From the Institute of Environmental Medicine, Division of Cardiovascular Epidemiology, Karolinska Institutet, Stockholm, Sweden in collaboration with the Rheumatology Unit, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

ROLE OF GENES AND ENVIRONMENT FOR THE DEVELOPMENT OF RHEUMATOID ARTHRITIS

- RESULTS FROM THE SWEDISH EIRA STUDY

Henrik Källberg

Stockholm 2009

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, Karolinska Institutet

ISBN 978-91-7409-284-4

Dedicated to:

GCAAGCACAoAGAUAAAGCGUAGAAAGAAGAGAAUAAAUGoAGAAUGoA
GAUAAAUGoAGAUUCGCAAGAUAAAUGGCAAUGAUGGCAUAACCAGCAC
CACCAGCAUAAUUCGCAAGAAUGoAGAUAAUUCGCAAGAUUCGCAAGAU
AAUGCCACGCAAGAUUAoACAACAGCA.

"If you thought that science was certain - well, that is just an error on your part."
- Richard P Feynman
"Science may set limits to knowledge, but should not set limits to imagination."
- Bertrand Russell
Det krävs ett helt nytt sätt att tänka för att lösa de problem vi skapat med det gamla sättet att tänka.
- Albert Einstein
"To understand this fully, one must transcend from the duality of 'for' and 'against' into one organic unity which is without distinctions."
- Bruce Lee
“Confusion is my name, ask me again I'll tell you the same.”
- Red Hot Chili Peppers

ABSTRACT

Rheumatoid arthritis (RA) is a complex disease with autoimmune features primarily causing destruction in the joints of the body. Knowledge regarding risk indicators and causes are increasing but so far only a few genetic and very few environmental risk factors have been consistently identified.

The overall aim of this thesis is to investigate interaction between genetic and environmental factors for different phenotypes of RA. Specific aims are to investigate:

1. gene-environment interaction between HLA-DRB1 SE alleles and smoking.
2. Gene-gene interaction between HLA-DRB1 SE and R620W PTPN22 alleles.
3. Interaction between alcohol consumption, smoking and HLA-DRB1 SE alleles.
4. Dose dependency between smoking and risk of developing RA with consideration taken to interaction between smoking and HLA-DRB1 SE alleles.

In all the above aims different phenotypes for RA as defined by presence or absence of antibodies to citrullinated protein antigens (ACPA+, ACPA- RA respectively) is considered.

This thesis is primarily based on data from a Swedish study named Epidemiological Investigation of Rheumatoid Arthritis (EIRA). EIRA is a population based case-control study which consists of information from incident cases and controls matched on age, sex and living area. Cases and controls were given the opportunity to fill in an extensive questionnaire and to provide a blood sample for genetic and serological analysis. In paper II additional studies from USA (the NARAC study) and the Netherlands (Leiden EAC) was used to investigate potential gene-gene interaction between HLA-DRB1 SE and R620W PTPN22 alleles. In paper III information from the Danish case-control study of rheumatoid arthritis (the CACORA study) in addition to EIRA, was used to investigate interaction between alcohol consumption, smoking and HLA-DRB1 SE alleles.

Smoking and alcohol consumption patterns and dosage were estimated through self-reported information in questionnaires. Interaction was primarily defined in terms of deviance from additivity of effects.

A strong interaction between smoking and HLA-DRB1 SE alleles was observed regarding risk of developing ACPA+ RA. The most pronounced interaction was observed for the combination of smoking and homozygosity for HLA-DRB1 SE alleles. Smoking was also associated, in a dose dependent manner with increased risks of developing ACPA+ RA with consideration taken to HLA-DRB1 SE alleles.

Interaction between HLA-DRB1 SE and R620W PTPN22 alleles regarding risk of developing ACPA+ RA was observed in three different studies (EIRA, NARAC and Leiden EAC). No associations between HLA-DRB1 SE or R620W PTPN22, and only a minor association for smoking regarding risk of ACPA- RA were observed.

