From the Rheumatology Unit, Department of Medicine,
Karolinska Institutet, Stockholm, Sweden

GENETIC PREDISPOSITIONS TO RHEUMATOID ARTHRITIS IN MALAYSIAN POPULATION

Chun Lai Too

杜 翠 麗

Stockholm 2012

To My Family

ABSTRACT

Genetic predisposition is a significant and fundamental determinant of susceptibility to rheumatoid arthritis (RA), a complex autoimmune disease with a painful and disabling condition. The vast majority of genetic studies in RA have been centered on populations of European descent with very few studies on East Asian populations.

In this thesis, we aimed to determine the genetic predisposition to RA in the Malaysian population residing in the South East Asian region. More specifically, we addressed the question of how far the identified RA risk loci in Europeans and East Asians can be translated across different populations. We undertook this investigation using the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA) case control study involving mainly the early RA cases, which comprised of Malay, Chinese and Indian ethnic groups. We showed that different HLA-DRB1 shared epitope (SE) alleles, which are common in Asian (i.e. DRB1*0405), but not in European populations conferred significantly increased risk of developing anti-citrullinated protein antibody (ACPA)-positive, but not ACPA-negative RA. With the preponderance of the DRB1*12 alleles in our study population, we demonstrated a novel protective effect of DRB1*1202 associated with ACPA-positive RA in Malay and Chinese populations.

The combination between genetic and environment factors is widely believed to be the major trigger of RA development. Our analysis of gene-environment interaction between smoking and HLA-DRB1 shared epitope (SE) alleles revealed a strong association with ACPA-positive RA and this interaction seem to apply between smoking and DRB1*0405 allele, which is common in Asian populations.

Polymorphisms in the peptidylarginine deiminase type IV (PADI4) gene have been repeatedly shown to associate with RA susceptibility in individuals of Asian descent, but weak or no association was observed in the European populations, despite of comparable risk allele frequency between these populations. We scrutinized the entire PADI locus including PADI1, PADI2, PADI3, PADI4 and PADI6 genes with a set of 320 single nucleotide polymorphisms (SNPs) for association with RA. Our findings revealed an association between PADI4 in the diverse populations from Malaysia. In addition, we also suggest a novel association in a PADI2 gene.

Approximately 40% of RA patients are diagnosed as having ACPA-negative disease. As yet, few validated risk alleles were associated exclusively with ACPA-negative RA. We investigated the association between the previously reported ACPA-negative-associated dendritic cell immunoreceptor (DCIR) polymorphisms and RA in four independent Asian populations from China and Malaysia. Our results provide evidence for an association between the DCIR variant and RA in non-European populations. We also confirmed the genetic effect of DCIR polymorphisms on RA risk particularly in ACPA-negative RA.

Taken together, this thesis provides evidence that no single population is sufficient for fully uncovering the risk variants underlying RA in all populations. Therefore, studies in diverse population could provide a better understanding of genetic architecture of RA especially in the RA susceptibility risk loci.

LIST OF PUBLICATIONS

I. **Shared Epitope Alleles Remain A Risk Factor for Anti-citrullinated Protein Antibody (ACPA)-Positive Rheumatoid Arthritis in Three Asian Ethnic groups**

    TOO CHUN-LAI, Leonid Padyukov, Jasbir Singh Dhaliwal, Emeli Lundström, Abqariyah Yahya, Nor Asiah Muhamad, Lars Klareskog, Lars Alfredsson, Per Tobias Larsson, Shahnaz Murad, for the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA) Study Group.

    PLoS One 2011, 6:e21069

II. **Smoking interacts with *HLA-DRB1* shared epitope in the development of anti-citrullinated protein antibody-positive rheumatoid arthritis: results from the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA)**

    CHUN LAI TOO*, Abqariyah Yahya*, Shahnaz Murad, Jasbir Singh Dhaliwal, Per Tobias Larsson, Nor Asiah Muhamad, Nor Aini Abdullah, Amal Nasir Mustafa, Lars Klareskog, Lars Alfredsson, Leonid Padyukov and Camilla Bengtsson, for MyEIRA study group.

    Arthritis Research & Therapy 2012, 14:R89

III. **Polymorphisms in peptidylarginine deiminases (PADI) associate with rheumatoid arthritis in diverse Asian populations: evidence from MyEIRA study and meta-analysis**

    CHUN LAI TOO, Shahnaz Murad, Jasbir Singh Dhaliwal, Per Larsson, Xia Jiang, Bo Ding, Lars Alfredsson, Lars Klareskog, and Leonid Padyukov.

    Arthritis Research & Therapy 2012, [Accepted]

IV. **A Replication Study Confirms the Association of Dendritic cell Immunoreceptor (DCIR) Polymorphisms with ACPA-Negative RA in a Large Asian Cohort**

    Jianping Guo*, Xinyu Wu*, **CHUN LAI TOO**, Fangrui Yin, Xiaolan Lu, Jing He, Ru Li, Xu Liu, Shahnaz Murad, Leonid Padyukov, Zhanguo Li.

    PLoS ONE 2012, 7:e41228

    *these authors contributed equally to this work.

List of publication and manuscripts not included in this thesis

I. **Increased Occurrence of IgG rheumatoid factor in Asian rheumatoid arthritis patients irrespective of ethnicity**

   CHUN LAI TOO, Johan Rönnelid, Yuslina Mat Yusoff, Jasbir Singh Dhaliwal, Nor Ashikin Jinah, Abqariyah Yahya, Heselynn Hussein, Wahinuddin Sulaiman, Per Tobias Larsson, and Shahnaz Murad

   Submitted manuscript

II. **Smoking is associated with an increased risk of developing ACPA-positive but not ACPA-negative rheumatoid arthritis in Asian populations: evidence from the Malaysian MyEIRA case-control study**

   Abqariyah Yahya, Camilla Bengtsson, **TOO CHUN LAI**, Per T Larsson, Amal Nasir Mustafa, Nor Aini Abdullah, Norasiah Muhamad, Heselynn Hussein, Lars Klareskog, Lars Alfredsson, Shahnaz Murad.

   Mod Rheumatol. 2012 Aug;22 (4):524-31

III. **Silica exposure is associated with an increased risk of developing ACPA-positive rheumatoid arthritis in an Asian population: evidence from Malaysian MyEIRA case-control study**

   Abqariyah Yahya, Camila Bengtsson, Per Larsson, **CHUN LAI TOO**, Amal Nasir Mustafa, Nor Aini Abdullah, Nor Asiah Muhamad, Lars Klareskog, Shahmaz Murad and Lars Alfredsson, for MyEIRA study group.

   Submitted manuscript

CONTENTS

LIST OF ABBREVIATIONS

ACPA	Anti-citrullinated protein/peptide antibodies
ACR	American College of Rheumatology
AP	Atributable proportion
Anti-CCP	Anti cyclic citrullinated peptides
DNA	Deoxyribonucleic acid
DCIR	Dendritic cell immunoreceptor
EIRA	Epidemiological investigation of rheumatoid arthritis
GWAS	Genome wide association study
HLA	Human leukocyte antigen
LD	Linkage disequilibrium
MAF	Minor allele frequency
MHC	Major histocompatibility complex
MyEIRA	Malaysian Epidemiological Investigation of Rheumatoid Arthritis
OR	Odds ratio
PADI4	Peptidylarginine deiminase type 4
PCA	Principal component analysis
PTPN22	Protein tyrosine phosphatase, non-receptor type 22
RA	Rheumatoid arthritis
RF	Rheumatoid factor
SE	Shared epitope
SNP	Single nucleotide polymorphism

1. INTRODUCTION

Have you ever thought about this: "we are different but basically the same" or "we are the same but different"?

Here, I will give a brief introduction on the genetics overview for the studied disease, rheumatoid arthritis (RA), in two heterogeneous populations: the Caucasians and the Asians. I will discuss the shared genetic risk loci as well as the population-specific risk loci between these populations and why it is important to perform disease association research in this field. A short overview of the genetic tools used in studies of human complex diseases will also be included.

1.1 Immune system, autoimmunity and autoimmune diseases

In order to understand autoimmune diseases such as RA, it is helpful to know how the immune system normally works in our body. The immune system is the body's means of protection against 'foreign' substances such as those carried by bacteria, viruses, fungi and parasites. It has the ability to recognize cells and tissues that are its own (self) as distinct from those that are not (non-self). The immune system generally protects rather than attack its own body tissues. The immune system can be categorized into innate immune system (non-specific) and adaptive immune system (specific). Furthermore, the immune system is built on a network of special cells, tissues and organs that work together to defend our body against foreign invaders.

Our defense mechanism begins with the physical barriers such as skin and epithelial lining. If these barriers are breached by the invaders known as pathogens, the macrophages (a type of phagocytic cells) will take the lead to control or destroy the pathogens. The macrophages express pattern recognition receptors (PRRs) such as the toll like receptors (TLRs) and C-type lectin-like receptor (CLRs), which recognize the conserved structures on the pathogens. They become activated and subsequently engulf and destroy the invading pathogens. Alternatively, macrophages can serve as antigen-presenting cells (APC) to process and present the pathogens during activation of the second line of defense, the adaptive immune system.

The adaptive immunity refers to the antigen-specific defense mechanisms characterized by T cell specific for antigen derived from the invading pathogens, and by B cells producing antibodies that bind to these antigens. During adaptive immune response, antigens are transported to lymphoid organs where they are recognized by naïve B lymphocytes and T lymphocytes. These activated B and T lymphocytes will then proliferate and differentiate into effector cells and work together to eliminate the invading pathogens.

Adaptive immunity relies on the ability of immune cells to distinguish between the host cells (self) and unwanted invaders (non-self). Each cell carries protein markers on its surface defined as major histocompatibility complexes (MHC). In normal or healthy conditions, our immune cells will not attack any other cells with markers identifying it as 'self'. However, the immune system sometimes goes awry and cannot distinguish between self-cells and non-self-cells. As a consequence of this, the immune cells attack the body itself. The misdirected

immune responses thus are designated as autoimmunity. Autoimmunity occurs through the presence of autoantibodies or T lymphocytes reactive with host antigens. Failure to maintain the control mechanisms of autoimmunity will lead to illness referred as autoimmune diseases.

Autoimmune diseases are phenotypically heterogeneous. Some are organ specific such as type 1 diabetes and multiple sclerosis, among others, and non-organ specific (or systemic) such as RA and systemic lupus erythematosus (SLE). Autoimmune diseases are one of the commonest human complex diseases triggered by the combination of genetic and environmental factors.

1.2 Rheumatoid arthritis

Rheumatoid arthritis (RA) is a complex autoimmune disease mainly involves the peripheral joints of hands and feet, which generally lead to a painful and disabling condition that affects one in every hundred people with an estimated heritability of approximately 60% [2] (*heritability is the proportion of total variation in the population that can be attributed to variation in genetic factor*). RA is more common among women and in older age groups.

The disease generally presents in a symmetrical (*both side of the body*) pattern, most often involving the hand joints. Therefore, joint destruction becomes the hallmark of RA. RA can affect the whole body, including several organs, and so is described as systemic disease. Progressive and irreversible joint damage is caused by the immune system attacking its own body tissues, particularly those lining the joints.

In RA, the immune system initially attacks the synovium (a type of tissue that produces fluid to lubricate and nourish the joint tissues). White blood cells move from the blood stream and invade the synovium and small blood vessels infiltrate the area. Consequently, the synovial membrane becomes thick and inflamed, resulting in unwanted tissue growth. The inflammation also involves the release of various biochemical substances that cause pain, swelling and joint damage. These substances can also damage the surrounding cartilage, bone, tendons and ligaments. Also when they enter the bloodstream, these substances can cause fatigue and a general feeling of being unwell. Gradually, the joint loses its shape and alignment and undergoes changes that are mostly irreversible.

For RA, almost all the evidence for autoimmune conditions is limited to the seropositive (rheumatoid factor (RF) and/or anti-citrullinated protein antibodies (ACPA) positive) subset of disease, whereas it is still unclear how much autoimmunity is associated with the seronegative RA. The combination between genetic and environmental factors is widely believed to be the major trigger of RA development.

Epidemiological studies revealed that RA is relatively common in almost all population of the world, albeit somewhat higher prevalence has been reported in northern Europe and north America (0.5-1.1%) as compared to Asia [3, 4]. In some geographical areas, the prevalence and incidence of RA varies across ethnic groups, with high RA prevalence among Pima Indians [5] and low in some areas of rural Africa [6].

The American Rheumatism Association developed classification criteria for RA with 91-94% sensitivity and 89% specificity (Table 1) [7]. However, the criteria are useful for established RA, but not sufficiently reliable when applied to patients with early arthritis. Thus the 2010 ACR/EULAR criteria (Table 2) were established to enable for early diagnosis and treatment of RA which often can improve the patient's quality of life [8].

Table 1. The conventional ACR 1987 criteria for RA

ACR 1987 Classification Criteria for RA
1. Morning stiffness (at least 1 hour)
2. Arthritis of three or more joint areas
3. Arthritis of hand joints (≥ 1 swollen joints)
4. Symmetrical arthritis
5. Rheumatoid nodules
6. Serum rheumatoid factor
7. Radiographic changes (erosions)
Four of these seven criteria must be present. Criteria 1-4 must have present for at least 6 weeks.

Table 2. The new 2010 ACR/EULAR classification criteria for RA

2010 ACR/EULAR Classification Criteriafor RA	
1. Joint involvement (0-5)	
• 1 large joint	(0)
• 2-10 large joints	(1)
• 1-3 small joints (large joints not counted)	(2)
• 4-10 small joints (large joints not counted)	(3)
• >10 joints (at least one small joint)	(5)
2. Serology (0-3)	
• Negative RF <u>AND</u> negative ACPA	(0)
• Low positive RF <u>OR</u> low positive ACPA	(2)
• High positive RF <u>OR</u> high positive ACPA	(3)
3. Symptom duration (0-1)	
• <6 weeks	(0)
• ≥6 weeks	(1)
4. Acute-phase reactants (0-1)	
• Normal CRP <u>AND</u> normal ESR	(0)
• Abnormal CRP <u>OR</u> abnormal ESR	(1)
Points are shown in paratheses. Cutpoint for rheumatoid arthritis 6 points or more. Patients can also be classified as having rheumatoid arthritis if (a) typical erosions; (b) long-standing disease previously sastifying the classification criteria.	

ACR=American College of Rheumatology; EULAR=European League Against Rheumatism; RF=rheumatoid factor; ACPA=anti-citrullinated antigens; CRP=C-reactive protein; ESR=erythrocyte sedimentation rate. ACR 1987 criteria were designed to classify established RA. The 2010 ACR/EULAR criteria are intended to classify both early and established disease.

2. GENETICS AND RHEUMATOID ARTHRITIS

Genetic predisposition is a significant and fundamental determinant of susceptibility to RA. A genetic predisposition means that an individual has a genetic susceptibility to developing a certain disease, but this does not mean that an individual inheriting a genetic susceptibility will definitely develop the disease. Extensive research has been carried out for the identification of genetic regions tagged by genetic variations referred to as single nucleotide polymorphisms (SNP). Combined data from recently conducted genome-wide scan (GWAS) meta-analyses revealed more than 45 confirmed and validated genetic regions associated with RA as at September 2012 (Figure 1 and Table 3) [9-13]. Many themes have emerged during the gene-hunting for RA risk loci:

- The possibility of genetically distinct subsets of RA,

- The existence of ethnic heterogeneity,

- The evidence of shared risk loci across multiple autoimmune diseases, also known as pleiotropy,

- The most popularly debated concept of 'missing heritability'

In this thesis, I have restricted my focus on the first two themes which were related to my current research.

2.1 Genetics and subsets of rheumatoid arthritis

There is overwhelming evidence that RA is not a single disease entity [14-17], but rather can be sub-classified on the basis of its serologic factors, such as rheumatoid factor (RF) and anti-citrullinated protein antibodies (ACPA). RF is an antibody against the Fc portion of IgG, while ACPA are autoantibodies against citrullinated protein that are formed by post translational conversion of arginine to citrulline by peptidylarginine deiminase (PADI) [18]. RF is, however not unique for RA patients as it can also be found in the elderly and certain infectious diseases. Conversely, ACPA are highly specific for RA and has been widely accepted as the hallmark of RA. Furthermore, ACPA-positive patient seem to have a more aggressive clinical course and more destructive disease [19-22].

The contribution of genetic risk factors for RA development varies among these subsets of RA. However, all the large-scale genetic studies to date have been biased toward ACPA-positive RA patients and/or anti-CCP positive patients in particular [23-26]. As a result, more genetic risk alleles have been reported in ACPA-positive RA as compared to ACPA-negative RA, both in human leukocyte antigen (HLA) and non-HLA genes regions. Despite this bias, it is clear that many validated RA risk alleles are more strongly, if not exclusively, associated with ACPA-positive RA [14], but not ACPA-negative RA [27-30]. The two best examples for association with ACPA-positive RA are the HLA-DRB1 shared epitope risk alleles and PTPN22 gene [31].

As yet, less validated risk alleles were associated exclusively with ACPA-negative RA. Although HLA-DRB1*03 has been reported to be associated with ACPA-negative disease [32, 33], not all studies have confirmed this association [34, 35]. Intriguingly, the Swedish researchers have identified the genetic modulators of ACPA-positivity in which HLA-DRB1*13 has dual role: it protects against ACPA-positive RA but in combination with DRB1*03, it inversely increases the risk of ACPA-negative disease [34]. Recently, our group compared the GWAS data for both RA subsets as characterized by ACPA status [16]. Our findings revealed that ACPA-positive and ACPA-negative RA display significant risk allele frequency differences which are mainly confined to the HLA region. Only one SNP close to RPS12P4 locus in chromosome 2 reached a p value of 2×10^{-6} and the locus can then be considered as a tentative candidate locus for ACPA-negative RA.

Meanwhile, it should be noted that the heritability of RA among twin pairs for ACPA-positive RA and ACPA-negative RA was found to be 68% and 66%, respectively [36]. However, the presence of SE explained 18% of the genetic variance in ACPA-positive individuals but it contributes only 2.4% of the variance in the ACPA-negative subset. The authors concluded that although genetic predisposition plays a significant role in susceptibility to both types of RA, the development of ACPA-negative disease depends on as yet unidentified non-shared epitope genetic factors. Thus, deciphering the genetic factors and pathogenesis of ACPA-negative RA remain a major challenge for genetic studies of RA.

