.

Low-dose methotrexate (MTX) treatment is used extensively as a disease-modifying anti-arthritic drug (DMARD) in RA although its precise mechanisms of action are not fully understood. It is known that in low-doses MTX enters the cells through the reduced folate carrier (RFC1, also known as SLC19A1) and directly inhibits enzymes of the folate metabolic pathway (Antony 1992; Brzezinska *et al.* 2000). Although used widely to treat RA its long-term use can be limited by toxicity. It has been estimated that about 10-30%

of patients with RA discontinue MTX because of side-effects (Alarcon *et al.* 1989). Most of the toxicity associated with MTX is thought to be linked to effects on folate metabolism. The evidence that polymorphisms in genes encoding the transport and enzymatic pathways necessary for MTX metabolism could be associated with increased efficacy or toxicity is variable and conflicting. For example, in a recent study conducted on 213 Slovenian RA patients examining the polymorphism in *RFC1* A80G, the GG genotype appeared to have the strongest impact on the overall MTX toxicity and on the risk of developing infections. In contrast, it did not influence the efficacy of MTX (Bohanec Grabar *et al.* 2008). The authors suggested that patients with *RFC1* 80GG genotype may develop MTX induced adverse events due to lower cellular uptake of the drug, as higher concentrations of polyglutamated MTX in erythrocytes have been observed in patients with *RFC1* AA genotype treated with low-dose MTX (Dervieux *et al.* 2004a,b). In the studies by Dervieux *et al. RFC1* the 80AA genotype was reported to be associated with more effective MTX therapy. However, contrary to this, in a separate study, a higher risk for elevated liver enzyme level and a higher probability of remission of RA was noted in patients with *RFC1* 80AA genotype (Drozdik *et al.* 2007). Furthermore, in three independent studies conducted on Japanese and Caucasian RA patients no association between the *RFC1* A80G polymorphism and MTX efficacy or toxicity was observed (Wessels *et al.* 2006; Takatori *et al.* 2006; Dervieux *et al.* 2006). Polymorphisms in other enzymes involved in folate pathway and their impact on MTX efficacy and toxicity have been widely studied, for example: 5,10-methylenetetrahydrofolate reductase (*MTHFR* C677T, A1298C) (van Ede *et al.* 2001; Taniguchi *et al.* 2007; Kuzawski *et al.* 2007), thymidylate synthase (TS - double and

triple tandem repeat in the promoter region) (Dervieux *et al.* 2004a; Weisman *et al.* 2006; Bohanec Grabar *et al.* 2008), methionine synthase (*MS* A2756G) and methionine synthase reductase (*MTRR* A66G) (Berkun Y *et al.* 2007; Dervieux et al. 2006; Bohanec Grabar *et al.* 2008). Similarly to the studies of *RFC1* studies, no consistent trends were seen for increased efficacy or toxicity associated with polymorphisms in these genes have been reported.

Over recent years RA treatment has been revolutionised by the development and introduction of TNF antagonists. Therapy with TNF blockers is expensive and bears substantial risks. Predictors of treatment response would therefore be useful for patient selection. However, up to present there are very few published reports that have investigated pharmacogenetic effects in association with anti- TNF and treatment response. The most promising data so far have been derived from studies into TNF promoter polymorphism at position -308G/A. An initial study conducted on French RA patients showed that patients with the -308 G/G genotype were responded better to infliximab (Mugnier *et al.* 2003). A subsequent study on Swiss RA, PsA and AS patients confirmed the result and extended the link between the polymorphism and clinical response in addition to infliximab, as well as to etanacept and adalimumab (Seitz *et al.* 2007). The presence of -308G allele being associated with better response to anti-TNF treatment was also supported by a meta-analysis of 6 studies, conducted on 311 RA patients (Lee *et al.* 2006).

A genome wide association study investigating different response to anti-TNF treatment in RA patients was also published very recently (Liu *et al.* 2008). The authors identified 16 genomic regions associated with response to anti-TNF. Only one of them localising in IL-10 promoter (rs1800896) has been previously associated with good response to etanercept (Padyukov *et al.* 2003). Unfortunately this study was conducted on a small heterogeneous RA patient population and the results are not therefore conclusive.

Although genetic studies have shed some light on RA treatment efficacy and toxicity there is still a long way to go before genetic testing in everyday clinical practice. The major drawback lies in the inconsistency of the reported results. This is due to various technical and methodological reasons: overall, the studies published so far are mostly retrospective, conducted on a small sample of RA patients and focused mainly on limited genes involved in DMARDs metabolic pathways. In addition it is difficult to measure clinical response. DAS28 is widely used as an indicator of RA disease activity and response to treatment. However it is a composite score consisting of clinical, laboratory and patients own views of the disease activity. A rather complex calculation of DAS28 also conceals relative impact of each component enabling to create situations where the score does not truly reflect current clinical state. Therefore further large-scale prospective studies with better clinical assessment methods are needed to confirm the previous results and investigate epigenetic interactions.

8.3 Conclusion

Recent developments have increased the knowledge of the genetic basis of RA with eleven susceptibility loci confirmed in the Northern Caucasian population. Extensive resequencing and fine-mapping work is currently being undertaken to characterise these novel loci. Many of the greatest challenges to be faced in the years ahead lie not so much in identification of the association signals themselves, but in defining the molecular mechanisms through which they influence disease risk and phenotype expression. In addition, further studies are needed to assess the epigenetic effects of these newly defined susceptibility markers as well as gene-environmental interactions. Better understanding of the genetic basis of RA may facilitate identification of genetic markers that are suitable pharmacological targets and hopefully, not in the very distant future, be essential in a clinical setting.

References

Abecasis GR, Cookson WO. GOLD--graphical overview of linkage disequilibrium. Bioinformatics. 2000;16:182-3.

Accolla RS. Human B cell variants immunoselected against a single Ia antigen subset have lost expression of several Ia antigen subsets. J Exp Med. 1983;157:1053-8.

Accolla RS, Jotterand-Bellomo M, Scarpellino L, Maffei A, Carra G, Guardiola J. aIr-1, a newly found locus on mouse chromosome 16 encoding a trans-acting activator factor for MHC class II gene expression. J Exp Med. 1986;164:369-74.

Agrawal S, Aggarwal A, Dabadghao S, Naik S, Misra R. Compound heterozygosity of HLA-DR4 and DR1 antigens in Asian Indians increases the risk of extra-articular features in rheumatoid arthritis. Br J Rheumatol. 1995;34:41-4.

Aho K, Koskenvuo M, Tuominen J, Kaprio J. Occurrence of rheumatoid arthritis in a nationwide series of twins. J Rheumatol. 1986;13:899-902.

Akkad DA, Jagiello P, Szyld P, Goedde R, Wieczorek S, Gross WL, Epplen JT. Promoter polymorphism rs3087456 in the MHC class II transactivator gene is not associated with susceptibility for selected autoimmune diseases in German patient groups. Int J Immunogenet. 2006;33:59-61.

Alarcón GS, Tracy IC, Blackburn WD Jr. Methotrexate in rheumatoid arthritis. Toxic effects as the major factor in limiting long-term treatment. Arthritis Rheum. 1989;32:671-6.

Alonso A, Sasin J, Bottini N, Friedberg I, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T. Protein tyrosine phosphatases in the human genome. Cell. 2004;117:699-711.

Anaya JM, Diethelm L, Ortiz LA, Gutierrez M, Citera G, Welsh RA, Espinoza LR.Pulmonary involvement in rheumatoid arthritis. Semin Arthritis Rheum. 1995;24:242-54.

Antony AC. The biological chemistry of folate receptors. Blood. 1992;79:2807-20.

Ardlie K, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. Nat Rev Genet. 2002; 3:299-309.

Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, *et al.* The American Rheumatism Association 1987

revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988 Mar;31:315-24.

Barrera P, Balsa A, Alves H, Westhovens R, Maenaut K, Cornelis F, Fritz P, Bardin T, de Almeida G, Lopes-Vaz A, Pascual Salcedo D, de la Concha EG, Radstake TR, van de Putte LB, Migliorini P, Prud'homme JF, Charron D, Spyropoulou M, Mendes A, Spaepen M, Martinez M, Lepage V, Stravopoulos C. Noninherited maternal antigens do not play a role in rheumatoid arthritis susceptibility in Europe. European Consortium on Rheumatoid Arthritis Families. Arthritis Rheum. 2000;43(4):758-64.

Barrera P, Fauré S, Prud'homme JF, Balsa A, Migliorini P, Chimenti D, Radstake TR, van de Putte LB, Pascual-Salcedo D, Westhovens R, Maenaut K, Alves H, Lopes-Vaz A, Stravopoulos C, Spyropoulou M, Fritz P, Bardin T, Charron D, Lepage V, Alibert, Martinez M, Cornélis F; European Consortium on Rheumatoid Arthritis Families (ECRAF). European genetic study on rheumatoid arthritis: is there a linkage of the interleukin-1 (IL-1), IL-10 or IL-4 genes to RA? Clin Exp Rheumatol. 2001 Nov-Dec;19:709-14.

Barton A, Jury F, Eyre S, Bowes J, Hinks A, Ward D, Worthington J. Haplotype analysis in simplex families and novel analytic approaches in a case-control cohort reveal no evidence of association of the CTLA-4 gene with rheumatoid arthritis. Arthritis Rheum. 2004;50:748-52.

Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Gibbons L, Plant D: WTCCC, Wilson AG, Marinou I, Morgan A, Emery P: YEAR Consortium, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Worthington J. 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 Gen. 2008 Aug;17:2274-9

Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Plant D, Gibbons LJ; Wellcome Trust Case Control Consortium; YEAR Consortium; BIRAC Consortium, Wilson AG, Bax DE, Morgan AW, Emery P, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Worthington J. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet. 2008 Oct;40:1156-9.

Barton NH, Keightley PD. Understanding quantitative genetic variation. Nat Rev Genet. 2002 ;3:11-21.

Bas S, Perneger TV, Mikhnevitch E, Seitz M, Tiercy JM, Roux-Lombard P, Guerne PA. Association of rheumatoid factors and anti-filaggrin antibodies with severity of erosions in rheumatoid arthritis. Rheumatology (Oxford).2000;39:1082-8.

Bateson W. Mendel's principles of heredity. Cambridge: Cambridge University Press, 1909.

References

Begovich AB, Caillier SJ, Alexander HC, Penko JM, Hauser SL, Barcellos LF, Oksenberg JR. The R620W polymorphism of the protein tyrosine phosphatase PTPN22 is not associated with multiple sclerosis. Am J Hum Genet. 2005;76:184-7.

Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batliwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ, Gregersen PK. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet. 2004 Aug;75:330-7.

Bellman R. Adaptive Control Processes. Pincerton, Pincerton University Press. 1961.

Berkun Y, Abou Atta I, Rubinow A, Orbach H, Levartovsky D, Aamar S, Arbel O, Dresner-Pollak R, Friedman G, Ben-Yehuda A. 2756GG genotype of methionine synthase reductase gene is more prevalent in rheumatoid arthritis patients treated with methotrexate and is associated with methotrexate-induced nodulosis. J Rheumatol. 2007;34:1664-9.

Billingham MEJ. Adjuvant arthritis: the first model. In Henderson B, Edwards JCW, Pettipher ER eds. Mechanisms and models in rheumatoid arthritis. London: Academic Press. 1995:389-410.

Bohanec Grabar P, Logar D, Lestan B, Dolžan V. Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism. Eur J Clin Pharmacol. 2008 Jul 8. [Epub ahead of print]

Boone DL, Turer EE, Lee EG, Ahmad RC, Wheeler MT, Tsui C, Hurley P, Chien M, Chai S, Hitotsumatsu O, McNally E, Pickart C, Ma A. The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. Nat Immunol. 2004 Oct;5:1052-60.

Boss JM, Jensen PE. Transcriptional regulation of the MHC class II antigen presentation pathway. Curr Opin Immunol. 2003;15:105-11.

Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004 Apr;36:337-8.

Bray NJ, Buckland PR, Owen, MJ, O'Donovan, M.C. Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet. 2003;113:149-153.

References

Bray NJ, Jehu L, Moskvina V, Buxbaum JD, Dracheva S, Haroutunian V, Williams J, Buckland PR, Owen MJ, O'Donovan MC. Allelic expression of APOE in human brain: effects of epsilon status and promoter haplotypes.Hum Mol Genet. 2004; 3:2885-2892.

Bromley M, Woolley DE. Histopathology of the rheumatoid lesion. Identification of cell types at sites of cartilage erosion. Arthritis Rheum. 1984;27:857-63.

Bronson PG, Criswell LA, Barcellos LF. The MHC2TA -168A/G polymorphism and risk for rheumatoid arthritis: a meta-analysis of 6861 patients and 9270 controls reveals no evidence for association. Ann Rheum Dis. 2008;67:933-6.

Brown MA, Wordsworth BP, and J. D. Reveille. Genetics of ankylosing spondylitis. Clin Exp Rheumatol. 2002;20:S43–9.

Brzezińska A, Wińska P, Balińska M. Cellular aspects of folate and antifolate membrane transport. Acta Biochim Pol. 2000;47:735-49.

Bucala R, Ritchlin C, Winchester R, Cerami A. (1991). Constitutive production of inflammatory and mitogenic cytokines by rheumatoid synovial fibroblasts. J Exp Med. 1991;173:569-74.

Buchs N, di Giovine FS, Silvestri T, Vannier E, Duff GW, Miossec P. IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: interaction with their plasma levels. Genes Immun. 2001;2:222-8.

Bukhari M, Lunt M, Harrison BJ, Scott DG, Symmons DP, Silman AJ. Rheumatoid factor is the major predictor of increasing severity of radiographic erosions in rheumatoid arthritis: results from the Norfolk Arthritis Register Study, a large inception cohort. Arthritis Rheum. 2002 Apr;46(4):906-12.

Byng MC, Whittaker JC, Cuthbert AP, Mathew CG, Lewis CM. SNP subset selection for genetic association studies. Ann Hum Genet. 2003 Nov;67:543-56.

Campion G, Maddison PJ, Goulding N, James I, Ahern MJ, Watt I, Sansom D. The Felty syndrome: a case-matched study of clinical manifestations and outcome, serologic features, and immunogenetic associations. Medicine (Baltimore). 1990 Mar;69:69-80.

Cantagrel A, Navaux F, Loubet-Lescoulié P, Nourhashemi F, Enault G, Abbal M, Constantin A, Laroche M, Mazières B. Interleukin-1beta, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis. Arthritis Rheum. 1999 Jun;42:1093-100.

Cantón I, Akhtar S, Gavalas NG, Gawkrodger DJ, Blomhoff A, Watson PF, Weetman AP, Kemp EH. A single-nucleotide polymorphism in the gene encoding lymphoid protein tyrosine phosphatase (PTPN22) confers susceptibility to generalised vitiligo. Genes Immun. 2005;6:584-7.

References

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004 Jan;74:106-20.

Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, Chang M, Catanese JJ, Leong DU, Ardlie KG, Kastner DL, Seldin MF, Criswell LA, Gregersen PK, Beasley E, Thomson G, Amos CI, Begovich AB. PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. Am J Hum Genet. 2005 Oct;77:567-81.

Carthy D, Ollier W, Papasteriades C, Pappas H, Thomson W. A shared HLA-DRB1 sequence confers RA susceptibility in Greeks. Eur J Immunogenet. 1993;20:391-8.

Chabaud M, Durand JM, Buchs N, Fossiez F, Page G, Frappart L, Miossec P. Human interleukin-17: a T cell-derived proinflammatory cytokine produced by the rheumatoid synovium. Arthritis Rheum. 1999 May;42:963-70.

Chabaud M, Lubberts E, Joosten L, van Den BW Miossec P. IL-17 derived from juxta-articular bone and synovium contributes to joint degradation in rheumatoid arthritis. Arthritis Res. 2001;3:168-77.

Chan SH, Lin YN, Wee GB, Koh WH, Boey ML. HLA class 2 genes in Singaporean Chinese rheumatoid arthritis. Br J Rheumatol. 1994;33(8):713-7.

Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum Hered. 2003;56:18-31.

Chen H, Wilkins LM, Aziz N, Cannings C, Wyllie DH, Bingle C, Rogus J, Beck JD, Offenbacher S, Cork MJ, Rafie-Kolpin M, Hsieh CM, Kornman KS, Duff GW. Single nucleotide polymorphisms in the human interleukin-1B gene affect transcription according to haplotype context. Hum Mol Genet. 2006 Feb;15:519-29.

Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. Genetics. 1995;139:1455-61.

Choi SJ, Rho YH, Ji JD, Song GG, Lee YH. Genome scan meta-analysis of rheumatoid arthritis. Rheumatology (Oxford). 2006;45:166-70.

Clayton D. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet. 1999 Oct;65:1170-77.

Clayton D. SNPHAP, a program for esimtaing frequencies of haplotypes of large numbers of diallelic markers from unphased genotype data from unrelated subjects. http//www-gene.cimr.cam.ac.uk/clayton/software/

References

Clayton D, Chapman J, Cooper J. The use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol. 2004 Dec;27:415-28.

Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J. A new method to test genetic models in HLA associated diseases: the MASC method. Ann Hum Genet. 1988 Jul;52:247-58.

Cleveland WS. Robust locally weighted regression and smoothing scatteplots. J Am Stat Assoc. 1979;88:7276-80.

Cloutier JF, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. J Exp Med. 1999;189:111-21.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science. 2004;305:869-72.

Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM. Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. Blood. 1999;93:2013-24

Cordell HJ, Barratt BJ, Clayton DG. Case/pseudo-control analysis in genetic association studies: a unified framework for detection of genotype and haplotype association, gene-gene and gene-environment interactions and parent-of-origin effects. Genet Epidemiol. 2004 Apr;26:167-85.

Cordell H, Clayton D. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet. 2000;67:146-54.

Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993 Aug;261:921-3.

Cornélis F, Fauré S, Martinez M, Prud'homme JF, Fritz P, Dib C, Alves H, Barrera P, de Vries N, Balsa A, Pascual-Salcedo D, Maenaut K, Westhovens R, Migliorini P, Tran TH, Delaye A, Prince N, Lefevre C, Thomas G, Poirier M, Soubigou S, Alibert O, Lasbleiz S, Fouix S, Bouchier C, Lioté F, Loste MN, Lepage V, Charron D, Gyapay G, Lopes-Vaz A, Kuntz D, Bardin T, Weissenbach J; ECRAF. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. Proc Natl Acad Sci USA. 1998 Sep;95:10746-50.

Côté JF, Chung PL, Théberge JF, Hallé M, Spencer S, Lasky LA, Tremblay ML. PSTPIP is a substrate of PTP-PEST and serves as a scaffold guiding PTP-PEST toward a specific dephosphorylation of WASP. J Biol Chem. 2002;25;277:2973-86.

References

Cox A, Camp NJ, Cannings C, di Giovine FS, Dale M, Worthington J, John S, Ollier WE, Silman AJ, Duff GW. Combined sib-TDT and TDT provide evidence for linkage of the interleukin-1 gene cluster to erosive rheumatoid arthritis. Hum Mol Genet. 1999 Sep;8:1707-13.

Crilly A, Maiden N, Capell HA, Madhok R. Predictive value of interleukin 1 gene polymorphism for surgery. Ann Rheum Dis. 2000;59:695-699.

Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W, Lee AT, Ortmann W, Behrens TW, Gregersen PK. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J_Hum Genet. 2005 Apr;76(4):561-71.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High resolution haplotype structure in the human genome. Nat Genet. 2001; 29:229-32.

Davidson D, Bakinowski M, Thomas ML, Horejsi V, Veillette A.Phosphorylation-dependent regulation of T-cell activation by PAG/Cbp, a lipid raft-associated transmembrane adaptor. Mol Cell Biol. 2003;23:2017-28.

de Carvalho A, Graudal H, Jørgensen B. Evaluation of the progression of rheumatoid arthritis. Significance of age at onset and sex. Acta Radiol Diagn (Stockh). 1980;21(4):545-50.

Deighton CM, Cavanagh G, Rigby AS, Lloyd HL, Walker DJ. Both inherited HLA haplotypes are important in predisposition to rheumatoid arthritis. Br J Rheumatol. 1993;32:893–898.

Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. Clin Genet. 1989;36:178-82.

Deleuran BW, Chu CQ, Field M, Brennan FM, Katsikis P, Feldmann M, Maini RN. Localization of interleukin-1 alpha, type 1 interleukin-1 receptor and interleukin-1 receptor antagonist in the synovial membrane and cartilage/pannus junction in rheumatoid arthritis. Br J Rheumatol. 1992;31:801-9.

del Rincon ID, Williams K, Stern MP, Freeman GL, Escalante A. High incidence of cardiovascular events in a rheumatoid arthritis cohort not explained by traditional cardiac risk factors. Arthritis Rheum. 2001;44:2737-45.

Dervieux T, Furst D, Lein DO, Capps R, Smith K, Walsh M, Kremer J. 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 Rheum. 2004 Sep;50:2766-74.

References

Dervieux T, Greenstein N, Kremer J. Pharmacogenomic and metabolic biomarkers in the folate pathway and their association with methotrexate effects during dosage escalation in rheumatoid arthritis. Arthritis Rheum. 2006;54:3095-103.

Dervieux T, Kremer J, Lein DO, Capps R, Barham R, Meyer G, Smith K, Caldwell J, Furst DE. Contribution of common polymorphisms in reduced folate carrier and gamma-glutamylhydrolase to methotrexate polyglutamate levels in patients with rheumatoid arthritis. Pharmacogenetics. 2004 Nov;14:733-9.

De Rycke L, Nicholas AP, Cantaert T, Kruithof E, Echols JD, Vandekerckhove B, Veys EM, De Keyser F, Baeten D. Synovial intracellular citrullinated proteins colocalizing with peptidyl arginine deiminase as pathophysiologically relevant antigenic determinants of rheumatoid arthritis-specific humoral autoimmunity._Arthritis Rheum. 2005 ;52:2323-30.

Dieudé P, Teixeira VH, Pierlot C, Cornélis F, Petit-Teixeira E; ECRAF Testing for linkage and association with rheumatoid arthritis a ptpn22 promoter polymorphism reported to be associated and linked with type 1 diabetes in the Caucasian population. Ann Rheum Dis. 2008;67:900-1.

di Giovine FS, Camp NJ, Cox A, Chaudhary AG, Sorrell JA, Crane A et al. Detection and population analysis of IL-1 and TNF gene polymorphisms. In: Balkwill F, editor. Cytokine molecular biology. Oxford (UK): Oxford University Press; 2000;p.21-46.