Alcohol consumption was inversely associated with risk of developing RA in a dose dependent manner in both EIRA and CACORA. Lack of alcohol consumption was also associated with interaction with smoking and HLA-DRB1 SE alleles regarding risk of developing ACPA+ RA.

Key words: Rheumatoid arthritis, HLA-DRB1, PTPN22, smoking, alcohol, case control, interaction.

LIST OF PUBLICATIONS

I. Klareskog L, Stolt P, Lundberg K, **KÄLLBERG H**, Bengtsson C, Grunewald J, Ronnelid J, Harris HE, Ulfgren AK, Rantapaa-Dahlqvist S, Eklund A, Padyukov L, Alfredsson L. 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 Rheum. 2006 Jan;54(1):38-46.

II. **KÄLLBERG H**, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH, Toes RE, Huizinga TW, Klareskog L, Alfredsson L; Epidemiological Investigation of Rheumatoid Arthritis study group. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet. 2007 May;80(5):867-75.

III. **HENRIK KÄLLBERG,** Søren Jacobsen, Camilla Bengtsson, Merete Pedersen, Leonid Padyukov, Peter Garred, Morten Frisch, Elizabeth W Karlson, Lars Klareskog, Lars Alfredsson. Alcohol consumption is associated with decreased risk of rheumatoid arthritis; Results from two Scandinavian case-control studies. Ann Rheum Dis. 2008 Jul 15. [Epub ahead of print]

IV. **HENRIK KÄLLBERG**, Bo Ding, Leonid Padyukov, Camilla Bengtsson, Johan Rönnelid, Lars Klareskog, Lars Alfredsson, EIRA study group. Smoking is a major preventable risk factor for RA;
Estimations of risk for ACPA-positive RA after various exposures to cigarette smoke. Manuscript

CONTENTS

LIST OF ABBREVIATIONS

ACPA	Antibodies to citrullinated protein antigens
ACPA+	Antibodies to citrullinated protein antigens positive
ACPA-	Antibodies to citrullinated protein antigens negative
ACR	American College of Rheumatology
Anti-CCP	antibodies toward cyclic citrullinated peptides
AP	Attributable proportion due to interaction
CI	Confidence interval
EIRA	Epidemiological Investigation of Rheumatoid Arthritis
Fc	Fragment constant region
HLA	Human leukocyte antigen
IgG	Immunoglobulin G
LD	Linkage disequilibrium
MHC	Major histocompatobility complex
OR	Odds ratio
PAD	Peptidylarginine deiminases
EF%	Excess fraction percent
PTPN22	protein tyrosine phosphatase non receptor 22
RA	Rheumatoid arthritis
RF	Rheumatoid factor
RF+	Rheumatoid factor positive
RF-	Rheumatoid factor negative
RR	Relativ risk
SAS	Statistical Analysis System
SE	Shared epitope
TNFα	Tumor necrosis factor alpha

1 INTRODUCTION

Rheumatoid arthritis was first described by name in the second half of the 17th century by Garrod [1-2], but has most likely been present without the name rheumatoid arthritis much longer than that [3]. Rheumatoid arthritis (RA) is a complex disease that is approximately two to three times more common among women than men. The disease is an autoimmune disease, which means that the immune system is reacting against tissues in the own body. RA is primarily affecting the joints in the body by destruction of cartilage and bone.. Later stages in the disease course can be associated with systemic manifestations such as pulmonary fibrosis, inflammatory serositis in the lung and heart. In addition patients with RA suffer from increased mortality in cardiovascular disease and co-morbidities such as lymphomas and lung cancer [4], in addition to the pain and impairment caused by the disease itself [5].

Little is known about the aetiology of RA, but both environmental and genetic factors are believed to be risk factors for developing RA.

Lately, evidence has emerged suggesting that the diagnosis “rheumatoid arthritis” is in fact a family name comprising several diseases with different aetiology but similar symptoms. One classical basis for a subdivision of RA is the presence of Rheumatoid Factor (RF). RF is a complex of anti-bodies associated with more severe RA. Recently, a subdivision based on another anti-body called anti-CCP (antibodies toward cyclic citrullinated peptides, later called ACPA, which is short for anti-citrullinated peptide antigens and antibodies to citrullinated protein antigens) has been suggested to be of more fundamental relevance because of the high specificity regarding RA [6-7].