In this chapter, I will focus on two examples of the genes: one is associated with ACPA-positive RA, namely the HLA-DRB1 gene, particular the SE risk alleles, and another associated with ACPA-negative RA, namely the DCIR gene, which has been studied in this thesis.
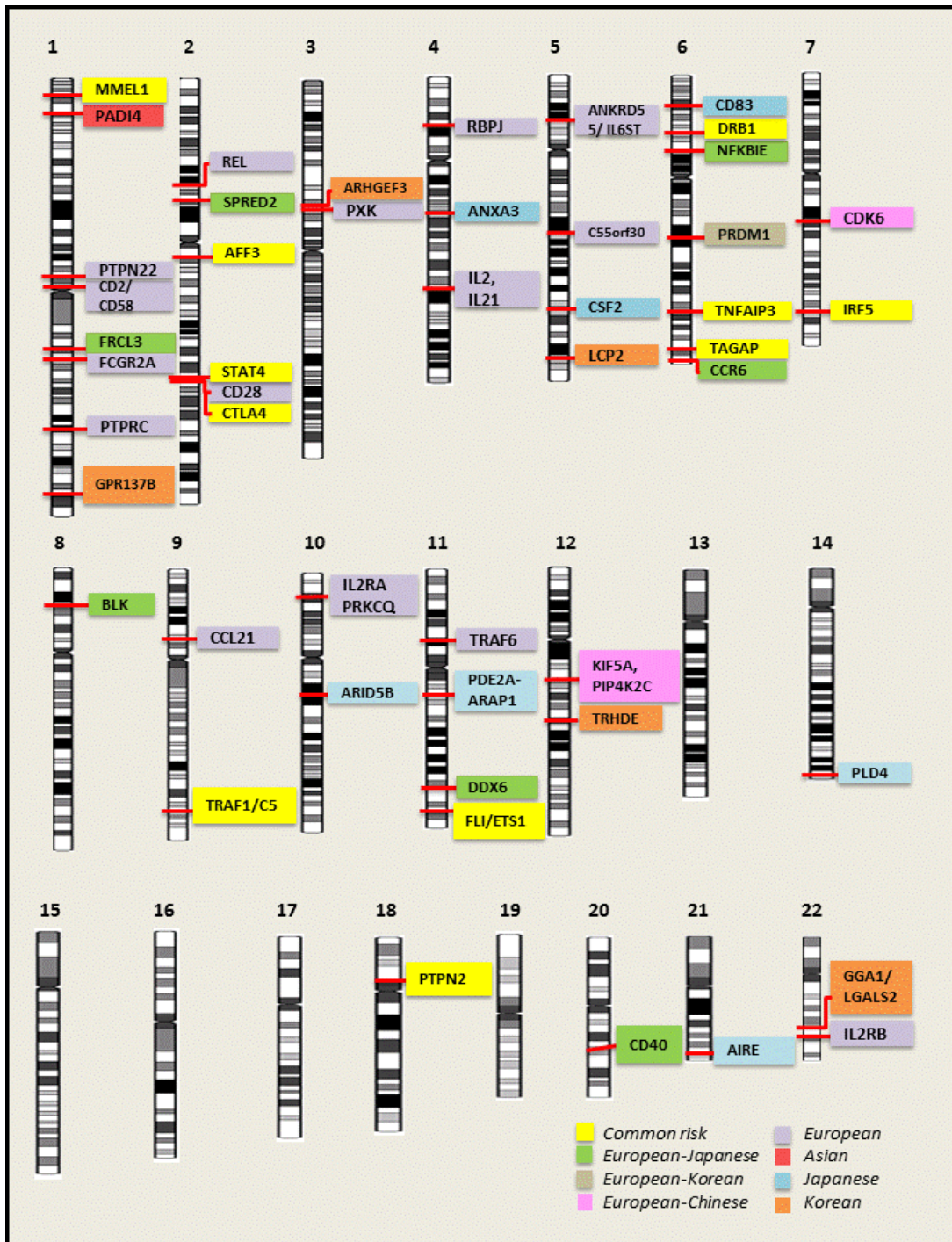
Figure 1. Chromosomal location for the validated RA susceptibility gene/gene loci in European and Asian populations

Table 3. List of validated RA susceptibility genes/gene loci in European and Asian populations

Chr	Gene(s)	SNP ID	Chromosomal location	RA association				Reference
				EUR	JP	KOR	CHI	
1	PTPN22	rs2476601	1p13	YES	#	#	#	[10-13]
1	CD2, CD58	rs11586238	1p13	YES	NO		NO	[10, 12, 13]
1	MMEL1/TNFRSF14	rs3890745	1p36	YES	YES	NO	YES	[10-13]
1	PADI4	rs2240340	1p36	NO	YES	YES	YES	[10-13]
1	FRCL3	rs3761959	1q22	YES	YES			[12, 13]
1	FCGR2A	rs12746613	1q23	YES	#			[12, 13]
1	PTPRC	rs10919563	1q31	YES	NO			[12, 13]
1	GPR137B	rs7537965	1q42-q43		YES			[11]
2	SPRED2	rs934734	2p14	YES	YES			[12, 13]
2	B3GNT2	rs11900673	2p15	NO	YES			[12, 13]
2	REL	rs13031237	2p16	YES	NO		NO	[10, 12, 13]
2	AFF3	rs10865035, rs11676922	2q11	YES	YES	YES		[11-13]
2	STAT4	rs7574865	2q32	YES	YES	YES	YES	[10-13]
2	CD28	rs1980422	2q33	YES	NO		NO	[10, 12, 13]
2	CTLA4	rs3087243	2q33	YES	YES		YES	[10, 12, 13]
3	PXK	rs13315591	3p14	YES	#			[12, 13]
3	ARHGEF3	rs2062583	3p14.3		YES			[11]
4	RBPJ	rs874040	4p15	YES	#			[12, 13]
4	ANXA3	rs2867461	4q21	NO	YES			[12, 13]
4	1L2, IL21	rs6822844	4q27	YES	#		NO	[10, 12, 13]
5	ANKRD55, IL6ST	rs6859212	5q11	YES	NO			[12, 13]
5	C5orf30	rs26232	5q21	YES	NO			[12, 13]
5	CSF2	rs657075	5q31	NO	YES			[12, 13]
5	LCP2	rs4867947	5q35.1			YES		[11]
6	NFKBIE	rs2233434	6p21.1	YES	YES			[12, 13]
6	DRB1	rs6910071 rs2157337	6p21.3	YES	YES	YES	YES	[10-13]
6	CD83	rs12529514	6p23		YES			[12]
6	PRDM1	rs548234	6q21	YES	NO	YES	NO	[10-13]
6	TNFAIP3	rs10499194 rs6920220 rs5029937	6q23	YES	YES	YES		[11-13]
6	TAGAP	rs394581	6q25	YES	YES	YES	YES	[11-13]
6	CCR6	rs3093023	6q27	YES	YES			[12, 13]

Table 3. List of validated RA susceptibility genes/gene loci in European and Asian populations *(continue)*

Chr	Gene(s)	SNP ID	Chromosomal location	RA association				References
				EUR	JP	KOR	CHI	
7	CDK6	rs10225965 rs2282983, rs102782993	7q21-q22	YES			YES	[10]
7	IRF5	rs10488631	7q32	YES	YES			[12, 13]
8	BLK	rs2736340	8p23	YES	YES			[12, 13]
9	CCL21	rs2812378 rs951005	9p13	YES	NO		NO	[10, 12, 13]
9	TRAF1,C5	rs3761847	9q33	YES	YES		YES	[10, 12, 13]
10	IL2RA	rs2104286 rs706778	10p15	YES	NO		NO	[10, 12, 13]
10	PRKCQ	rs4750316	10p15	YES	NO		NO	[10, 12, 13]
10	ARID5B	rs10821944	10q21	NO	YES			[12, 13]
11	TRAF6	rs540386	11p12	YES	NO			[12, 13]
11	PDE2A-ARAP1	rs3781913	11q13	NO	YES			[12, 13]
11	DDX6	rs10892279	11q23.3	YES	YES			[12, 13]
11	FLI/ETS1	rs4936059	11q24	YES	YES	YES		[11-13]
12	KIF5A, PIP4K2C	rs1678542	12q13	YES	NO		YES	[10, 12, 13]
12	TRHDE	rs12831974	12q15-q21			YES		[11]
14	PLD4	rs2841277	14q32	NO	YES			[12, 13]
18	PTPN2	rs2847297	18p11	YES	YES	YES		[11-13]
20	CD40	rs4810485	20q13	YES	YES		NO	[10, 12, 13]
21	AIRE	rs2075876	21q22.3	NO	YES			[12, 13]
22	IL2RB	rs3218253	22q12	YES	NO		NO	[10, 12, 13]
22	GGA1/LGALS2	rs4547623	22q13.31, 22q13.1			YES		[11]

Chr=chromosome; EUR = European; JP= Japanase; KOR= Korean, CHI= Han Chinese; #: monomorphic risk variant

9

2.1.1 MHC, HLA-DRB1 gene, shared epitope hypothesis and rheumatoid arthritis

The major histocompatibility complex (MHC) has been consistently associated with RA across different populations. The MHC genes located on chromosome 6p21.3, extends over 3.6 Mb [1]. The MHC is a highly gene dense region containing ~200 defined genes, mainly involved in immune function [37]. The MHC region is divided into three main regions: MHC class I, class II and class III (Figure 2).

The class I region, at the telomeric end of the MHC, contains the human leukocyte antigen (HLA) class I genes: HLA-A, HLA-B and HLA-C and encodes the HLA class I molecules. These molecules are expressed on all nucleated cells. They present antigens to CD8$^+$ T cells and are involved in the natural killer (NK) cell mediate immune response.

In the HLA class II region are the HLA-DR, HLA-DP, and HLA-DQ loci; encoding the α- and β-chains of the various HLA class II molecules. These HLA class II molecules are expressed as heterodimers on the cell surface of antigen presenting cells (APC), such as dendritic cells, macrophages and B lymphocytes. They present antigenic peptides to CD4+ T cells.

On the chromosome, the class III region is lies between the class I and class II regions, encodes proteins with immune-response-related functions (other than direct antigen presentation) such as complement, etc.
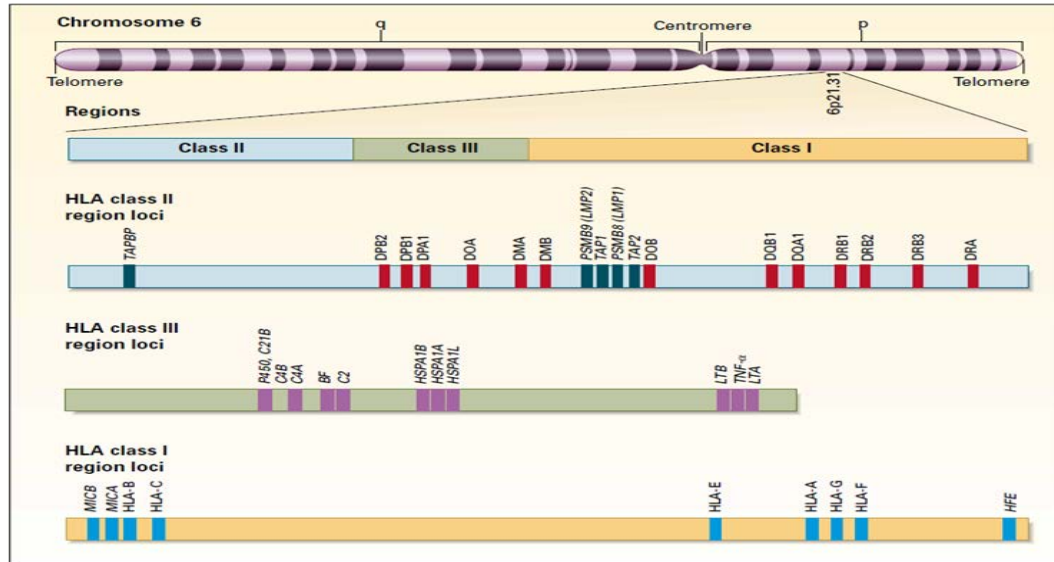

Figure 2. Location and organization of the HLA complex on chromosome 6. *The complex is conventionally divided into three regions: I, II, and III. Each region contains numerous loci (genes), only some of which are shown. Of the class I and II genes, only the expressed genes are depicted. Class III genes are not related to class I and class II genes structurally or functionally (Adapted from Klien J et al, [1], copyright Massachusetts Medical Society).*

It has been estimated that HLA genes account for 30% to 50% of overall genetic susceptibility to RA [38-41]. Within the HLA locus, the strongest association is linked to the HLA-DRB1 gene, which encodes the β chain of the class II molecule referred to as HLA-DR. The first association between RA and HLA-Dw4 was reported in 1976 by Peter Stastny [42]. A decade later, Gregersen et al. advanced the hypothesis that multiple RA risk alleles within the HLA-DRB1 gene shared a conserved amino acid sequence at position 70 to 74 ([70]QRRAA[74], [70]RRRAA[74] or [70]QKRAA[74]) in the third hypervariable region of DR beta 1 chain, designated as 'shared epitope' (SE) and the risk alleles are termed as SE alleles [43].

The mechanism underlying the SE and RA association is insufficiently understood. The common hypotheses attribute it to the presentation of arthritogenic antigens [44] or alteration in peptide affinity or T cell repertoire selection [45, 46], which are promoting autoreactive adaptive immune responses. Other hypothetical explanations for the association between RA and SE include molecular mimicry of the SE by microbial proteins, increased T-cells senescence induced by SE-containing HLA molecules, and potential proinflammatory signaling function that is unrelated to the role of SE in antigen recognition [47, 48].

The RA-associated HLA-DRB1 SE risk alleles for RA are believed to differ between ethnic groups. For example HLA-DRB1*0401 and HLA-DRB1*0404 are the commonest alleles in RA patients of European ancestry [49-51], whereas DRB1*0405 is prevalent in Asian populations [52-56]. These observations are largely due to differential distribution of HLA-DRB1 alleles among ethnicities. Interestingly, a remarkable high frequency of HLA-DRB1*0901 allele (7-15%) was also observed in individuals of Asian ancestry, but is less common in Europeans. Although the HLA-DRB1*0901 allele does not belong to the SE-containing DRB1 risk alleles, its association with RA susceptibility was evident in Asian populations [56-58].

ACPA has been widely believed to be clinically important [59] and several studies investigating the relationship between ACPA and HLA-DRB1 have revealed that the association between DRB1 SE alleles and RA was restricted to ACPA-positive RA, but not ACPA-negative RA, both in European and Asian populations [58-62].

Although HLA-DRB1 alleles, particularly the SE risk alleles explain much of the risk due to the MHC region, there is growing evidence that additional non-SE risk loci within the MHC were associated with RA susceptibility [63-67]. Recently, the dense SNP genotyping across the MHC provided continued support for additional alleles i.e. HLA-DPB1 gene and the European ancestral haplotype of A1-B8-DR3 [63, 68-71]. Additionally, a very recent GWAS performed on a large combined samples of European ancestry provided supporting evidence that three amino acid position (11, 71 and 74) and additional amino acid positions in HLA-B (at position 9) and HLA-DP (at position 9) conferred risk to ACPA-positive RA. These amino acids were found to be located in the peptide-binding grooves in HLA molecules [72]. Another recent report from Japanese population also suggested an independent association between RA and HLA loci other than the HLA-DRB1 alleles (HLA-DP, HLA-B and HLA-C) [73].

As yet, there is compelling evidence from recent genetic discoveries suggesting the SE hypothesis alone is insufficient to explain the contribution of the HLA genes to RA

susceptibility. Furthermore, genetic variants underlying RA are likely to be multiple, each with a relatively small effect, but act as in a concert (i.e. gene-gene interactions) or with environmental influence (i.e. gene-environment interactions) leading to clinical disease.

Considerable efforts have been undertaken to unravel the genetic contribution to RA and these efforts have led to the discovery of an extensive list of RA risk loci outside the MHC region such as the PTPN22 gene, the PADI4 and other loci, both in patients of European and Asian ancestries.

2.1.2 The DCIR gene and rheumatoid arthritis

Overall, approximately 40% of RA patients are diagnosed as having ACPA-negative disease [74]. They hold an important clue to the RA mechanism and are as important as the ACPA-positive RA. Little is known about the genetic susceptibility to ACPA-negative RA [16]. Recent studies of oil-induced model of inflammatory arthritis in rats in our group provided supportive evidence of an association of resistance to disease with variant in antigen-presenting lectin-like receptor gene complex (APLEC), including the dendritic cell immunoreceptor (DCIR) [28, 75-77].

The DCIR gene, located on human chromosome 12p13 is characterized by the carbohydrate recognition domain (CRD) and signaling through an immunoreceptor tyrosine-based inhibitory (ITIM) [78]. Four DCIR SNP markers were initially found to be significantly associated with ACPA-negative RA with modest effects (estimated ORs 1.21 to 1.27), but not ACPA-positive RA in a Swedish population [28]. Additional validation is therefore warranted in other populations as to unravel the nature of variation that potentially underlies susceptibility to RA, particularly in the ACPA-negative subgroup.

2.2 Ethnic heterogeneity in rheumatoid arthritis risk loci

Over the past several years, there has been an explosion of new data on genetic risk factors for RA. This progress has resulted from the expansion of case-control samples collections and the development of commercial genotyping platforms such as the SNP-based genome wide scan or GWAS, which allow for production of hundreds of millions of genotypes in a rapid and cost-effective manner [9-13, 16, 23-25, 43, 79-96]. It is important to highlight that almost all the GWAS experiments and many other large-scale genetic studies have been conducted among individuals with self-reporting as 'white' European ancestry [23, 79, 91, 95, 97]. Therefore, the majority of literature has been centered on the risk-associated alleles among the patients of European ancestry. Very few large studies have been performed involving individuals of Asian ancestry (largely limited to Japanese population) [9, 12, 80, 81, 85], which lead to identification of the PADI4 and CD244 as risk factors for RA in this population.

Interestingly, several genes including HLA-DRB1, STAT4, TNFAIP3, TAGAP, CTLA4 and NKFBIE [12, 13] have been validated as shared susceptibility genes among populations of

European and Asian ancestries, while other RA susceptible genes seemed to be limited to specific ethnic populations as presented in Figure 1 and Table 2.

2.2.1 Rheumatoid arthritis genetics in Europeans

The majority of RA genetic studies undertaken by far have been focused on populations of European ancestry, showing remarkable consistency in their findings. Many novel RA risk loci have been detected recently through large-scale meta-analysis studies (*a statistical analysis tool that can be used by combining data from several studies that have been conducted on similar research aim/hypothesis*). The major goal of such meta-analysis is to increase the power of a study and ultimately identify novel association of modest effect. An example was demonstrated in the recent meta-analyses performed by Stahl and colleagues which had included a total of 12,307 seropositive (RF or ACPA) RA patients and 28,975 European controls recruited from four different geographical regions: Sweden, North America, UK and The Netherlands. The study successfully confirmed 24 previously established RA risk loci and additionally introduced seven novel RA risk loci (CCR6, IL6ST, RF5, PXK, RBPJ, and SPRED2) to the existing RA risk loci list. As a result, the number of robustly associated RA risk loci has increased to 34 in the European populations [13].

2.2.2 Rheumatoid arthritis genetics in Asians

Multi-ancestry studies on validated RA susceptibility loci have proven the presence of both population-specific and shared genetic components of RA [15, 98]. Unlike the European studies, most of the Asian studies were conducted independently in individual population such as Japanese, Korean and Han Chinese. These three ethnic groups are by far considered as the major ethnic groups from East Asia region.

Most of the RA risk loci identified in Asian populations were based on the findings obtained from the Japanese studies. Very recently, Okada and colleagues reported a GWAS meta-analysis performed in Japanese population including 9,351 RA patients (approximately 80% of individuals were with ACPA-positive RA) and 38,575 controls from five different centers in Japan. The researchers identified nine loci newly associated with RA at genome-wide significance level ($P<5.0 \times 10^{-8}$).These gene loci included B3GNT2, ANXA3, CSF2, CD83, NKFBIE, ARID5B, PDE2A-ARAP1, PLD4 and PTN2. In addition, the authors also conducted a multi-ancestry comparative analysis together with the data from previously pooled meta-analysis from the European populations [13]. The results provided convincing evidence of ~30% overlapping RA genetic risk factors between the Asian and European populations' e.g. 14 loci from a total of 46 loci were shared. However, while all the 46 investigated RA risk variants were polymorphic in the European subjects; six loci were actually monomorphic in the Japanese population including the PTPN22 variant. Interestingly, only two of the newly detected RA risk loci in the Japanese study e.g. ARID5B and PTPN2 were associated with RA in the European populations.

In Korean RA population, the GWAS data revealed that only MHC region and PADI4 reached the genome-wide significance level. The shared genetic susceptibility was observed in four regions namely STAT4, BLK, AFF3 and CCL2 between the Korean and European patients with RA. Freudenberg and co-workers reported that few RA genetic risk factors may be specific to the Korean population after performing the validation in an independent cohort [11].

Danoy and colleagues used a set of known European RA-associated risk variants from 19 distinct gene regions to replicate the associations in another Asian RA population, the Han Chinese from China. The study included 2,737 RA patients and controls and reported that only two genes loci, MMEL1 and CTLA4 were associated with Han Chinese RA patients, while six other loci showed nominal association with RA. It is important to mention that in this replication study, the authors confirmed that no association was observed between the PTPN22 genetic variant and RA in the Han Chinese population.

Taken together the GWAS data from different populations, it is evident now that there are substantial genetic differences between the European and Asian populations. However, one should also consider that populations of European and Japanese ancestries have dominated the majority of the GWAS and RA association studies in the literature. It is therefore important to perform further dissection of RA genetic determinants across other ethnic groups in the world as to clarify the RA genetic heterogeneity between different ethnic populations and eventually provide further insight into the RA pathogenesis on the ethnicity-dependent manner.

In summary, it is now obvious that there is heterogeneity among the different RA subsets defined by ACPA status and also across different ethnic populations. The differences are best reflected by the major genetic determinant of predisposition to RA, the HLA-DRB1 alleles, which were discussed exclusively in earlier chapter under section 2.1.1. The most replicated non-HLA associations are also good examples of the ethnic differences in RA risk factors: PTPN22 in European populations and PADI4 in Asian populations, which will be discussed further in the next chapters.

2.2.3 The PTPN22 gene and rheumatoid arthritis

Protein tyrosine phosphatase non-receptor type 22 or PTPN22 gene was first reported as a non-HLA RA risk factor in European populations, after an initial finding of association with the related autoimmune disease type 1 diabetes (T1D) in 2004 [88, 99]. Since then the association with RA has been consistently documented in multiple ethic populations of European descent [24, 26, 31, 100-108].