Donner H, Braun J, Seidl C, Rau H, Finke R, Ventz M, Walfish PG, Usadel KH, Badenhoop K. Codon 17 polymorphism of the cytotoxic T lymphocyte antigen 4 gene in Hashimoto's thyroiditis and Addison's disease. J Clin Endocrinol Metab. 1997;82:4130-2.

Drozdzik M, Rudas T, Pawlik A, Gornik W, Kurzawski M, Herczynska M. Reduced folate carrier-1 80G>A polymorphism affects methotrexate treatment outcome in rheumatoid arthritis. Pharmacogenomics J. 2007;7:404-7.

Eastgate JA, Symons JA, Wood NC, Grinlinton FM, di Giovine FS, Duff GW. Correlation of plasma interleukin 1 levels with disease activity in rheumatoid arthritis. Lancet. 1988;2:706-9.

Eaves L, Meyer J. Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet. 1994;24(5):443-55.

Eitan Y, Kashi Y. Direct micro-haplotyping by multiple double PCR amplifications of specific allele (MD-PASA). Nucleic Acids Res. 2002 Jun;30:e62.

El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA. The role of interleukin-1 polymorphisms in the pathogenesis of gastric cancer. Nature. 2001;412:99.

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Shahab A, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JN, Yu Y, Ruan Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsdóttir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu

References

B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007 Jun;447:799-816.

Evans TI, Han J, Singh R, Moxley G. The genotyping distribution of shared epitope DRB1 alleles suggests a recessive mode of inheritance for rheumatoid arthritis disease-susceptibility genes. Arthritis Rheum. 1995;38:1754-1761.

Eyre S, Barton A, Shephard N, Hinks A, Brintnell W, MacKay K, Silman A, Ollier W, Wordsworth P, John S, Worthington J. Investigation of susceptibility loci identified in the UK rheumatoid arthritis whole-genome scan in a further series of 217 UK affected sibling pairs. Arthritis Rheum. 2004 Mar;50:729-35.

Eyre S, Bowes J, Spreckley K, Potter C, Ring S, Strachan D, Worthington J, Barton A. Investigation of the MHC2TA gene, associated with rheumatoid arthritis in a Swedish population, in a UK rheumatoid arthritis cohort. Arthritis Rheum. 2006;54:3417-22.

Fallin D, Schork N. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation–maximization algorithm for unphased diploid genotype data. Am J Hum Genet. 2000;67:947-59.

Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, Myers RH, Pericak-Vance MA, Risch N, van Duijn CM. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA. 1997;278:1349-56.

Feuk L, Marshall CR, Wintle RF, Schere SW. Structural variants: changing the landscape of chromosomes and design of disease studies. Hum Mol Genet. 2006;15:R57-R66.

Firestein GS, Alvaro-Gracia JM, Maki R. Quantitative analysis of cytokine gene expression in rheumatoid arthritis. J Immunol. 1990;144:3347-53.

Fisher R. The correlation between relative on the supposition of Mendelian inheritance. Trans R Soc Edin. 1918; 52:399-433.

Fisher SA, Lanchbury JS, Lewis CM. Meta-analysis of four rheumatoid arthritis genome-wide linkage studies: confirmation of a susceptibility locus on chromosome 16. Arthritis Rheum. 2003;48:1200-6.

Friese MA, Jones EY, Fugger L. MHC II molecules in inflammatory diseases: interplay of qualities and quantities. Trends Immunol. 2005;26:559-61.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002 ;296:2225-9.

References

Gabriel SE, Crowson CS, O'Fallon WM. Comorbidity in arthritis. J Rheumatol. 1999;26:2475-9.

Gabriel SE, Crowson CS, Kremers HM, Doran MF, Turesson C, O'Fallon WM, Matteson EL . Survival in rheumatoid arthritis: a population-based analysis of trends over 40 years. Arthritis Rheum. 2003;48:54-8.

Gabriel SE. The epidemiology of rheumatoid arthritis. Rheum Dis Clin North Am. 2001 ;27:269-81

Gao XJ, Brautbar C, Gazit E, Segal R, Naparstek Y, Livneh A, Stastny P. A variant of HLA-DR4 determines susceptibility to rheumatoid arthritis in a subset of Israeli Jews. Arthritis Rheum. 1991;34:547-51.

Genevay S, di Giovine FS, Perneger TV, Silvestri T, Stingelin S, Duff G, Guerne PA. Association of interleukin-4 and interleukin-1B gene variants with Larsen score progression in rheumatoid arthritis. Arthritis Rheum. 2002;47:303-9.

Génin E, Babron MC, McDermott MF, Mulcahy B, Waldron-Lynch F, Adams C, Clegg DO, Ward RH, Shanahan F, Molloy MG, O'Gara F, Clerget-Darpoux F. Modelling the major histocompatibility complex susceptibility to RA using the MASC method. Genet Epidemiol. 1998;15:419-430.

Girbal-Neuhauser E, Durieux JJ, Arnaud M, Dalbon P, Sebbag M, Vincent C, Simon M, Senshu T, Masson-Bessiere C, Jolivet-Reynaud C, Jolivet M, Serre G. The epitopes targeted by the rheumatoid arthritis-associated antifilaggrin autoantibodies are posttranslationally generated on various sites of (pro)filaggrin by deimination of arginine residues. J Immunol. 1999;162:585-94.

Gjorloff-Wingren A, Saxena M, Williams S, Hammi D, Mustelin T. Characterisation of TCR-induced receptor-proximal signaling events negatively regulated by the protein tyrosine phosphatase. Eur.J. Immunol. 1999;29:3845-54

Glant TT, Mikecz K, Arzoumanian A, Poole AR. Protetoglycan incuded arthritis in BALB/c mice. Clinical features and histopathology. Arthritis Rheum. 1987;30:201-12.

Gonzalez-Juanatey C, Testa A, Garcia-Castelo A, Garcia-Porrua C, Llorca J, Gonzalez-Gay MA. Active but transient improvement of endothelial function in rheumatoid arthritis patients undergoing long-term treatment with anti-tumor necrosis factor alpha antibody. Arthritis Rheum. 2004;51:447-50.

Gonzalez-Lopez L, Gamez-Nava JI, Jhangri GS, Ramos-Remus C, Russell AS, Suarez-Almazor ME. Prognostic factors for the development of rheumatoid arthritis and other connective tissue diseases in patients with palindromic rheumatism. J Rheumatol. 1999 ;26:540-5.

References

Gorman JD, David-Vaudey E, Pai M, Lum RF, Criswell LA. Lack of association of the HLA-DRB1 shared epitope with rheumatoid nodules: an individual patient data meta-analysis of 3,272 Caucasian patients with rheumatoid arthritis. Arthritis Rheum. 2004 Mar;50(3):753-62.

Gorman JD, Lum RF, Chen JJ, Suarez-Almazor ME, Thomson G, Criswell LA. Impact of shared epitope genotype and ethnicity on erosive disease: a meta-analysis of 3,240 rheumatoid arthritis patients. Arthritis Rheum. 2004 Feb;50:400-12.

Goronzy JJ, Matteson EL, Fulbright JW, Warrington KJ, Chang-Miller A, Hunder GG, Mason TG, Nelson AM, Valente RM, Crowson CS, Erlich HA, Reynolds RL, Swee RG, O'Fallon WM, Weyand CM. Prognostic markers of radiographic progression in early rheumatoid arthritis. Arthritis Rheum. 2004;50(1):43-54.

Goronzy JJ, Weyand CM. T-cell regulation in rheumatoid arthritis. Curr Opin Rheumatol. 2004;16:212-217.

Goronzy JJ, Bartz-Bazzanella P, Hu W, Jendro MC, Walser-Kuntz DR, Weyand CM. Dominant clonotypes in the repertoire of peripheral CD4+ T cells in rheumatoid arthritis. J Clin Invest. 1994;94:2068-76.

Gossec L, Bettembourg-Brault I, Pham T, Dougados M. HLA DRB1*01 and DRB1*04 phenotyping does not predict the need for joint surgery in rheumatoid arthritis. A retrospective quantitative evaluation of 300 French patients. Clin Exp Rheumatol. 2004 Jul-Aug;22(4):462-4.

Gravallese EM, Galson DL, Goldring SR, Auron PE. The role of TNF-receptor family members and other TRAF-dependent receptors in bone resorption. Arthritis Res. 2001;3:6-12.

Gray-McGuire C, Moser KL, Gaffney PM, Kelly J, Yu H, Olson JM, Jedrey CM, Jacobs KB, Kimberly RP, Neas BR, Rich SS, Behrens TW, Harley JB. Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. Am J Hum Genet. 2000;67:1460-9.

Gregersen PK, Lee HS, Batliwalla F, Begovich AB.PTPN22: setting thresholds for autoimmunity. Semin Immunol. 2006;18:214-23.

Gregersen PK, Silver J, Winchester RJ. Genetic susceptibility to rheumatoid arthritis and human leukocyte antigen class II polymorphism. The role of shared conformational determinants. Am J Med. 1988;85:17-9.

Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987;30:1205-13.

References

Grennan DM, Sanders PA, Thomson W, Dyer PA.Rheumatoid arthritis: inheritance and association with other autoimmune diseases. Dis Markers. 1986;4:157-62.

Guasch JF, Bertina RM, Reitsma PH. Five novel intragenic dimorphisms in the human interleukin-1 genes combine to high informativity. Cytokine. 1996;8:598-602.

Guillemin F, Billot L, Boini S, Gerard N, Ødegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J.Rheumatol. 2005 May;32(5):778-86.

Hakala M, Ilonen J, Reijonen H, Knip M, Koivisto O, Isomäki H. No association between rheumatoid arthritis and insulin dependent diabetes mellitus: an epidemiologic and immunogenetic study. J.Rheumatol. 1992 Jun;19(6):856-8.

Hall FC, Weeks DE, Camilleri JP, Williams LA, Amos N, Darke C, Gibson K, Pile K, Wordsworth BP, Jessop JD. Influence of the HLA-DRB1 locus on susceptibility and severity in rheumatoid arthritis. QJM. 1996;89:821-9.

Hall SK, Perregaux DG, Gabel CA, Woodworth T, Durham LK, Huizinga TW, Breedveld FC, Seymour AB Correlation of polymorphic variation in the promoter region of the interleukin-1 beta gene with secretion of interleukin-1 beta protein. Arthritis Rheum. 2004;50:1976-83.

Harney SM, Vilariño-Güell C, Adamopoulos IE, Sims AM, Lawrence RW, Cardon LR, Newton JL, Meisel C, Pointon JJ, Darke C, Athanasou N, Wordsworth BP, Brown MA. Fine mapping of the MHC Class III region demonstrates association of AIF1 and rheumatoid arthrtitis. Rheumatology (Oxford). 2008 Oct. In press.

Harney S, Newton J, Milicic A, Brown MA, Wordsworth BP. Non-inherited maternal HLA alleles are associated with rheumatoid arthritis. Rheumatology (Oxford) 2003;42:171-4.

Harvey JN, Craney L, Kelly D. Estimation of the prevalence of diagnosed diabetes from primary care and secondary care source data: comparison of record linkage with capture-recapture analysis. J Epidemiol Community Health. 2002 Jan;56(1):18-23.

Harvey J, Lotze M, Stevens MB, Lambert G, Jacobson D. Rheumatoid arthritis in a Chippewa band. I. Pilot screening study of disease prevalence. Arthritis Rheum. 1981 May;24:717-21.

Hattori M, Buse JB, Jackson RA, Glimcher L, Dorf ME, Minami M, Makino S, Moriwaki K, Kuzuya H, Imura H, *et al*. The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. Science. 1986 Feb;231:733-5.

References

Hazes JM, Silman AJ. Review of UK data on the rheumatic diseases--2. Rheumatoid arthritis. Br J Rheumatol. 1990 Aug;29:310-2.

Hermann R, Lipponen K, Kiviniemi M, Kakko T, Veijola R, Simell O, Knip M, Ilonen J. Lymphoid tyrosine phosphatase (LYP/PTPN22) Arg620Trp variant regulates insulin autoimmunity and progression to type 1 diabetes. Diabetologia. 2006;49:1198-208.

Heward JM, Allahabadia A, Armitage M, Hattersley A, Dodson PM, Macleod K, Carr-Smith J, Daykin J, Daly A, Sheppard MC, Holder RL, Barnett AH, Franklyn JA, Gough SC. The development of Graves' disease and the CTLA-4 gene on chromosome 2q33. J Clin Endocrinol Metab. 1999;84:2398-401.

Hill WG, Weir BS. Maximum-likelihood estimation of gene location by linkage disequilibrium. Am J Hum Genet. 1994;54:705-14.

Hinks A, Eyre S, Barton A, Thomson W, Worthington J. Investigation of genetic variation across the protein tyrosine phosphatase gene in patients with rheumatoid arthritis in the UK. Ann Rheum Dis. 2007 May;66(5):683-6.

Hinks A, Barton A, John S, Bruce I, Hawkins C, Griffiths CE, Donn R, Thomson W, Silman A, Worthington J. Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: Further support that PTPN22 is an autoimmunity gene. Arthritis Rheum. 2005;52(6):1694-1699.

Hochberg MC. The application of genetic epidemiology to systemic lupus erythematosus. J Rheumatol. 1987;14:867-9.

Hoeve MA, Savage ND, de Boer T, Langenberg DM, de Waal Malefyt R, Ottenhoff TH, Verreck FA. Divergent effects of IL-12 and IL-23 on the production of IL-17 by human T cells. Eur J Immunol. 2006;36:661-70.

Hoving JL, Buchbinder R, Hall S, Lawler G, Coombs P, McNealy S, Bird P, Connell D. A comparison of magnetic resonance imaging, sonography, and radiography of the hand in patients with early rheumatoid arthritis. J Rheumatol. 2004;31:663-75.

Hüffmeier U, Steffens M, Burkhardt H, Lascorz J, Schürmeier-Horst F, Ständer M, Kelsch R, Baumann C, Küster W, Mössner R, Reich K, Wienker TF, Traupe H, Reis A. Evidence for susceptibility determinant(s) to psoriasis vulgaris in or near PTPN22 in German patients. J Med Genet. 2006 Jun;43:517-22.

Hwand SY, Kim HY, Expression of IL-17 homologs and their receptors in the synovial cells of rheumatoid arthritis patients. Mol Cells. 2005;19:180-4.

Iikuni N, Ikari K, Momohara S, Tomatsu T, Hara M, Yamanaka H, Okamoto H, Kamatani N. MHC2TA is associated with rheumatoid arthritis in Japanese patients. Ann Rheum Dis. 2007;66:274-5.

Jang E, Kim HR, Cho SH, Paik DJ, Kim JM, Lee SK, Youn J. Prevention of spontaneous arthritis by inhibiting homeostatic expansion of autoreactive $CD4^+$ T cells in the K/BxN mouse model. Arthritis Rheum. 2006 Feb;54:492-8.

Jaakkola E, Herzberg I, Laiho K, Barnardo MC, Pointon JJ, Kauppi M, Kaarela K, Tuomilehto-Wolf E, Tuomilehto J, Wordsworth BP, Brown MA. Finnish HLA studies confirm the increased risk conferred by HLA-B27 homozygosity in ankylosing spondylitis. Ann Rheum Dis. 2006;65:775-80.

Jagiello P, Aries P, Arning L, Wagenleiter SE, Csernok E, Hellmich B, Gross WL, Epplen JT. The PTPN22 620W allele is a risk factor for Wegener's granulomatosis. Arthritis Rheum. 2005;52:4039-43.

James ES, Harney S, Wordsworth BP, Cookson WO, Davis SJ, Moffatt MF. PDCD1: a tissue-specific susceptibility locus for inherited inflammatory disorders. Genes Immun. 2005;6:430-7.

Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Etzel C, Damle A, Xiao X, Chen D, Lum RF, Monteiro J, Kern M, Criswell LA, Albani S, Nelson JL, Clegg DO, Pope R, Schroeder HW Jr, Bridges SL Jr, Pisetsky DS, Ward R, Kastner DL, Wilder RL, Pincus T, Callahan LF, Flemming D, Wener MH, Gregersen PK; North American Rheumatoid Arthritis Consortium. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis Rheum. 2003;48:906-16.

Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, Monteiro J, Khalili H, Lee A, Lundsten R, Begovich A, Bugawan T, Erlich H, Elder JT, Criswell LA, Seldin MF, Amos CI, Behrens TW, Gregersen PK. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. Am J Hum Genet. 2002 Sep;71:585-94.

Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, Kern M, Criswell LA, Albani S, Nelson JL, Clegg DO, Pope R, Schroeder HW Jr, Bridges SL Jr, Pisetsky DS, Ward R, Kastner DL, Wilder RL, Pincus T, Callahan LF, Flemming D, Wener MH, Gregersen PK. A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. Am J Hum Genet. 2001 ;68:927-36.

Johansson M, Arlestig L, Hallmans G, Rantapää-Dahlqvist S. PTPN22 polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid arthritis and has a specificity of 100% for the disease. Arthritis Res Ther. 2006;8:R19.

John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. Am J Hum Genet. 2004;75:54-64.

References

Johnsen AK, Plenge RM, Butty V, Campbell C, Dieguez-Gonzalez R, Gomez-Reino JJ, Shadick N, Weinblatt M, Gonzalez A, Gregersen PK, Benoist C, Mathis D. A broad analysis of IL1 polymorphism and rheumatoid arthritis. Arthritis Rheum. 2008 Jul;58:1947-57.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet. 2001 Oct;29:233-7.

Jorgenson E, Witte JS. Coverage and power in genomewide association studies. Am J Hum Genet. 2006;78:884-8.

Jouvenne P, Chaudhary A, Buchs N, Giovine FS, Duff GW, Miossec P. Possible genetic association between interleukin-1alpha gene polymorphism and the severity of chronic polyarthritis. Eur Cytokine Netw. 1999;10:33-6.

Jovanovic DV, Martel-Pelletier J, di Battista JA, Mineau F, Jolicoeur FC, Benderdour M, Pelletier JP. Stimulation of 92kd gelatinase (matrix metalloproteinase 9) production by interleukin-17 in human monocyte/macrophages: a possible role in rheumatoid arthritis. Arthritis Rheum. 2000 May;43:1134-44.

Julià A, Ballina J, Cañete JD, Balsa A, Tornero-Molina J, Naranjo A, Alperi-López M, Erra A, Pascual-Salcedo D, Barceló P, Camps J, Marsal S. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. Arthritis Rheum. 2008 Aug;58:2275-86.

Kang YM, Zhang X, Wagner UG, Yang H, Beckenbaugh RD, Kurtin PJ, Goronzy, JJ, Weyand CM. CD8 T cells are required for the formation of ectopic germinal centers in rheumatoid synovitis. J Exp Med. 2002 May;195:1325-36.

Kannan P, Buettner R, Chiao PJ, Yim SO, Sarkiss M, Tainsky MA. N-ras oncogene causes AP-2 trabcriptional self-interference, which leads to transformation. Genes Dev. 1994 Jun;8:1258-69.

Katz Y, Ndiv O, Beer Y. Interleukin-17 enchances tumor necrosis factor alpha-induced synthesis of inteleukins 1, 6 and 8 in skin and synovial fibroblasts: a possible role as a "fine tuning cytokine" in inflammation processes. Arthritis Rheum. 2001;44:2176-84.

Kawasaki E, Awata T, Ikegami H, Kobayashi T, Maruyama T, Nakanishi K, Shimada A, Uga M, Kurihara S, Kawabata Y, Tanaka S Systematic search for single nucleotide polymorphisms in a lymphoid tyrosine phosphatase gene (PTPN22): association between a promoter polymorphism and type 1 diabetes in Asian populations. Am J Med Genet A. 2006;140:586-93.

References

Khanna D, Ranganath VK, Fitzgerald J, Park GS, Altman RD, Elashoff D, Gold RH, Sharp JT, Furst DE, Paulus HE; Western Consortium of Practicing Rheumatologists. Increased radiographic damage scores at the onset of seropositive rheumatoid arthritis in older patients are associated with osteoarthritis of the hands, but not with more rapid progression of damage. Arthritis Rheum. 2005;52:2284-92.

Kim HY, Kim TG, Park SH, Lee SH, Cho CS, Han H. Predominance of HLA-DRB1*0405 in Korean patients with rheumatoid arthritis. Ann Rheum Dis. 1995;54:988-90.

Knight JC. Regulatory polymorphisms underlying complex disease traits. Mol Med. 2005;83:97-109.

Koch O, Kwiatowski DP, Udalova IA. Context-specific functional effects of IFNGR1 promoter polymorphism. Hum Mol Genet. 2006;15:1475-1481.

Kong EK, Prokunina-Olsson L, Wong WH, Lau CS, Chan TM, Alarcon-Riquelme M, Lau YL. A new haplotype of PDCD1 is associated with rheumatoid arthritis in Hong Kong Chinese. Arthritis Rheum. 2005;52:1058-62.

Kong YY, Feige U, Sarosi I, Bolon B, Tafuri A, Morony S, Capparelli C, Li J, Elliott R, McCabe S, Wong T, Campagnuolo G, Moran E, Bogoch ER, Van G, Nguyen LT, Ohashi PS, Lacey DL, Fish E, Boyle WJ, Penninger JM. (1999). Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. Nature. 1999;402:304-9.

Kornman KS, Crane A, Wang HY, di Giovine FS, Newman MG, Pirk FW, Wilson TG Jr, Higginbottom FL, Duff GW. The interleukin-1 genotype as a severity factor in adult periodontal disease. J Clin Periodontol. 1997;24:72-7.

Kotake S, Udagawa N, Takahashi N, Matsuzaki K, Itoh K, Ishiyama S, Saito S, Inoue K, Kamatani N, Gillespie MT, Martin TJ, Suda T. IL-17 in synovial fluid from patients with rheumatoid arthritis is a potent stimulator of osteoclastogenesis. J Clin Invest. 1999 May;103:1345-52.

Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet. 2001;27:234-6.

Kunkel HG, Williams RC. Rheumatoid arthritis. Annu Rev Med. 1964;15:37-52.