Due to the complexity and different features of RA, it is natural to consider both environmental and genetic risk factors, as well as serological types of the disease simultaneously, when searching for, or testing different hypotheses regarding disease mechanisms. This thesis is an attempt to integrate biological-(genetic and serological) and environmental-, information, in order to search for and test hypotheses regarding factors involved in the disease mechanisms of RA and to investigate potential interaction effects between these factors. The scope of this thesis relates to a recent description of the new era in biomedical sciences:

"The reductionist approach has successfully identified most of the components and many of the interactions but, unfortunately, offers no convincing concepts or methods to understand how system properties emerge...the pluralism of causes and effects in biological networks is better addressed by observing, through quantitative measures, multiple components simultaneously and by rigorous data integration with mathematical models"[8]

So, what does the RA puzzle look like and which parts are associated with each other? First, I will describe some basic concepts and previous research findings that could help the reader to understand things more easily.

1.1 THE IMMUNE SYSTEM

Since most of the RA cases have features of autoimmune disease, a brief description of the immune system is helpful to understand some of the mechanisms that are involved in the disease.
The immune system can be subdivided into two major parts, the innate immune system and the adaptive immune system. The innate immune system is the first protection or alert system against foreign substances. The other major part of the immune system is the adaptive or acquired immune system which is, among other things, responsible for creating an immunological memory against pathogens. The adaptive system consists of T-cells, including cells which have the capacity to kill other cells, such as virus infected cells (T-killer cells), T cells which can activate other cells, such as B cells (T-helper cells), and T cells that can regulate the immune system (T regulatory cells). B cells can differentiate into plasma cells that produce anti-bodies toward distinct antigens. Anti-bodies are then used to facilitate immunological reactions such as the complement system or by enhancing recognition to phagocyting cells.
Many reactions in the immune system are mediated through certain signal substances called cytokines. Cytokines are proteins affecting cells in many different ways. Some of the cytokines are promoting inflammatory response. TNF-α and interleukin-6 (IL-6) are examples of cytokines that promotes inflammatory reactions [9]. These two cytokines are also target substances for drugs used to treat RA.

1.2 AUTOIMMUNITY

An autoimmune disease is a disease where the immune system is reacting against tissues in the own body. In RA, autoimmunity is exemplified by the activation and proliferation of cells in the synovial membranes and by the destruction of adjacent cartilage and bone. In RA as well as in other autoimmune diseases, it has long remained unclear how environmental and genetic factors may interact in causing immune reactions that may contribute to the development of inflammation and destruction.

1.3 RHEUMATOID ARTHRITIS DIAGNOSIS

The diagnosis of Rheumatoid arthritis is given according to seven criteria, as proposed by the guidelines from ACR in year 1987(American college of rheumatology) [10].
The criteria are as follows:

1. Morning stiffness
2. Symmetric arthritis
3. Arthritis in ≥ 3 joints
4. Presence of rheumatic nodules
5. Radiographic changes
6. Presence of serum Rheumatoid Factor
7. Arthritis in hand joints

A person that fulfils at least four of these seven criterions is given the RA diagnosis. This way of defining RA is currently challenged as some of the criteria, including appearance of Radiographic changes and rheumatoid nodules are not present at the time of optimal early diagnosis and initiation of therapy. Additionally, rheumatoid factors are present only in a subset of RA patients (see more below). Thus, work is ongoing in collaboration between the European and American rheumatology associations in order

to produce new diagnostic criteria that will be more useful for today´s clinical research and clinical practice.