The PTPN22 gene is located on chromosome 1p13 and encodes an intracellular tyrosine phosphatase [109]. The best associated genetic variant rs2476601, which affects amino acid 620, is an arginine (R) to tryptophan (W) missense polymorphism that may alter the function of protein [110, 111].

The frequency of the associated PTPN22 risk variant rs2476601 differs among European individuals, showing a gradient of decreasing from northern to southern Europe i.e. from 12.5% in the Swedish and Finnish to 2.5-7.4% in the Spanish and Italian populations,

respectively [112]. Interestingly, this allele is rarely found or absent in African American and in populations of Asian descent [10, 12, 113-115]. Although there were several attempts to find different SNPs in the PTPN22 gene that may be associated n with RA in the non-European populations, there were no evidence of association with RA, both with haplotype analysis and re-sequencing of this region [113, 114].

PTPN22 gene is now widely accepted as a pleiotropic gene (i.e. one gene influences multiple phenotypic traits). In addition to RA, the PTPN22 gene was associated with a number of autoimmune diseases, including T1D, SLE, Grave's disease, generalized vitiligo, Hashimoto thyroiditis, Myasthenia gravis, and Addison's disease [115-124]. Strikingly, there was no evidence of association with multiple sclerosis, inflammatory bowel disease and celiac disease [125-130]. These contrasting patterns of association are likely to reflect fundamental similarities and differences in the mechanism underlying the pathogenesis of these disorders.

The PTPN22 risk variant is generally found to be associated with almost exclusively with the ACPA-positive RA, but not ACPA-negative RA [24]. The combination of PTPN22 variant and anti-CCP antibodies illustrated 100% specificity for diagnosing RA [103]. Also, this genetic variant appears to contribute risk of developing RA mainly in individuals with SE-positive [24, 131, 132].

In summary, the robust and reproducible genetic association of PTPN22 variant with RA in European populations, but not in Asian populations suggesting PTPN22 is associated with RA only in specific populations.

2.2.4 The PADI4 gene and rheumatoid arthritis

Apart from the long-established HLA-DRB1 risk alleles associated with RA, the peptidylarginine deiminase type 4 (PADI4) gene, residing outside the HLA locus, was identified in a Japanese population as a second risk factor for RA in 2003 through a large-scale linkage disequilibrium SNPs mapping experiment [81].

PADI4 gene is located on chromosome 1p36 and is a member of the PADI member encoding an enzyme that converts the arginine to citrulline (also known as citrullination, a type of post-translational modification) (Figure 3) [133]. The physiological role of citrullination is not well understood, but believed to be connected with inflammation and cell death. Citrullinated proteins are found in sites of inflammation, including lining and sub lining cells of inflamed rheumatoid joints [59, 134, 135]. In the initial case-control association study, the authors also demonstrated that a PADI4 haplotype consists of multiple coding SNPs affected the stability of RNA transcript, suggesting increased expression and function of PADI4 may be involved in RA pathogenesis.

The PADI4 polymorphisms have been repeatedly shown to associate with RA susceptibility in patients of Asian ancestry [10, 11, 136-147], but their contribution to RA susceptibility in European populations was weak and remains inconclusive [24, 148-155]. In the largest UK study totaling 19,000 individuals, no association was found for PADI4 variant rs2240340 (as

well as 56 other PADI4 SNPs) [150]. A very recent published large-scale GWAS meta-analysis between individuals of Asian descent from Japan (9,349 RA cases and 38,575 controls) and individuals of European descent (5,277 RA cases and 20,169 controls)revealed that the genome-wide significant association of PADI4 variant rs766499 was evident in the Japanese RA population with an estimated OR of 1.17 (95% CI 1.11 to 1.24), whereas nominal association was observed in the European populations with an estimated OR of 1.09 (95% CI 1.03 to 1.15, P_{GWAS}=0.0022) [12].

Although the association between the PADI4 gene and RA in the Asian populations has been convincingly determined, as yet, it is debatable to what extent the PADI4 genotype is associated with RA in European populations. Evidence from previous studies revealed that the risk variant frequency of RA susceptible SNPs in Asian individuals e.g. the rs2240340 variant was comparable to those in Europeans, thus, the genetic background is believed not the best explanation for the lack of association in patients of European descent. It is possible that the as-yet-unidentified environmental exposure may modulate the biological effect of PADI4 variant in different populations. Despite the ethnic differences observed in the PADI4 association with RA, antibodies to citrullinated peptide antigens (ACPA) are universally recognized and widely believed as an important serologic marker in RA, the pathway of citrullination by PADI, could, therefore be a common pathological pathway in RA even in European populations.
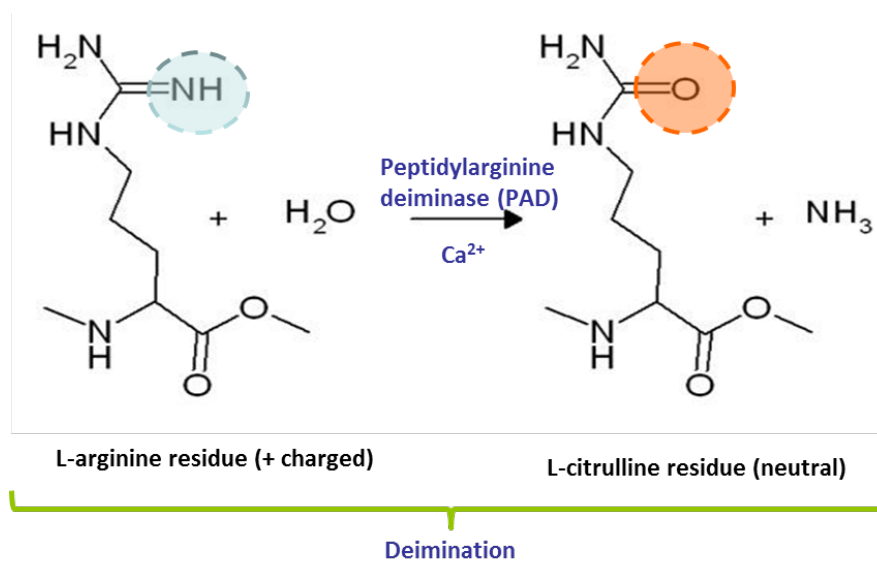

Figure 3. Conversion of arginine to citrulline by peptidylarginine deiminase (PAD) in the presence of calcium ions i.e. deamination or citrullination. *A post translational calcium-dependent enzymatic process results in the loss of one positive charge for every arginine residue to neutral citrulline. Amended from Klareskog L et al. Annu. Rev.Immunol.2008.*

3. ENVIRONMENTAL FACTORS AND RHEUMATOID ARTHRITIS

In addition to the vast data on novel genetic risk factors, an expanding catalogue of environmental exposures have been implicated in RA susceptibility [156-160], prevalent among these is cigarette smoking. Other factors that have been associated with the development of RA including reproductive/hormonal factors such as oral contraceptive use; air pollution, dietary factors including vitamin D and antioxidant intake; alcohol consumption and certain infections [159-171].

3.1 Smoking and rheumatoid arthritis

Smoking is by far representing the most prominent and consistently replicated environmental risk for developing RA [49, 51, 62, 132, 147, 172-174]. Several studies reported increased RA risk associated with cumulative smoking exposure [175, 176], denoting the importance of extent to the exposure in risk of RA development.

Recent studies revealed that the gene-environment interaction between HLA-DRB1 SE risk alleles and smoking conferred risk of RA development and restricted to only seropositive i.e. RF-positive and/or ACPA-positive RA subset, but not seronegative RA [19, 51, 62, 132, 140, 174]. An article published by our group recently exhibited that interaction occurred with HLA-DRB1*04 group as well as DRB1*01/*10 groups, implying that regardless of fine specificity, all the HLA-DRB1 SE alleles strongly interacted with smoking in the development of ACPA-positive RA in the Swedish Caucasian population [51].

Furthermore, research on the antigenic specificity of ACPA in gene-environment interaction has characterized a subpopulation of ACPA-positive RA, who exhibited anti-citrullinated α-enolase (CEP1) specificity (43-63%). The authors revealed that combined effect of the SE, PTPN22 and smoking showed strong interaction with CEP1, presenting extremely high OR of 37 for RA risk [177].

Although RA is relatively common in almost all populations of the world, smoking prevalence rates differed highly among the populations. For example, a recent epidemiological survey has shown that smoking prevalence was generally higher in men from Asia countries than in the European countries except in Sweden. Moreover, the gender gap in smoking rate was particularly large in most Asian countries i.e. Korea, Japan, Indonesia and China [178]. Therefore, the attribution of smoking to the development of RA may differ among populations. A recent report from Swedish EIRA study revealed that smoking was estimated to be responsible for 35% of the ACPA-positive RA. Additionally, 55% of ACPA-positive RA was attributable to smoking in individuals carrying two copies of the HLA-DRB1 SE risk alleles [179]. These findings indicated that smoking may affect disease risk differently in different individuals and smoking attribution to RA is important at the population level. Given that only one third of cases of ACPA-positive RA were attributed to smoking in this study, other environment and genetic factors may modify the effect of smoking on the development of RA in different populations including the Swedish population.

4. GENETIC MARKERS AND APPROACHES

Complex diseases are often referred to as multifactorial disorders because of the involvement of multiple genes and environmental exposures together with the complicated interactions between them. Understanding the genetic principal including the genetic markers selection, various study designs and also the genetic approaches, may thus help in planning and designing the experimental studies for unraveling the underlying etiology of RA.

4.1 Genetic markers

A genetic marker is a genetic variant that could be used to study the qualitative or quantitative traits. Genetic markers associated with certain diseases can be detected in the DNA samples and can thus be used to determine whether an individual is at risk of developing disease. The commonly used genetic markers include microsatellites and single nucleotide polymorphisms (SNPs).

4.1.1 Microsatellites

Microsatellites or SSR (single sequence repeats) or STR (single tandem repeats) are among the commonly used markers in genetic studies. They consist of a stretch of DNA with a few nucleotides long i.e. 2 to 6 basepair (bp) repeated several time in tandem e.g. CACACACACACACA. Microsatellites are relatively small in size, and can therefore easily be analyzed by PCR and gel electrophoresis.

4.1.2 Single nucleotide polymorphisms (SNPs)

A SNP is a single base mutation in DNA. SNPs are the simplest form of genetic variants, usually biallelic, and they are also the most common source of genetic polymorphism in the human genome. SNPs are present throughout the genome. They are highly abundant and widely distributed i.e. one SNP in every 1000bp, in human genome [180]. SNPs have become the popular and standard genetic marker used to identify associated alleles in different diseases. SNPs can be found across different regions within a particular gene with different roles. For example, a SNP may affect the protein structure if it is in the exon region; in promoter region, it may affect transcription activation; it may also influences the alternative splicing when it is in the intron region or even influences the mRNA stability if it is found in the 3' or 5' untranslated region (UTR). They may also be in linkage disequilibrium with the disease susceptibility locus.

4.2 Genetic approaches and strategies

Different genetic approaches and strategies have been effectively applied to identify risk loci in complex diseases including RA. The following description will be centered on two major approaches known as linkage analysis and association mapping and two strategies: GWAS and candidate gene approach.

4.2.1 Linkage analysis

Traditionally, the search for a disease gene begins with the family-based linkage analysis. Genetic linkage studies are usually performed to identify the estimate location of a gene relative to its variations designated as genetic marker with known position. This method has been phenomenally successful in mapping Mendelian variants, which are usually rare single gene disorders [181, 182]. Genetic linkage analysis is depends on cosegregation of chromosomal regions with a phenotypic trait within families. For most common autoimmune diseases such as RA, familial aggregation is rather modest, and therefore genetic linkage analysis has quite low statistical power to detect the chromosomal regions with shared genetic risk within families.

4.2.2 Association mapping
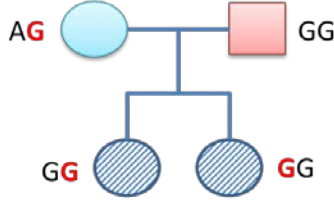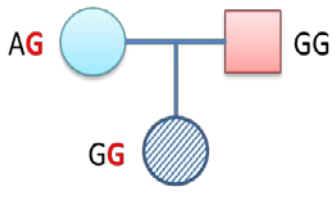
The basic idea of high density SNPs-based association analysis is to assess association between SNPs and a disease phenotype in a population. In this approach, the genetic causal variants are generally unknown and cannot be observed. Thus, the disease-gene association is identified through disease-marker associations, with the assumption that the genetic variants are in high LD with causal disease variants. Therefore, LD holds the main key in genetic association mapping. Association mapping approach can be implemented in case-control study design, both in family-based design (i.e. related controls) or non-family based design (i.e. unrelated controls derived from general population). Briefly, the transmission disequilibrium test (TDT) has been proposed as a family-based association test design, which tests for the presence of genetic linkage between a genetic marker within a family and this study design is not affected by population structure. The case control studies which form the majority of published report will be discussed separately in following chapter.

4.2.2.1 Case control study design

Case control studies are extensively used and the most powerful study design. A case control study compares the frequency of SNP alleles in two well-defined groups of individuals: cases that have been diagnosed with the disease under study, and controls, who are either known to be unaffected, or who have been randomly selected from the population. It is noteworthy that both methods of selections for the controls form a valid study. This study design also referred to as the non-family based design and is appropriate for studying relatively common SNP alleles that confer a modest or small effect on disease risk. An increased frequency of a SNP allele or genotype in cases when compared to the control subjects indicates that the presence of the SNP allele may associated with the disease risk. However, this study design could be confounded by the population stratification. For example, a good match between the genetic background of cases and controls is crucial for ensuring any genetic difference between them is related to the investigated disease, and not due to bias sampling. Taken together, careful selection of cases and controls is obligatory to warrant homogeneous genetic background, i.e. to perform the case control study within narrowly defined ethnic groups and avoid possible stratification [183]. Nonetheless, the limitation derived from population stratification can be detected and corrected by the means of bioinformatics computational methodology [184, 185].

In summary, although genetic linkage and association mapping shared the similar idea of 'recombination', these approaches have different characteristics for identifying the trait/disease loci with certain strengths and limitation as presented in Table 4 [186].

Table 4. Characteristics between linkage analysis and association mapping*

Characteristic		
Property of mapping approach	Linkage analysis	Association mapping
Data type studied	Relatives	Singleton patients and controls
Objective	Identify a single gene with large effect on phenotype	Identify common susceptibility variants underlying disease
Relevant parameter	Recombination fraction	Association statistic
Range of effect detected (linkage or association)	Long (≤5Mb)	Short (≤100kb)
Number of markers required for genome wide coverage	Moderate (500-1000)	Large (>100,000)
Statistics used	Cumbersome (requires tailor-made likelihood methods)	Elegant, can use the range of classical statistical tools
Dealing with correlated markers	Pose problems in presence of non-genotyped individuals	Can be handled efficiently
Biological basic of approach	Observe (or infer) recombination in pedigree data	Exploit unobserved recombination event in past generation
Dealing with allele heterogeneity	Not a problem	Reduces power
Detecting genotyping errors	Potentially detected as Mendelian inconsistencies	Potentially detected only in family data, but not in unrelated case control data
Most suitable application	Rare, dominant trait	Common trait

Adapted from Ott et al, 2011 Nature Genetics [186].

4.2.3 Genome-wide association studies (GWAS)

The GWAS approach is a hypothesis-free study design with advantage that no prior information about the genetic risk factors is required. We can assess tens or hundreds of thousands of SNP polymorphisms simultaneously by this approach. The basic concept of GWAS is based on 'common disease-common variants' model. For instance, since the disease is common, its presence may arise from a set of predisposing allele at multiple loci, each of which is itself common in the population. With the completion of the International HapMap Phase I and Phase II project [187, 188], ~3.2 million genotyping data in four populations of different ethnic origin (Caucasians of Northern and Western European origin, Japanese from Tokyo, Han Chinese from Beijing and Yoruba from Nigeria) were released into the public domain. HapMap provides information about the linkage disequilibrium (LD) patterns across the genome, thus allow the investigators underpin the design of SNP marker panels for GWAS. Most of the GWAS studies employ 50,000 or more SNPs across the genome and each addressing a separate hypothesis. For this reason, the statistical significance levels must be corrected for multiple testing. An overall p value of $<5\times10^{-7}$ is now widely accepted as compelling evidence of true association. The success and popularity of this approach has been marked by an online catalogue which contains 1,617 published GWAS reports at $p\leq5\times10^{-8}$ for 249 traits as at the third quarter of 2011 and is still ongoing [189]. A summary of these findings gives insight to the current genetic landscape of RA.

4.2.4 Candidate gene approach

Candidate gene studies have been a mainstay for human genetic studies for several decades. Candidate gene analysis is a hypothesis-driven approach, in which the candidate genes can be selected from biological pathway that harbors other previously associated risk loci. Depending on the size of the candidate region and cost implications, the investigator can choose to include all potentially functional SNPs residing in predicted regulatory regions, splice sites, intergenic sequence, introns or coding exons in marker selection. A good example of this is the 9p33 region which contains complement component 5(C5) and TNF-receptor-associated factor 1(TRAF1) genes. The first association between this region and experimental arthritis was observed in a genome-scan of mice [190]. Based on these findings, Kurreeman et al went on to investigate this disease locus in RA in a case control study population and found evidence for association with RA [25]. The authors concluded that TRAF1/C5 region increases the susceptibility to and severity of RA, possibly by influencing the structure function, and/or expression levels of TRAF1/C5.

4.2.5 Population stratification

Case control association studies assume that any difference in SNP genotypes between cases and control is due to their difference in disease status, but not any difference in their genetic background. However, the genotypic differences between cases and controls could sometime generate as a result of the differences between population ancestries, instead of

any effect on the disease risk [191]. This phenomenon is generally referred as population stratification. Population stratification can lead to 1) the presence of spurious associations between the SNP and disease, and 2) lack of replication across many association studies. For example even if a well-designed population case control study carefully draws the cases and controls from the same population, a hidden fine-scale genetic substructure within the single population (or the unintentional inclusion of individuals from other population) cannot be rule out. Furthermore, confounding occurs when the population substructure is not equally distributed between the cases and control groups. Therefore, after giving careful consideration to matching cases and controls on population origins [192], investigators should always examine and characterize the potential population stratification during the quality control process. Efforts should than be made to remove or reduce the effect of population stratification by removing the individuals of divergent ancestry. It is possible to correct for unknown population stratification using the entire set of genotyped SNP markers. This is commonly done using the genomic control approach [185, 193] and principal component approach [184].

4.2.5.1 Principal component analysis (PCA)

The most commonly used method for detecting and subsequently removing individuals with large-scale differences in ancestry is principal component analysis (PCA) [184, 194]. PCA is a multivariate statistical method used to produce several uncorrelated variables termed principal components, from a data matrix containing observations across a number of potentially correlated variables. The principal components are calculated to that the first PCA for as much variation occurs in the data in a single component, following by the second component and so on. A common set of approximately 50,000 independent markers must be used for the model-building. Region of extended high LD i.e. the HLA region should be excluded before analysis because these can substantially influence the PCA model [184].

4.2.6 Meta-analysis

Meta-analysis is a statistical technique for combining the findings from independent studies, which address a set of related research hypotheses. Meta-analysis can increase effective sample size as to improve the statistical power of analysis. One of the main advantages is that it can reduce the probability that random error will produce spurious association. Meta-analysis overcomes the problem of small sample size and inadequate statistical power in genetic studies of complex diseases. Apart from this, meta-analysis also evaluates the existence of heterogeneity. The presence or absence of heterogeneity influences the subsequent method of analysis. For example, if heterogeneity is absent, then the analysis employs the fixed effect modeling. This assumes the size of genetic effect is the same (fixed) across all studies investigated and the variation between the studies occurs by chance. Conversely, random effect model assumes that the genetic effect really does vary between studies. Such models tend to increase the variance of the summary measure making it more difficult to obtain significant outcomes [195, 196].

5. AIMS OF THE THESIS

The main goal of this thesis was to identify the genetic risk factors that predispose to rheumatoid arthritis in the multiethnic Malaysian population and to compare the results between the three major ethnic groups i.e. Malay, Chinese and Indian.