Kurki P, Aho K, Palosuo T, Heliovaara M. Immunopathology of rheumatoid arthritis. Antikeratin antibodies precede the clinical disease. Arthritis Rheum. 1992;35:914-7.

Kurzawski M, Pawlik A, Safranow K, Herczynska M, Drozdzik M. 677C>T and 1298A>C MTHFR polymorphisms affect methotrexate treatment outcome in rheumatoid arthritis. Pharmacogenomics. 2007;8:1551-9.

Kwoh CK, Venglish C, Lynn AH, Whitley DM, Young E, Chakravarti A. Age, sex, and the familial risk of rheumatoid arthritis. Am J Epidemiol. 1996;144:15-24.

Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Baechler EC, Batliwalla FM, Novitzke J, Williams AH, Gillett C, Rodine P, Graham RR, Ardlie KG, Gaffney PM, Moser KL, Petri M, Begovich AB, Gregersen PK, Behrens TW. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet. 2004 Sep;75:504-7.

Lacey DL, Timms E, Tan HL, Kelley MJ, Dunstan CR, Burgess T, Elliott R, Colombero A, Elliott G, Scully S, Hsu H, Sullivan J, Hawkins N, Davy E, Capparelli C, Eli A, Qian YX, Kaufman S, Sarosi I, Shalhoub V, Senaldi G, Guo J, Delaney J, Boyle WJ.(1998).Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. Cell. 1998;93:165-76.

Lanchbury JS, Jaeger EE, Sansom DM, Hall MA, Wordsworth P, Stedeford J, Bell JI, Panayi GS. Strong primary selection for the Dw4 subtype of DR4 accounts for the HLA-DQw7 association with Felty's syndrome. Hum Immunol. 1991;32:56-64.

Lander ES. The new genomics: global views of biology. Science. 1996;274:536-9.

Lawrence J Rheumatoid arthritis—nature or nurture? Ann Rheum Dis 1970;29:357.

Lee AT, Li W, Liew A, Bombardier C, Weisman M, Massarotti EM, Kent J, Wolfe F, Begovich AB, Gregersen PK. The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. Genes Immun. 2005;6(2):129-33.

Lee HS, Remmers EF, Le JM, Kastner KL, Bae SC, Gregersen PK. Association of STAT4 with rheumatoid arthritis in the Korean population. Mol Med. 2007;13:455-60.

Lee KA, Ki CS, Kim HJ, Sohn KM, Kim JW, Kang WK, Rhee JC, Song SY, Sohn TS. Novel interleukin 1beta polymorphism increased the risk of gastric cancer in a Korean population. J Gastroenterol. 2004;39:429-33.

Lee YH, Rho YH, Choi SJ, Ji JD, Song GG, Nath SK, Harley JB. The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis. Rheumatology (Oxford). 2007 Jan;46:49-56.

Lee YH, Rho YH, Choi SJ, Ji JD, Song GG. Association of TNF-alpha -308 G/A polymorphism with responsiveness to TNF-alpha-blockers in rheumatoid arthritis: a meta-analysis._Rheumatol Int. 2006;27:157-61.

Liu C, Batliwalla F, Li W, Lee A, Roubenoff R, Beckman E, Khalili H, Damle A, Kern M, Furie R, Dupuis J, Plenge RM, Coenen MJ, Behrens TW, Carulli JP, Gregersen PK. Genome-wide association scan identifies candidate polymorphisms associated with

differential response to anti-TNF treatment in rheumatoid arthritis. Mol Med. 2008;14:575-81.

Lei C, Dongqing Z, Yeqing S, Oaks MK, Lishan C, Jianzhong J, Jie Q, Fang D, Ningli L, Xinghai H, Daming R. Association of the CTLA-4 gene with rheumatoid arthritis in Chinese Han population. Eur J Hum Genet. 2005 Jul;13:823-8.

Lewontin R. The interaction of selection and linkage I. General considerations. Genetics. 1964;49:49-67.

Lie BA, Viken MK, Odegård S, van der Heijde D, Landewé R, Uhlig T, Kvien TK. Associations between the PTPN22 1858C->T polymorphism and radiographic joint destruction in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Ann Rheum Dis. 2007;66:1604-9.

Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. Allelic variation in gene expression is common in the human genome. Genome Research 2003;13:1855-1862.

Lopez-Ben R, Bernreuter WK, Moreland LW, Alarcon GS. Ultrasound detection of bone erosions in rheumatoid arthritis: a comparison to routine radiographs of the hands and feet. Skeletal Radiol. 2004;33:80-4.

Lynn AH, Kwoh CK, Venglish CM, Aston CE, Chakravarti A. Genetic epidemiology of rheumatoid arthritis. Am J Hum Genet. 1995;57:150-159.

MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, Silman AJ.Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. 2000;43:30–37.

MacGregor AJ, Riste LK, Hazes JM, Silman AJ. (1994). Low prevalence of rheumatoid arthritis in black-Caribbeans compared with whites in inner city Manchester. Ann Rheum Dis. 1994;53:293-7.

MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, Barrett J, Lee D, White S, John S, Brown MA, Bell J, Silman A, Ollier W, Wordsworth P, Worthington J. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. Arthritis Rheum. 2002;46:632-9.

Mackay TF, Lyman RF, Jackson MS. Effects of P element insertions on quantitative traits in Drosophila melanogaster. Genetics. 1992;130:315-32.

Marson A, Kretschmer K, Frampton GM, Jacobsen ES, Polansky JK, MacIsaac KD, Levine SS, Fraenkel E, von Boehmer H, Young RA. Foxp3 occupancy and regulation of key target genes during T-cell stimulation. Nature. 2007;445:931-935.

Martín MC, Oliver J, Urcelay E, Orozco G, Gómez-Garcia M, López-Nevot MA, Piñero A, Brieva JA, de la Concha EG, Nieto A, Martín J. The functional genetic variation in the PTPN22 gene has a negligible effect on the susceptibility to develop inflammatory bowel disease. Tissue Antigens. 2005 Oct;66:314-7.

Martínez A, Sánchez-Lopez M, Varadé J, Mas A, Martín MC, de Las Heras V, Arroyo R, Mendoza JL, Díaz-Rubio M, Fernández-Gutiérrez B, de la Concha EG, Urcelay E. Role of the MHC2TA gene in autoimmune diseases. Ann Rheum Dis. 2007 Mar;66:325-9.

Masson-Bessière C, Sebbag M, Girbal-Neuhauser E, Nogueira L, Vincent C, Senshu T, Serre G. The major synovial targets of the rheumatoid arthritis-specific antifilaggrin autoantibodies are deiminated forms of the alpha- and beta-chains of fibrin. J Immunol. 2001;166:4177-84.

Masson-Bessiere C, Sebbag M, Durieux JJ, Nogueira L, Vincent C, Girbal-Neuhauser E, Durroux R, Cantagrel A, Serre G. In the rheumatoid pannus, anti-filaggrin autoantibodies are produced by local plasma cells and constitute a higher proportion of IgG than in synovial fluid and serum. Clin Exp Immunol. 2000;119:544-52.

Masternak K, Muhlethaler-Mottet A, Villard J, Zufferey M, Steimle V, Reith W. CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. Genes Dev. 2000;14:1156-66.

Mattey DL, Hassell AB, Dawes PT, Cheung NT, Poulton KV, Thomson W, Hajeer AH, Ollier WE. Independent association of rheumatoid factor and the HLA-DRB1 shared epitope with radiographic outcome in rheumatoid arthritis. Arthritis Rheum. 2001;44:1529-33.

Mattey DL, Hassell AB, Plant MJ, Cheung NT, Dawes PT, Jones PW, Thomson W, Poulton KV, Hajeer AH, Ollier WE. The influence of HLA-DRB1 alleles encoding the DERAA amino acid motif on radiological outcome in rheumatoid arthritis. Rheumatology (Oxford). 1999;38:1221-7.

Matthews RJ, Bowne DB, Flores E, Thomas ML. Characterization of hematopoietic intracellular protein tyrosine phosphatases: description of a phosphatase containing an SH2 domain and another enriched in proline-, glutamic acid-, serine-, and threonine-rich sequences. Mol Cell Biol. 1992;12:2396-405.

Matthys P, Vermeire K, Mitera T, Heremans H, Huang S, Billiau A. Anti-IL-12 antibody prevents the development and progression of collagen-induced arthritis in IFN-gamma receptor-deficient mice. Eur J Immunol. 1998;28:2143-51.

Meyer JM, Evans TI, Small RE, Redford TW, Han J, Singh R, Moxley G. HLA-DRB1 genotype influences risk for and severity of rheumatoid arthritis. J Rheumatol. 1999 May;26(5):1024-34.

References

Michalatos-Beloin S, Tishoff SA, Bentley KL, Kidd KK, Ruano G. Molecular haplotyping of genetic markers 10 kb apart by allele-specific long range PCR. Nucleic Acids Res. 1996 Dec;24:4841-3.

Michou L, Lasbleiz S, Rat AC, Migliorini P, Balsa A, Westhovens R, Barrera P, Alves H, Pierlot C, Glikmans E, Garnier S, Dausset J, Vaz C, Fernandes M, Petit-Teixeira E, Lemaire I, Pascual-Salcedo D, Bombardieri S, Dequeker J, Radstake TR, Van Riel P, van de Putte L, Lopes-Vaz A, Prum B, Bardin T, Dieudé P, Cornélis F; European Consortium on Rheumatoid Arthritis Families. Linkage proof for PTPN22, a rheumatoid arthritis susceptibility gene and a human autoimmunity gene. Proc Natl Acad Sci USA. 2007 Jan;104:1649-54.

Milicic A, Brown MA, Wordsworth BP. Polymorphism in codon 17 of the CTLA-4 gene (+49 A/G) is not associated with susceptibility to rheumatoid arthritis in British Caucasians. Tissue Antigens. 2001 Jul;58(1):50-4.

Milkiewicz P, Pache I, Buwaneswaran H, Liu X, Coltescu C, Heathcote EJ, Siminovitch KA. The PTPN22 1858T variant is not associated with primary biliary cirrhosis. Tissue Antigens. 2006;67:434-7.

Miossec P. Interleukin-17 in rheumatoid arthritis: if T cells were to contribute to inflammation and destruction through synergy. Arthritis Rheum. 2003;48:594-601.
Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955;7:277-318.
Morel L, Croker BP, Blenman KR, Mohan C, Huang G, Gilkeson G, Wakeland EK. Genetic reconstitution of systemic lupus erythematosus immunopathology with polycongenic murine strains. Proc Natl Acad Sci USA. 2000;97:6670-5.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. Nature. 2004;12;430:743-7.

Mugnier B, Balandraud N, Darque A, Roudier C, Roudier J, Reviron D. Polymorphism at position -308 of the tumor necrosis factor alpha gene influences outcome of infliximab therapy in rheumatoid arthritis. Arthritis Rheum. 2003 Jul;48(7):1849-52.
Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J. 1997;16:2851-60.

Neff MM, Turk E and Kalishman M. Web-based Primer Design for Single Nucleotide Polymorphism Analysis. Trends in Genetics. 2002:18;613-615.

Neidel J, Schulze M, Lindschau J. Association between degree of bone-erosion and synovial fluid-levels of tumor necrosis factor alpha in the knee-joints of patients with rheumatoid arthritis. Inflamm Res. 1995;44:217-21.

References

Nelson JL. Microchimerism in human health and disease. Autoimmunity. 2003;36(1):5-9.

Nelson MR, Kardia SL, Ferrell RE, Sing CF. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. Genome Res. 2001;11:458-70.

Neumann E, Barnum SR, Tarner IH, Echols J, Fleck M, Judex M, Kullmann F, Mountz JD, Scholmerich J, Gay S, Muller-Ladner U: Local production of complement proteins in rheumatoid arthritis synovium. Arthritis Rheum. 2002 Apr;46:934-945.

Newman WG, Zhang Q, Liu X, Walker E, Ternan H, Owen J, Johnson B, Greer W, Mosher DP, Maksymowych WP, Bykerk VP, Keystone EC, Amos CI, Siminovitch KA. Rheumatoid arthritis association with the FCRL3 -169C polymorphism is restricted to PTPN22 1858T-homozygous individuals in a Canadian population. Arthritis Rheum. 2006 Dec;54:3820-7.

Newton J, Brown MA, Milicic A, Ackerman H, Darke C, Wilson JN, Wordsworth BP, Kwiatkowski D. The effect of HLA-DR on susceptibility to rheumatoid arthritis is influenced by the associated lymphotoxin alpha-tumor necrosis factor haplotype. Arthritis Rheum. 2003 Jan;48(1):90-6.

Nicklin MJ, Barton JL, Nguyen M, FitzGerald MG, Duff GW, Kornman K. A sequence-based map of the nine genes of the human interleukin-1 cluster. Genomics. 2002;79:718-25.

Nicklin MJ, Weith A, Duff GW. A physical map of the region encompassing the human interleukin-1 alpha, interleukin-1 beta, and interleukin-1 receptor antagonist genes. Genomics. 1994;19:382-4.

Nicola PJ, Maradit-Kremers H, Roger VL, Jacobsen SJ, Crowson CS, Ballman KV, Gabriel SE. The risk of congestive heart failure in rheumatoid arthritis: a population-based study over 46 years. Arthritis Rheum. 2005;52:412-20.

Nielsen C, Barington T, Husby S, Lillevang ST. Expression of human PTPN22 alleles. Genes Immun. 2007;8:131-137.

Nishimura H, Okazaki T, Tanaka Y, Nakatani K, Hara M, Matsumori A, Sasayama S, Mizoguchi A, Hiai H, Minato N, Honjo T. Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice. Science. 2001;291:319-22.

Nishimura H, Nose M, Hiai H, Minato N, Honjo T. (1999). Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor. Immunity. 1999;11:141-51.

Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, Larrad

MT, Rios MS, Chow CC, Cockram CS, Jacobs K, Mijovic C, Bain SC, Barnett AH, Vandewalle CL, Schuit F, Gorus FK, Tosi R, Pozzilli P, Todd JA.The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Belgian Diabetes Registry. Hum Mol Genet. 1996;5(7):1075-80.

O'Connell J. Zero-recombinant haplotyping: application of fine mapping using SNPs. Genet Epidemiol. 2000;19:S64-S70.

O'Doherty C, Hawkins S, Rooney M, Vandenbroeck K. The MHC2TA-168A/G and +1614G/C polymorphisms and risk for multiple sclerosis or chronic inflammatory arthropathies._Tissue Antigens. 2007;70:247-51.

Orozco G, Pascual-Salcedo D, López-Nevot MA, Cobo T, Cabezón A, Martín-Mola E, Balsa A, Martín J. Auto-antibodies, HLA and PTPN22: susceptibility markers for rheumatoid arthritis. Rheumatology (Oxford). 2008;47:138-41.

Orozco G, García-Porrúa C, López-Nevot MA, Raya E, González-Gay MA, Martín J. Lack of association between ankylosing spondylitis and a functional polymorphism of PTPN22 proposed as a general susceptibility marker for autoimmunity. Ann Rheum Dis. 2006 ;65:687-8.

Orozco G, Robledo G, Linga Reddy MV, García A, Pascual-Salcedo D, Balsa A, González-Gay MA, Eimon A, Paira S, Scherbarth HR, Pons-Estel BA, Peterssen IF, Alarcón-Riquelme M, Martin J. Study of the role of a functional polymorphism of MHC2TA in rheumatoid arthritis in three ethnically different populations. Rheumatology (Oxford). 2006;45:1442-4.

Orozco G, Sánchez E, González-Gay MA, López-Nevot MA, Torres B, Cáliz R, Ortego-Centeno N, Jiménez-Alonso J, Pascual-Salcedo D, Balsa A, de Pablo R, Nuñez-Roldan A, González-Escribano MF, Martín J. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum. 2005 Jan;52:219-24.

Padyukov L, Lampa J, Heimbürger M, Ernestam S, Cederholm T, Lundkvist I, Andersson P, Hermansson Y, Harju A, Klareskog L, Bratt J. Genetic markers for the efficacy of tumour necrosis factor blocking therapy in rheumatoid arthritis. Ann Rheum Dis. 2003;62:526-9.

Paimela L, Palosuo T, Aho K, Lukka M, Kurki P, Leirisalo-Repo M, von Essen R. Association of autoantibodies to filaggrin with an active disease in early rheumatoid arthritis. Ann Rheum Dis. 2001;60:32-5.

Paleolog EM. Angiogenesis in rheumatoid arthritis. Arthritis Res. 2002;4:S81-S90.

Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T., Brändström H. A survey of genetic and epigenetic variation affecting human gene expression. Physiol Genomics. 2004;16:184-193.

Peltomaa R, Leirisalo-Repo M, Helve T, Paimela L. Effect of age on 3 year outcome in early rheumatoid arthritis. J Rheumatol. 2000;27:638-43.

Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coblyn J, Weinblatt ME, Gabriel SB, Purcell S, Klareskog L, Gregersen PK, Shadick NA, Daly MJ, Altshuler D. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet. 2007 Dec;39:1477-82.

Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D, Gregersen PK, Klareskog L, Rioux JD. Replication of putative candidate-gene association with rheumatoid arthritis in >4,000 samples from North America and Sweden: Association of susceptibility with PTPN22, STLA4 and PADI4. Am J Hum Genet. 2005 Dec;77:1044-60.

Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klareskog L, Gregersen PK. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med. 2007 Sep;357:1199-209.

Ploski R, Mellbye OJ, Ronningen KS, Forre O, Thorsby E. Seronegative and weakly seropositive rheumatoid arthritis differ from clearly seropositive rheumatoid arthritis in HLA class II associations. J Rheumatol. 1994;21:1397-402.

Prescott NJ, Fisher SA, Onnie C, Pattni R, Steer S, Sanderson J, Forbes A, Lewis CM, Mathew CG. A general autoimmunity gene (PTPN22) is not associated with inflammatory bowel disease in a British population. Tissue Antigens. 2005;66:318-20.

Prins JB, Todd JA, Rodrigues NR, Ghosh S, Hogarth PM, Wicker LS, Gaffney E, Podolin PL, Fischer PA, Sirotina A, *et al*. Linkage on chromosome 3 of autoimmune diabetes and defective Fc receptor for IgG in NOD mice. Science. 1993 Apr;260:695-8.

Pongratz R, Buchinger W, Semlitsch G, Meister E, Nadler K, Rainer F. Increased occurrence of autoimmune thyroiditis in patients with chronic rheumatoid arthritis. Acta Med Austriaca. 2000;27(2):58-60.

Qin S, Rottman JB, Myers P, Kassam N, Weinblatt M, Loetscher M, Koch AE, Moser B, Mackay CR. The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. J Clin Invest. 1998;101:746-54.

Qu H, Tessier MC, Hudson TJ, Polychronakos C. Confirmation of the association of the R620W polymorphism in the protein tyrosine phosphatase PTPN22 with type 1 diabetes in a family based study. J Med Genet. 2005;42:266-70.

Rantapaa-Dahlqvist S, de Jong BA, Berglin E, Hallmans G, Wadell G, Stenlund H, Sundin U, van Venrooij WJ. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum. 2003;48:2741-9.

Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. J Rheumatol. 1998 Nov;25(11):2094-107.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME Global variation in copy number in the human genome. Nature. 2006 Nov;444:444-54.

Reith W, LeibundGut-Landmann S, Waldburger JM. Regulation of MHC class II gene expression by the class II transactivator. Nat Rev Immunol. 2005 Oct;5(10):793-806.

Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PL, Le JM, Lee JS, Batliwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WV, Amos CI, Criswell LA, Seldin MF, Kastner DL, Gregersen PK. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med. 2007 Sep;357:977-86.

Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Engl J Med. 2002;347:1557-65.

Rigby AS, MacGregor AJ, Thomson G. HLA haplotype sharing in rheumatoid arthrtitis sibships: risk estimates subdivided by proband genotype. Gen Epidemiol. 1998;15:403-418.

Rigby AS, Silman AJ, Voelm L, Gregory JC, Ollier WE, Khan MA Nepom GT, Thomson G. Investigating the HLA component in rheumatoid arthritis: and additive

(dominant) mode of inheritance is rejected, a recessive mode is preferred. Gen Epidemiol. 1991;8:53-75.

Rigby AS, Voelm L, Silman AJ. Epistatic modeling in rheumatoid arthritis: an application of the Risch theory._Genet Epidemiol. 1993;10(5):311-20.

Risch N, Merikangas K.The future of genetic studies of complex human diseases. Science. 1996;13;273:1516-7.

Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science. 1995;268:1584-9.

Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet. 1990;46:222–228.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer._Am J Hum Genet. 2001;69:138-47.

Ritchlin C. Fibroblast biology. Effector signals released by the synovial fibroblast in arthritis. Arthritis Res. 2000;2:356-60.

Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Population-based re-sequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet. 2007;39:513-6.

Rueda B, Núñez C, Orozco G, López-Nevot MA, de la Concha EG, Martin J, Urcelay E. C1858T functional variant of PTPN22 gene is not associated with celiac disease genetic predisposition. Hum Immunol. 2005;66:848-52.

Sabatti C, Service S, Freimer N. False Discovery Rate in Linkage and Association Genome Screens for Complex Disorders. Genetics. 2003 Jun;164:829-833.

Sambrook J, Russell DW. Commonly Used Techniques in Molecular Cloning, Volume 3, 3rd edition. Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA. 2001.

Sanchez B, Moreno I, Magarino R, Garzon M, Gonzalez MF, Garcia A, Nunez-Roldan A.HLA-DRw10 confers the highest susceptibility to rheumatoid arthritis in a Spanish population. Tissue Antigens. 1990;36:174-6.

Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest. 1998;101:273-81.

Schellekens GA, Visser H, de Jong BA, van den Hoogen FH, Hazes JM, Breedveld FC, van Venrooij WJ. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrllinated peptide. Arthritis Rheum. 2000;43:155-163.

Schuierer M, Hilger-Eversheim K, Dobner T, Bosserhoff AK, Moser M, Turner J, Crossley M, Buettner R. Induction of AP-2α expression by adenoviral infection involved inactivation of AP-2rep trancritptional corepressor CtBP1. J Biol Chem. 2001 Jul;276:27944-9.

Seitz M, Wirthmüller U, Möller B, Villiger PM. The -308 tumour necrosis factor-alpha gene polymorphism predicts therapeutic response to TNFalpha-blockers in rheumatoid arthritis and spondyloarthritis patients. Rheumatology (Oxford). 2007;46:93-6.