1.3.1 Anti bodies and diversity of Rheumatoid arthritis

Recently, many studies have pinpointed (explicitly and implicitly) that the diagnosis RA is in fact a collection of many different diseases with different phenotypes and different aetiology, as well as different prognostic outcomes. One traditional way of separating RA into subsets is based on the presence or absence of rheumatoid factor (RF). RF is a so called autoantibody, which means that it is an antibody which primary target is tissues in the own body. RF is an antibody directed towards the Fc section in the IgG, which is also an antibody. IgG and RF form an immune complex where the Fc (the Fragment constant region that binds to cells and complement proteins) section in the IgG is the target. RF is however present also in many other conditions than RA including diseases such as Sjögrens syndrome, hepatitis and Systematic lupus erythematosus. As pointed out under the introduction section, a recently discovered anti body has gained interest as being more specific for RA compared to RF. This is anti bodies toward citrullinated peptides, anti-CCP (antibodies toward cyclic citrullinated peptides, or more recently called ACPA, which is short for anti-citrullinated peptide antigens). ACPA is an anti body targeting citrullinated peptides which is amino acid sequences that contain the amino acid citrulline. Citrulline is not coded for by DNA. Instead it is a result of a posttranslational modification of the amino acid arginine that is carried out by enzymes called peptidylarginine deiminases (PADs). ACPA have been associated with higher sensitivity regarding RA as compared to RF [6, 7]. Presence of ACPA is considered to be stable over time in this sub group of RA patients and has also been observed to predict RA in different RA populations. [7, 11-13]

1.4 ENVIRONMENTAL EXPOSURES AND RHEUMATOID ARTHRITIS

There are not many studies carried out with the aim to identify environmental risk factors for developing RA. Traditionally occupations were used as proxy estimators of environmental exposures and risk of RA. A few studies have reported increased risks for RA among employees in certain occupations such as concrete workers, farmers and hairdressers, to mention a few [14-17]. In addition to certain occupations, environmental exposures such as smoking, parity, exposure to mineral oil, exposure to silica have been associated with increased risks for RA [18-24].

1.4.1 Smoking

Smoking is the most established environmental risk factor for developing RA. The first article on smoking as a risk factor for RA was published as late as 1987 [18]. Smoking has since then been associated with certain sub types of RA such as RF-positive RA. Smoking has also been associated with a dose response relationship regarding increased risk of RA. [24-27]
Unfortunately, not all studies included stratified analysis by presence of RF or ACPA. Several, recently published studies have found evidence for smoking being a risk factor for RA characterized by presence of RF or ACPA [28-33]. Smoking is associated, in a dose dependent manner, with increased risk for developing RA, especially for rheumatoid factor positive RA [25-27]. There are also increasing evidence for smoking

being associated with the presence of anti-bodies toward citrullinated peptides (ACPA) [28, 31-33]. For a more thorough review on smoking and RA I recommend the thesis by Patrik Stolt [34]
By summing up the evidence for smoking as a risk factor for RA, it seems that smoking fulfils many of signs of a causal effect such as strength, consistency, temporality and biological gradient as proposed by Hill [35]

1.4.2 Alcohol

Alcohol or ethanol consumption is associated with an overall deprivation of the activity in the immune system [36]. It has also been suggested that alcohol inhibit production of TNF-α through nuclear regulatory factor -κβ (NF-κβ) regulation in monocytes [37].
In earlier studies on humans investigating potential associations between alcohol and risk of developing RA , two studies reported no increased nor decreased risk for developing RA [38-39] and three studies reported a protective effect [24, 31, 40]. Unfortunately, only one of these studies investigated ACPA specific effects [31] and the other study indicating a decreased risk for RA due to alcohol consumption pointed out, in contrast to the majority of studies, a protective effect of smoking [40]. In addition to these findings an experimental study on male DBA/1 mice found a decreased incidence of destructive arthritis in mice exposed to non-hepatotoxic levels of alcohol compared to non exposed mice [41].