More specifically, the aims of this thesis were:

- To determine the frequencies of the HLA-DRB1 shared epitope (SE) alleles and to test the hypothesis whether the selective association between ACPA-positive RA and certain HLA-DRB1 alleles is valid in the Malaysian patients with RA

- To investigate the gene-environment interaction between smoking and HLA-DRB1 SE alleles (including those that are rare in European populations) in different subsets of RA defined by ACPA status, in a multiethnic population of Asian descent

- To explore the association between PADI4 polymorphism and RA risk in the Malaysian population residing in South East Asia with the goal of elucidating generalizability of association in Asian populations.

- To identify the possible association of DCIR polymorphism with RA subsets in four independent Asian populations and further validate the ACPA-negative RA association, originally detected in European population, in Asian cohort.

6. METHODOLOGICAL CONSIDERATIONS

This thesis is mainly based on a multicenter case control study entitled the Malaysian Epidemiological Investigation of Rheumatoid arthritis (MyEIRA) similar to the Swedish EIRA study. The study was established in August, 2005 by Allergy and Immunology Research Center and Biostatistics and Epidemiology Unit, Institute for Medical Research, Kuala Lumpur, Malaysia in collaboration with Rheumatology Unit, Department of Medicine and Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. Together with genetic risk factors investigation of RA, epidemiological study for environmental risk factors for RA was initiated and will complement my thesis research.

6.1 Malaysian background

Malaysia is a country in South East Asia and consists of two geographical regions divided by the South China Sea (latitude 2°30'N, longitude 112°30'E) with a population of 28.3 million people, comprising of multiethnic groups (Malay 63.1%, Chinese 24.6%, Indian 7.3%, others 5%)(http://www.statistics.gov.my/portal/download_Population/files/census2010)(Figure 4). History has shown that Malaysia's geographical position and location in between the great civilizations of Europe and the Middle East to the West and China and Japan to the East. Thus, Malaysia is a natural meeting spot of trade routes and ultimately formed a multi-racial and multicultural nation.
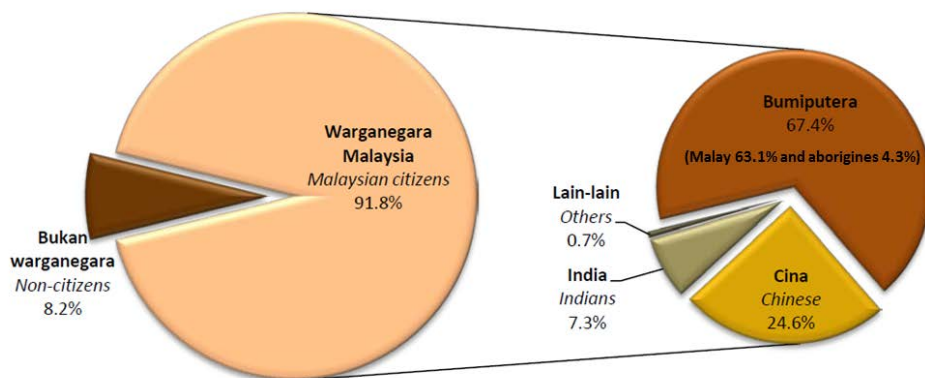
Figure 4. Percentage distribution of the population by ethnic group, Malaysia, 2010. *Source of data: http://www.statistics.gov.my/portal/download_Population/files/census2010.*

6.2 Study populations

The MyEIRA study is a sister study of the Swedish EIRA [197], consisting of mainly early RA defined according to the American College of Rheumatology (ACR) 1987 criteria [7]. RA cases were identified from eight rheumatology centers in Peninsular Malaysia. Controls were randomly selected and matched to cases for gender, age and residential area. In our genetic analyses, we further matched the cases and controls by ethnic groups i.e. Malay, Chinese, Indian and others.

Blood samples were collected from all the participants. Information about environmental exposures was collected by means of face-to-face interviews by well-trained interviewers with a modified questionnaire based on the Swedish EIRA questionnaire. Clinical data were also recorded. In total 1,260 cases and 1,625 controls were recruited between August 2005 and December 2009.

In paper IV, we also used information from the Han Chinese case control study including 1,193 cases (both early RA and established RA) and 1,278 unrelated controls.

Detailed information regarding the selection and exclusion criteria for the MyEIRA study can be found in the respective paper/ manuscript included in this thesis and in our smoking exposure article [198].

6.3 Ethical approvals

All participants gave their written consent. The MyEIRA study was approved by the Medical Research and Ethics Committee, Ministry of Health, Malaysia (KKM/JEPP/02 Jld.1 (86); (14)dlm.KKM/NIHSEC/08/0804/MRG-2005-12) and Stockholm Regional Ethics Committee, Sweden (2012/1381-31/1). Whereas, the Han Chinese case control study was approved by the Medical Ethics Committee in Peking University People's Hospital.

6.4 Laboratory considerations

In this section, I have selected to discuss about the cells isolation as the source of DNA and the Luminex application used for HLA typing. Both methods are routinely performed in our histocompatibility and immunogenetic laboratory at IMR at Kuala Lumpur in Malaysia.

6.4.1 Ficoll-hypaque density gradient separation

We used the density gradient centrifugation technique for separating the buffy-coat enriched cells (mainly monocytes and lymphocytes). The basis for this cell separation assay is the differential migration of cells during centrifugation according to their buoyant density,

which results in the separation of different cell types into distinct layers [199, 200]. Briefly, the diluted blood sample was layered onto a Ficoll-sodium metrizoate gradient of specific density i.e. 1.077 g/ml, following centrifugation highly purified buffy-coat with enriched lymphocytes and monocytes are harvested from the plasma-Ficoll interface. This method has been found to be rapid, simple and robust with very low contamination of erythrocytes (i.e. in which the presence of erythrocytes may hamper the DNA extraction process, therefore affecting the yield and quality of DNA). A brief protocol is presented in Figure 5. The DNA was extracted using the commercial QIAamp DNA Blood Mini kits from QIAGEN (Hilden, Germany). The concentration and quality of DNA were assessed by optical density at 260/280nm with NanoDrop™ 1000 spectrophotometer (Thermo Fisher Scientific, U.S.A).

6.4.2 The Luminex application to HLA genotyping

Luminex technology is a new flow cytometry technology enabling us to analyze numerous reactions in a single tube or well. It is a multiplexed data acquisition and analysis platform of microsphere-based assays that performs simultaneous measurements of up to 100 different analytes.

HLA typing using Luminex is a reverse polymerase chain reaction sequence specific oligonucleotide (PCR-SSO) system which involves PCR amplification of targeted regions within the MHC Class I or Class II regions with group specific primer. The primers used for amplification are biotinylated, and the PCR product is then denatured and allowed to hybridize to complementary DNA probes conjugated to up to 100 fluorescently code beads. Bound amplicon is detected by labeling with a Streptavidin-Phycoerytherin (SAPE) conjugate, with Streptavidin binding to the biotin used to label the primers and phycoerytherin serving as the reporter dye for the presence of bound amplicon. Subsequently, the Luminex platform is used to identify any SAPE bound to the beads. The observed reaction patterns are used to assign HLA type. Positive and negative control beads are used to quality control the typing assay. The flow of this method is illustrated in Figure 6.

The combinations of speed and reproducibility form the main advantages of Luminex as used for HLA typing. This technique is fairly robust and requires very little DNA i.e. 1µl of 20ng/µl DNA with purity between 1.65-1.80. The Luminex technique is suitable for both high and low throughput laboratories as the entire process takes place in a single well of 96 well PCR plate and thus, different loci can be process on the same tray. The use of advance probe technology reduces the number of ambiguities usually encountered in HLA typing. Furthermore, as compared to PCR-SSP, Luminex offer a safer method with the elimination of the use of agarose gel electrophoresis and its associated use of ethidium bromide.
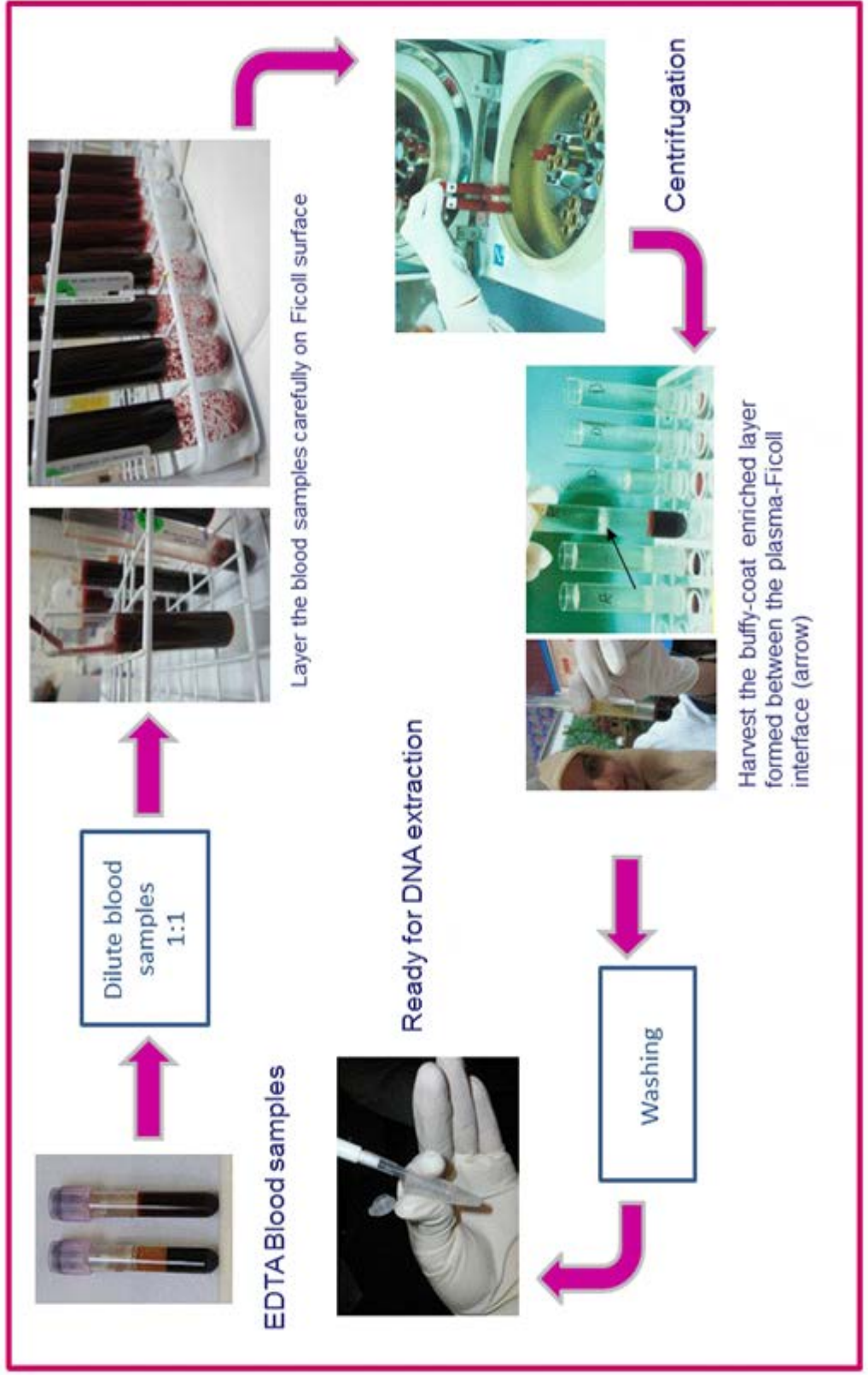
Figure 5. Flow chart showing the Ficoll-hypaque density gradient separation.
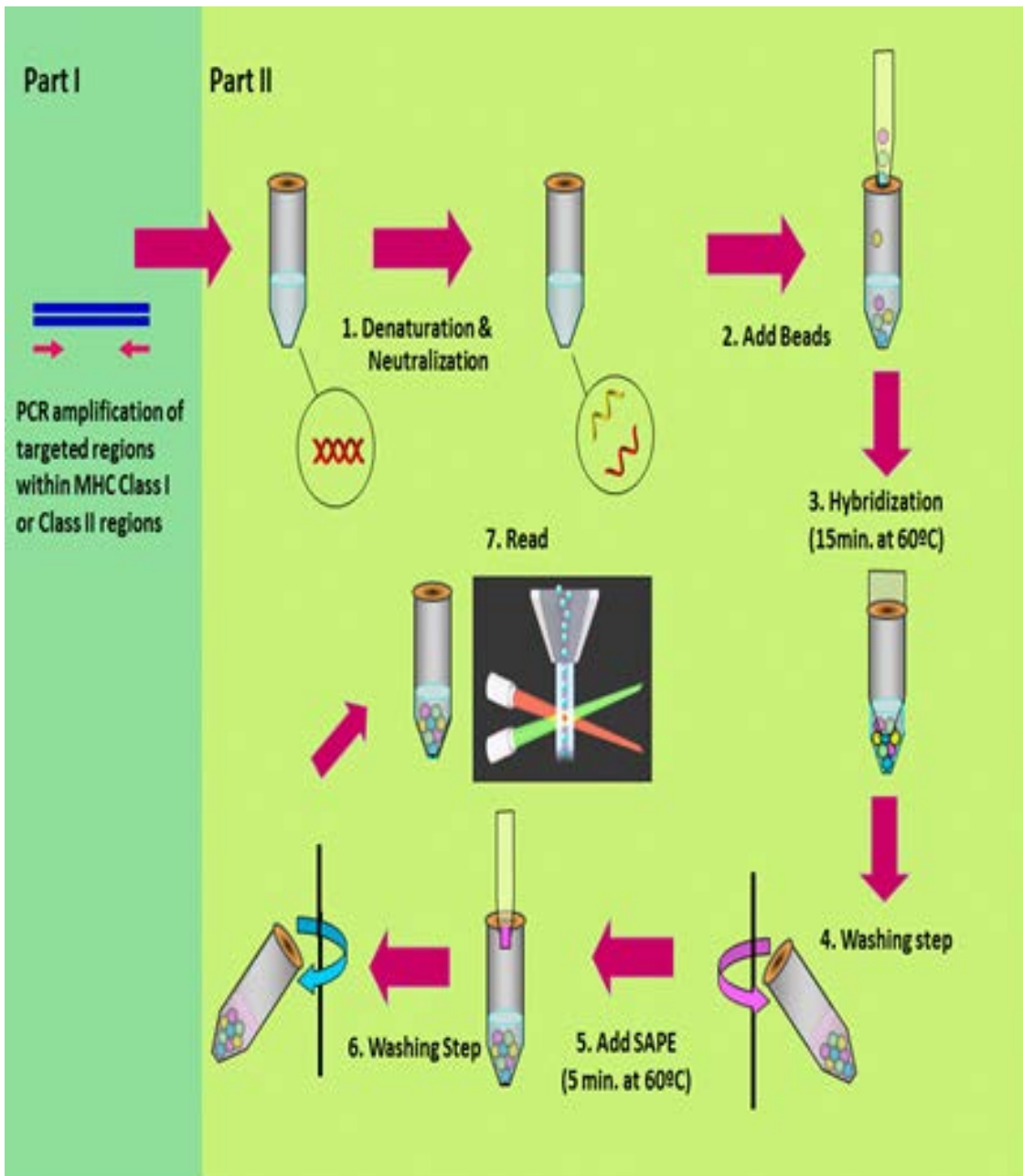
EDTA Blood samples

Dilute blood samples 1:1

Layer the blood samples carefully on Ficoll surface

Centrifugation

Harvest the buffy-coat enriched layer formed between the plasma-Ficoll interface (arrow)

Washing

Ready for DNA extraction

Figure 6. Flow chart showing the protocol for HLA genotyping using Luminex technology.

7. RESULTS AND DISCUSSIONS

The knowledge on disease pathology and risk assignment in RA is mainly based on studies of European populations and limited to a few studies of East Asian populations from Japan, Korea and China. These East Asian populations are generally referred to as the Asian populations in the published literature.

Malaysia is a unique multiethnic country in the South East Asia region, representing the genetic diversity across multiple large populations of Asian descent i.e. Malay, Chinese and Indian. In Malaysia, the three major ethnic groups are living side by side thereby sharing many environmental factors. There is low degree of genetic mixing between the groups, which allows the comparison between genetically divergent groups partly sharing environmental influences.

The MyEIRA study is the first large Asian case control study, performed in Malaysia and even in South East Asia region, involving mainly early RA patients, who are matched epidemiologically to their controls. With the following results, we aimed to elucidate the genetic predisposition to RA in this population and associate the similarities and differences in disease risk with other ethnic populations in the world.

Only the main results from Paper I – Paper IV are presented and discussed in this section. Additional information can be found in the complete paper under the Paper I – Paper IV section of this thesis.

7.1 HLA-DRB1 alleles association

7.1.1 Evaluation in subsets of rheumatoid arthritis (PAPER I)

Compelling evidence from the past few years has indicated that HLA-DRB1 shared epitope (SE) risk alleles are associated only with a subset of RA that is characterized by the presence of ACPA or rheumatoid factor [60, 62, 132, 174]. Recently, a GWAS performed in ACPA-positive RA and ACPA-negative RA further support the idea that distinct risk factors operate in these subsets of RA defined by ACPA status [16], and therefore these subsets should be regarded as two separate disease entities and studied separately in both genetic and functional studies of RA pathophysiology.

Overall, our study demonstrated that HLA-DRB1 SE-positivity confers higher risk in ACPA-positive RA regardless of the ethnicity as presented in Table 5, thus providing supporting evidence for the theory that the selective association between ACPA-positive RA and HLA-DRB1 SE alleles as observed in Europeans is also valid in non-European populations. In addition, our findings also revealed that the occurrence of ACPA was linked to the presence of HLA-DRB1 SE alleles in a gene-dose dependent manner, consistent with those in individuals of European descent [62, 174].

In RA patients, specific subtypes of DRB1 alleles have been consistently reported as increased in many populations, however, the subtypes associated with RA differs between ethnic groups. In populations of European ancestry for instance, DRB1*0401 and DRB1*0404 and DRB1*0101 are the most common alleles associated with RA susceptibility, while DRB1*0405 has a preponderance in Asian populations [34, 52, 55, 201]. Our current data corroborated that the frequency of DRB1*0405 was increased sizably in the Malay and Chinese patient with ACPA-positive RA. Limited increased frequency of DRB1*0405 was noticed in Indian ethnic groups, both in cases and controls, as compared with other ethnic groups. The findings may be explained by the dominance of DRB1*0403 allele in the Indian population. Interestingly, while DRB1*1001 was rarely present in the European populations [34, 35], it was significantly associated with increased risk of developing ACPA-positive RA in all three ethnic groups. Conversely, DRB1*0101 was infrequently found in the Malaysian population.

We demonstrated a novel inverse association of DRB1*1202 alleles and ACPA-positive RA in Malay and Chinese ethnic groups, independent of DRB1 SE alleles. This association was also observed in Malay patients with ACPA-negative RA. It is worth noting that DRB1*1202 is a common allele found in Asian populations, but rare in European populations (Figure 7) [202]. Fascinatingly, a very recent publication from Japanese population identified DRB1*1201 as novel RA susceptibility allele in ACPA-negative RA. The authors also provided evidence that combination of DRB1*1201 and DRB1*0901 exhibited remarkable association in this RA subset [203]. On the contrary, our findings showed that DRB1*1201 did not associate with ACPA-negative RA but rather with ACPA-positive RA in the Malay population. It would be interesting to perform meta-analysis for found association of RA with different DRB1*12 subtypes in different oriental populations, including the Japanese and our cohort.