Seltman H, Roeder K, Devlin B. Transmission/disequilibrium test meets measured haplotype analysis: family based association analysis guided by evolution of haplotypes. Am J Hum Genet. 2001;68: 1250-63.

Seyler TM, Park YW, Takemura S, Bram RJ, Kurtin PJ, Goronzy JJ, Weyand CM. BLys and APRIL in rheumatoid arthritis. J Clin Invest. 2005 Nov;115:3083-92.

Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA pooling: a tool for large-scale association studies. Nat Rev Genet. 2002 Nov;3:862-71.

Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann Hum Genet. 1995;59:97-105.

Sharfe N, Dadi HK, Shahar M, Roifman CM. Human immune disorder arising from mutation of the alpha chain of the interleukin-2 receptor. Proc Natl Acad Sci USA. 1997;94:3168-71.

Shiozawa S, Hayashi S, Tsukamoto Y, Goko H, Kawasaki H, Wada T, Shimizu K, Yasuda N, Kamatani N, Takasugi K, Tanaka Y, Shiozawa K, Imura S. Identification of the gene loci that predispose to rheumatoid arthritis. Int Immunol. 1998;10:1891-5.

Shiozawa S, Tanaka Y, Fujita T, Tokuhisa T. Destructive arthritis without lymphocyte infiltration in H2-c-fos transgenic mice. J Immunol. 1992;148:3100-4.

Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, Ollier WE. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. Br J Rheumatol. 1993;32:903-7.

Silman AJ, Ollier WE, Bubel MA. Autoimmune thyroid disease and thyroid autoantibodies in rheumatoid arthritis patients and their families. Br J Rheumatol. 1989; 28:18-21.

Simon M, Girbal E, Sebbag M, Gomès-Daudrix V, Vincent C, Salama G, Serre G.

References

The cytokeratin filament-aggregating protein filaggrin is the target of the so-called "antikeratin antibodies," autoantibodies specific for rheumatoid arthritis. J Clin Invest. 1993;92:1387-93.

Sinsheimer J, Palmer C, Woodward J. Detecting genotype combinations that increase risk for disease: maternal-foetal genotype incompatibility test. Genet Epidemiol. 2003;24:1-13.

Smyth D, Cooper JD, Collins JE, Heward JM, Franklyn JA, Howson JM, Vella A, Nutland S, Rance HE, Maier L, Barratt BJ, Guja C, Ionescu-Tîrgoviste C, Savage DA, Dunger DB, Widmer B, Strachan DP, Ring SM, Walker N, Clayton DG, Twells RC, Gough SC, Todd JA. Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. Diabetes. 2004;53:3020-3.

Snyder MR, Nannini C, Crowson C, Matteson EL, Homburger HA. The T allele of the PTPN22 1858C/T polymorphism is associated with markers of increased disease severity in Rheumatoid Arthritis. Abstract No 1052. Arthritis Rheum. 2007;56:S432.

Sokka TM, Kaarela K, Möttönen TT, Hannonen PJ Conventional monotherapy compared to a "sawtooth" treatment strategy in the radiographic progression of rheumatoid arthritis over the first eight years. Clin Exp Rheumatol. 1999;17:527-32.

Spencer SD, Dowbenko D, Cheng J, Li WL, Brush J, Utzig S, Simanis V, Lasky LA PSTPIP a tyrosine phosphorylated cleavage furrow-associated protein that is a substrate for a PEST tyrosine phosphatase, J. Cell Biol. 1997;138:845–860.

Stastny P. Mixed lymphocyte cultures in rheumatoid arthritis. J Clin Invest. 1976;57:1148-57.

Steer S, Lad B, Grumley JA, Kingsley GH, Fisher SA. Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: evidence for an early onset/disease severity effect. Arthritis Rheum. 2005;52(1):358-60.

Steimle V, Otten LA, Zufferey M, Mach B. Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). Cell. 1993;75:135-46.

Stephens M, Smith N, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001;68:978-89.

Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature. 1994;368:215-21.

References

Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC. Choosing haplotype tagging SNPs based on unphased genotype data using a preliminary sample of unrelated subjects with an example form the Multiethnic Cohort Study. Hum Hered. 2003;55:27-36.

Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, Nagasaki M, Nakayama-Hamada M, Kawaida R, Ono M, Ohtsuki M, Furukawa H, Yoshino S, Yukioka M, Tohma S, Matsubara T, Wakitani S, Teshima R, Nishioka Y, Sekine A, Iida A, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet. 2003;34:395-402.

Svendsen AJ, Holm NV, Kyvik K, Petersen PH, Junker P. Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. BMJ. 2002;324:264-6.

Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, Jagodic M, Lobell A, Khademi M, Börjesson O, Lindgren CM, Lundman P, Brookes AJ, Kere J, Luthman H, Alfredsson L, Hillert J, Klareskog L, Hamsten A, Piehl F, Olsson T. MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. Nat Genet. 2005;37:486-94.

Szabo SJ, KimST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A novel transcription factor, T-bet , directs Th1 lineage commitment. Cell. 2000;100:655-69.

Szkudlarek M, Narvestad E, Klarlund M, Court-Payen M, Thomsen HS, Ostergaard M. Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination. Arthritis Rheum. 2004;50:2103-12.

Takatori R, Takahashi KA, Tokunaga D, Hojo T, Fujioka M, Asano T, Hirata T, Kawahito Y, Satomi Y, Nishino H, Tanaka T, Hirota Y, Kubo T. ABCB1 C3435T polymorphism influences methotrexate sensitivity in rheumatoid arthritis patients. Clin Exp Rheumatol. 2006;24:546-54.

Takemura S, Braun A, Crowson C, Kurtin P J, Cofield RH, O'Fallon WM, Goronzy JJ, Weyand, CM. Lymphoid neogenesis in rheumatoid synovitis. J Immunol. 2001 Jul;167:1072-1080.

Taneja V, Giphart MJ, Verduijn W, Naipal A, Malaviya AN, Mehra NK. Polymorphism of HLA-DRB, -DQA1, and -DQB1 in rheumatoid arthritis in Asian Indians: association with DRB1*0405 and DRB1*1001. Hum Immunol. 1996;46(1):35-41.

Taniguchi A, Urano W, Tanaka E, Furihata S, Kamitsuji S, Inoue E, Yamanaka M, Yamanaka H, Kamatani N. 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. Pharmacogenet Genomics. 2007;17:383-90.

Thomas D, Clayton D. Betting odds and genetic associations. J Natl Cancer Inst. 2004;96:421-23.

Thomas E, Hay EM, Hajeer A, Silman AJ. Sjogren's syndrome: a community-based study of prevalence and impact. Br J Rheumatol. 1998;37:1069-76.

Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN, Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P; YEAR Consortium, Carter A, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Strachan D, Worthington J. Rheumatoid arthritis association at 6q23. Nat Genet. 2007;39:1431-3.

Thomson W, Pepper L, Payton A, Carthy D, Scott D, Ollier W, Silman A, Symmons D. Absence of an association between HLA-DRB1*04 and rheumatoid arthritis in newly diagnosed cases from the community. Ann Rheum Dis. 1993 ;52(7):539-41.

Timms AE, Crane AM, Sims AM, Cordell HJ, Bradbury LA, Abbott A, Coyne MR, Beynon O, Herzberg I, Duff GW, Calin A, Cardon LR, Wordsworth BP, Brown MA. The interleukin 1 gene cluster contains a major susceptibility locus for ankylosing spondylitis. Am J Hum Genet. 2004;75:587-95.

Ting JP, Trowsdale J. Genetic control of MHC class II expression. Cell. 2002;109 Suppl:S21-33.

Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet. 2003;35:341-8.

Tolusso B, Pietrapertosa D, Morelli A, de Santis M, Gremese E, Farina G, Carniello SG, del Frate M, Ferraccioli G. IL-1B and IL-1RN gene polymorphisms in rheumatoid arthritis: relationship with protein plasma levels and response to therapy. Pharmacogenomics. 2006 Jul;7:683-95.

Turesson C, Schaid DJ, Weyand CM, Jacobsson LT, Goronzy JJ, Petersson IF, Sturfelt G, Nyhall-Wahlin BM, Truedsson L, Dechant SA, Matteson EL . The impact of HLA-DRB1 genes on extra-articular disease manifestations in rheumatoid arthritis. Arthritis Res Ther. 2005;7:R1386-93.

Turesson C, O'Fallon WM, Crowson CS, Gabriel SE, Matteson EL.Extra-articular disease manifestations in rheumatoid arthritis: incidence trends and risk factors over 46 years. Ann Rheum Dis. 2003;62:722-7.

References

Turesson C, O'Fallon WM, Crowson CS, Gabriel SE, Matteson EL. Occurrence of extraarticular disease manifestations is associated with excess mortality in a community based cohort of patients with rheumatoid arthritis. J Rheumatol. 2002;29:62-7.

Turesson C, Weyand CM, Matteson EL Genetics of rheumatoid arthritis: Is there a pattern predicting extraarticular manifestations? Arthritis Rheum. 2004;51:853-63.
Turesson C, Jacobsson L, Bergstrom U. Extra-articular rheumatoid arthritis: prevalence and mortality. Rheumatology (Oxford) 1999;38:668-74.

Vallejo AN, Nestel AR, Schirmer M, Weyand CM, Goronzy JJ. Aging-related deficiency of CD28 expression in CD4+ T cells is associated with the loss of gene-specific nuclear factor binding activity. J Biol Chem. 1998;273:8119-29.

van Aken J, Schmedders M, Feuerstein G, Kollek R. Prospects and limits of pharmacogenetics: the thiopurine methyl transferase (TPMT) experience. Am J Pharmacogenomics. 2003;3:149-55.

van der Heijde DM, van Leeuwen MA, van Riel PL, van de Putte LB. Radiographic progression on radiographs of hands and feet during the first 3 years of rheumatoid arthritis measured according to Sharp's method (van der Heijde modification). J Rheumatol. 1995;22:1792-6.

van der Heijde DM, van Riel PL, van Rijswijk MH, van de Putte LB. Influence of prognostic features on the final outcome in rheumatoid arthritis: a review of the literature. Semin Arthritis Rheum. 1988;17:284-92.

van der Horst-Bruinsma IE, Hazes JM, Schreuder GM, Radstake TR, Barrera P, van de Putte LB, Mustamu D, van Schaardenburg D, Breedveld FC, de Vries RR. Influence of non-inherited maternal HLA-DR antigens on susceptibility to rheumatoid arthritis. Ann Rheum Dis. 1998;57:672-5.

Vandiedonck C, Capdevielle C, Giraud M, Krumeich S, Jais JP, Eymard B, Tranchant C, Gajdos P, Garchon HJ. Association of the PTPN22*R620W polymorphism with autoimmune myasthenia gravis. Ann Neurol. 2006;59:404-7.

van Ede AE, Laan RF, Blom HJ, Huizinga TW, Haagsma CJ, Giesendorf BA, de Boo TM, van de Putte LB. The C677T mutation in the methylenetetrahydrofolate reductase gene: a genetic risk factor for methotrexate-related elevation of liver enzymes in rheumatoid arthritis patients. Arthritis Rheum. 2001;44:2525-30.

van Eden W, Hogervorst EJ, Hensen EJ, van der Zee R, van Embden JD, Cohen IR. A cartilage mimicking T-cell epitope on a 65K mycobacterial heat shock protein: adjuvant arthritis as a model for human rheumatoid arthritis. Cur Top Microbiol Immunol. 1989;145:27-43.

References

Vang T, Miletic AV, Arimura Y, Tautz L, Rickert RC, Mustelin T. Protein tyrosine phosphatases in autoimmunity. Annu Rev Immunol. 2008;26:29-55.

Vang T, Congia M, Macis MD, Musumeci L, Orrú V, Zavattari P, Nika K, Tautz L, Taskén K, Cucca F, Mustelin T, Bottini N. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet. 2005;37:1317-9.

van Gaalen FA, van Aken J, Huizinga TW, Schreuder GM, Breedveld FC, Zanelli E, van Venrooij WJ, Verweij CL, Toes RE, de Vries RR. Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. Arthritis Rheum. 2004;50:2113-21.

van Oene M, Wintle RF, Liu X, Yazdanpanah M, Gu X, Newman B, Kwan A, Johnson B, Owen J, Greer W, Mosher D, Maksymowych W, Keystone E, Rubin LA, Amos CI, Siminovitch KA. Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. Arthritis Rheum. 2005 Jul;52:1993-8.

van Schaardenburg D, Hazes JM, de Boer A, Zwinderman AH, Meijers KA, Breedveld FC Outcome of rheumatoid arthritis in relation to age and rheumatoid factor at diagnosis. J Rheumatol. 1993;20:45-52.

Vawter MP, Atz ME, Rollins BL, Cooper-Casey KM, Shao L, Byerley WF. Genome scans and gene expression microarrays converge to identify gene regulatory loci relevant in schizophrenia. Hum Genet. 2006;119:558-70.

Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, Ball SG, James RA, Quinton R, Perros P, Pearce SH. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab. 2004 Nov;89:5862-5.

Viken MK, Amundsen SS, Kvien TK, Boberg KM, Gilboe IM, Lilleby V, Sollid LM, Forre OT, Thorsby E, Smerdel A, Lie BA. Association analysis of the 1858C>T polymorphism in the PTPN22 gene in juvenile idiopathic arthritis and other autoimmune diseases. Genes Immun. 2005;6(3):271-3.

Vlak T. Incidence of rheumatoid nodule in Dalmatia: similarities and differencies among populations. Arch Med Res. 2003;34:56-9.

Vyse TJ, Todd JA.Genetic analysis of autoimmune disease. Cell. 1996;85:311-8.

Waase I, Kayser C, Carlson PJ, Goronzy JJ, Weyand CM. Oligoclonal T cell proliferation in patients with rheumatoid arthritis and their unaffected siblings. Arthritis Rheum. 1996;39:904-13.

Wacholder S, Chanock S, Garcia-Closas M, El Ghormili L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst. 2004;96:434-42.

Wagenleiter SE, Klein W, Griga T, Schmiegel W, Epplen JT, Jagiello P. A case-control study of tyrosine phosphatase (PTPN22) confirms the lack of association with Crohn's disease. Int J Immunogenet. 2005;32:323-4.

Wagner UG, Koetz K, Weyand CM, Goronzy JJ.Perturbation of the T cell repertoire in rheumatoid arthritis. Proc Natl Acad Sci USA. 1998 Nov;95(24):14447-52.

Wajapeyee N, Somasundraram K Cell cycle arrest and apoptosis induction by activator protein 2α (AP-2α) and the role of p53 and p21WAF1/CIP1 in AP-2α –mediated growth inhibition. J Biol Chem. 2003;278:52093-101.

Wall J, Pritchard J. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet. 2003; 4: 587-97.

Wallin J, Hillert J, Olerup O, Carlsson B, Strom H. Association of rheumatoid arthritis with a dominant DR1/Dw4/Dw14 sequence motif, but not with T cell receptor β chain gene alleles or haplotypes. Arhtritis Rheum. 1991;34:1416-1424.

Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet. 2005;6:109-18.

Weinberg C, Wilcox A, Lie R. A log-linear approach to casepoarent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet. 1998;62:969-78.

Weisman MH, Furst DE, Park GS, Kremer JM, Smith KM, Wallace DJ, Caldwell JR, Dervieux T. Risk genotypes in folate-dependent enzymes and their association with methotrexate-related side effects in rheumatoid arthritis. Arthritis Rheum. 2006;54:607-12.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 hared controls. Nature. 2007 Jun;447:661-678.

Wertz IE, O'Rourke KM, Zhou H, Eby M, Aravind L, Seshagiri S, Wu P, Weismann C, Baker R, Boone DL, Ma A, Koonin EV, Dixit VM. De-ubiquitination and ubiquitin ligase domains of A20 downregulate ND-kappaB signalling. Nature. 2004 Aug;430:694-9.

Wesoly J, Hu X, Thabet MM, Chang M, Uh H, Allaart CF, Toes RE, Houwing-Duistermaat JJ, Begovich AB, Huizinga TW.The 620W allele is the PTPN22 genetic

variant conferring susceptibility to RA in a Dutch population. Rheumatology (Oxford). 2007;46:617-21.

Wessels JA, de Vries-Bouwstra JK, Heijmans BT, Slagboom PE, Goekoop-Ruiterman YP, Allaart CF, Kerstens PJ, van Zeben D, Breedveld FC, Dijkmans BA, Huizinga TW, Guchelaar HJ. Efficacy and toxicity of methotrexate in early rheumatoid arthritis are associated with single-nucleotide polymorphisms in genes coding for folate pathway enzymes. Arthritis Rheum. 2006;54:1087-95.

Weyand CM, Goronzy JJ. Ectopic germinal center formation in rheumatoid synovitis. Ann N Y Acad Sci. 2003;987:140-9.

Weyand CM, Goronzy JJ. HLA-DRB1 alleles as severity markers in RA. Bull Rheum Dis. 1994;43:5-8.

Weyand CM, Hicok KC, Conn DL, Goronzy JJ. The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. Ann Intern Med. 1992;117:801-806.

Weyand CM, Schmidt D, Wagner U, Goronzy JJ. The influence of sex on the phenotype of rheumatoid arthritis. Arthritis Rheum 1988;41:817-22.

Willkens RF, Nepom GT, Marks CR, Nettles JW, Nepom BS. Association of HLA-Dw16 with rheumatoid arthritis in Yakima Indians. Further evidence for the "shared epitope" hypothesis. Arthritis Rheum.1991;34:43-7.

Wolf CR, Smith G. Pharmacogenetics. Br Med Bull. 1999;55(2):366-86.

Wooley PH, Luthra HS, Stuart JM, David CS. Type II collagen induced arthritis in mice. Major histocompatibility complex linkage and antibody correlates. J Exp Med. 1981;154:688-700.

Wordsworth BP, Bell JI. The immunogenetics of rheumatoid arthritis. Springer Semin Immunopathol. 1992;14:59–78.

Wordsworth BP, Lanchbury JS, Sakkas LI, Welsh KI, Panayi GS, Bell JI.(1989) HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. Proc Natl Acad Sci USA. 1989;86:10049-53.

Wu H, Cantor RM, Graham DS, Lingren CM, Farwell L, Jager PL, Bottini N, Grossman JM, Wallace DJ, Hahn BH, Julkunen H, Hebert LA, Rovin BH, Birmingham DJ, Rioux JD, Yu CY, Kere J, Vyse TJ, Tsao BP. Association analysis of the R620W polymorphism of protein tyrosine phosphatase PTPN22 in systemic lupus erythematosus families: increased T allele frequency in systemic lupus erythematosus patients with autoimmune thyroid disease. Arthritis Rheum. 2005 Aug;52:2396-402.

Yamanishi Y, Boyle DL, Green DR, Keystone EC, Connor A, Zollman S. p53 tumor suppressor gene mutations in fibroblast-like synoviocytes from erosion synovium and non-erosion synovium in rheumatoid arthritis. Arthritis Res Ther. 2005;7:R12-8.

Yan H, Yuan W, Velculescu VE, Vogelstein B. and Kinzler KW Allelic variation in human gene expression. Science. 2002;297:1143.

Yazdani-Biuki B, Brickmann K, Wohlfahrt K, Mueller T, März W, Renner W, Gutjahr M, Langsenlehner U, Krippl P, Wascher TC, Paulweber B, Graninger W, Brezinschek HP. The MHC2TA -168A>G gene polymorphism is not associated with rheumatoid arthritis in Austrian patients. Arthritis Res Ther. 2006;8:R97.

Zeggini E, Rayner W, Morris AP, Hattersley AT, Walker M, Hitman GA, Deloukas P, Cardon LR, McCarthy MI. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. Nat Genet. 2005;37:1320-2.

Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, Lu Y, Denoeud F, Antonarakis SE, Snyder M, Ruan Y, Wei CL, Gingeras TR, Guigó R, Harrow J, Gerstein MB. Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res. 2007 Jun;17:839-851.

Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes Immun. 2005;6:459-61.

Zvaifler NJ The immunopathology of joint inflammation in rheumatoid arthritis. Adv Immunol 1973;16:265-336.

Appendices

Appendix A

Nuffield NHS
Orthopaedic Centre
NHS Trust
Windmill Rd
Oxford
OX3 7LD

Professor B. Paul Wordsworth
Professor of Rheumatology
Claire Farrar
Research nurse
Dr. Pille Harrison
DPhil student / Clinical research fellow
Tel: 01865 227982
E-mail: pille.harrison@ndos.ox.ac.uk

To participate in this research please read and sign the form below:

Genetics and Clinical Determinants of Rheumatoid Arthritis
(OxREC Number C02.032)

Researchers:

- Professor B. Paul Wordsworth
- Dr. Pille Harrison
- Claire Farrar

Please initial box

1.I confirm that I have read the information sheet, dated 23/02/05 (version 3) about the above study. I understand why the research is being done and any possible risks involved ☐

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

3. I give permission for my medical records to be looked at and information taken From them to be analysed in strict confidence by responsible people from the Oxford ☐

research team.

4. I consent to my GP being informed of my participation in this study
.

5. I understand that the results of the research will be periodically published in the Arthritis Research Campaign newsletter and on the Oxford group website.

6. I understand that I will not benefit financially if this research leads to the development of a new treatment or medical test.

8. I understand how to contact the research team if need to.

9. I have kept and signed a copy of the consent form.

10. I consent to participate in this study.

Name of Participant …… Date……Signature
……

Name of Researcher …….. Date ……..Signature
……

OXAGEN INFORMATION

Individual number

Oxagen number

Bloods for Oxagen ○ Yes ○ No

Date sent to Oxagen

DNA received back ○ Yes ○ No

DNA concentration

Date of Birth

Gender

ACR Criteria:

1. BMS > 1 HOUR ○ Yes ○ No

2 SIMULTANEOUS ARTHRITIS OF 3 OR MORE JOINT AREAS (PIP, MCP, WRISTS, ELBOWS, KNEES, ANKLES AND MTP) ○ Yes ○ No

3 ARTHRITIS OF HAND JOINTS ○ Yes ○ No

4 SYMMETRIC ARTHRITIS ○ Yes ○ No

5 RHEUMATOID NODULES ○ Yes ○ No

6 RHEUMATOID FACTOR ○ Yes ○ No

7 XRAY CHANGES ○ Yes ○ No

-> At least 4 of these criteria must be present
-> Criteria 1-4 must be present for at least 6 weeks

Ethnicity

Maternal Grand Father	○ English	○ Irish	○ Scottish	○ Welsh	○ Northern European	○ Other Caucasian
Maternal Grand Mother	○ English	○ Irish	○ Scottish	○ Welsh	○ Northern European	○ Other Caucasian
Paternal Grand Father	○ English	○ Irish	○ Scottish	○ Welsh	○ Northern European	○ Other Caucasian
Paternal Grand Mother	○ English	○ Irish	○ Scottish	○ Welsh	○ Northern European	○ Other Caucasian

CLINICAL INFORMATION

Individual number

Oxagen number

Age of Disease Onset

Duration of disease

Nodules ever ○ Yes ○ No

Type 1 Diabetes ○ Yes ○ No

Thyroid disease ○ Yes ○ No

Primary biliary cirrhosis ○ Yes ○ No

Coeliac disease ○ Yes ○ No

IBD ○ Yes ○ No

Vitiligo ○ Yes ○ No

Autoimmune hepatitis ○ Yes ○ No

Pernicious anaemia ○ Yes ○ No

Sjogrens ○ Yes ○ No

Vasculitis ○ Yes ○ No

Joint replacement surgery ○ Yes ○ No

Type of surgery

Age at surgery

FAMILY HISTORY

MOTHER ALIVE ○ Yes ○ No

FATHER ALIVE ○ Yes ○ No

FAM HX-RA ○ Yes ○ No

DRUG THERAPY

Individual number

Oxagen number

NSAIDS

NSAIDS SE'S

Current RA drug treatment

CONTACT SHEET

Date recruited

Individual number

Oxagen number

Hospital number

Title

First Name

Surname

Address 1

Address 2

Address 3

Address 4

Postcode

Telephone Home

Date of Birth

GP Name

GP Address 1

GP Address 2

GP Address 3

GP Address 4

GP Postcode

Comments

Appenidx B

Table B.1 *DRB1* allele counts in RA patients.

HLA-DRB1 type	All patients N=756 (%)	Female RA N= 542 (%)	Male RA N= 214 (%)
*01	186 (12)	128 (12)	58 (13.5)
*03, 08, 11, 13,14	328 (22)	238 (22)	90 (21)
*04	633 (42)	444 (41)	189 (44)
*07	133 (9)	104 (10)	29 (8)
*09	30 (2)	24 (2)	6 (1)
*10	14 (1)	10 (1)	4 (1)
*12	20 (1)	18 (2)	2 (0.5)
*15, 16	165 (11)	116 (10)	49 (11)
	All patients N= 633 (%)	Female RA N=444 (%)	Male RA N=189 (%)
*0401	417 (66)	300 (68)	117 (62)
*0404, *0408	168 (27)	111 (25)	57 (30)
*0405	18 (3)	14 (3.5)	4 (2)
*0406,*0407, *0402	26 (4)	15 (3.5)	11 (6)

Table B.2 The mean percentile severity score (and range) in RA patients with and without each *HLA-DRB1* marker. P values were derived from a Mann-Whitney U test.

Markers	No	Mean (range) Present	Absent	P value
DRB1*01	47	52 (2-98)	50 (0-99)	0.77
DRB1*03, *08, *11, *13, *14	88	50 (2-98)	50 (0-99)	0.88
DRB1*04	135	50 (0-100)	51(2-98)	0.75
DRB1*07	26	60 (3-99)	49(0-99)	0.05
DRB1*09	7	38(8-73)	50(0-99)	0.23
DRB1*10	5	58 (2-95)	50(0-99)	0.54
DRB1*12	5	53(13-94)	50(0-99)	0.88
DRB1*15, *16	41	47(3-94)	51(0-99)	0.51

Table B.3 Radiographic severity scores, mean and range in *HLA-DRB1*04* homozygotes, heterozygotes and non *DRB1*04* subjects.

Table B.3 Radiographic severity scores, mean and range in *HLA-DRB1*04* homozygotes, heterozygotes and non *DRB1*04* subjects.

	Mean	Range	No	P value
DR4/DR4	50.62	(2-99)	28	
DR4/DRX	50.49	(2-99)	107	0.55
DRX/DRX	51.28	(2-98)	71	

Table B.4 Number of subjects having selected MHC variants with mild or severe disease P values are Fisher's exact probabilities.

Marker	Mild	Severe	P value
DRB1*01	10	10	1.0
DRB1*03, *08, *11, *13, *14	19	14	0.50
DRB1*04	27	25	1.0
DRB1*07	5	11	0.1
DRB1*15, *16	6	6	1.0

Table B.5 Comparison of seven hospital based RA cohorts used in GWAS validation study. The Oxford cohort is presented in bold. Addabted from Thomson *et al.* Nat.Genet.2007;39:1431-3

Cohort	**Total N**	**Manchester N**	**Aberdeen N**	**Leeds N**	**Sheffield N**	**London N**	**Oxford N**	**WTCCC N**
Controls	3531	832	825	422	929	-	**523**	2938
RA Cases	3962	1051	131	844	927	276	**733**	1860
Age at onset	48.8±16.0	47.6±17.2	47.9±12.2	58.9±14.0	43.9±14.9	47.6±13.9	**48.2±14.4**	46.3±14.6
Case Character	**Total N (%)**	**Manchester N (%)**	**Aberdeen N (%)**	**Leeds N (%)**	**Sheffield N (%)**	**London N (%)**	**Oxford N (%)**	**WTCCC N (%)**
Gender								
M	1106 (28.2)	286 (27.2)	40 (30.5)	270 (32.0)	248 (27.5)	55 (20.6)	**207 (28.2)**	468 (25.2)
F	2822 (71.8)	765 (72.8)	91 (69.5)	574 (68.0)	654 (72.5)	212 (79.4)	**526 (71.8)**	1392 (74.8)
RF								
Negative	1039 (27.9)	280 (26.9)	30 (24.0)	243 (33.6)	226 (27.1)	96 (35.9)	**164 (22.4)**	285 (16.3)
Positive	2684 (72.1)	762 (73.1)	95 (76.0)	480 (66.4)	607 (72.9)	171 (64.1)	**569 (77.6)**	1461 (83.7)
Anti-CCP								
Negative	731 (32.0)	364 (37.8)	-	167 (38.7)	200 (22.6)	-	-	271 (20.6)
Positive	1550 (67.9)	599 (62.2)	-	265 (61.3)	686 (77.4)	-	-	1042 (79.4)
SE(no.copies)								
0	766 (25.1)	256 (25.6)	-	96 (27.1)	186 (21.2)	58 (22.0)	**170 (31.0)**	364 (20.2)
1	1506 (49.5)	521 (52.0)	-	184 (52.0)	408 (46.5)	129 (48.9)	**264 (48.1)**	893 (49.5)
2	773 (25.4)	224 (22.4)	-	74 (20.9)	283 (32.3)	77 (29.1)	**115 (20.9)**	547 (30.3)

Appendix C

Table C.1 Primer sequences for *PTPN22* C1858T (rs2476601) polymorphism genotyping.

Primer sequences	Mg (mM)	$T_mC°$	Ensyme	Size
For 5'CATGCTATTGCTCTGCT Rev 5'ATGTTGCTTCAACGGAATTT	2.0	56	*Xcm*I Cuts T	241/171 bp

The PCR conditions were as follows: 95°C - 10min, 35 × (95°C - 30s, 56°C - 30s, 72° - 30s), 72°C - 10min. *Xcm*I digestion products were resolved on 3% agarose gel stained with ethidium bromide and visualised under ultraviolet light.

Figure C.1. Agarose gel view of RFLP genotyping of *PTPN22* C1858T.

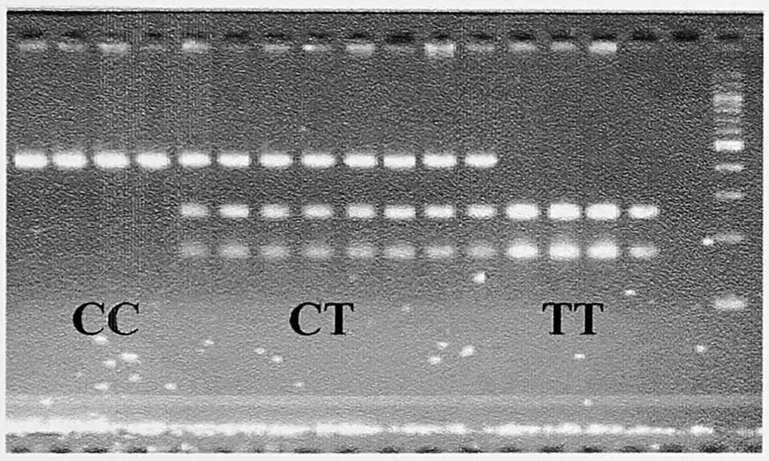


Figure C.2 An example of the electropherogram of *PTPN22* C1858T polymorphism.



Appendix D

Figure D.1 Overview of 138 kb genomic region on 1p13 demonstrating overlapping of *HIPK1-203* transcript with *PTPN22*, *AP4B1* and *DCLRE1B*.

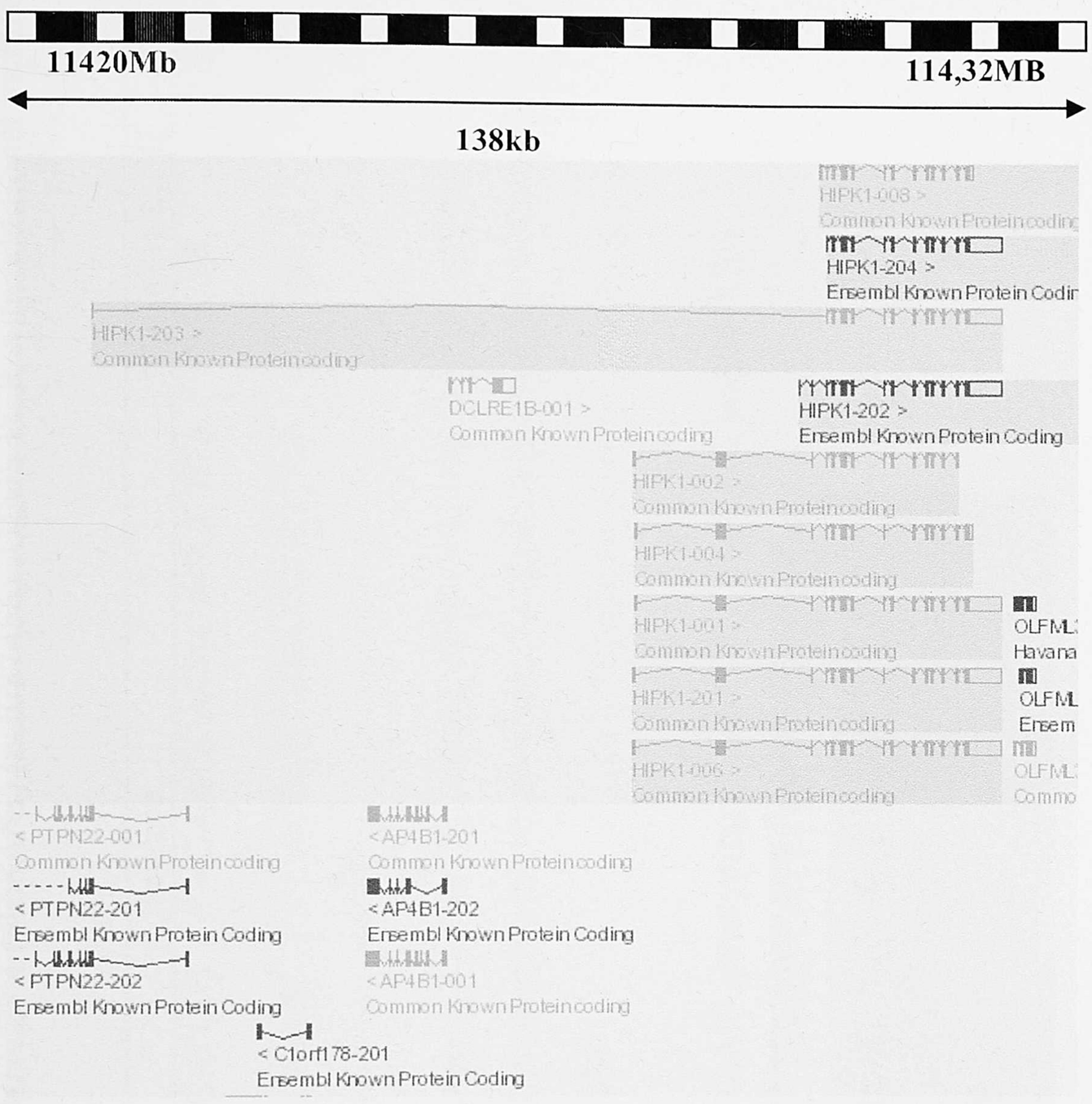


Table E.1 Primer sequences for *MHC2TA* (-168A/G) (rs3087456) ARMS PCR with additional mismatch underlined in bold.

Allele	Forward	Reverse	Mg(mM) T_m°C	Size
A	5'GTGAAATTAATTTCAGAG GTG**G**A	5'AGAAGCACACAGCCTCA TCA	1.25 58	213
G	5'ATGATACTGGCCCCATCC TG	5'CACTCCCTTAAGCCCTC C**A**C	1.5 60	204

The PCR conditions were: 95°C 10min, 35 x (95°C – 30sec, 58°C 30sec for A allele and 60°C for G allele, 72°C 30sec), 72°C – 10 min. PCR products were resolved on 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light.

Figure E.1 Agarose gel view of ARMS genotyping of *MHC2TA* -168*A/G (rs3087456)

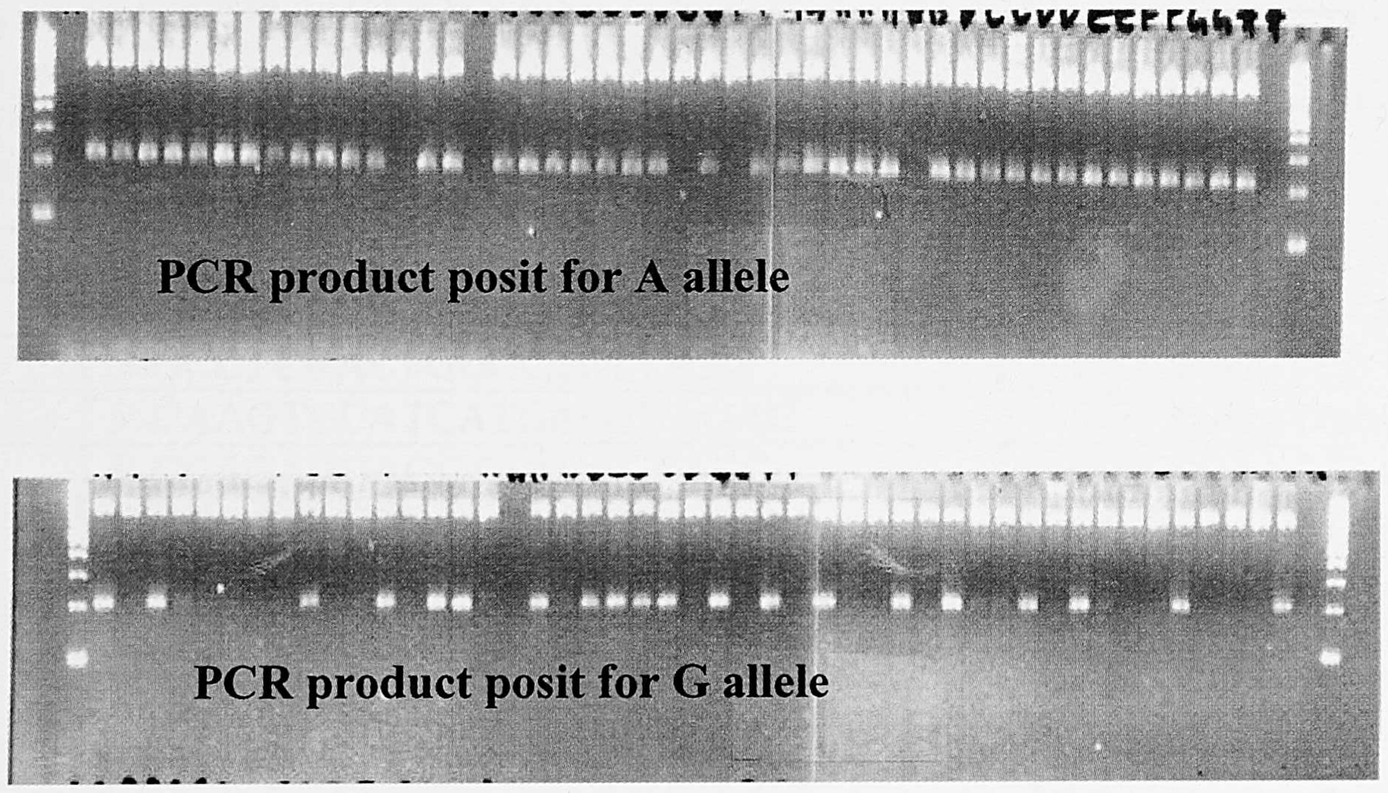


Table E.2 Primer sequences for *MHC2TA* sequencing * Exon 11 present in CIITA-001, CIITA-002 and CIITA-004 transcripts. **Exon 18 in CIITA-003 transcripts, exon 20 in CIITA-001 and CIITA-002 transcripts

The PCR conditions were as follows: 95°C - 10min, 35 × (95°C - 30s, 60°C - 30s, 72° - 30s), 72°C - 10min

Exon	Forward	Reverse
1	5'-AGTGTGGTAAAATTAGAGGG	5'CACAGACTCCTGTTCCCATC
2	5'-AGAGGTGCTGAATGAGYGCT	5'-CAGAGAGCGTCCCACAGAC
3	5'-GTCTGTGGGACGCTCTCTG	5'-TCCGCATCCCCCTTGCAATG
4	5'-CTGGAGTGTCAGTGTTCATTC	5'-GACCAACTCACACAATCAATG
5	5'-GTGCCCTAATACCTGACGAC	5'-GCTTTTAAAGGGCCTCCCAAC
6	5'-GGTTGGAATGTTAGAGCGAG	5'-TCTACCCAAAGAAGGGCAGG
7	5'-CAGCCTTATTCGTTCATCAGC	5'-GACCACCTCAGGAATCTCTG
8	5'-CAAGGTTCCAGGAGCTAGAT	5'-GATTCTCTAGCTGTCCATTC
9	5'-GAGTTAGATACAGCCTCAGG	5'-AGTGTTAACACCCAGTGCAG
11*	5'-GAGGTGGAGTCAGGAGGATC	5'-TGAGCCTCAACATCCTCATC
11*	5'-GAGGTGGAGTCAGGAGGATC	5-GATCCTGCAGGCCATAGGCATC
11*	5'-GATCTGGTGCAGGCCAGGCT	5'-CATGGAGAAGCCGGACAGCTC
11*	5'-GATGCCTATGGCCTGCAGGATC	5'-CTGGTCCTCCTGTAGGGTACT
11*	5'-GAGCTGTCCGGCTTCTCCATG	5'-AGGCTGGAAGATCAGCCCAG
11*	5'-AGTACCCTACAGGAGGACCA	5'-ACCTGAAACGGGTGACACAG
11*	5'-CTGGGCTGATCTTCCAGCCT	5'-TGAGCCTCAACATCCTCATC
10	5'-ACCACTGTGTGAGTTGTGAG	5'-CGAGTGCAGTGAGAGACTCT
11	5'-GCACCGCACCCAGACWGATC	5'-GTTAASCAGCAAACGGGTCCA
12	5'-GCCTCAACTGGCCTCTTATG	5'-TGTGGCGTCATCTCATCACT
13	5'-CAAGTCCATCATGAGTTAGAC	5'-GGTTGGATGAGGGARTAAAG
14	5'-CCACAGGACCTAACAGTATC	5'-GTGACAAGTGCTTACATAGAC
15-16	5'-GGCCTATGTAAGTACTTGGC	5'-GGTGGATGTCAGCAATTTCAT
17	5'-CAGAACCATAAAGGAATCTCG	5'-GATCTCTAAGCCTGACCACA
18**	5'-TGCTTGACAAGTCTCCTGCT	5'-CTATTCCGTAGTGGCACTGAG
18**	5'-TGCTTGACAAGTCTCCTGCT	5'-TGCCTCTGTCACTCCCTCAGG
18**	5'-TCCAGGCTGTATCCCATGAG	5'-GCCAGACATGATGGTGGGTG
18**	5'-CCAGAGGGAGTGACAGAGGCA	5'-GGTTGCCCATCAGTGTCTTC
18**	5'-CACCCACCATCATGTCTGGC	5'-CACTATGATTTGGGGTGGCT
18**	5'-GAAGACACTGATGGGCAACC	5'-CTATTCCGTAGTGGCACTGAG

Figure E.1 Electropherogram showing a G→ T transition, presenting an intronic SNP rs6498130.