Table 5. Risk of developing ACPA-positive or ACPA-negative RA among carriers of SE (Single, double or any) in the MyEIRA study population by ethnicity

	No SE*		Single SE			Double SE			Any SE		
	Cases	Controls	Cases	Controls	OR (95% CI)	Cases	Controls	OR (95% CI)	Cases	Controls	OR (95% CI)
Malay											
ACPA-positive	140	600	100	79	5.29 (3.74–7.48)	16	4	17.14(5.64–52.07)	116	83	5.85(4.18–8.18)
ACPA-negative	119		30		1.87 (1.18–2.97)	0		1.78 (1.12–2.82)	30		1.82(1.15–2.89)
Chinese											
ACPA-positive	72	147	54	17	6.49 (3.51–11.98)	6	0	0.00(-)**	60	17	7.20 (3.92–13.23)
ACPA-negative	62		9		1.26 (0.53–2.97)	1		0.00(-)#	10		4.74 (0.16–143.20)
Indian											
ACPA-positive	88	148	98	56	2.75 (1.81–4.16)	14	5	4.71 (1.64–13.52)	112	61	2.89 (1.93–4.34)
ACPA-negative	65		29		1.10 (0.65–1.87)	2		1.09 (0.65–1.82)	31		0.91 (0.17–4.82)

*cases and controls without SE alleles were used as reference group; **Fisher exact two-tailed p value, p=0.0015; #Fisher exact two-tailed p value, p=0.3, NS.
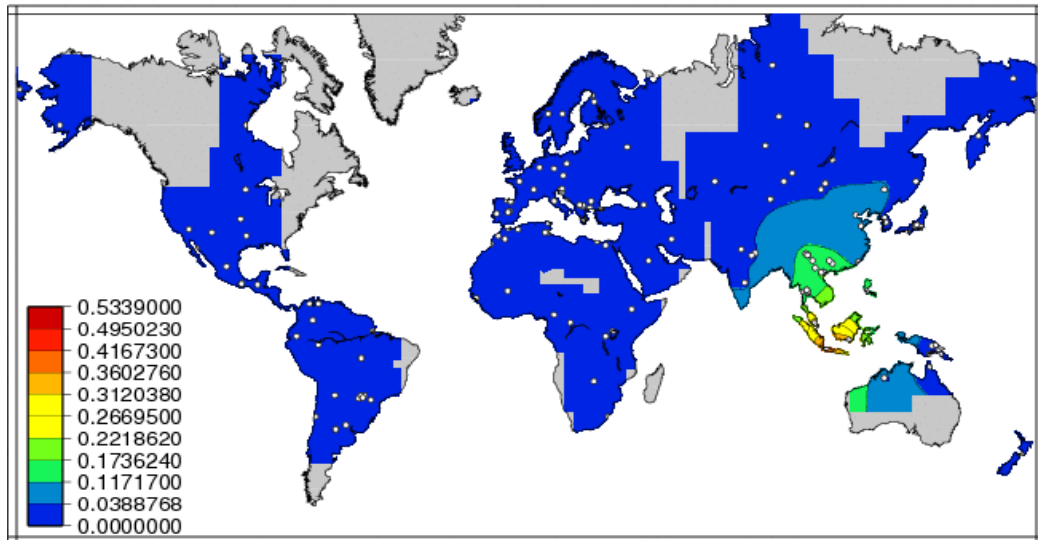
Figure 7. World-wide distribution of DRB1*1202 allele. *Adapted from Solberg et al, Hum Immunol, 2008 (www.pypop.org/popdata)*

7.1.2 Interaction with smoking (PAPER II)

Until recently, the role of genes, environment and immunity as risk factors for RA has mainly been studied independently. However, growing evidence revealed that these parameters indeed do not act exclusively independently but rather interact with each other contributing to RA development [49, 51, 62, 132, 157, 159, 174, 175, 177].

In Paper II, we investigated the gene-environment interactions in major subsets of RA dichotomized by ACPA status, taking into consideration smoking as the well-established environmental risk factors along with the presence of DRB1 SE alleles as the best-known genetic risk factor. We could compellingly demonstrate that smoking and the SE alleles were associated with increased risk of developing ACPA-positive RA, but not ACPA-negative RA (Figure 8). Strikingly, the combination of smoking and SE alleles reveals very high risk for ACPA-positive RA (OR=25.6, 95% CI, 10.4 to 63.4).
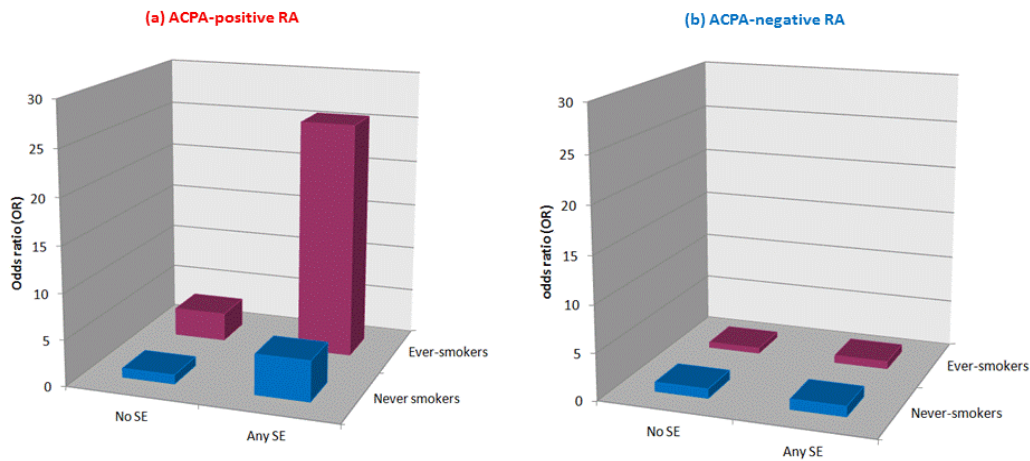
Figure 8. Odds ratios of developing ACPA-positive and ACPA-negative RA for different combinations of smoking and SE alleles. *The odds ratios of developing (a) ACPA-positive RA (b) ACPA-negative RA were compared with those who never smoked and have no SE alleles.*

The interaction between smoking and the presence of SE alleles was termed as departure from the additivity of effect [204] and was quantified by calculating the attributable proportion due to interaction (AP) together with 95% CI. Briefly, the AP is the proportion of the incidence among individuals exposed to two interacting factors, being attributable to the interaction per se i.e. the combined effect exceeding the sum of their independent effect (Figure 9).
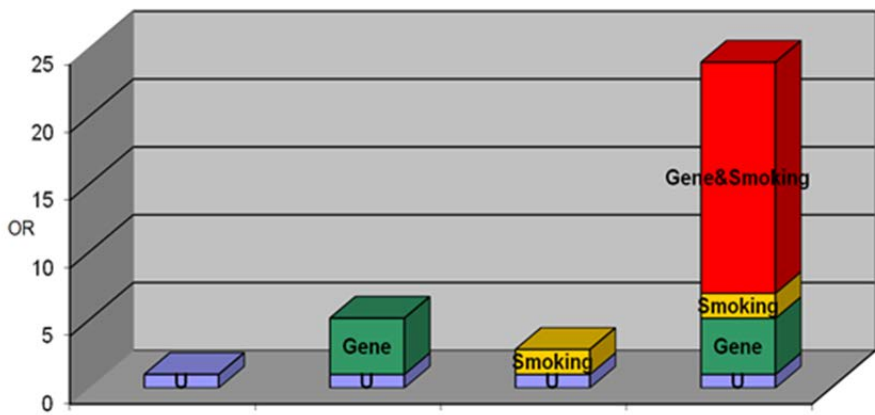
Figure 9. A schematic representation of increased risk due to additive interaction. *The attributable proportion (AP) is the percentage extra risk (in red), not explained by the sum of separate risks (i.e. gene and smoking), of the total risk.*

Our group has previously reported that regardless of fine specificity, all DRB1 SE alleles strongly interact with smoking in conferring an increased risk of ACPA-positive RA in patients of European descent [51]. Applying the same case control study design and analysis strategy, we could demonstrate that ever smokers carrying the Asian common DRB1 SE allele i.e. DRB1*0405 had a 12.9-fold increased risk for developing ACPA-positive RA, though the interaction between smoking and DRB1*0405 allele was not statistically significant (AP =0.4; 95% CI, -0.1 to 0.9).

7.2 Generalizability of PADI4 gene association (PAPER III)

Previous GWAS linkage studies of RA sibling pairs have shown that a susceptibility locus at 1p36 which confers risk to RA [205-208]. The gene region contains clusters of peptidylarginine deiminase (PADI) genes encoding enzymes, which are responsible for converting the peptidylarginine to peptidylcitrulline in the presence of calcium ions, thus, may change the conformation and functional properties of target proteins after citrullination [209]. In addition, a series of PADI activities including expression of PADI enzymes, protein citrullination and production of ACPA occurs in the synovium of RA patients, pinpointing its importance in disease pathogenesis [210]. Compelling evidence revealed that association at PADI4 gene locus is largely limited to RA patients of Asian descent [81, 137-140, 142-147, 152].

In this study, we aimed to determine if the association between PADI4 polymorphisms and RA risk could be generalized for different oriental populations, outside of the East Asian region. Additionally, we also searched for possible RA associations across the PADI locus including PADI1, PADI2, PADI3, PADI4 and PADI6 genes.

Given that association in case control studies can be potentially confounded by population stratification if the two study groups are poorly matched for genetic ancestry, also, the PADI4 gene association with RA was believed to be ethnic-dependent, we undertook the study carefully throughout the investigation of PADI association with RA within different ethnic groups in the study. In our analysis, we performed the principle component analyses (PCA) for cases and ethnically matched controls for the Malay, Chinese and Indian populations separately to detect and correct for possible population stratification (Figure 10) [184]. No substantial overlap between the groups was observed.

Accordingly, we convincingly validated the association between PADI4 polymorphisms and RA risk in multiethnic populations of Asian descent. Meta-analysis of our current data and the previous published data from Asian populations on PADI4 SNP rs2240340 variant [81, 137, 143, 145, 152, 211], provides strong evidence supporting PADI4 as an RA susceptibility gene diverse Asian populations (Figure 11). Hitherto, our study exhibited a possible novel association between PADI2 SNP rs1005753 variant and RA in the multiethnic Malaysian population (Figure 12).

It is also important to mention in this thesis that our single point analyses of variations across the various genes in PADI locus revealed modest effect size in RA population and the peak association varies between different ethnic groups (Figure 13).
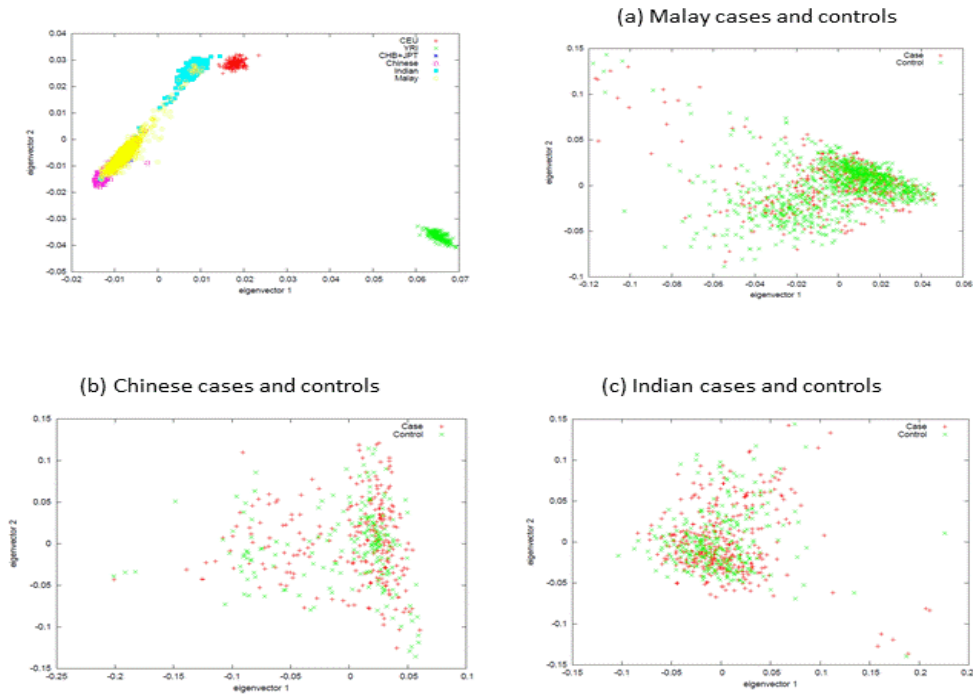

Figure 10. Principal component analysis (PCA) plots of genetic diversity across HapMap data and MyEIRA study population. *PCA plots of genetic diversity for each of the case control study in each ethnic group are presented in (a) Malay (b) Chinese and (c) Indian. Cases and controls for each ethnic group represented by red and green, respectively.*
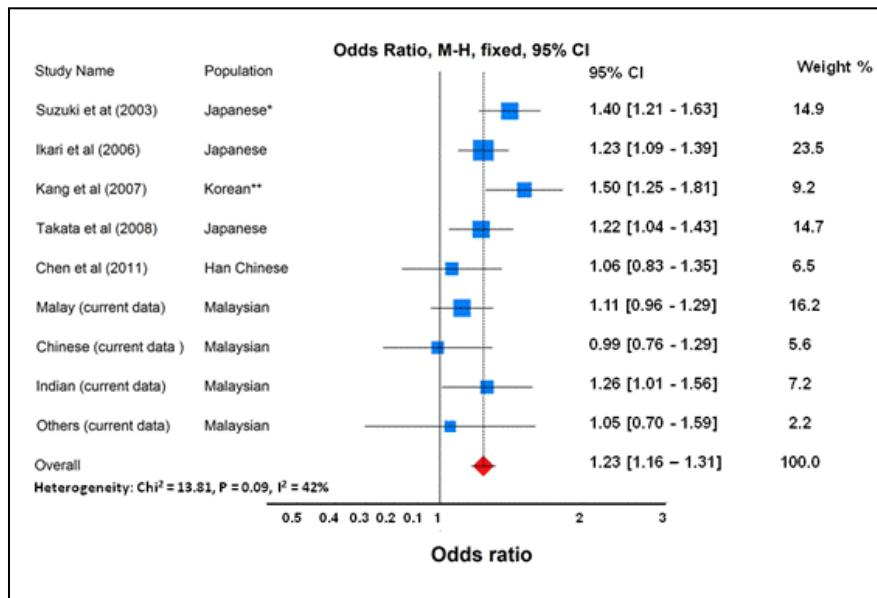
Figure 11. Meta-analysis of allelic association between PADI4 rs2240340 variants and risk of developing RA in Asian populations.
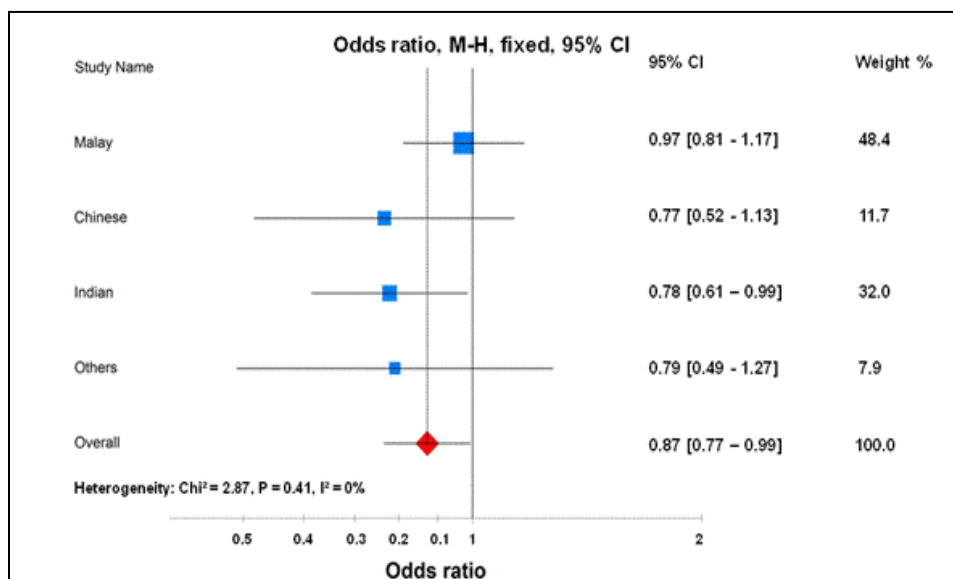
Figure 12. Meta-analysis of the allelic association between PADI2 rs1005753 variants and risk of developing RA in the MyEIRA study population
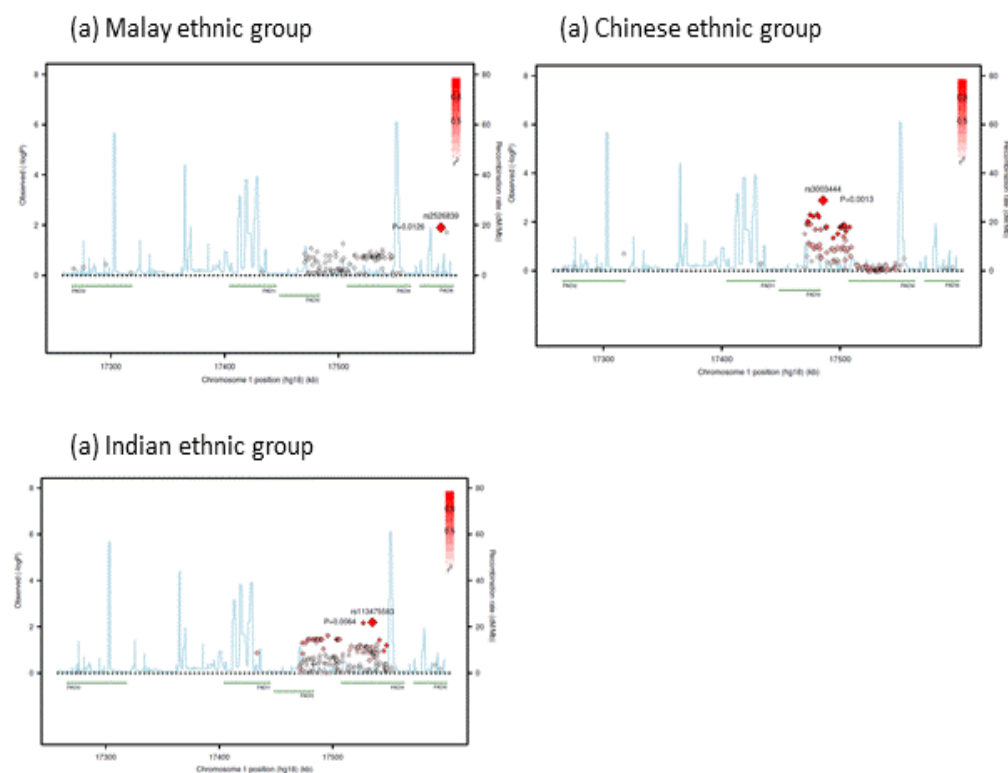
Figure 13. Regional association plots with recombination rate on the PADI genes for the three major ethnic groups from MyEIRA study. *The graphs are centered on the most significant SNP in each ethnic group.*

7.3 DCIR polymorphisms and ACPA-negative RA (PAPER IV)

Genetic factors contribute to the development of ACPA-negative RA as much as ACPA-positive RA [36] however, little is known about the ACPA-negative RA susceptibility alleles of HLA or non-HLA genes.

Dendritic cell immunoreceptor (DCIR) was mapped by Dr Johnny Lorentzen in our group, to a quantitative trait locus in a congenic rat strain protected from oil-induced arthritis. Subsequently, the DCIR variants were found to be associated with ACPA-negative RA in the Swedish population [28]. In this paper, we conducted a large case control study involving four independent Asian populations with the aim of elucidating the initially reported association of DCIR polymorphism with RA. While a consistent significant association between DCIR polymorphism and ACPA-negative RA was replicated in the Han Chinese population, we did not find any association between DCIR polymorphism and RA in all the three major ethnic groups from Malaysia. Nevertheless, when we performed meta-analysis using our present data together with the initial published data, we could demonstrate a significant association with ACPA-negative RA under a fixed effect model (OR=1.17; 95% CI 1.06 to 1.30), despite a somewhat significant heterogeneity across different ethnic groups (I^2=67%, p=0.02) (Figure 14).

Failure to replicate DCIR association in one ethnic group but not the other in not unforeseen and this finding may reflect an important clue about genetic architecture underlying different populations across the world [212]. For instance, a recent study documented population stratification within the Han Chinese population from China, with the main observed clusters corresponding roughly to northern Han Chinese and southern Han Chinese [213]. In this present study, our findings provide support for evidence that the Han Chinese individuals were mainly from the northern part of China, historically the Malaysian Chinese were mostly from southern China. This is a possible reason for the inconsistent association observed between these two seemingly homogeneous groups.