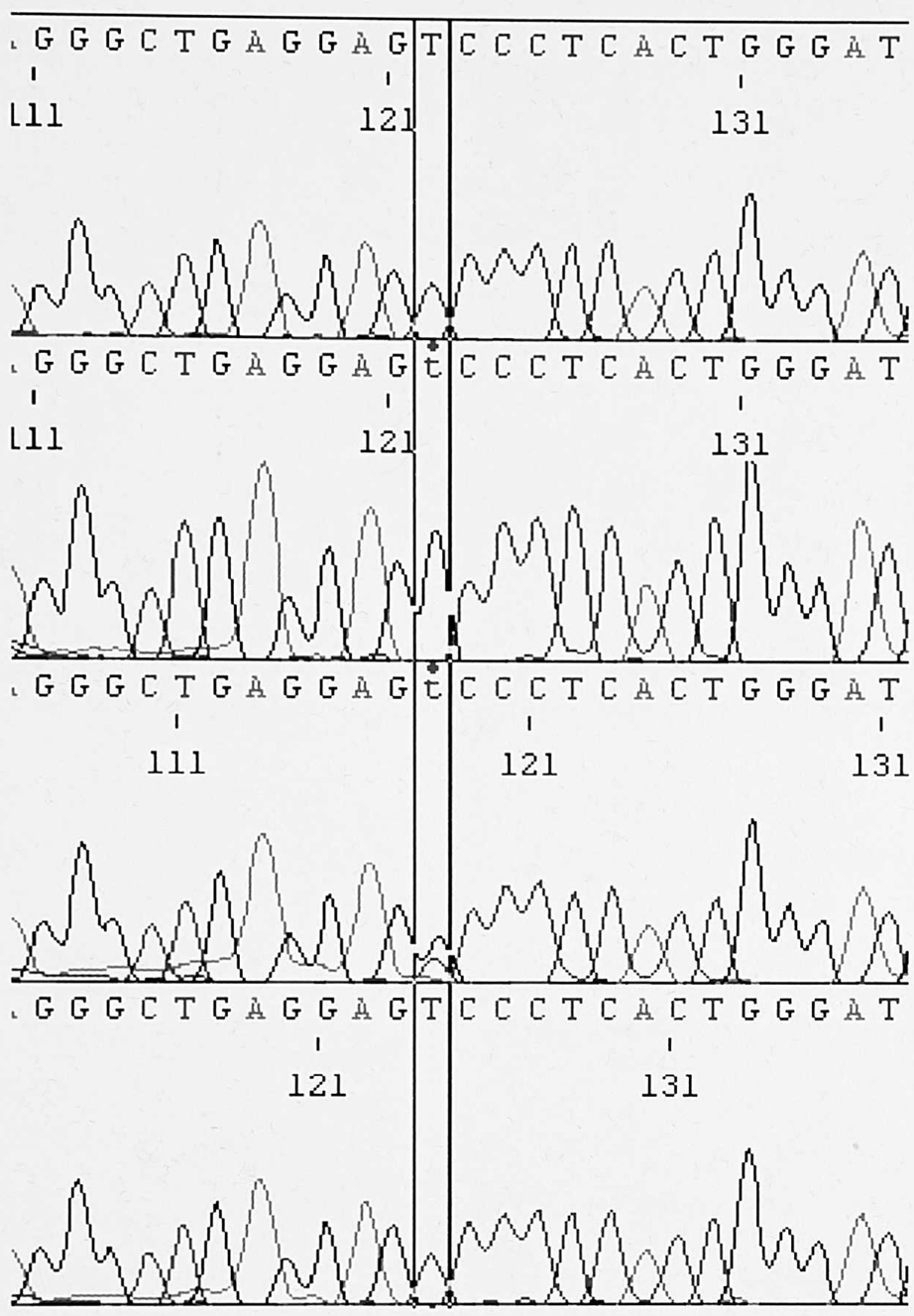


Table E.3 Primer sequences and PCR conditions for additional *MHC2TA* genotyping (additional mismatch is in bold).

SNP	dbSNP ID	Type	Primers	Mg	T_m °C	Method:	Product size
2	rs12928665*	A/G	For 5'- GAAACCATTCTGAATTGGGGATG Rev 5'- CATGCTGAGAGTTAGTGACACA CTGCCCGAAAGTCTGACTTCTAGAA**G**CT	2.5	60	RFLP Hind III	Cuts A 192/50 bp
3	rs8043545*	G/C	For 5'- GCTCCATGCAACTTAACTCC Rev G 5'- GGGAAATCAGGTGTCGCC**C**C Rev C 5' - GGGAAATCAGGTGTCGCC**C**G	1.5	60	ARMS	298 bp
4	rs6498119*	T/C	For C 5'-GACCCAGAAAGCTATCTG**G**C For T 5'-GACCCAGAAAGCTATCTG**G**T Rev 5'-CTACTACCCAGGAATCGATG	2.5	60	ARMS	319 bp
5	rs6498121	T/G	For 5'-CTGGAGTGTCAGTGTTCATTC Rev 5'-GCTTTTAAAGGGCCTCCCAAC	1.5	60	RFLP Rsa I	Cuts T 356/274 bp
6	rs4781016	C/A	For 5'-CTTGGGAGGTGGATAAGGAG Rev 5'-CTTGCATAGCACCTGTGTTTG	2.0	60	RFLP Dpn II	Cuts A 320/85 bp
7	rs45545642	C/T	For 5'- TTCATTCACTTATTTGATTATGCC ATCATTTCACCATTCACTCCAC**G**ACT Rev 5- TATTGGGATTGTGTGAGATG	2.0	60	RFLP Hinf I	Cuts C 263/50 bp
8	rs4774	G/C	For 5'-GATCTGGTGCAGGCCAGGCT Rev G 5' -CGCTTCCAGCTCCTCGAA**T**C Rev C 5'- CGCTTCCAGCTCCTCGAA**T**G	1.0	60	ARMS	429 bp
9	rs13330686	C/T	For 5'- GCCTCCCGCCCGCTGCCTGGG AGCCCTACTCGGGCCATCGGCGGCTG**T**CT Rev 5'- GAGCTCCTGTACCACGTGCTG	1.5	60	RFLP BsmA I	Cuts C 145/50 bp
10	rs2229320	G/A	For 5'-AGTACCCTACAGGAGGACCA Rev 5'-GAGCTCCTGTACCACGTGCTG	1.5	60	RFLP Msp I	Cuts G 377/93 bp
11	rs2228238	G/A	For 5'-ACCACTGTGTGAGTTGTGAG Rev 5'-CGAGTGCAGTGAGAGACTCT	1.5	65	RFLP TspRI	Cuts A 303/210 bp

12	rs45481191	C/T	For 5'-ACCACTGTGTGAGTTGTGAG Rev C 5'-GCCCAGAAAGAGGGCATG**T**G Rev T 5'- GCCCAGAAAGAGGGCATG**T**A	1.5	60	ARMS	483 bp
13	rs4781019	C/T	For 5'-GCACCGCACCCAGACWGATC Rev 5'- GTTAASCAGCAAACGGGTCCA	1.5	60	RFLP Fsp I	Cuts C 253/134 bp
14	rs4781020*	G/A	For 5'- CTTTAGCAGGGAGCCAATG Rev 5'- GCCAGCAGTGGGTATTTTGT	1.5	60	RFLP Taq$^{\alpha}$I	Cuts G 120/88 bp
15	rs7500869	C/A	For 5'- TCAACATCTGGGTGCAGTGCT GTGATCATAGCTCACTGCAGCCTCG**G**ATT Rev 5'- TGTGGCGTCATCTCATCACT	2.0	56	RFLP Hinf I	Cuts C 215/50 bp
16	rs6498130	G/T	For 5'-CTGCTCTCAGCAATTCAGCTG Rev 5'-CC**A**TCAGCTAGCGTCCTGGGG CCATCTGACCCTTCCCATCCCAGTGTGGG	2.5	60	RFLP Hinf I	Cuts T 268/50 bp
17	rs9788916*	G/A	For 5' – CGCCCAGTTCAACATGATTT Rev 5'- GGGACGGAAGGACTAAGGAG	1.5	60	RFLP Msp I	Cuts G 213/100 bp
18	rs7201430	C/T	For 5'-GGCCTATGTAAGTACTTGGC Rev C 5'-AGCTCTCGGCTCCACGT**A**G Rev T 5'-AGCTCTCGGCTCCACGT**A**A	2.0	56	ARMS	214 bp
19	rs4781024	G/A	For A 5'- GATGGCCAAATAATTTAC**A**A For G 5'- GATGGCCAAATAATTTAC**A**G Rev 5'- GATCTCTAAGCCTGACCACA	1.5	56	ARMS	363 bp
20	rs11074940	C/T	For C 5- CCGGGCCACAGAGAAAGT**C**C For T 5'- CCGGGCCACAGAGAAAGT**C**T Rev 5'- GCTGTTTCCCTGCTAAGGGC	1.5	60	ARMS	237 bp
21	rs34538398	G/A	For 5'- CTCAAACTCTTGACCTCAGG Rev 5'- GAGTCCCATAGACAAAAGCAG	1.5	60	RFLP BstN I	Cuts A 120/79 bp
22	rs1139564	T/C	For C 5'-CTTAGCAGGGAAACAGCTA**C**C For T 5'- CTTAGCAGGGAAACAGCTA**C**T ComRev. 5'-CACTATGATTTGGGGTGGCT	1.5	60	ARMS	218 bp

RFLP- restriction fragment length polymorphism, ARMS- amplification-refractory mutation system

Appendix F

Table F.1 Primer sequences and PCR conditions for *IL-1A* (-889 C/T, rs17561), *IL-1B* (-511 A/G, rs16944), *IL-1B* (-1464 C/G, rs1143623), and *IL-1B* (3737 C/T, rs4848306) polymorphisms genotyping.

SNP	SNP ID	Primer sequence	Mg C° (mM)	T_m	Ens	Size
IL-1A -889C/A*	Rs 1800587	For 5'-GGAGATTGGGAGAAAGGAAG GCATGGATTTTTACATATGAGCCTTC**C**ATG Rev 5'- CTGCCTCAGCCTCCTGAATAG	1.5	60	Nco I cutsG	208/50 bp
IL-1B -511 A/G	Rs 16944	For 5'-GTTTAGGAATCTTCCCACTT Rev 5'- TGGCATTGATCTGGTTCATC	1.5	60	Ava I cuts G	192/113 bp
IL-1B -1464 C/G	Rs 1143623	For 5'-ATGGGTGAATGGGAATTTGAC Rev 5'- AAATCAGAAGGCTGCTTGGA	1.5	60	BsaJI cuts C	106/72 bp
IL-1B -3737 G/A	Rs 4848306	For A 5'-GTACCTGAAGCGGCTGCAGC**A**A Rev A 5'- CAGTACATGTGCATCAGGAG For G 5'- GTACCCACACCTGTCTTTGT Rev G 5'-CAGGGAGGAGAATGGAATG**G**C	1.5 2.0	60 58		220 bp 200 bp

*Due to technical reasons rs1800587 was genotyped instead of IL-1A (-889) rs17561. Two markers are in complete LD (r^2=1). Primer sequences for RFLP and ARMS PCR with additional mismatch indicated in bold.

For the RFLP, the PCR conditions were as follows: 95°C 10min, 35 × (95°C 30s, 60°C 30s, 72° 30s), 72°C 10min.
For the ARMS, the PCR conditions were: 95°C 10min, 35 x (95°C 30s, 60°C 30s for A allele and 56°C for G allele, 72°C 30s), 72°C 10 min.

PCR products were resolved on 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light.

Table F.2 Allele and genotyping counts for *IL-1A, IL-1B*(-511), *IL-1B*(-1464) and *IL-1B*(-3737) in RA patients stratified by HLA-DRB1 SE, HLA-DRB1*04 subtypes, RF and clinical parameters.

IL-1A	1	2	P value, OR 95%CI	11	12	22	P value
SE present	786	346	P=0.67 OR=0.94	266	254	46	P=0.32
SE absent	245	101	95%CI 0.65-1.05	89	67	17	
DRB1*04 pos	658	288	P=0.87	223	212	38	
DRB1*04 neg	373	159	OR=0.97, 95%CI 0.77-1.24	132	109	25	P=0.55
RF present	800	342	P=0.67 OR=1.07	275	250	46	P=0.80
RF absent	214	98	95%CI 0.81-1.41	73	68	15	
Nodules present	271	121	P=0.69	88	95	13	
Nodules absent	739	311	OR=0.94 95%CI 0.73-1.22	260	219	46	P=0.24
JR present	231	91	P=0.46	82	67	12	
JR absent	773	341	OR=1.12, 95%CI 0.85-1.50	263	247	47	P=0.72

IL-1B (-511)	1	2	P value, OR 95%CI	11	12	22	P value
SE present	794	346	P=0.33 OR=1.15	272	250	48	P=0.54
SE absent	228	114	95%CI 0.88-1.50	75	78	18	
DRB1*04 pos	667	287	P=0.31	230	207	40	
DRB1*04 neg	355	173	OR=1.13, 95%CI 0.89-1.43	117	121	26	P=0.54
RF present	786	358	P = 0.94 OR=1.01	260	266	46	P=0.06
RF absent	215	99	95%CI 0.77-1.32	78	59	20	
Nodules present	263	121	P=0.93	86	91	15	
Nodules absent	733	331	OR= 0.98 95%CI 0.76-1.27	250	233	49	P=0.67
JR present	212	110	P=0.21 OR=0.84	69	74	18	P=0.34
JR absent	781	339	95%CI 0.64-1.10	266	249	45	

IL-1B-(1464)	1	2	P value, OR 95%CI	11	12	22	P value
SE present	862	274	P=0.60 OR=1.09	329	204	35	P=0.49
SE absent	254	88	95%CI 0.82-1.45	92	70	9	
DRB1*04 pos	723	229	P=0.64 OR=1.07	273	177	26	P=0.76
DRB1*04 neg	393	133	95%CI 0.83-1.38	148	97	18	
RF present	860	286	P=0.70 OR=0.93	321	218	34	P=0.69
RF absent	235	73	95%CI 0.69-1.26	91	53	10	
Nodules present	302	92	P=0.53	114	74	9	
Nodules absent	786	264	OR=1.10 95%CI 0.83-1.46	295	196	34	P=0.66
JR present	240	84	P=0.60 OR=0.92	86	68	8	P=0.42
JR absent	842	270	95%CI 0.68-1.23	320	202	34	

IL-1B (-3737)	1	2	P value, OR 95%CI	11	12	22	P value
SE present	628	530	P=0.73 OR=0.95	157	282	124	P=0.41
SE absent	182	146	95%CI 0.74-1.29	54	74	36	
DRB1*04 pos	497	449	P=0.31 OR=0.89	130	237	106	P=0.45
DRB1*04 neg	281	227	95%CI 0.72-1.11	81	119	54	
RF present	640	526	P>0.99 OR1.00	160	286	120	P=0.22
RF absent	164	136	95%, CI 0.76-1.31	50	64	36	
Nodules present	212	176	P=0.60	56	100	38	
Nodules absent	545	485	OR=1.07 95%CI 0.84-1.36	149	247	119	P=0.57
JR present	172	140	P=0.53 OR=1.09	50	72	34	P=0.57
JR absent	583	519	95%CI 0.84-1.42	154	275	122	

Table F.3 Most frequent extended *IL1A-IL1B* four marker haplotypes and their association with 756 RA patients and 625 healthy controls generated by PHASE v 2.

	Haplotype structure	No. of RA	No. of HC	OR95%CIp
2	**1112**	572	443	1.11 (0.95-1.30) 0.19
7	**1221**	341	316	0.86 (0.72-1.00) 0.09
9	**2111**	274	226	1.00 (0.82-1.22) 0.97
10	**2112**	127	99	1.06 (0.81-1.41) 0.65
Remaining haplotypes		198	166	0.98 (0.78-1.24) 0.89

Publications

Papers

P Harrison, JJ Pointon, K Chapman, A Roddam, BP Wordsworth. Interleukin-1 promoter region polymorphism role in rheumatoid arthritis: A meta-analysis of IL-1B -511A/G variant reveals association with rheumatoid arthritis. Rheumatology (Oxford). 2008 Oct 4

P Harrison, JJ Pointon, C Farrar, BP Wordsworth. *MHC2TA* promoter polymorphism (-168*G/A, rs3087456) is not associated with susceptibility to rheumatoid arthritis in British Caucasian rheumatoid arthritis patients. Rheumatology (Oxford). 2007 ;46:409-11

P Harrison, J J Pointon, C Farrar, M A Brown, B P Wordsworth.Effects of *PTPN22 C1858T* polymorphism on susceptibility and clinical characteristics of British Caucasian rheumatoid arthritis patients. Rheumatology (Oxford). 2006 ;45:1009-11

Invited review

P. Harrison P, BP. Wordsworth. Rheumatoid Arthritis and Related Arthropathies. Chapter 26,Genomics and Clinical Medicine, Oxford University Press 2008

Abstracts

P Harrison, L Southam, J Pointon, K Chapman, A Sabokbar, BP Wordsworth.Differential allelic expression of Protein Thyrosine Phosphatase Non-receptor 22 in Rheumatoid Arthritis Patients. American College of Rheumatology Annual Scientific Meeting. October 2008

P Harrison, K Chapman, J Pointon, Y Zhang, B Wordsworth.Interleukin-1 Gene Cluster Promoter Polymorphisms in Caucasian Rheumatoid Arthritis Population: A Meta-Analysis and Effects on Clinical Parameters. EULAR 2008, June 2008

P Harrison, K Chapman, J Pointon, Y Zhang, B Wordsworth.The *MHC2TA* and Risk for Rheumatoid Arthritis: a Study of a Three Marker Haplotype and its Association with *HLA-DRB1* Shared Epitope. EULAR 2008, June 2008

P Harrison, K Chapman, J Pointon, C Farrar, A Harin, C Jess, B P Wordsworth. Effects of *Interleukin-1* polymorphisms on susceptibility and clinical characteristics of British Caucasian rheumatoid arthritis patients. British Society of Rheumatology Annual Meeting. April 2008

P Harrison, K Chapman, J Pointon, C Farrar, A Harin, C Jess, B P Wordsworth.The influence of genetic factors on radiographic outcome in rheumatoid arthritis. British Society of Rheumatology Annual Meeting. April 2008

Publications

P Harrison, JJ Pointon, C Farrar, V Ziel, C Jess, A Harin, BP Wordsworth Relevance of *PTPN22* R620W polymorphism on susceptibility in British Caucasian rheumatoid arthritis patients and effects on clinical characteristics. MRS / Academy / RCP Clinical Scientists in Training Meeting. February 2007

P Harrison, JJ Pointon, C Farrar, V Ziel, C Jess, A Harin, MA Brown, BP Wordsworth *HLA-DRB1* influence on anti-TNF treatment in British Caucasian rheumatoid arthritis patients.. British Society of Rheumatology Annual Meeting. May 2006

P Harrison, J J Pointon, C E Farrar, M A Brown, B P Wordsworth.The *PTPN22 R620W* Polymorphism - Effects on Susceptibility and Clinical Features on British Caucasian Rheumatoid Arthritis Patients. American College of Rheumatology 69th Annual Scientific Meeting. November 2005

Rheumatology 2006;45:1009–1011
Advance Access publication 20 February 2006

doi:10.1093/rheumatology/kei250

Effects of *PTPN22* C1858T polymorphism on susceptibility and clinical characteristics of British Caucasian rheumatoid arthritis patients

P. Harrison, J. J. Pointon, C. Farrar, M. A. Brown and B. P. Wordsworth

Objectives. To confirm the association of a functional single-nucleotide polymorphism (SNP), C1858T (rs2476601), in the *PTPN22* gene of British Caucasian rheumatoid arthritis (RA) patients and to evaluate its influence on the RA phenotype.

Methods. A total of 686 RA patients and 566 healthy volunteers, all of British Caucasian origin, were genotyped for C1858T polymorphism by PCR–restriction fragment length polymorphism assay. Data were analysed using SPSS software and the χ^2 test as applicable.

Results. The *PTPN22* 1858T risk allele was more prevalent in the RA patients (13.9%) compared with the healthy controls (10.3%) ($P=0.008$, odds ratio 1.4, 95% confidence interval 1.09–1.79). The association of the T allele was restricted to those with rheumatoid factor (RF)-positive disease ($n=524$, 76.4%) ($P=0.004$, odds ratio 1.5, 95% confidence interval 1.1–1.9). We found no association between *PTPN22* and the presence of the HLA-DRB1 shared epitope or clinical characteristics.

Conclusions. We confirmed the previously reported association of *PTPN22* with RF-positive RA, which was independent from the HLA-DRB1 genotype.

KEY WORDS: PTPN22, Rheumatoid arthritis, HLA, Risk allele, Rheumatoid factor positive.

Rheumatoid arthritis (RA) is a destructive inflammatory condition affecting about 1% of the UK population. A genetic component of susceptibility to RA is well established and around 30% of it has been attributed to the *HLA-DRB1* locus [1, 2]. Association between a functional single-nucleotide polymorphism (SNP) in the coding region of the gene *PTPN22* and RA in Caucasians was first described by Begovich *et al.* and has been replicated by several other groups in the US, UK, Spanish and Dutch RA patient populations [3–8]. *PTPN22* encodes the lymphocyte-specific phosphatase (Lyp), which is an important potential factor in modifying signalling through the T-cell receptor (TCR) [9]. The presence of the *PTPN22* 1858 T/T genotype increases the risk of RA more than 2-fold. However, little is known about its effects on severity and co-association with other autoimmune diseases.

We studied the distribution of the *PTPN22* C1858T polymorphism in our large sample of British Caucasian RA patients for two reasons: firstly, to confirm the strength of the association with RA; and secondly, to evaluate epistatic interactions with *HLA-DRB1* alleles.

Patients and methods

Patients

We studied 686 British Caucasian RA patients (488 female and 198 male), all satisfying the 1987 American College of Rheumatology diagnostic criteria, attending the Nuffield Orthopaedic Centre, Oxford. Written informed consent was obtained from all subjects and the study was approved by the Oxford Research Ethics Committee. DNA samples from 566 British Caucasian healthy blood donors were used as controls.

Methods

A structural questionnaire was used to identify demographic characteristics and clinical features of the disease, including age of onset, treatment, personal or family history of autoimmune diseases, and surgical history. Patients who were positive for rheumatoid factor (RF) by the particle agglutination test at two different time points from the review of the case notes were considered to have seropositive disease. Where these data were not available, RF was measured using nephelometry at the time of recruitment; a titre $\geq$40 IU/ml was considered positive. The presence of changes consistent with RA was estimated from standard hand radiographs. These were available for 661 patients (96%). Co-association of insulin-dependent diabetes mellitus (IDDM) was present in 10 (1.5%) patients and 53 (7.7%) had autoimmune thyroid pathology. Genomic DNA was extracted from whole blood. Cases and controls were genotyped for C1858T polymorphism by RFLP (restriction fragment length polymorphism analysis) of PCR products. The C$\rightarrow$T transition at codon 620 creates in the T allele a restriction site for *Xcm*I (New England Biolabs). For DNA amplification, the forward primer 5′-CAT GCT ATT GCT CTG CT-3′ and the reverse primer 5′-ATG TTG CTT CAA CGG AAT TT-3′ were used. The PCR conditions were as follows: 95°C for 10 min; 35 cycles of 95°C for 30 s, 56°C for 30 s and 72°C for 30 s; and 72°C for 10 min. *Xcm*I digestion products were resolved on 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light. To confirm the genotypes obtained by RLFP, we performed direct sequencing for *PTPN22* C1858T variants on representative samples from each genotype (Applied Biosystems 3100, Japan). Limited *HLA-DRB1* typing of RA cases using eight group-specific PCR amplification reactions were performed on all RA cases for

Botnar Research Centre, University of Oxford, Nuffield Orthopaedic Centre, Oxford OX3 7LD, UK.

Received 3 July 2005; revised version accepted 11 November 2005.

Correspondence to: P. Harrison, Botnar Research Centre, University of Oxford, Nuffield Orthopaedic Centre, Oxford OX3 7LD, UK. E-mail: pille.harrison@ndos.ox.ac.uk

© The Author 2006. Published by Oxford University Press on behalf of the British Society for Rheumatology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

*DRB1*01–*16* typing [10]. Statistical analysis for association and Hardy–Weinberg equilibrium were evaluated using the χ^2 test. *P* values less than 0.05 were considered significant. The results were also expressed as odds ratios (OR) with or without 95% confidence intervals (CI). The Bonferroni correction was used to control for multiple comparisons where appropriate. To analyse continuous variables, the software package SPSS 13.0 for Windows was used.

Results

The *PTPN22* genotypes were in Hardy–Weinberg equilibrium in cases and controls. The genotyping data are shown in Table 1. The risk allele (1858T) was more prevalent in RA patients (14%) than in healthy controls (10%) ($P=0.008$, OR 1.4, 95% CI 1.1–1.8). Of interest, while the T allele was significantly increased in the RF-positive cases (14%), its prevalence (11%) in the RF-negative cases did not differ significantly from that in controls.

In the RA group overall, 536 (78%) patients had one or more copies of *HLA-DRB1* shared epitope (SE) alleles. The proportion of patients in the RF-positive subset carrying the SE did not differ from the general RA population (80 and 78%, respectively). There was no association between SE alleles and *PTPN22* 1858T (data not shown).

The different *PTPN22* genotypes were equally distributed among female and male patients. The median age of RA onset for the entire group was 48 yr and did not differ significantly between the various *PTPN22* genotypes.

We extended our study by stratifying RA cases by clinical features (data not shown). No association between *PTPN22* and the presence of RA nodules or radiographic changes was elicited. Hip or knee joint replacement during the first 15 yr of disease duration was more common in the *PTPN22* 1858T allele group ($P=0.04$, OR 1.4, 95% CI 1.0–2.0), although this was not significant after applying Bonferroni correction for the number of comparisons made.

Finally, we compared the effect of this *PTPN22* polymorphism on the co-occurrence of RA and either IDDM or autoimmune thyroid disease. IDDM was more common in the *PTPN22* C1858T group ($P=0.04$, OR 2.7, 95% CI 1.1–6.9), but only a weak (non-significant) trend was seen with autoimmune thyroid disease ($P=0.2$, OR 1.4, 95% CI 0.8–2.4).

Discussion

Our data confirm previously reported associations of the functional *PTPN22* 1858T polymorphism with RA, but this effect was limited to RF-positive RA cases. This result is similar to those of other studies conducted on different RA patient groups, although associations with RF-negative RA have also been reported [6, 11].

PTPN22 appears to function primarily by modifying T-cell receptor signalling. The presence of a risk allele does not lead to the development of autoimmune disease *per se*. It has been suggested that subtle changes, especially in T-cell function, might, in part, influence B-cell differentiation and consequently have an effect on autoantibody production and the development of humoral autoimmunity. This is based on the observation that the *PTPN22* risk allele has been associated with multiple autoimmune diseases, where autoantibodies have been an established feature.

The frequency of the *PTPN22* 1858T allele varies in different ethnic populations. Begovich *et al.* [3] found that the risk allele frequency in all control Caucasians studied was 8.7%. The T allele was virtually absent in Han Chinese and Africans, whereas in African Americans and Mexican Americans the risk allele was detected in intermediate frequencies (2.4 and 3.5%, respectively). In our control population, the frequency of the 1858T risk allele was 10.3%, similar to that in another UK study, indicating that this allele is relatively common in Caucasians [6]. This study did not reveal additional effects arising from homozygosity in the *PTPN22* 1858T risk allele. However, our ability to detect subtle differences was limited by the relatively low frequency of the T allele and the small number of TT homozygotes.

Little is known about the effects of the *PTPN22* risk allele on the clinical course of RA. Orozco *et al.* [7] studied 826 Spanish RA patients and suggested differences in extra-articular disease in patients carrying the CC genotype, although these were not robust after correction for multiple comparisons. In our study, *PTPN22* status did not generally associate with specific clinical characteristics of RA. However, the increased frequency of the 1858T allele in our patients with early hip or knee replacement (within 15 yr of disease onset) is worth further study. While such end-stage arthritis may be considered a marker for more aggressive, destructive inflammation, it would be preferable to confirm this in a prospective study of erosive changes over time.

An excess of various other autoimmune diseases is well recognized in RA patients and their relatives. For example, thyroid dysfunction and IDDM have been found in higher frequencies among RA patients and their family members [12, 13]. Studies have shown that co-association of autoimmune illnesses might be slightly higher in patients carrying the risk allele [14, 15]. In our RA population, the prevalence of thyroid pathology and IDDM was slightly higher than in the population (7.7 *vs* 1% and 1.5 *vs* 0.3%, respectively). We found only a weak predisposition to IDDM in the *PTPN22* 1858T group ($P=0.04$, OR 2.7, 95% CI 1.1–6.9) and also a weak (non-significant) trend was seen with autoimmune thyroid disease. Since associations have already been described between *PTPN22* 1858T and IDDM and Graves disease independently of RA, we feel these trends are likely to be relevant but that our study was underpowered to detect them reliably.

TABLE 1. Genotyping data and allele frequencies

	Allele frequency			Genotype frequency			
	C No. (%)	T No. (%)	OR, 95% CI	CC No. (%)	CT No. (%)	TT No. (%)	OR, CT/TT *vs* CC
RA cases ($n=686$)	1182 (86.2)	190 (13.9)	1.4 (1.1–1.8)	507 (73.9)	168 (24.5)	11 (1.6)	1.5 (1.1–1.9)
			$P=0.008$				$P=0.004$
Controls ($n=566$)	1015 (89.7)	117 (10.3)		457 (80.7)	101 (17.9)	8 (1.4)	
RF^-($n=149$)	265 (88.9)	33 (11.1)	1.1 (0.7–1.6)	117 (78.5)	31 (20.8)	1 (0.7)	1.2 (0.7–1.8)
			$P=0.7$*				$P=0.5$*
RF^+ ($n=524$)*	897 (85.6)	151 (14.4)	1.5 (1.1–1.9)	381 (72.7)	135 (25.8)	8 (1.5)	1.6 (1.2–2.1)
			$P=0.004$*				$P=0.001$*
RF^+ *vs* RF^-			1.4 (0.9–2.0)				1.6 (0.9–2.1)
			$P=0.1$				$P=0.2$

*Compared with healthy controls.

In summary, we have confirmed the *PTPN22* 1858T association with RA that is restricted to RF-positive patients and independent of HLA genotype. This genetic association may have importance in helping to understand the pathogenesis of RA and related autoimmune diseases.

Rheumatology	Key messages
	• The association of *PTPN22* 1858T with RA is restricted to RF-positive patients and is independent of the HLA genotype.

The authors have declared no conflicts of interest.

Reference

1. MacGregor AJ, Snieder H, Rigby AS *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30–7.
2. Rigby AS, Silman AJ, Voelm L *et al.* Investigating the HLA component in rheumatoid arthritis: an additive (dominant) mode of inheritance is rejected, a recessive mode is preferred. Genet Epidemiol 1991;8:153–75.
3. Begovich AB, Carlton VE, Honigberg LA *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.
4. Lee AT, Li W, Liew A, Bombardier C *et al.* The PTPN22 R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. Genes Immun 2005;6:129–33.
5. Steer S, Lad B, Grumley JA, Kingsley GH, Fisher SA. Association of R602W in a protein tyrosine phosphatase gene with a high risk of rheumatoid arthritis in a British population: evidence for an early onset/disease severity effect. Arthritis Rheum 2005; 52:358–60.
6. Hinks A, Barton A, John S *et al.* Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an autoimmunity gene. Arthritis Rheum 2005;52:1694–9.
7. Orozco G, Sanchez E, Gonzalez-Gay MA *et al.* Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum 2005; 52:219–24.
8. Zhernakova A, Eerligh P, Wijmenga C, Barrera P, Roep BO, Koeleman BP. Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population. Genes Immun 2005;6:459–61.
9. Cloutier JF, Veillette A. Association of inhibitory tyrosine protein kinase p50csk with protein tyrosine phosphatase PEP in T cells and other hemopoietic cells. EMBO J 1996;15:4909–18.
10. Jaakkola E. Investigating genetic determinants of ankylosing spondylitis. [Dissertation]. Oxford: Open University, 2003:153–155.
11. Simkins HM, Merriman ME, Highton J *et al.* Association of the PTPN22 locus with rheumatoid arthritis in a New Zealand Caucasian cohort. Arthritis Rheum 2005;52:2222–5.
12. Silman AJ, Ollier WE, Bubel MA. Autoimmune thyroid disease and thyroid autoantibodies in rheumatoid arthritis patients and their families. Br J Rheumatol 1989;28:18–21.
13. Grennan DM, Sanders PA, Thomson W, Dyer PA. Rheumatoid arthritis: inheritance and association with other autoimmune diseases. Dis Markers 1986;4:157–62.
14. Criswell LA, Pfeiffer KA, Lum RF *et al.* Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 2005;76:561–71.
15. Wu H, Cantor RM, Graham DS *et al.* Association analysis of the R620W polymorphism of protein tyrosine phosphatase PTPN22 in systemic lupus erythematosus families: increased T allele frequency in systemic lupus erythematosus patients with autoimmune thyroid disease. Arthritis Rheum 2005;52:2396–402.

Rheumatology 2007;46:409–411
Advance Access publication 18 August 2006

doi:10.1093/rheumatology/kel300

Concise Report

MHC2TA promoter polymorphism (−168*G/A, rs3087456) is not associated with susceptibility to rheumatoid arthritis in British Caucasian rheumatoid arthritis patients

P. Harrison, J. J. Pointon, C. Farrar[1], A. Harin[1] and B. P. Wordsworth

Objective. To investigate the association of a single-nucleotide polymorphism (SNP) in the promoter region of *MHC2TA* gene (−168*G/A, rs3087456), which has previously been described in a Swedish rheumatoid arthritis (RA) cohort, in British Caucasian RA patients.

Methods. We genotyped 733 RA patients and 613 healthy controls for *MHC2TA* −168*G/A SNP by amplification-refractory mutation system (ARMS). Data were analysed using SPSS version 13.0 software and the chi-square test was applied where appropriate.

Results. The *MHC2TA* −168*G/A SNP was not associated with increased susceptibility to RA in our patients. Stratifying the patients according to the presence or absence of rheumatoid factor (RF) showed the SNP to be more common in RF negative patients, but this did not reach statistical significance.

Conclusion. We did not confirm the previously reported association of this *MHC2TA* polymorphism with RA in our UK population despite its ethnic similarities with the Swedish population in which it was first described.

KEY WORDS: *MHC2TA*, Rheumatoid arthritis, HLA, Rheumatoid factor negative.

Introduction

Rheumatoid arthritis (RA) is a classic example of a common complex disease with a multifactorial aetiology. It is a clinically heterogeneous condition, the seropositive erosive form of which has a substantially higher sibling recurrence risk than forms of the mild non-erosive disease [1]. The strongest association is with *HLA-DRB1* alleles, estimated at around 30% of the total genetic effect [2]. It is likely that several other genes are also involved, but only *PTPN22*, encoding a protein tyrosine phosphatase important in regulating T-cell activation, has been reliably and reproducibly implicated to date [3, 4]. *HLA-DRB1* is one of a number of genes within the class II region of the major histocompatibility complex (MHC), expression of which is predominantly regulated by the MHC class II transactivator (CIITA). CIITA production is dependent on the transcription of its gene *MHC2TA*, which is further controlled by four functional promoters [5]. Recently a Swedish group described a single-nucleotide polymorphism (SNP) (−168*A/G, rs3087456) in the type III promoter of *MHC2TA* associated with susceptibility to RA, multiple sclerosis and myocardial infarction [6]. This result was not replicated in a smaller German population [7]. Since the power to replicate positive findings is heavily influenced by the sample size, particularly where the genetic effect is weak. Our aim was to study the rs3087456 polymorphism in a large sample of British Caucasian RA patients.

Method

Patients with RA ($n=733$) were recruited from the Nuffield Orthopaedic Centre in Oxford. All RA patients were Caucasians living in the UK and satisfied the 1987 American College of Rheumatology criteria. The controls ($n=613$) were healthy British Caucasian blood donors from the same geographical area. All subjects gave informed, written consents, and approval was granted by the Oxford Research Ethics Committee. Medical records were examined to determine the rheumatoid factor (RF) status of each patient. Patients who were positive at two different time points were considered to have seropositive disease. Where the data was not available, RF was measured using nephelometric technique and a titre over 40 IU/ml was considered positive.

Genomic DNA was extracted from whole blood. Cases and controls were genotyped for *MHC2TA* −168*G/A polymorphism using amplification-refractory mutation system (ARMS) with additional mismatch (underlined). The primer sequences used were: for allele A, forward 5′-GTGAAATTAATTTCAG AGGTGGA and reverse 5′-AGAAGCACACAGCCTCATCA; for allele G, forward 5′-ATGATACTGGCCCCATCCTG and reverse 5′-CACTCCCTTAAGCCCTCCAC. The PCR conditions were as following: 95°C for 10 min, 35 cycles of 95°C for 30 s, 58°C for 30 s for allele A or 60°C for allele G, 72°C for 30 s, followed by 72°C for 10 min. PCR products were resolved on 3% agarose gel stained with ethidium bromide and visualized under ultraviolet light. All samples were genotyped in duplicates and

University of Oxford Institute of Musculoskeletal Sciences, Botnar Research Centre and [1]Nuffield Department of Orthopaedic Surgery Nuffield Orthopaedic Centre, Oxford, UK.

Submitted 11 June 2006; revised version accepted 14 July 2006.

Correspondence to: Pille Harrison, Botnar Research Centre, University of Oxford, Nuffield Orthopaedic Centre, Oxford OX3 7LD, UK.
E-mail: pille.harrison@ndos.ox.ac.uk

© The Author 2006. Published by Oxford University Press on behalf of the British Society for Rheumatology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

TABLE 1. Genotyping data and allele frequencies

	Allele frequency			Genotyping frequency Codominant model				Genotyping frequency G-dominant model		
	A (%)	G (%)	OR, 95% CI	AA (%)	AG (%)	GG (%)	*P*-value	AA (%)	AG + GG (%)	OR, 95%CI
RA cases $n = 733$	1082 (74)	384 (26)	1.0 (0.9–1.1) $P = 0.7$	404 (55)	274 (37)	55 (8)	0.4	404 (55)	329 (45)	1.0 (0.8–1.2) $P = 0.9$
Controls $n = 613$	914 (75)	312 (25)		336 (55)	242 (39)	35 (6)		336 (55)	277 (45)	
RF neg[a] $n = 159$	222 (70)	96 (30)	0.8 (0.6–1.0)[a] $P = 0.08$	78 (49)	66 (42)	15 (9)	0.2	78 (49)	81 (51)	1.3 (0.9–1.8) $P = 0.2$
RF pos[a] $n = 559$	834 (75)	284 (25)	1.0 (0.8–1.2)[a] $P = 1.0$	315 (56)	204 (37)	40 (7)	0.4	315 (56)	244 (44)	0.9 (0.7–1.2) $P = 0.6$
RF pos *vs* RF neg			1.3 (1.0–1.7) $P = 0.08$				0.2			0.8 (0.5–1.1) $P = 0.1$

[a]*vs* healthy controls.

positive and negative controls were included. Statistical analysis for association and Hardy–Weinberg equilibrium were evaluated using chi-square test. *P*-values <0.05 were considered significant. For continuous variables statistical software package SPSS version 13.0 for Windows was used.

Results

All the genotypes were in Hardy–Weinberg equilibrium, and there was no association of the *MHC2TA* −168*G/A with susceptibility to RA (Table 1). When we stratified our RA population for the presence of RF, the minor allele frequency for RF positive patients was 0.25 (similar to healthy controls) compared with 0.30 for RF negative patients, but this did not reach statistical significance.

No significant associations were observed when the patients were stratified for gender, age of onset or *HLA-DRB1* genotype.

Discussion

It has become evident that distinct autoimmune diseases can share common genetic polymorphisms, a classic example for which is in *PTPN22*. This gene encodes a protein tyrosine phosphatase essential for T-cell signalling. The R620W variant has now been associated with several autoimmune diseases, including type I diabetes, RA and systemic lupus erythematosus [8–10]. Another possible example of a genetic variant common to several diseases was recently suggested: a promoter SNP in *MHC2TA* was associated with RA, multiple sclerosis and myocardial infarction in Swedes. Including our study, there have been two subsequent independent studies assessing the *MHC2TA* −168*A/G polymorphism in German and British RA patient populations. The Swedish, German and British study groups were all reasonably powerful statistically, containing 1288, 319 and 733 RA patients, compared with 709, 463 and 613 healthy controls, respectively. In contrast to the initial association of *MHC2TA* −168*A/G SNP with RA in the Swedish RA population, we could not confirm this in our British RA population, and this is consistent with the recent German report [7]. The likely explanation for this apparent difference between the three populations may lie in the minor allele frequencies found in different control populations. It is striking that the minor allele frequencies are very similar in the RA patients from each population (24.4% in Swedes, 25% in German and 26% in British) but in contrast, it is only 20.5% in the Swedish controls compared with 26% in the Germans and 25% in the British. Such a difference in the control samples could have occurred by chance, generating a spurious association with RA in the Swedes. It seems relatively unlikely, given the very similar allele frequencies in all but the Swedish control group, that this is a true ethnic-related genetic effect. Furthermore, if the minor allele frequency observed in the Swedish controls were spuriously low, it might also account for the association seen with the two other diseases. In particular, it could account for the surprising association between *MHC2TA* and myocardial infarction, which defies an obvious biological explanation. In the original Swedish study, a surprisingly low proportion (62.5%) of the RA group were RF positive compared with our sample. *MHC2TA* may, therefore, be more relevant to the aetiology of seronegative RA, and we did find a small (non-significant) difference between our seropositive and negative patients. Replication of the initial results in a further independent Swedish population is, therefore, desirable.

In their analysis of the Swedish population, Swanberg *et al.* [6] indicated a significant effect for the GG genotype in a codominant model of susceptibility. This equates with an excess of 21 patients with the GG genotype in the RA group compared with predictions ($n = 62$) based on the genotype frequency in controls (4.8%). This would be a tiny patient subgroup (constituting only 1.6% of their study population) compared with other known RA associations, such as HLA-DRB1.

Swanberg *et al.* [6] indicated that the putative disease-associated polymorphism caused a modest (<50%) reduction in IFN-γ-induced *MHC2TA* and HLA-DR mRNA expression in peripheral blood cells. Since HLA-DR molecules are believed to be important in the pathogenesis of RA, it is tempting to conclude that factors involved in regulating their expression might also be involved in the aetiology of the disease. CIITA expression in lymphoid cells is almost exclusively regulated by type III promoter in contrast to cells of myeloid origin, where CIITA expression is regulated by promoter I or induced by INF-γ. Promoter IV is responsible for CIITA expression in non-professional antigen-presenting cells and thymic epithelial cells and is strongly influenced by INF-γ [11]. It is also interesting that statins (HMG-CoA reductase inhibitors), by inhibiting *MHC2TA* gene activation, also repress INF-γ induced MHC class-II expression [12]. It remains to be seen whether there may be potential future benefit from suppressing such MHC class II expression that can be harnessed in the treatment of diseases where MHC class II-dependent T-cell activation is required. Thus, *MHC2TA* and its regulation are attractive candidates for therapeutic manipulation in RA, but we suggest that genetic variation at this locus is unlikely to play a significant role in its genetic aetiology.

Rheumatology	Key message
	• Polymorphism in the third promoter of *MHC2TA* gene is not associated with increased susceptibility to RA in British Caucasian patients.

Acknowledgements

The authors are grateful to the patients for their cooperation. The authors specifically want to thank Dr J. David and Dr P. Bowness for the access to the patients, and nurses V. Ziel and C. Jess for their assistance in sample collections.

The authors have declared no conflicts of interest.

References

1. Wolfe F, Kleinheksel SM, Khan MA. Prevalence of familial occurrence in patients with rheumatoid arthritis. Br J Rheumatol 1988;27:150–2.
2. Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. Clin Genet 1989;36:178–82.
3. Hinks A, Barton A, John S *et al.* Association between the PTPN22 gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: further support that PTPN22 is an autoimmunity gene. Arthritis Rheum 2005;52:1694–9.
4. Orozco G, Sanchez E, Gonzalez-Gay MA *et al.* Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. Arthritis Rheum 2005;52:219–24.
5. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J 1997;16:2851–60.
6. Swanberg M, Lidman O, Padyukov L *et al.* MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. Nat Genet 2005;37:486–94.
7. Akkad DA, Jagiello P, Szyld P *et al.* Promoter polymorphism rs3087456 in the MHC class II transactivator gene is not associated with susceptibility for selected autoimmune diseases in German patient groups. Int J Immunogenet 2006;33:59–61.
8. Begovich AB, Carlton VE, Honigberg LA *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.
9. Bottini N, Musumeci L, Alonso A *et al.* A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004;36:337–8.
10. Wu H, Cantor RM, Graham DS *et al.* Association analysis of the R620W polymorphism of protein tyrosine phosphatase PTPN22 in systemic lupus erythematosus families: increased T allele frequency in systemic lupus erythematosus patients with autoimmune thyroid disease. Arthritis Rheum 2005;52:2396–402.
11. LeibundGut-Landmann S, Waldburger JM, Krawczyk M *et al.* Mini-review: specificity and expression of CIITA, the master regulator of MHC class II genes. Eur J Immunol 2004;34:1513–25.
12. Kwak B, Mulhaupt F, Myit S, Mach F. Statins as a newly recognized type of immunomodulator. Nat Med 2000;6:1399–402.