Abundant DCIR expression was observed in the synovial fluid and tissue of RA patients, but not detected in the joints of healthy individuals [214]. We showed an increased DCIR expression level in RA patients, compared with the healthy controls. In RA patients, the increased DCIR expression observed was associated with individuals carrying the risk variant (i.e. TC or CC genotype) regardless of their ACPA status. Although the DCIR expression was insignificantly associated with the ACPA-negative disease, this may possibly be due to the loss of statistical power. It is noteworthy that a recent study investigating the differential expression of DCIR revealed no difference of DCIR mRNA expression between RA cases and controls [215]. The possible explanation could be linked to the study design i.e. selection criteria of RA cases and or the sample size used.
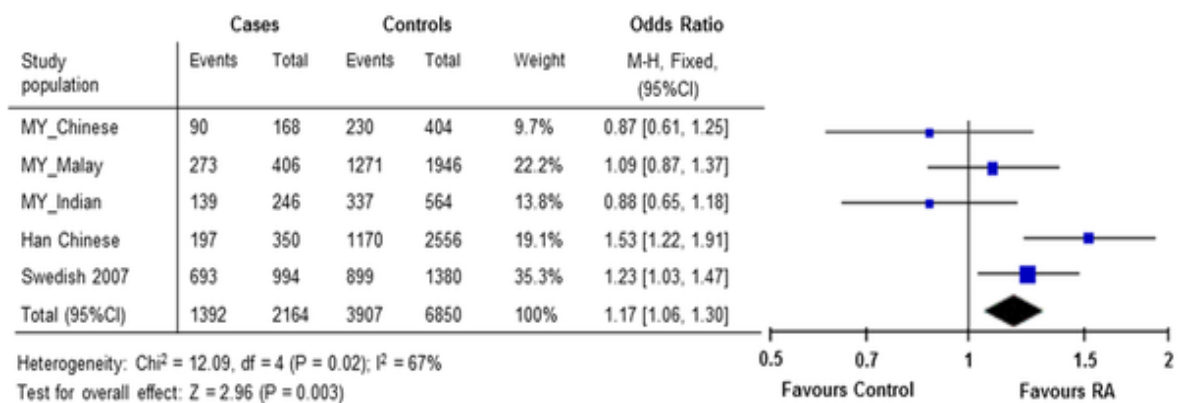
Study population	Cases		Controls		Weight	Odds Ratio M-H, Fixed, (95%CI)
	Events	Total	Events	Total		
MY_Chinese	90	168	230	404	9.7%	0.87 [0.61, 1.25]
MY_Malay	273	406	1271	1946	22.2%	1.09 [0.87, 1.37]
MY_Indian	139	246	337	564	13.8%	0.88 [0.65, 1.18]
Han Chinese	197	350	1170	2556	19.1%	1.53 [1.22, 1.91]
Swedish 2007	693	994	899	1380	35.3%	1.23 [1.03, 1.47]
Total (95%CI)	1392	2164	3907	6850	100%	1.17 [1.06, 1.30]

Heterogeneity: Chi² = 12.09, df = 4 (P = 0.02); I² = 67%
Test for overall effect: Z = 2.96 (P = 0.003)

Figure 14. Meta-analysis of rs2377422 for ACPA-negative RA across different ethnic groups.

8. CONCLUDING REMARKS

The vast majority of genetic studies in RA have centered on populations of European descent and/or a few from East Asian populations. The fact is that RA is caused by interplay of genetic variation and environmental factors. Genetic studies have revealed a substantial divergence of genetic variation across population in term of allele frequency; linkage disequilibrium and haplotype structure [216-218]. Taken together, the current findings on RA risk may provide an incomplete picture of the genetic basis of the disease.

In this thesis, our main goal is to determine the genetic predisposition to RA in a multiethnic Malaysian population residing in South East Asia. More specifically, we would like to address the questions as of how far the identified RA risk factors in both the Europeans and East Asian can be translated across different ethnic populations. Overall, our data further support different genetic backgrounds that were associated with the distinct subsets of RA characterized by ACPA status. We demonstrated that different DRB1 SE alleles, that are common in Asian (i.e. DRB1*0405), but not European populations confers significant increased risk of developing ACPA-positive RA, but not ACPA-negative RA in Malaysian population regardless of ethnicity. In addition, our analysis revealed a considerable heterogeneity of association signals between our population and those of European descent and East Asian descent. An example is the DRB1 SE risk alleles seemed to be shared, but the risk signal often appeared to be population-specific i.e. DRB1*0401, DRB1*0404 and DRB1*0101 in Europeans, DRB1*0405 and DRB1*0901 in Japanese populations and DRB1*0405 and DRB1*1001 in Malaysian population. This phenomenon also observed for the protective alleles i.e. we demonstrated a novel association between the DRB1*1202 and ACPA-positive RA in Malay and Chinese populations, while DRB1*13 alleles were documented in European patients with ACPA-positive RA. The possibility of analyzing these relationships further in different ethnic groups will be of major importance for our understanding of which immune reaction contributes to RA in different subsets of RA from different parts of the world.

By scrutinizing the gene-environment interaction between DRB1 SE alleles and smoking in the development of RA, we show that the risk of developing ACPA-positive RA is associated with a strong gene-environment interaction between DRB1 SE alleles and smoking in the Malaysian population. Notably, this interaction seem to apply to smoking and DRB1*0405 allele, which is common in Asian populations, but not in populations of European descent.

There is no doubt about the existence of the population-specific risk variants associated with RA, as it was clearly illustrated previously by the PTPN22 RA risk gene in European populations and PADI4 gene in Asian populations. However, to what extent the association particularly between PADI4 gene and RA could be demonstrated in other oriental populations' remains to be determined. In this thesis, we scrutinized the entire PADI locus including PADI1, PADI2, PADI3, PADI4 and PADI6 with a set of 320 SNPs markers for association with RA. Our findings revealed an association between PADI4 and RA in the diverse Asian populations from Malaysia and suggest an additional novel association in a PADI2 gene. The latter finding is awaiting replication in independent cohort or in other ethnic populations.

As yet, very few validated risk alleles associated with ACPA-negative RA, though their heritability was comparable to that of ACPA-positive disease. We investigated the association between the previously reported ACPA-negative-related DCIR polymorphisms and RA in four independent Asian populations, with special focus on the ACPA-negative RA subset. Our data provide evidence for association between the DCIR rs2377422 variant and RA in non-European populations. We also confirmed the influence of DCIR polymorphisms on RA susceptibility especially on ACPA-negative RA.

9. FUTURE DIRECTIONS

In the case of genetic studies of RA, it remains a great challenge if "we are different (i.e. by ethnicity) but basically the same (i.e. disease phenotype) or "we are the same (i.e. sharing common risk loci) but different (i.e. population-specific risk signals). However, I think both are valid for this thesis.

An impressive list of validated and highly suggestive non-MHC RA risk is growing after GWAS became available. However, these non-HLA risk alleles explain only 5% of the genetic burden of RA [79], indicating that additional non-MHC risk alleles remain to be discovered.

Existing genotyping technology captures most common variants that are believed to be ancient origin and shared among different populations. Nevertheless, RA risk can also be due to rare variants, which are more likely to be population-specific and could carry a greater risk. New technologies are required to detect the rare variants. For example, the whole genome sequencing that may help to facilitate detailed analysis of rare and structural variants. This will undoubtedly lead to the further identification and refinement of causal variation and thereafter will help to resolve some of the controversy surrounding the concept of 'missing heritability'. It is worth noting that the rare variants are individually difficult to identify, but, once detected, they can collectively make a substantial contribution to the genetic risk underlying RA in the near future.

While it is important to design a dense mapping genotyping array to capture more common variants as well as rare variants, it is also vital to conduct such mapping in non-European populations. This is because most of the genetic platforms today i.e. GWAS are designed for the optimal use in people of European ancestry, thus the outcomes may give an incomplete picture of the genetic makeup of RA. Moreover, the current approach limits us to globally assess the heterogeneity association signals between populations of European descent and Asian descent.

The RA risk variants identified so far have small or moderate risk effects and explain a small part of RA heritability as discussed earlier. Thus, when comparing the results of genetic studies between different populations, we are actually comparing a small fraction of the total genetic risk presence in each population. Larger case control cohorts will then be needed to detect such small effects. Yet, there are challenges which lie in collecting a larger cohort of homogeneous samples from a variety of different populations, as well as conducting genetic studies in a variety of well-characterized populations. These challenges however, could be possibly resolved by establishing international collaboration consortia as we are currently practicing and are reflected in this thesis. A meta-analysis of the pooled data from different populations could also increase the statistical powers for detecting modest effect exhibited in RA susceptibility genes.

In order to complete the understanding of genetic predisposition to RA, it is essential to study other types of polymorphisms that are also contributing to genomic variation i.e. copy number polymorphisms and repeat elements. Further dissection of genetic and environmental determinants and their interactions including gene-gene and gene-

environment interactions will help clarify genetic heterogeneity between different ethnic populations.

Accordingly, once the causal variant(s) have been identified at a given locus, functional studies such as on the regulation of gene expression and protein structure, location and function could be designed and followed.

The ultimate goal of genetic association studies is nevertheless, to translate the findings into clinical practice. For example identification of associated variants suggests biological pathway, which might lead to the intervention of therapeutic targets. It may be possible to develop a more personalized approach to disease management from the knowledge of a patient's genetic background.

Finally, it may be possible to use the information about genetic and environmental susceptibility factors for RA to identify high-risk groups in the general population with the aim of develop preventive strategies.

10. SAMMANFATTNING PÅ SVENSKA

Svensk titel: Genetisk predisposition för reumatoid artrit (ledgångsreumatism) i Malaysia

Har du någonsin tänkt på det här: "Vi är olika men i grunden lika som bär" eller "vi är lika som bär men olika"

Min avhandling handlar om likheterna och olikheter mellan europeiska (kaukasiska) populationer och asiatiska fram för allt malaysisk population och riken att utveckla RA. Vi vet att både genetiska och omgivningsfaktorer påverkar risken att få RA. Vi vet att immunologiska faktorer som vita blodkroppar och speciella antikroppar är avgörande, men hur ser sambandet ut mellan de immunologiska mekanismerna som orsakar sjukdomen och de faktorer som medför risk att utveckla reumatoid artrit **(**RA)?

De fyra delarbetena som ingår i avhandling behandlar just detta.

I delarbete 1 visar jag att speciella genvarianter (alleler) kallad shared epitope alleler, som är viktig för reglering av immunsystemet, ökar risken att få RA hos en grupp av patienter som har antikroppsmarkören ACPA, anti-citrullinated protein antibody. Man finner samma gen i europeiska som i malaysiska populationer men det är olika genvarianter (alleler) som är överrepresenterade i Malaysia jämfört med i Europa.

I delarbete 2 visar jag att de genvarianter som vi studerade i delarbete 1 den så kallade shared epitope alleler interagerar med rökning. Personer som har riskvariant och dessutom rökar har en mångfaldigt ökad risk för att utveckla ACPA-positivt sjukdomen. I motsats, ökar denna kombination inte risken för ACPA-negativt sjukdomen.

I delarbete 3 visar jag att det finns speciella genvarianter skilda från shared epitope som ökar risk för RA. Min studie visar att genvarianter i PADI-gener som inblandad i citrullinering av proteiner ökar risken för sjukdomen. Det som är speciellt här är att dessa genvarianter har hittills bara visats vara en riskfaktor för RA i asiatisk befolkning och inte i europeisk befolkning.

I delarbete 4 har jag tillsammans med kinesiska forskare visat genvarianter som kan påverka risken för att utveckla ACPA-negativt sjukdomen. De genvarianter i så kallade DCIR-genen som påverkar makrofagaktivering ökar alltså risken av ACPA-negativt sjukdomen. DCIR-genvarianter är däremot ger inte en risk av ACPA-positivt RA.

Sammanfattningsvis visar min avhandling på att det finns en avgörande koppling mellan genetiska faktorer jag har studerat i HLA-DRB1, PADI och DCIR gener, och omgivningsfaktor, jag har studerat rökning. Samtliga studerade faktorer har betydelse för immunsystemets aktivering. Långt flera gener och omgivningsfaktorer har betydelse för utveckling av sjukdom.

Det är mycket intressant att se att många av faktorerna är gemensamma för människor i olika delar världen–vi är lika. Men att det också finns skillnader i hur sjukdomen /sjukdomarna uppstår – vi är olika.

Med kunskapen om likheter och skillnader kommer vi att bättre kunna förhindra sjukdomar som ledgångsreumatism och specialdesigna behandlingar som på ett vetenskapligt sätt passar olika populationer och till slut passar varje individ.

11. ACKNOWLEDGEMENTS

12. REFERENCES

1. Klein J, Sato A: **The HLA system. First of two parts**. *The New England journal of medicine* 2000, **343**(10):702-709.
2. 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 and rheumatism* 2000, **43**(1):30-37.
3. Alamanos Y, Voulgari PV, Drosos AA: **Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review**. *Seminars in arthritis and rheumatism* 2006, **36**(3):182-188.
4. Tobon GJ, Youinou P, Saraux A: **The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis**. *Journal of autoimmunity* 2010, **35**(1):10-14.
5. Del Puente A, Knowler WC, Pettitt DJ, Bennett PH: **High incidence and prevalence of rheumatoid arthritis in Pima Indians**. *American journal of epidemiology* 1989, **129**(6):1170-1178.
6. Silman AJ, Ollier W, Holligan S, Birrell F, Adebajo A, Asuzu MC, Thomson W, Pepper L: **Absence of rheumatoid arthritis in a rural Nigerian population**. *The Journal of rheumatology* 1993, **20**(4):618-622.
7. 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 and rheumatism* 1988, **31**(3):315-324.
8. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD *et al*: **2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative**. *Annals of the rheumatic diseases* 2010, **69**(9):1580-1588.
9. Myouzen K, Kochi Y, Okada Y, Terao C, Suzuki A, Ikari K, Tsunoda T, Takahashi A, Kubo M, Taniguchi A *et al*: **Functional Variants in NFKBIE and RTKN2 Involved in Activation of the NF-kappaB Pathway Are Associated with Rheumatoid Arthritis in Japanese**. *PLoS genetics* 2012, **8**(9):e1002949.
10. Danoy P, Wei M, Johanna H, Jiang L, He D, Sun L, Zeng X, Visscher PM, Brown MA, Xu H: **Association of variants in MMEL1 and CTLA4 with rheumatoid arthritis in the Han Chinese population**. *Annals of the rheumatic diseases* 2011, **70**(10):1793-1797.
11. Freudenberg J, Lee HS, Han BG, Shin HD, Kang YM, Sung YK, Shim SC, Choi CB, Lee AT, Gregersen PK *et al*: **Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci**. *Arthritis and rheumatism* 2011, **63**(4):884-893.
12. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl EA, Kurreeman FA, Nishida N *et al*: **Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population**. *Nature genetics* 2012, **44**(5):511-516.
13. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FA, Zhernakova A, Hinks A *et al*: **Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci**. *Nature genetics* 2010, **42**(6):508-514.
14. van der Helm-van Mil AH, Huizinga TW: **Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets**. *Arthritis research & therapy* 2008, **10**(2):205.
15. Kurreeman F, Liao K, Chibnik L, Hickey B, Stahl E, Gainer V, Li G, Bry L, Mahan S, Ardlie K *et al*: **Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records**. *American journal of human genetics* 2011, **88**(1):57-69.
16. Padyukov L, Seielstad M, Ong RT, Ding B, Ronnelid J, Seddighzadeh M, Alfredsson L, Klareskog L: **A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis**. *Annals of the rheumatic diseases* 2011, **70**(2):259-265.
17. Daha NA, Toes RE: **Rheumatoid arthritis: Are ACPA-positive and ACPA-negative RA the same disease?** *Nature reviews Rheumatology* 2011, **7**(4):202-203.

18. Makrygiannakis D, af Klint E, Lundberg IE, Lofberg R, Ulfgren AK, Klareskog L, Catrina AI: **Citrullination is an inflammation-dependent process**. *Annals of the rheumatic diseases* 2006, **65**(9):1219-1222.

19. Forslind K, Ahlmen M, Eberhardt K, Hafstrom I, Svensson B: **Prediction of radiological outcome in early rheumatoid arthritis in clinical practice: role of antibodies to citrullinated peptides (anti-CCP)**. *Annals of the rheumatic diseases* 2004, **63**(9):1090-1095.

20. Kastbom A, Strandberg G, Lindroos A, Skogh T: **Anti-CCP antibody test predicts the disease course during 3 years in early rheumatoid arthritis (the Swedish TIRA project)**. *Annals of the rheumatic diseases* 2004, **63**(9):1085-1089.

21. Ronnelid J, Wick MC, Lampa J, Lindblad S, Nordmark B, Klareskog L, van Vollenhoven RF: **Longitudinal analysis of citrullinated protein/peptide antibodies (anti-CP) during 5 year follow up in early rheumatoid arthritis: anti-CP status predicts worse disease activity and greater radiological progression**. *Annals of the rheumatic diseases* 2005, **64**(12):1744-1749.

22. 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 and rheumatism* 2004, **50**(7):2113-2121.

23. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN *et al*: **Rheumatoid arthritis association at 6q23**. *Nature genetics* 2007, **39**(12):1431-1433.

24. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D *et al*: **Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4**. *American journal of human genetics* 2005, **77**(6):1044-1060.

25. Kurreeman FA, Padyukov L, Marques RB, Schrodi SJ, Seddighzadeh M, Stoeken-Rijsbergen G, van der Helm-van Mil AH, Allaart CF, Verduyn W, Houwing-Duistermaat J *et al*: **A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis**. *PLoS medicine* 2007, **4**(9):e278.

26. Wesoly J, van der Helm-van Mil AH, Toes RE, Chokkalingam AP, Carlton VE, Begovich AB, Huizinga TW: **Association of the PTPN22 C1858T single-nucleotide polymorphism with rheumatoid arthritis phenotypes in an inception cohort**. *Arthritis and rheumatism* 2005, **52**(9):2948-2950.

27. Skinningsrud B, Lie BA, Husebye ES, Kvien TK, Forre O, Flato B, Stormyr A, Joner G, Njolstad PR, Egeland T *et al*: **A CLEC16A variant confers risk for juvenile idiopathic arthritis and anti-cyclic citrullinated peptide antibody negative rheumatoid arthritis**. *Annals of the rheumatic diseases* 2010, **69**(8):1471-1474.

28. Lorentzen JC, Flornes L, Eklow C, Backdahl L, Ribbhammar U, Guo JP, Smolnikova M, Dissen E, Seddighzadeh M, Brookes AJ *et al*: **Association of arthritis with a gene complex encoding C-type lectin-like receptors**. *Arthritis and rheumatism* 2007, **56**(8):2620-2632.

29. Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Ronnelid J, Klareskog L, Huizinga TW *et al*: **Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis**. *Arthritis and rheumatism* 2007, **56**(7):2202-2210.

30. Wang C, Kokkonen H, Sandling JK, Johansson M, Seddighzadeh M, Padyukov L, Rantapaa-Dahlqvist S, Syvanen AC: **Preferential association of interferon regulatory factor 5 gene variants with seronegative rheumatoid arthritis in 2 Swedish case-control studies**. *The Journal of rheumatology* 2011, **38**(10):2130-2132.

31. 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 and immunity* 2005, **6**(2):129-133.

32. Verpoort KN, van Gaalen FA, van der Helm-van Mil AH, Schreuder GM, Breedveld FC, Huizinga TW, de Vries RR, Toes RE: **Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis**. *Arthritis and rheumatism* 2005, **52**(10):3058-3062.

33. Irigoyen P, Lee AT, Wener MH, Li W, Kern M, Batliwalla F, Lum RF, Massarotti E, Weisman M, Bombardier C *et al*: **Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles**. *Arthritis and rheumatism* 2005, **52**(12):3813-3818.

34. Lundstrom E, Kallberg H, Smolnikova M, Ding B, Ronnelid J, Alfredsson L, Klareskog L, Padyukov L: **Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis**. *Arthritis and rheumatism* 2009, **60**(4):924-930.

35. van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, Houwing-Duistermaat JJ, Verduijn W, Nordang GB, Alfredsson L, Klareskog L *et al*: **Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations**. *Arthritis and rheumatism* 2010, **62**(5):1236-1245.

36. van der Woude D, Houwing-Duistermaat JJ, Toes RE, Huizinga TW, Thomson W, Worthington J, van der Helm-van Mil AH, de Vries RR: **Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis**. *Arthritis and rheumatism* 2009, **60**(4):916-923.

37. Milner CM, Campbell RD: **Genetic organization of the human MHC class III region**. *Frontiers in bioscience : a journal and virtual library* 2001, **6**:D914-926.

38. Imboden JB: **The immunopathogenesis of rheumatoid arthritis**. *Annual review of pathology* 2009, **4**:417-434.

39. Bowes J, Barton A: **Recent advances in the genetics of RA susceptibility**. *Rheumatology (Oxford)* 2008, **47**(4):399-402.

40. Seldin MF, Amos CI, Ward R, Gregersen PK: **The genetics revolution and the assault on rheumatoid arthritis**. *Arthritis and rheumatism* 1999, **42**(6):1071-1079.