Rheumatology 2008;47:1768–1770
Advance Access publication 4 October 2008

doi:10.1093/rheumatology/ken374

Concise Report

Interleukin-1 promoter region polymorphism role in rheumatoid arthritis: a meta-analysis of *IL-1B*-511A/G variant reveals association with rheumatoid arthritis

P. Harrison[1], J. J. Pointon[1], K. Chapman[1], A. Roddam[2] and B. P. Wordsworth[1]

Objectives. IL-1 has a central role mediating inflammation and joint destruction in RA. Single nucleotide polymorphisms (SNPs) and haplotype structure in the promoter region can modulate *IL-1* function. This study examined the effects of four common promoter SNPs in the *IL-1* region on susceptibility and clinical characteristics of RA in British Caucasian patients and assessed the risk of RA by meta-analysis of published studies.

Methods. Using PCR-based methods, 756 RA patients and 625 healthy controls (HCs) were genotyped for *IL-1A* (−889 C/A, rs17561), *IL-1B* (−511 A/G, rs16944), *IL-1B* (−1464 C/G, rs1143623) and *IL-1B* (−3737 G/A, rs4848306) SNPs. Further meta-analysis was performed for *IL-1B* (−511 A/G) incorporating 3712 RA patients and 2331 HC from six association studies.

Results. The *IL-1B* (−1464 C/G) G allele was found to be less common in the RA group [$P = 0.01$; odds ratio (OR) 1.24; 95% CI 1.04, 1.48]. There was no association between *IL-1* SNPs and the presence of *HLA-DRB1* shared epitope, RF or clinical characteristics. Meta-analysis revealed statistically significant association between *IL-1B* (−511 A/G) and RA ($P = 0.02$; pooled OR 1.13; 95% CI 1.02, 1.26).

Conclusions. There may be a protective effect in RA from the *IL-1B* (−1464 C/G) G variant. No direct association between the polymorphisms studied and clinical severity characteristics were observed. Further meta-analysis revealed *IL-1B* (−511 A/G) to be associated with increased susceptibility to RA.

KEY WORDS: Rheumatoid arthritis, Meta-analysis, Interleukin-1, Interleukin-1A, Interleukin-1B, Association study, Major histocompatibility complex, Radiographic severity, Genetics, Polymorphism.

Introduction

Cytokines are small regulatory proteins that act as intercellular mediators in the generation and control of immune and inflammatory responses. In RA, two main cytokines, TNF and IL-1, are particularly involved. The most studied of the IL-1 proteins are IL-1α and IL-1β. They function as agonist molecules that activate target cells by binding to IL-1 receptor. The production and activity of IL-1α and IL-1β are regulated by transcriptional and post-translational mechanisms as well as from a natural competitive antagonist. The *IL-1* gene cluster (comprising 10 related genes) maps to 2q12. Both *IL-1A* (MIM 147760) and *IL-1B* (MIM 147720) may have arisen through gene duplications, and hence encode similar protein products [1]. Strong linkage disequilibrium (LD) has been demonstrated across this gene cluster in Caucasian populations [2].

A genetic linkage study in European RA families provided suggestive evidence of linkage of RA to the *IL-1* region [3]. The reported associations of certain *IL-1* gene cluster polymorphisms with susceptibility and severity, especially in the more erosive form of RA, have been variable [4–8]. These studies have generally been conducted on relatively small RA patient samples, apart from the one recent exception [8]. In the latter, the authors performed a comprehensive analysis of *IL-1A* and *IL-1B* of a large number of North American and Spanish RA populations but identified no major contribution to RA from these genes.

We examined four common single nucleotide polymorphisms (SNPs) and haplotypes in the *IL-1A* and *IL-1B* promoter regions and their association with susceptibility and severity with RA in well-characterized British Caucasian patients. We performed meta-analysis of published data on *IL-1B-511* to evaluate the effect of this polymorphism on susceptibility to RA among Caucasians.

Patients and methods

Patients and study design

A total of 756 British Caucasian RA patients (542 females, 72%) of Northern European descent attending the Nuffield Orthopaedic Centre in Oxford, UK, were recruited. All satisfied the 1987 ACR criteria for RA and mean age of disease onset was 48 yrs. The controls ($n = 625$) were healthy British Caucasian blood donors from the same geographical area. All subjects gave informed written consent and approval was granted by the Oxford Research Ethics Committee.

Medical records were examined for the RF status of each patient. Patients who were positive at two different time-points were considered seropositive. Where data were not available, RF was measured using nephelometric technique and a titre over 40 IU/ml was considered positive. Overall, 586 patients (78%) deemed to be RF positive.

Severity of RA was determined by the presence of rheumatoid nodules or the requirement of major joint replacement as established from medical records. The association with radiographic outcome was examined in 206 RA patients with minimum disease duration of 3 yrs. Hand radiographs were scored using Rau–Ratingen method [9].

Genotyping

Genomic DNA was extracted from whole blood. Cases and healthy controls (HCs) were genotyped for *IL-1A* (−889 C/A, rs17561), *IL-1B* (−511 A/G, rs16944), *IL-1B* (−1464 C/G, rs1143623) and *IL-1B* (−3737 G/A, rs4848306) SNPs by PCR-based methods (details available on request). Limited *HLA-DRB1* and further *DRB1*04* subtyping was performed as published [10]. All samples were genotyped in duplicates, and positive and

[1]Institute of Musculoskeletal Sciences and [2]Cancer Research UK Epidemiology Unit, University of Oxford, Oxford, UK.

Submitted 1 July 2008; revised version accepted 18 August 2008.

Correspondence to: P. Harrison, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Road, Oxford OX3 7LD, UK. E-mail: pille.harrison@ndos.ox.ac.uk

© The Author 2008. Published by Oxford University Press on behalf of the British Society for Rheumatology. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

TABLE 1. Distribution of *IL-1A* (-889 C/A), *IL-1B* (-511 A/G), *IL-1B* (-1464 C/G) and *IL-1B* (-3737 G/A) alleles and genotypes among RA patients and HC

	Allele counts (%)			Genotype counts (%)				Genotype counts (%)		
	1	2	*P*-value, OR 95% CI	11	12	22	*P*-value	11	12 + 22	*P*-value, OR 95% CI
IL-1A-889										
RA	1031 (70)	447 (30)	P=0.93, OR = 1.01	355 (48)	321 (43)	63 (9)	P = 0.91	355 (48)	384 (52)	P = 0.80, OR = 1.03
HC	841 (70)	367 (30)	95% CI 0.85, 1.19	286 (47)	269 (45)	49 (8)		286 (47)	318 (53)	95% CI 0.84, 1.28
IL-1B-511[a]										
RA	1022 (69)	460 (31)	P = 0.14, OR = 1.13	347 (47)	328 (44)	66 (9)	P = 0.20	347 (47)	394 (53)	P = 0.39, OR = 1.10
HC	796 (66)	404 (34)	95% CI 0.96, 1.33	267 (45)	262 (44)	71 (11)		267 (45)	333 (55)	95% CI 0.88, 1.37
IL-1B-1464										
RA	1116 (76)	362 (24)	P = 0.01, OR = 1.24	421 (57)	274 (37)	44 (6)	P = 0.05	421 (57)	318 (43)	P = 0.03, OR = 1.28
HC	874 (71)	352 (29)	95% CI 1.04, 1.48	312 (51)	250 (41)	51 (8)		312 (51)	301 (49)	95% CI 1.02, 1.59
IL-1B-3737										
RA	778 (54)	676 (46)	P = 0.18, OR = 0.90	211 (29)	356 (49)	160 (22)	P = 0.26	211 (29)	516 (71)	P = 0.10, OR = 0.82
HC	669 (56)	523 (44)	95% CI 0.77, 1.05	198 (33)	273 (46)	125 (21)		198 (33)	398 (67)	95% CI 0.65, 1.05

[a]Alleles were named according to their frequency in the general population i.e. allele 1 being the frequent or more common allele, allele 2 the rare one

negative controls were included. The *HLA-DRB1* shared epitope (SE) was defined by the presence of one or two copies of the *HLA DRB1*01,*0401, *0404, *0405, *0408* and **10* alleles.

Statistical analysis

Case–control association analysis. Statistical analysis for association and Hardy–Weinberg equilibrium were evaluated using χ^2-test. $P < 0.05$ was considered statistically significant. Corrections for multiple testing were applied as necessary [11]. For continuous variables, SPSS version 15.0 for Windows (SPSS Inc., IL, USA) was used. Haplotype estimations and frequency comparison were performed using PHASE v. 2 (available at http://www.stat.washington.edu/stephens/software.html) and CLUMP software (available at http://linkage.rockefeller.edu/soft/clump.html) [12, 13].

Analysis of radiological progression. To allow comparison of radiographic scores in RA patients with different disease duration, we made use of the almost linear relationship between the radiographic score and the square root of disease duration (SQRD) [14]. To ensure that the predicted severity scores on all percentiles remained within the real range of 0–140, a logit transformation of radiographic score {ln [score/(140−score)]} was entered as the dependent variable with SQRD as independent variable. The final regression equation was: logit score = (0.66 × SQRD) − 4.10. Associations between particular *IL-1* gene cluster polymorphism and radiographic severity score were assessed by comparing the frequencies of *IL-1* SNPs in subjects with mild (score <20) or severe disease (score >80) using Fisher's exact probability test.

Meta-analysis

Search strategy and study selection. Meta-analysis was conducted for *IL-1B* (-511 A/G) SNPs based on the PubMed database using the terms 'rheumatoid arthritis' and 'interleukin-1'. Publications on Caucasian populations in English language were included. Study selection criteria were: (i) RA patients fulfilled 1987 ACR classification criteria; (ii) controls were drawn from the same geographic area and ethnic background as patients; (iii) authors provided original genotype frequencies; and (iv) RA patient and control groups did not overlap between studies. Six studies met the selection criteria [5, 7, 8, 15–17]; all were original articles with case–control study design. The meta-analysis included 3712 RA patients and 2331 controls.

Evaluation of the statistical association. Though previous studies describe the *IL-1B*-511 SNP as a G→A change, in the dbSNP database the 'A' allele is ancestral, defining the variant allele as 'G'. However, the 'G' allele occurs more commonly (allele frequencies among HC are 0.67). In the meta-analysis, we compared the variant 'G' allele with ancestral A as well as recessive (G/G *vs* G/A + AA) and dominant (G/G + G/A *vs* AA) models for the G allele. Genotype-specific risks such as G/G *vs* A/A and G/A *vs* A/A were also analysed. Statistical analyses were performed using StatsDirect software. Single odd ratios (OR) and 95% CIs were estimated for each study. Two-tailed *P*-values were calculated using the χ^2-test of independence. Between-study heterogeneity was assessed with Cochran's *Q*-test statistics. The pooled OR and 95% CI for the variant allele in RA patients against HC were calculated from fixed-effects model, as no significant heterogeneity was observed. Forest plots were generated in StatsDirect to display the OR and 95% CI.

Genetic power calculations for the meta-analysis were done using software program Quanto v1.2.3 (available at http://hydra.usc.edu/gxe) assuming a disease prevalence of 1% and 'G' allele frequency of 0.67. Under log-additive model, we calculated the effect of 0.05 with OR of 1.2 (average OR of all combined studies).

Results

Case–control association study

The genotyping results and allele counts for RA patients and HCs are presented in Table 1. *IL-1B* (−1464 C/G) 'G' allele was less common among RA patients (P = 0.01; OR 1.24; 95% CI 1.04, 1.48). The minor allele frequencies for cases and controls were 24 and 29%, respectively. The association remained significant after applying corrections for false discovery rate.

Haplotype analysis revealed a global difference in haplotype frequencies between the RA and HC (P = 0.01). This effect was entirely due to the *IL-1B* (−1464 C/G) and there was no single over-represented haplotype.

Each of the polymorphisms was investigated in the RA cohort according to RF, SE status and presence of high-risk *HLA-DRB1**04 (DRB1*0401, 0404, 0405 or 0408) or a low-risk allele (DRB1*01, 0406, 0407 and *10). No significant difference was observed between the groups (data not shown).

Analysis according to the severity features of RA did not reveal any independent association with the presence of nodules, requirement for joint replacement or radiographic progression (data not shown).

Meta-analysis

Genotyping data in six studies adhered to Hardy–Weinberg equilibrium. No significant heterogeneity was observed between the studies (moment-based estimate of between studies variance = 0, I^2 = 0). The meta-analysis of the *IL-1B* (−511 A/G) showed significant association of the G/G genotype (P = 0.02; OR 1.13; 95% CI 1.02, 1.26) (Fig. 1) and variant allele G (P = 0.02; OR 1.10; 95% CI 1.02, 1.19) with RA. The genotype-specific analysis showed a significant dose effect between the G/G *vs* A/A (P = 0.04; OR 1.20; 95% CI 1.01, 1.44) but not between G/G *vs* GA or G/A *vs* AA genotypes (data not shown).

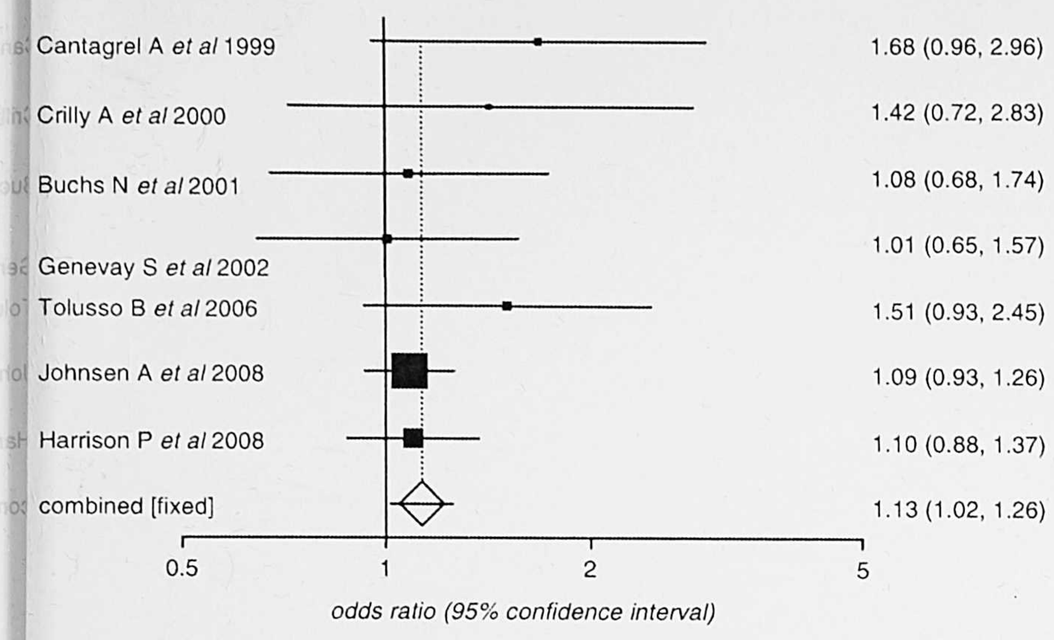


FIG. 1. OR and 95% CI from studies examining the association between *IL-1B* (−511) G/G *vs* G/A + AA genotypes (OR and 95% CI for each study is shown by a black square on a horizontal line and the summary OR and 95% CI shown by a dotted vertical line extending from a white diamond. The size of the black box is proportional to the percentage weight of each study in the analysis).

Discussion

We have demonstrated a novel association of the *IL-1B* (−1464 C/G) SNP with RA in a Caucasian population. This association should be replicated in an independent cohort in order to confirm the result. In our data set, *IL-1B* (−1464 C/G) is in near complete LD with *IL-1B* (−511 A/G) ($D' > 0.90$, $r^2 > 0.80$). However, the minor allele frequencies in our cohort for *IL-1B* (−511 A/G) and *IL-1B* (−1464 C/G) are dissimilar (0.31 and 0.24 for RA and 0.34 and 0.29 for controls, respectively), indicating that independent genotyping is warranted in order to investigate the *IL-1B* (−1464 C/G) association. It remains unknown whether the *IL-1B* (−1464 C/G) SNP is the 'causal' variant. This polymorphism has been shown to exhibit allele-specific difference in gene expression as *IL-1B*-1464G has shown weaker promoter activity than *IL-1B*-1464C on electromobility gel shift assay [18]. This promoter region harbours a core-binding motif for GATA-binding family proteins and *IL-1B* (−1464 C/G) could affect protein binding [18]. It is possible that the real 'protective' variant is elsewhere and exhibits different LD with *IL-1B* (−511 A/G) and *IL-1B* (−1464 C/G). Further studies investigating genetic variations across whole gene regulatory regions and more detailed functional analysis are essential.

Although both IL-1α and IL-1β mRNA have been detected in RA synovial fluid, it is IL-1α that has been found to be strongly expressed at the cartilage–pannus junction [19]. In a study of affected sib pair RA families, Cox *et al.* [4] found four *IL-1* gene cluster markers associated with erosive RA, particularly in sib pair families with 0 probability of sharing two *HLA-DRB1* alleles identical by descent. Evidence of less severe disease progression after 20 yrs in Swiss RA patients carrying one or two copies of the *IL-1B-511* minor allele has also been reported [7]. Using several different analyses, we were unable to demonstrate any significant relationship between *IL-1* polymorphisms and radiographic progression. Similar conclusions were reached in the recent study of Johnsen *et al.* [8] where no association was observed with erosive disease in a large sample of the Brigham Rheumatoid Arthritis Sequential Study cohort.

Meta-analysis allows us to maximize the sample size, increase statistical power and provide more robust findings. The meta-analysis presented here is the largest to have investigated the effects of *IL-1B* (−511) in RA. We demonstrated a significant association of the GG genotype and 'G' allele with RA. The power of this meta-analysis was over 90% to detect modest effect (OR 1.2 at a significance level of 0.05). However, the magnitude of the genetic risk associated with this polymorphism is relatively small and further studies on a much larger RA sample are desirable to confirm the true association.

In summary, the role of the genetic factors in RA is complex and the magnitude of the effects is likely to vary greatly. We have demonstrated an association of *IL-1B* (−511) with RA patients of Caucasian origin by conducting a meta-analysis. The potential protective influence of *IL-1B* (−1464) found in this study will require further confirmation in other populations.

Rheumatology key messages

- There may be a protective effect in RA from the *IL-1B* (−1464 C/G) polymorphism.
- Meta-analysis of *IL-1B* (−511 A/G) revealed an association with increased susceptibility to RA.

Acknowledgements

We would like to thank the many patients who participated in this study and our colleagues Dr Paul Bowness, Dr Joel David and Dr Raashid Luqmani for allowing us access to their patients.

Funding: This study was funded by the University of Oxford.

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Nicklin MJ, Barton JL, Nguyen M, FitzGerald MG, Duff GW, Kornman K. A sequence-based map of the nine genes of the human interleukin-1 cluster. Genomics 2002;79:718–25.
2 Nicklin MJ, Weith A, Duff GW. A physical map of the region encompassing the human interleukin-1 alpha, interleukin-1 beta, and interleukin-1 receptor antagonist genes. Genomics 1994;19:382–4.
3 Barrera P, Faure S, Prud'homme JF *et al.* European Consortium on Rheumatoid Arthritis Families (ECRAF). European genetic study on rheumatoid arthritis: is there a linkage of the interleukin-1 (IL-1), IL-10 or IL-4 genes to RA? Clin Exp Rheumatol 2001;19:709–14.
4 Cox A, Camp NJ, Cannings C *et al.* Combined sib-TDT and TDT provide evidence for linkage of the interleukin-1 gene cluster to erosive rheumatoid arthritis. Hum Mol Genet 1999;8:1707–13.
5 Cantagrel A, Navaux F, Loubet-Lescoulie P *et al.* Interleukin-1beta, interleukin-1 receptor antagonist, interleukin-4, and interleukin-10 gene polymorphisms: relationship to occurrence and severity of rheumatoid arthritis. Arthritis Rheum 1999;42:1093–100.
6 Jouvenne P, Chaudhary A, Buchs N, Giovine FS, Duff GW, Miossec P. Possible genetic association between interleukin-1alpha gene polymorphism and the severity of chronic polyarthritis. Eur Cytokine Netw 1999;10:33–6.
7 Genevay S, di Giovine FS, Perneger TV *et al.* Association of interleukin-4 and interleukin-1B gene variants with Larsen score progression in rheumatoid arthritis. Arthritis Rheum 2002;47:303–9.
8 Johnsen AK, Plenge MR, Butty V *et al.* A broad analysis of IL1 polymorphism and rheumatoid arthritis. Arthritis Rheum 2008;58:1947–57.
9 Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. J Rheumatol 1998;25:2094–107.
10 Jaakkola E, Herzberg I, Laiho K *et al.* Finnish HLA studies confirm the increased risk conferred by HLA-B27 homozygosity in ankylosing spondylitis. Ann Rheum Dis 2006;65:775–80.
11 Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. Genetics 2003;164:829–33.
12 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.
13 Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann Hum Genet 1995;59:97–105.
14 de Carvalho A, Graudal H, Jørgensen B. Evaluation of the progression of rheumatoid arthritis. Significance of age at onset and sex. Acta Radiol Diagn 1980;21:545–50.
15 Crilly A, Maiden N, Capell HA, Madhok R. Predictive value of interleukin 1 gene polymorphism for surgery. Ann Rheum Dis 2000;59:695–9.
16 Buchs N, di Giovine FS, Silvestri T, Vannier E, Duff GW, Miossec P. IL-1B and IL-1Ra gene polymorphisms and disease severity in rheumatoid arthritis: interaction with their plasma levels. Genes Immun 2001;2:222–8.
17 Tolusso B, Pietrapertosa D, Morelli A *et al.* IL-1B and IL-1RN gene polymorphisms in rheumatoid arthritis: relationship with protein plasma levels and response to therapy. Pharmacogenomics 2006;7:683–95.
18 Lee KA, Ki CS, Kim HJ *et al.* Novel interleukin 1beta polymorphism increased the risk of gastric cancer in a Korean population. J Gastroenterol 2004;39:429–33.
19 Deleuran BW, Chu CQ, Field M *et al.* Localization of interleukin-1 alpha, type 1 interleukin-1 receptor and interleukin-1 receptor antagonist in the synovial membrane and cartilage/pannus junction in rheumatoid arthritis. Br J Rheumatol 1992;31:801–9.