41. Deighton CM, Walker DJ, Griffiths ID, Roberts DF: **The contribution of HLA to rheumatoid arthritis**. *Clinical genetics* 1989, **36**(3):178-182.

42. Stastny P: **Mixed lymphocyte cultures in rheumatoid arthritis**. *The Journal of clinical investigation* 1976, **57**(5):1148-1157.

43. Gregersen PK, Silver J, Winchester RJ: **The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis**. *Arthritis and rheumatism* 1987, **30**(11):1205-1213.

44. Wucherpfennig KW, Strominger JL: **Selective binding of self peptides to disease-associated major histocompatibility complex (MHC) molecules: a mechanism for MHC-linked susceptibility to human autoimmune diseases**. *The Journal of experimental medicine* 1995, **181**(5):1597-1601.

45. Bhayani HR, Hedrick SM: **The role of polymorphic amino acids of the MHC molecule in the selection of the T cell repertoire**. *J Immunol* 1991, **146**(4):1093-1098.

46. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD: **Defining the role of the MHC in autoimmunity: a review and pooled analysis**. *PLoS genetics* 2008, **4**(4):e1000024.

47. De Almeida DE, Ling S, Pi X, Hartmann-Scruggs AM, Pumpens P, Holoshitz J: **Immune dysregulation by the rheumatoid arthritis shared epitope**. *J Immunol* 2010, **185**(3):1927-1934.

48. Weyand CM, Goronzy JJ: **The role of infectious agents in the etiopathogenesis of chronic rheumatic diseases**. *Verhandlungen der Deutschen Gesellschaft fur Innere Medizin* 1990, **96**:319-324.

49. van der Helm-van Mil AH, Verpoort KN, le Cessie S, Huizinga TW, de Vries RR, Toes RE: **The HLA-DRB1 shared epitope alleles differ in the interaction with smoking and predisposition to antibodies to cyclic citrullinated peptide**. *Arthritis and rheumatism* 2007, **56**(2):425-432.

50. Mackie SL, Taylor JC, Martin SG, Wordsworth P, Steer S, Wilson AG, Worthington J, Emery P, Barrett JH, Morgan AW: **A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population**. *Genes and immunity* 2012, **13**(2):120-128.

51. Lundstrom E, Kallberg H, Alfredsson L, Klareskog L, Padyukov L: **Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated**

protein antibody-positive rheumatoid arthritis: all alleles are important. *Arthritis and rheumatism* 2009, **60**(6):1597-1603.

52. Liu SC, Chang TY, Lee YJ, Chu CC, Lin M, Chen ZX, Liu HF, Dang CW, Chang SC, Lee CS *et al*: **Influence of HLA-DRB1 genes and the shared epitope on genetic susceptibility to rheumatoid arthritis in Taiwanese**. *The Journal of rheumatology* 2007, **34**(4):674-680.

53. Xue Y, Zhang J, Chen YM, Guan M, Zheng SG, Zou HJ: **The HLA-DRB1 shared epitope is not associated with antibodies against cyclic citrullinated peptide in Chinese patients with rheumatoid arthritis**. *Scandinavian journal of rheumatology* 2008, **37**(3):183-187.

54. Lin L, Chen Y, Xiao Z, Huang S, Yang Z: **The association of HLA-DRB1 alleles with rheumatoid arthritis in the Chinese Shantou population: a follow-up study**. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 2007, **85**(2):227-238.

55. Lee HS, Lee KW, Song GG, Kim HA, Kim SY, Bae SC: **Increased susceptibility to rheumatoid arthritis in Koreans heterozygous for HLA-DRB1*0405 and *0901**. *Arthritis and rheumatism* 2004, **50**(11):3468-3475.

56. Wakitani S, Imoto K, Murata N, Toda Y, Ogawa R, Ochi T: **The homozygote of HLA-DRB1*0901, not its heterozygote, is associated with rheumatoid arthritis in Japanese**. *Scandinavian journal of rheumatology* 1998, **27**(5):381-382.

57. Chan SH, Lin YN, Wee GB, Koh WH, Boey ML: **HLA class 2 genes in Singaporean Chinese rheumatoid arthritis**. *British journal of rheumatology* 1994, **33**(8):713-717.

58. Furuya T, Hakoda M, Ichikawa N, Higami K, Nanke Y, Yago T, Kobashigawa T, Tokunaga K, Tsuchiya N, Kamatani N *et al*: **Differential association of HLA-DRB1 alleles in Japanese patients with early rheumatoid arthritis in relationship to autoantibodies to cyclic citrullinated peptide**. *Clinical and experimental rheumatology* 2007, **25**(2):219-224.

59. Klareskog L, Ronnelid J, Lundberg K, Padyukov L, Alfredsson L: **Immunity to citrullinated proteins in rheumatoid arthritis**. *Annual review of immunology* 2008, **26**:651-675.

60. Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D, Schreuder GM, Wener M, Breedveld FC, Ahmad N *et al*: **Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins**. *Arthritis and rheumatism* 2005, **52**(11):3433-3438.

61. Auger I, Sebbag M, Vincent C, Balandraud N, Guis S, Nogueira L, Svensson B, Cantagrel A, Serre G, Roudier J: **Influence of HLA-DR genes on the production of rheumatoid arthritis-specific autoantibodies to citrullinated fibrinogen**. *Arthritis and rheumatism* 2005, **52**(11):3424-3432.

62. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, Ronnelid J, Harris HE, Ulfgren AK, Rantapaa-Dahlqvist S *et al*: **A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination**. *Arthritis and rheumatism* 2006, **54**(1):38-46.

63. Kochi Y, Yamada R, Kobayashi K, Takahashi A, Suzuki A, Sekine A, Mabuchi A, Akiyama F, Tsunoda T, Nakamura Y *et al*: **Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences**. *Arthritis and rheumatism* 2004, **50**(1):63-71.

64. Mulcahy B, Waldron-Lynch F, McDermott MF, Adams C, Amos CI, Zhu DK, Ward RH, Clegg DO, Shanahan F, Molloy MG *et al*: **Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis**. *American journal of human genetics* 1996, **59**(3):676-683.

65. Singal DP, Li J, Lei K: **Genetics of rheumatoid arthritis (RA): two separate regions in the major histocompatibility complex contribute to susceptibility to RA**. *Immunology letters* 1999, **69**(3):301-306.

66. Zanelli E, Jones G, Pascual M, Eerligh P, van der Slik AR, Zwinderman AH, Verduyn W, Schreuder GM, Roovers E, Breedveld FC *et al*: **The telomeric part of the HLA region predisposes to rheumatoid arthritis independently of the class II loci**. *Human immunology* 2001, **62**(1):75-84.

67. Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, Monteiro J, Khalili H, Lee A, Lundsten R *et al*: **Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis**. *American journal of human genetics* 2002, **71**(3):585-594.

68. Ding B, Padyukov L, Lundstrom E, Seielstad M, Plenge RM, Oksenberg JR, Gregersen PK, Alfredsson L, Klareskog L: **Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region**. *Arthritis and rheumatism* 2009, **60**(1):30-38.

69. Vignal C, Bansal AT, Balding DJ, Binks MH, Dickson MC, Montgomery DS, Wilson AG: **Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci**. *Arthritis and rheumatism* 2009, **60**(1):53-62.

70. Lee HS, Lee AT, Criswell LA, Seldin MF, Amos CI, Carulli JP, Navarrete C, Remmers EF, Kastner DL, Plenge RM *et al*: **Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus**. *Mol Med* 2008, **14**(5-6):293-300.

71. Michou L, Croiseau P, Petit-Teixeira E, du Montcel ST, Lemaire I, Pierlot C, Osorio J, Frigui W, Lasbleiz S, Quillet P *et al*: **Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis**. *Arthritis research & therapy* 2006, **8**(3):R79.

72. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J *et al*: **Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis**. *Nature genetics* 2012, **44**(3):291-296.

73. Mitsunaga S, Suzuki Y, Kuwana M, Sato S, Kaneko Y, Homma Y, Narita A, Kashiwase K, Okudaira Y, Inoue I *et al*: **Associations between six classical HLA loci and rheumatoid arthritis: a comprehensive analysis**. *Tissue antigens* 2012, **80**(1):16-25.

74. van Venrooij WJ, van Beers JJ, Pruijn GJ: **Anti-CCP antibodies: the past, the present and the future**. *Nature reviews Rheumatology* 2011, **7**(7):391-398.

75. Guo JP, Verdrengh M, Tarkowski A, Lange S, Jennische E, Lorentzen JC, Harris RA: **The rat antigen-presenting lectin-like receptor complex influences innate immunity and development of infectious diseases**. *Genes and immunity* 2009, **10**(3):227-236.

76. Backdahl L, Guo JP, Jagodic M, Becanovic K, Ding B, Olsson T, Lorentzen JC: **Definition of arthritis candidate risk genes by combining rat linkage-mapping results with human case-control association data**. *Annals of the rheumatic diseases* 2009, **68**(12):1925-1932.

77. Guo JP, Backdahl L, Marta M, Mathsson L, Ronnelid J, Lorentzen JC: **Profound and paradoxical impact on arthritis and autoimmunity of the rat antigen-presenting lectin-like receptor complex**. *Arthritis and rheumatism* 2008, **58**(5):1343-1353.

78. Kanazawa N: **Dendritic cell immunoreceptors: C-type lectin receptors for pattern-recognition and signaling on antigen-presenting cells**. *Journal of dermatological science* 2007, **45**(2):77-86.

79. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, Gianniny L, Korman BD, Padyukov L, Kurreeman FA *et al*: **Common variants at CD40 and other loci confer risk of rheumatoid arthritis**. *Nature genetics* 2008, **40**(10):1216-1223.

80. Kochi Y, Yamada R, Suzuki A, Harley JB, Shirasawa S, Sawada T, Bae SC, Tokuhiro S, Chang X, Sekine A *et al*: **A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities**. *Nature genetics* 2005, **37**(5):478-485.

81. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, Suzuki M, Nagasaki M, Nakayama-Hamada M, Kawaida R, Ono M *et al*: **Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis**. *Nature genetics* 2003, **34**(4):395-402.

82. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, Catanese JJ, Xie G, Stahl EA, Chen R *et al*: **Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk**. *Nature genetics* 2009, **41**(12):1313-1318.

83. Kurreeman FA, Daha NA, Chang M, Catanese JJ, Begovich AB, Huizinga TW, Toes RE: **Association of IL2RA and IL2RB with rheumatoid arthritis: a replication study in a Dutch population**. *Annals of the rheumatic diseases* 2009, **68**(11):1789-1790.

84. Daha NA, Kurreeman FA, Marques RB, Stoeken-Rijsbergen G, Verduijn W, Huizinga TW, Toes RE: **Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis**. *Arthritis and rheumatism* 2009, **60**(5):1255-1260.

85. Suzuki A, Yamada R, Kochi Y, Sawada T, Okada Y, Matsuda K, Kamatani Y, Mori M, Shimane K, Hirabayashi Y *et al*: **Functional SNPs in CD244 increase the risk of rheumatoid arthritis in a Japanese population**. *Nature genetics* 2008, **40**(10):1224-1229.

86. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M *et al*: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis**. *Nature genetics* 2007, **39**(12):1477-1482.

87. Lee HS, Remmers EF, Le JM, Kastner DL, Bae SC, Gregersen PK: **Association of STAT4 with rheumatoid arthritis in the Korean population**. *Mol Med* 2007, **13**(9-10):455-460.

88. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM *et al*: **A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis**. *American journal of human genetics* 2004, **75**(2):330-337.

89. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, Franke B, Franke L, Posthumus MD, van Heel DA, van der Steege G *et al*: **Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases**. *American journal of human genetics* 2007, **81**(6):1284-1288.

90. Barton A, Eyre S, Ke X, Hinks A, Bowes J, Flynn E, Martin P, Wilson AG, Morgan AW, Emery P *et al*: **Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes**. *Human molecular genetics* 2009, **18**(13):2518-2522.

91. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, Plant D, Gibbons LJ, Wilson AG, Bax DE *et al*: **Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13**. *Nature genetics* 2008, **40**(10):1156-1159.

92. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, Seldin MF, Criswell LA, Plenge RM, Holers VM *et al*: **REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis**. *Nature genetics* 2009, **41**(7):820-823.

93. Orozco G, Hinks A, Eyre S, Ke X, Gibbons LJ, Bowes J, Flynn E, Martin P, Wilson AG, Bax DE *et al*: **Combined effects of three independent SNPs greatly increase the risk estimate for RA at 6q23**. *Human molecular genetics* 2009, **18**(14):2693-2699.

94. Orozco G, Eyre S, Hinks A, Ke X, Wilson AG, Bax DE, Morgan AW, Emery P, Steer S, Hocking L *et al*: **Association of CD40 with rheumatoid arthritis confirmed in a large UK case-control study**. *Annals of the rheumatic diseases* 2010, **69**(5):813-816.

95. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR *et al*: **TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study**. *The New England journal of medicine* 2007, **357**(12):1199-1209.

96. El-Gabalawy HS, Robinson DB, Daha NA, Oen KG, Smolik I, Elias B, Hart D, Bernstein CN, Sun Y, Lu Y *et al*: **Non-HLA genes modulate the risk of rheumatoid arthritis associated with HLA-DRB1 in a susceptible North American Native population**. *Genes and immunity* 2011, **12**(7):568-574.

97. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F *et al*: **STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus**. *The New England journal of medicine* 2007, **357**(10):977-986.

98. Kochi Y, Suzuki A, Yamada R, Yamamoto K: **Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis**. *Nature reviews Rheumatology* 2010, **6**(5):290-295.

99. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M *et al*: **A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes**. *Nature genetics* 2004, **36**(4):337-338.

100. Harrison P, Pointon JJ, Farrar C, Brown MA, Wordsworth BP: **Effects of PTPN22 C1858T polymorphism on susceptibility and clinical characteristics of British Caucasian rheumatoid arthritis patients**. *Rheumatology (Oxford)* 2006, **45**(8):1009-1011.

101. 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 and rheumatism* 2005, **52**(6):1694-1699.

102. Hinks A, Worthington J, Thomson W: **The association of PTPN22 with rheumatoid arthritis and juvenile idiopathic arthritis**. *Rheumatology (Oxford)* 2006, **45**(4):365-368.

103. Johansson M, Arlestig L, Hallmans G, Rantapaa-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 research & therapy* 2006, **8**(1):R19.

104. Orozco G, Sanchez E, Gonzalez-Gay MA, Lopez-Nevot MA, Torres B, Caliz R, Ortego-Centeno N, Jimenez-Alonso J, Pascual-Salcedo D, Balsa A *et al*: **Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus**. *Arthritis and rheumatism* 2005, **52**(1):219-224.

105. Pierer M, Kaltenhauser S, Arnold S, Wahle M, Baerwald C, Hantzschel H, Wagner U: **Association of PTPN22 1858 single-nucleotide polymorphism with rheumatoid arthritis in a German cohort: higher frequency of the risk allele in male compared to female patients**. *Arthritis research & therapy* 2006, **8**(3):R75.

106. Seldin MF, Shigeta R, Laiho K, Li H, Saila H, Savolainen A, Leirisalo-Repo M, Aho K, Tuomilehto-Wolf E, Kaarela K *et al*: **Finnish case-control and family studies support PTPN22 R620W polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis**. *Genes and immunity* 2005, **6**(8):720-722.

107. Simkins HM, Merriman ME, Highton J, Chapman PT, O'Donnell JL, Jones PB, Gow PJ, McLean L, Pokorny V, Harrison AA *et al*: **Association of the PTPN22 locus with rheumatoid arthritis in a New Zealand Caucasian cohort**. *Arthritis and rheumatism* 2005, **52**(7):2222-2225.

108. 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 and immunity* 2005, **6**(6):459-461.

109. Hill RJ, Zozulya S, Lu YL, Ward K, Gishizky M, Jallal B: **The lymphoid protein tyrosine phosphatase Lyp interacts with the adaptor molecule Grb2 and functions as a negative regulator of T-cell activation**. *Experimental hematology* 2002, **30**(3):237-244.

110. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, Buckner JH: **Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes**. *J Immunol* 2007, **179**(7):4704-4710.

111. Bottini N, Vang T, Cucca F, Mustelin T: **Role of PTPN22 in type 1 diabetes and other autoimmune diseases**. *Seminars in immunology* 2006, **18**(4):207-213.

112. Gregersen PK, Lee HS, Batliwalla F, Begovich AB: **PTPN22: setting thresholds for autoimmunity**. *Seminars in immunology* 2006, **18**(4):214-223.

113. Lee HS, Korman BD, Le JM, Kastner DL, Remmers EF, Gregersen PK, Bae SC: **Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations**. *Arthritis and rheumatism* 2009, **60**(2):364-371.

114. Ikari K, Momohara S, Inoue E, Tomatsu T, Hara M, Yamanaka H, Kamatani N: **Haplotype analysis revealed no association between the PTPN22 gene and RA in a Japanese population**. *Rheumatology (Oxford)* 2006, **45**(11):1345-1348.

115. Mori M, Yamada R, Kobayashi K, Kawaida R, Yamamoto K: **Ethnic differences in allele frequency of autoimmune-disease-associated SNPs**. *Journal of human genetics* 2005, **50**(5):264-266.

116. Skinningsrud B, Husebye ES, Gervin K, Lovas K, Blomhoff A, Wolff AB, Kemp EH, Egeland T, Undlien DE: **Mutation screening of PTPN22: association of the 1858T-allele with Addison's disease**. *European journal of human genetics : EJHG* 2008, **16**(8):977-982.

117. Provenzano C, Ricciardi R, Scuderi F, Maiuri MT, Maestri M, La Carpia F, Sferrazza A, Marino M, Leone L, Lucchi M *et al*: **PTPN22 and myasthenia gravis: replication in an Italian population and meta-analysis of literature data**. *Neuromuscular disorders : NMD* 2012, **22**(2):131-138.

118. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W *et al*: **Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes**. *American journal of human genetics* 2005, **76**(4):561-571.

119. Canton 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 and immunity* 2005, **6**(7):584-587.

120. Reddy MV, Johansson M, Sturfelt G, Jonsen A, Gunnarsson I, Svenungsson E, Rantapaa-Dahlqvist S, Alarcon-Riquelme ME: **The R620W C/T polymorphism of the gene PTPN22 is associated with SLE independently of the association of PDCD1**. *Genes and immunity* 2005, **6**(8):658-662.

121. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Baechler EC, Batliwalla FM *et al*: **Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE**. *American journal of human genetics* 2004, **75**(3):504-507.

122. Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, Ball SG, James RA, Quinton R, Perros P *et al*: **The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease**. *The Journal of clinical endocrinology and metabolism* 2004, **89**(11):5862-5865.

123. Zheng W, She JX: **Genetic association between a lymphoid tyrosine phosphatase (PTPN22) and type 1 diabetes**. *Diabetes* 2005, **54**(3):906-908.

124. Ladner MB, Bottini N, Valdes AM, Noble JA: **Association of the single nucleotide polymorphism C1858T of the PTPN22 gene with type 1 diabetes**. *Human immunology* 2005, **66**(1):60-64.

125. Rueda B, Nunez C, Orozco G, Lopez-Nevot MA, de la Concha EG, Martin J, Urcelay E: **C1858T functional variant of PTPN22 gene is not associated with celiac disease genetic predisposition**. *Human immunology* 2005, **66**(7):848-852.

126. 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**(4):318-320.

127. Martin MC, Oliver J, Urcelay E, Orozco G, Gomez-Garcia M, Lopez-Nevot MA, Pinero A, Brieva JA, de la Concha EG, Nieto A *et al*: **The functional genetic variation in the PTPN22 gene has a negligible effect on the susceptibility to develop inflammatory bowel disease**. *Tissue antigens* 2005, **66**(4):314-317.

128. De Jager PL, Sawcer S, Waliszewska A, Farwell L, Wild G, Cohen A, Langelier D, Bitton A, Compston A, Hafler DA *et al*: **Evaluating the role of the 620W allele of protein tyrosine phosphatase PTPN22 in Crohn's disease and multiple sclerosis**. *European journal of human genetics : EJHG* 2006, **14**(3):317-321.

129. Matesanz F, Rueda B, Orozco G, Fernandez O, Leyva L, Alcina A, Martin J: **Protein tyrosine phosphatase gene (PTPN22) polymorphism in multiple sclerosis**. *Journal of neurology* 2005, **252**(8):994-995.

130. 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**. *American journal of human genetics* 2005, **76**(1):184-187.

131. Kokkonen H, Johansson M, Innala L, Jidell E, Rantapaa-Dahlqvist S: **The PTPN22 1858C/T polymorphism is associated with anti-cyclic citrullinated peptide antibody-positive early rheumatoid arthritis in northern Sweden**. *Arthritis research & therapy* 2007, **9**(3):R56.

132. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH, Toes RE, Huizinga TW, Klareskog L, Alfredsson L: **Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis**. *American journal of human genetics* 2007, **80**(5):867-875.

133. Nakashima K, Hagiwara T, Ishigami A, Nagata S, Asaga H, Kuramoto M, Senshu T, Yamada M: **Molecular characterization of peptidylarginine deiminase in HL-60 cells induced by retinoic acid and 1alpha,25-dihydroxyvitamin D(3)**. *The Journal of biological chemistry* 1999, **274**(39):27786-27792.

134. Masson-Bessiere 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**(6):4177-4184.

135. Cantaert T, De Rycke L, Bongartz T, Matteson EL, Tak PP, Nicholas AP, Baeten D: **Citrullinated proteins in rheumatoid arthritis: crucial...but not sufficient!** *Arthritis and rheumatism* 2006, **54**(11):3381-3389.

136. Suzuki M, Miyagi J, Kuribayashi M, Negishi E, Ueno K, Moriya H: **Evaluation of allele frequencies in the PADI4 gene and anti-cyclic citrullinated peptide antibodies of patients with rheumatoid arthritis in a Japanese population**. *Annals of the rheumatic diseases* 2006, **65**(10):1399-1400.

137. Takata Y, Inoue H, Sato A, Tsugawa K, Miyatake K, Hamada D, Shinomiya F, Nakano S, Yasui N, Tanahashi T *et al*: **Replication of reported genetic associations of PADI4, FCRL3, SLC22A4 and RUNX1 genes with rheumatoid arthritis: results of an independent Japanese population and evidence from meta-analysis of East Asian studies**. *Journal of human genetics* 2008, **53**(2):163-173.

138. Prasad P, Kumar A, Gupta R, Juyal RC, Thelma BK: **Caucasian and Asian specific rheumatoid arthritis risk loci reveal limited replication and apparent allelic heterogeneity in north Indians**. *PloS one* 2012, **7**(2):e31584.

139. Lee YH, Rho YH, Choi SJ, Ji JD, Song GG: **PADI4 polymorphisms and rheumatoid arthritis susceptibility: a meta-analysis**. *Rheumatology international* 2007, **27**(9):827-833.

140. Kochi Y, Thabet MM, Suzuki A, Okada Y, Daha NA, Toes RE, Huizinga TW, Myouzen K, Kubo M, Yamada R *et al*: **PADI4 polymorphism predisposes male smokers to rheumatoid arthritis**. *Annals of the rheumatic diseases* 2011, **70**(3):512-515.

141. Hou S, Gao GP, Zhang XJ, Sun L, Peng WJ, Wang HF, Ge XJ, Huang W, Sun YH: **PADI4 polymorphisms and susceptibility to rheumatoid arthritis: a meta-analysis**. *Modern rheumatology / the Japan Rheumatism Association* 2012.

142. Cheng J, Zhang H, Zhuang C, Liu R: **Peptidylarginine deiminase type 4 and methyl-CpG binding domain 4 polymorphisms in Chinese patients with rheumatoid arthritis**. *The Journal of rheumatology* 2012, **39**(6):1159-1165.

143. Chen R, Wei Y, Cai Q, Duan S, Ren D, Shen J, He D, Fang M, Lv K, Cheng N *et al*: **The PADI4 gene does not contribute to genetic susceptibility to rheumatoid arthritis in Chinese Han population**. *Rheumatology international* 2011, **31**(12):1631-1634.

144. Panati K, Pal S, Rao KV, Reddy VD: **Association of single nucleotide polymorphisms (SNPs) of PADI4 gene with rheumatoid arthritis (RA) in Indian population**. *Genes & genetic systems* 2012, **87**(3):191-196.

145. Kang CP, Lee HS, Ju H, Cho H, Kang C, Bae SC: **A functional haplotype of the PADI4 gene associated with increased rheumatoid arthritis susceptibility in Koreans**. *Arthritis and rheumatism* 2006, **54**(1):90-96.

146. Cha S, Choi CB, Han TU, Kang CP, Kang C, Bae SC: **Association of anti-cyclic citrullinated peptide antibody levels with PADI4 haplotypes in early rheumatoid arthritis and with shared epitope alleles in very late rheumatoid arthritis**. *Arthritis and rheumatism* 2007, **56**(5):1454-1463.

147. Bang SY, Han TU, Choi CB, Sung YK, Bae SC, Kang C: **Peptidyl arginine deiminase type IV (PADI4) haplotypes interact with shared epitope regardless of anti-cyclic citrullinated peptide antibody or erosive joint status in rheumatoid arthritis: a case control study**. *Arthritis research & therapy* 2010, **12**(3):R115.

148. Harney SM, Meisel C, Sims AM, Woon PY, Wordsworth BP, Brown MA: **Genetic and genomic studies of PADI4 in rheumatoid arthritis**. *Rheumatology (Oxford)* 2005, **44**(7):869-872.

149. Barton A, Bowes J, Eyre S, Spreckley K, Hinks A, John S, Worthington J: **A functional haplotype of the PADI4 gene associated with rheumatoid arthritis in a Japanese population is not associated in a United Kingdom population**. *Arthritis and rheumatism* 2004, **50**(4):1117-1121.

150. Burr ML, Naseem H, Hinks A, Eyre S, Gibbons LJ, Bowes J, Wilson AG, Maxwell J, Morgan AW, Emery P *et al*: **PADI4 genotype is not associated with rheumatoid arthritis in a large UK Caucasian population**. *Annals of the rheumatic diseases* 2010, **69**(4):666-670.

151. Martinez A, Valdivia A, Pascual-Salcedo D, Lamas JR, Fernandez-Arquero M, Balsa A, Fernandez-Gutierrez B, de la Concha EG, Urcelay E: **PADI4 polymorphisms are not associated with rheumatoid arthritis in the Spanish population**. *Rheumatology (Oxford)* 2005, **44**(10):1263-1266.

152. Iwamoto T, Ikari K, Nakamura T, Kuwahara M, Toyama Y, Tomatsu T, Momohara S, Kamatani N: **Association between PADI4 and rheumatoid arthritis: a meta-analysis**. *Rheumatology (Oxford)* 2006, **45**(7):804-807.

153. Hoppe B, Haupl T, Gruber R, Kiesewetter H, Burmester GR, Salama A, Dorner T: **Detailed analysis of the variability of peptidylarginine deiminase type 4 in German patients with rheumatoid arthritis: a case-control study**. *Arthritis research & therapy* 2006, **8**(2):R34.

154. Farago B, Talian GC, Maasz A, Magyari L, Horvatovich K, Kovacs B, Cserep V, Kisfali P, Kiss CG, Czirjak L *et al*: **Prevalence of functional haplotypes of the peptidylarginine deiminase citrullinating enzyme gene in patients with rheumatoid arthritis: no influence of the presence of anti-citrullinated peptide antibodies**. *Clinical and experimental rheumatology* 2007, **25**(4):523-528.

155. Caponi L, Petit-Teixeira E, Sebbag M, Bongiorni F, Moscato S, Pratesi F, Pierlot C, Osorio J, Chapuy-Regaud S, Guerrin M *et al*: **A family based study shows no association between rheumatoid arthritis and the PADI4 gene in a white French population**. *Annals of the rheumatic diseases* 2005, **64**(4):587-593.

156. Costenbader KH, Gay S, Alarcon-Riquelme ME, Iaccarino L, Doria A: **Genes, epigenetic regulation and environmental factors: which is the most relevant in developing autoimmune diseases?** *Autoimmunity reviews* 2012, **11**(8):604-609.

157. Karlson EW, Deane K: **Environmental and gene-environment interactions and risk of rheumatoid arthritis**. *Rheumatic diseases clinics of North America* 2012, **38**(2):405-426.

158. Hart JE, Kallberg H, Laden F, Bellander T, Costenbader KH, Holmqvist M, Klareskog L, Alfredsson L, Karlson EW: **Ambient air pollution exposures and risk of rheumatoid arthritis: results from the Swedish EIRA case-control study**. *Annals of the rheumatic diseases* 2012.

159. Stolt P, Yahya A, Bengtsson C, Kallberg H, Ronnelid J, Lundberg I, Klareskog L, Alfredsson L: **Silica exposure among male current smokers is associated with a high risk of developing ACPA-positive rheumatoid arthritis**. *Annals of the rheumatic diseases* 2010, **69**(6):1072-1076.

160. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, Frisch M, Karlson EW, Klareskog L, Alfredsson L: **Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies**. *Annals of the rheumatic diseases* 2009, **68**(2):222-227.

161. Costenbader KH, Feskanich D, Holmes M, Karlson EW, Benito-Garcia E: **Vitamin D intake and risks of systemic lupus erythematosus and rheumatoid arthritis in women**. *Annals of the rheumatic diseases* 2008, **67**(4):530-535.

162. Kerr GS, Sabahi I, Richards JS, Caplan L, Cannon GW, Reimold A, Thiele GM, Johnson D, Mikuls TR: **Prevalence of vitamin D insufficiency/deficiency in rheumatoid arthritis and associations with disease severity and activity**. *The Journal of rheumatology* 2011, **38**(1):53-59.

163. Cerhan JR, Saag KG, Merlino LA, Mikuls TR, Criswell LA: **Antioxidant micronutrients and risk of rheumatoid arthritis in a cohort of older women**. *American journal of epidemiology* 2003, **157**(4):345-354.

164. Wegner N, Wait R, Sroka A, Eick S, Nguyen KA, Lundberg K, Kinloch A, Culshaw S, Potempa J, Venables PJ: **Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis**. *Arthritis and rheumatism* 2010, **62**(9):2662-2672.

165. Mangat P, Wegner N, Venables PJ, Potempa J: **Bacterial and human peptidylarginine deiminases: targets for inhibiting the autoimmune response in rheumatoid arthritis?** *Arthritis research & therapy* 2010, **12**(3):209.

166. Lundberg K, Wegner N, Yucel-Lindberg T, Venables PJ: **Periodontitis in RA-the citrullinated enolase connection**. *Nature reviews Rheumatology* 2010, **6**(12):727-730.

167. Dissick A, Redman RS, Jones M, Rangan BV, Reimold A, Griffiths GR, Mikuls TR, Amdur RL, Richards JS, Kerr GS: **Association of periodontitis with rheumatoid arthritis: a pilot study**. *Journal of periodontology* 2010, **81**(2):223-230.

168. Doran MF, Crowson CS, O'Fallon WM, Gabriel SE: **The effect of oral contraceptives and estrogen replacement therapy on the risk of rheumatoid arthritis: a population based study**. *The Journal of rheumatology* 2004, **31**(2):207-213.

169. Pikwer M, Bergstrom U, Nilsson JA, Jacobsson L, Berglund G, Turesson C: **Breast feeding, but not use of oral contraceptives, is associated with a reduced risk of rheumatoid arthritis**. *Annals of the rheumatic diseases* 2009, **68**(4):526-530.

170. Brown JM, Pfau JC, Pershouse MA, Holian A: **Silica, apoptosis, and autoimmunity**. *Journal of immunotoxicology* 2005, **1**(3):177-187.

171. Hart JE, Yanosky JD, Puett RC, Ryan L, Dockery DW, Smith TJ, Garshick E, Laden F: **Spatial modeling of PM10 and NO2 in the continental United States, 1985-2000**. *Environmental health perspectives* 2009, **117**(11):1690-1696.

172. Karlson EW, Lee IM, Cook NR, Manson JE, Buring JE, Hennekens CH: **A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals**. *Arthritis and rheumatism* 1999, **42**(5):910-917.

173. Vessey MP, Villard-Mackintosh L, Yeates D: **Oral contraceptives, cigarette smoking and other factors in relation to arthritis**. *Contraception* 1987, **35**(5):457-464.

174. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L: **A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis**. *Arthritis and rheumatism* 2004, **50**(10):3085-3092.

175. Karlson EW, Chang SC, Cui J, Chibnik LB, Fraser PA, De Vivo I, Costenbader KH: **Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis**. *Annals of the rheumatic diseases* 2010, **69**(1):54-60.

176. Costenbader KH, Feskanich D, Mandl LA, Karlson EW: **Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women**. *The American journal of medicine* 2006, **119**(6):503 e501-509.

177. Mahdi H, Fisher BA, Kallberg H, Plant D, Malmstrom V, Ronnelid J, Charles P, Ding B, Alfredsson L, Padyukov L *et al*: **Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis**. *Nature genetics* 2009, **41**(12):1319-1324.

178. OECD: **Tobacco comsumption among adults, in Health at a Glance 2011:OECD Indicators, OECD Publishing**. *http;//dxdoiorg/101787/health_glance-2011-16-en* 2011.

179. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, Klareskog L, Alfredsson L: **Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke**. *Annals of the rheumatic diseases* 2011, **70**(3):508-511.

180. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL *et al*: **A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms**. *Nature* 2001, **409**(6822):928-933.

181. Peltonen L, Perola M, Naukkarinen J, Palotie A: **Lessons from studying monogenic disease for common disease**. *Human molecular genetics* 2006, **15 Spec No 1**:R67-74.

182. Risch NJ, Bressman SB, deLeon D, Brin MF, Burke RE, Greene PE, Shale H, Claus EB, Cupples LA, Fahn S: **Segregation analysis of idiopathic torsion dystonia in Ashkenazi Jews suggests autosomal dominant inheritance**. *American journal of human genetics* 1990, **46**(3):533-538.

183. Cardon LR, Bell JI: **Association study designs for complex diseases**. *Nature reviews Genetics* 2001, **2**(2):91-99.

184. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies**. *Nature genetics* 2006, **38**(8):904-909.

185. Devlin B, Roeder K: **Genomic control for association studies**. *Biometrics* 1999, **55**(4):997-1004.

186. Ott J, Kamatani Y, Lathrop M: **Family-based designs for genome-wide association studies**. *Nature reviews Genetics* 2011, **12**(7):465-474.

187. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM *et al*: **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 2007, **449**(7164):851-861.

188. **A haplotype map of the human genome**. *Nature* 2005, **437**(7063):1299-1320.

189. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(23):9362-9367.

190. Ji H, Gauguier D, Ohmura K, Gonzalez A, Duchatelle V, Danoy P, Garchon HJ, Degott C, Lathrop M, Benoist C *et al*: **Genetic influences on the end-stage effector phase of arthritis**. *The Journal of experimental medicine* 2001, **194**(3):321-330.

191. Cardon LR, Palmer LJ: **Population stratification and spurious allelic association**. *Lancet* 2003, **361**(9357):598-604.

192. Zondervan KT, Cardon LR: **Designing candidate gene and genome-wide case-control association studies**. *Nature protocols* 2007, **2**(10):2492-2501.

193. Devlin B, Roeder K, Wasserman L: **Genomic control, a new approach to genetic-based association studies**. *Theoretical population biology* 2001, **60**(3):155-166.

194. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS genetics* 2006, **2**(12):e190.

195. Ioannidis JP, Patsopoulos NA, Evangelou E: **Heterogeneity in meta-analyses of genome-wide association investigations**. *PloS one* 2007, **2**(9):e841.

196. Ioannidis JP, Patsopoulos NA, Evangelou E: **Uncertainty in heterogeneity estimates in meta-analyses**. *BMJ* 2007, **335**(7626):914-916.

197. Bengtsson C, Berglund A, Serra ML, Nise L, Nordmark B, Klareskog L, Alfredsson L, Eira Study G: **Non-participation in EIRA: a population-based case-control study of rheumatoid arthritis**. *Scandinavian journal of rheumatology* 2010, **39**(4):344-346.

198. Yahya A, Bengtsson C, Lai TC, Larsson PT, Mustafa AN, Abdullah NA, Muhamad N, Hussein H, Klareskog L, Alfredsson L *et al*: **Smoking is associated with an increased risk of developing ACPA-positive but not ACPA-negative rheumatoid arthritis in Asian populations: evidence from the Malaysian MyEIRA case-control study**. *Modern rheumatology / the Japan Rheumatism Association* 2012, **22**(4):524-531.

199. Ting A, Morris PJ: **A technique for lymphocyte preparation from stored heparinized blood**. *Vox sanguinis* 1971, **20**(6):561-563.

200. Boyum A: **Separation of leukocytes from blood and bone marrow. Introduction**. *Scandinavian journal of clinical and laboratory investigation Supplementum* 1968, **97**:7.

201. Newton JL, Harney SM, Wordsworth BP, Brown MA: **A review of the MHC genetics of rheumatoid arthritis**. *Genes and immunity* 2004, **5**(3):151-157.

202. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G: **Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies**. *Human immunology* 2008, **69**(7):443-464.

203. Terao C, Ohmura K, Kochi Y, Ikari K, Maruya E, Katayama M, Shimada K, Murasawa A, Honjo S, Takasugi K *et al*: **A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects**. *Annals of the rheumatic diseases* 2011, **70**(12):2134-2139.

204. Greenland SRK, TL L: **Measures of effect and measures of association in modern epidemiology**. *3 Edition Edited by Rothman KJ, Greenland S, Lash TL Philadelphia: Lippincott-Raven* 2008:10.

205. Shiozawa S, Hayashi S, Tsukamoto Y, Goko H, Kawasaki H, Wada T, Shimizu K, Yasuda N, Kamatani N, Takasugi K *et al*: **Identification of the gene loci that predispose to rheumatoid arthritis**. *International immunology* 1998, **10**(12):1891-1895.

206. MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, Barrett J, Lee D, White S, John S *et al*: **Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom**. *Arthritis and rheumatism* 2002, **46**(3):632-639.

207. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, Kern M, Criswell LA, Albani S, Nelson JL *et al*: **A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases**. *American journal of human genetics* 2001, **68**(4):927-936.

208. Cornelis F, Faure S, Martinez M, Prud'homme JF, Fritz P, Dib C, Alves H, Barrera P, de Vries N, Balsa A *et al*: **New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(18):10746-10750.

209. Tarcsa E, Marekov LN, Mei G, Melino G, Lee SC, Steinert PM: **Protein unfolding by peptidylarginine deiminase. Substrate specificity and structural relationships of the natural

substrates trichohyalin and filaggrin**. *The Journal of biological chemistry* 1996, **271**(48):30709-30716.

210.   Yamada R, Suzuki A, Chang X, Yamamoto K: **Citrullinated proteins in rheumatoid arthritis**. *Frontiers in bioscience : a journal and virtual library* 2005, **10**:54-64.

211.   Ikari K, Kuwahara M, Nakamura T, Momohara S, Hara M, Yamanaka H, Tomatsu T, Kamatani N: **Association between PADI4 and rheumatoid arthritis: a replication study**. *Arthritis and rheumatism* 2005, **52**(10):3054-3057.

212.   Greene CS, Penrod NM, Williams SM, Moore JH: **Failure to replicate a genetic association may provide important clues about genetic architecture**. *PloS one* 2009, **4**(6):e5639.

213.   Xu S, Yin X, Li S, Jin W, Lou H, Yang L, Gong X, Wang H, Shen Y, Pan X *et al*: **Genomic dissection of population substructure of Han Chinese and its implication in association studies**. *American journal of human genetics* 2009, **85**(6):762-774.

214.   Eklow C, Makrygiannakis D, Backdahl L, Padyukov L, Ulfgren AK, Lorentzen JC, Malmstrom V: **Cellular distribution of the C-type II lectin dendritic cell immunoreceptor (DCIR) and its expression in the rheumatic joint: identification of a subpopulation of DCIR+ T cells**. *Annals of the rheumatic diseases* 2008, **67**(12):1742-1749.

215.   Ronninger M, Eklow C, Lorentzen JC, Klareskog L, Padyukov L: **Differential expression of transcripts for the autoimmunity-related human dendritic cell immunoreceptor**. *Genes and immunity* 2008, **9**(5):412-418.

216.   Evans DM, Cardon LR: **A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations**. *American journal of human genetics* 2005, **76**(4):681-687.

217.   Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK: **A worldwide survey of haplotype variation and linkage disequilibrium in the human genome**. *Nature genetics* 2006, **38**(11):1251-1260.

218.   Adeyemo A, Rotimi C: **Genetic variants associated with complex human diseases show wide variation across multiple populations**. *Public health genomics* 2010, **13**(2):72-79.