1.5 GENETIC RISK FACTORS AND RHEUMATOID ARTHRITIS

Family studies on twins and siblings have found a greater concordance between monozygotic twins, siblings in terms of RA manifestations compared to the concordance between dizygotic twins and non siblings, respectively [42-43]. These studies indicate that the genetic contribution is an important factor regarding RA incidence. A number of genes have been suggested to be associated with increased risks for developing RA [44-46]. Strong evidence exists that the HLA-DRB1 gene in the Major histocompatobility complex (MHC) region with loci in chromosome six plays an important role in the development of RA [47- 49]. The gene is coding for the human leukocyte antigen molecule (HLA) which play an important role in the immune system as a presenter of foreign substances as well as autogenic substances (peptides). The combination of DRB1*0401 and DRB1*0404 alleles in the HLA-DRB1 region have been found to be strongly associated with RA [50]. Gregersen et al. presented a hypothesis, where some alleles in the HLA-DRB1 locus were shown to have great influence on the beta chain in the MHC class II molecule [51]. These alleles are commonly known as the Shared epitope alleles or simply SE alleles.
HLA-DRB1 genotypes have also been associated with more severe disease as compared to those without these genotypes. [52- 53]
The R620W allele of the PTPN22 (protein tyrosine phosphatase non receptor 22) gene is another genetic factor that quite recently have been found to be associated to RA and also been replicated in different populations [54-56]. The PTPN22 gene is located on chromosome one, in contrast to HLA-DRB1 which is located within the MHC loci in chromosome 6. The R620W allelic form of PTPN22 has, in addition to RA, also been associated with type I diabetes [56]. The function of the PTPN22 gene is to encode for a protein, Lyp, which is a negative regulator of T-cell activation. The allelic form of the PTPN22 gene, substitute the amino acid arginine to tryptophan in the protein, Lyp.

Regarding ACPA status, both the HLA-DRB1 SE and R620W PTPN22 alleles are primarily associated with ACPA positive RA, indicating that there are specific disease mechanisms for this RA phenotype [55-58]

1.6 INTERACTION EFFECTS AND RHEUMATOID ARTHRITIS

So far, there are not many studies carried out in order to investigate presence of gene-environment or gene-gene interactions regarding the risk to develop RA. A strong interaction between smoking and presence of SE genes has been observed in relation to risk of developing RA, especially for developing RF+ or ACPA+ RA [28, 32, 58-62]. Article II [28] in this thesis was the first to present an interaction between smoking and HLA-DBR1 SE alleles regarding risk of developing ACPA+ RA. Additionally, a case-only report from US reported a mixed picture regarding interaction between smoking and HLA-DRB1 SE alleles [63].
When estimating interaction one should have in mind that there are different definitions of interaction. When investigating the potential interaction between factors in relation to risk of disease it is therefore important to be aware of what implications the different definitions have. Different definitions of interaction are presented under the methods section.

2 AIMS

The overall aim of this PhD project is to identify environmental-, and genetic risk factors and to investigate their potential interactions regarding the risk of developing RA. More specified aims of this project are to:

- Further explore the influence from smoking on risk of RA by investigating the potential interaction between smoking and HLA-DRB1 SE alleles and the R620W polymorphic form of PTPN22, respectively, with regard to the risk of developing ACPA+/ACPA- RA. And to investigate potential dose-dependent relationship between smoking and ACPA+/ACPA- RA with consideration taken to the presence of HLA-DRB1 SE alleles

- Investigate the potential gene-gene interaction between the HLA-DRB1 SE alleles and the R620W polymorphic form of PTPN22 with consideration taken to smoking and ACPA status.

- Investigate if alcohol consumption is associated with decreased risk for developing RA with consideration taken to ACPA status and to investigate its possible interaction with smoking and SE alleles.

3 METHODS AND MATERIAL

This thesis is mainly based on an epidemiological study called EIRA (Epidemiological Investigation of Rheumatoid Arthritis).
In paper II we also used information from two other studies, NARAC (North American Rheumatoid Arthritis Consortium) and Leiden EAC (Leiden Early Arthritis Clinic). Furthermore, in paper III we also used information from a Danish case-control study called the CACORA study (the CAse-COntrol of Rheumatoid Arthritis study).

3.1 THE EIRA STUDY

EIRA is a population based case-control study that is being carried out in the south and middle of Sweden. The EIRA study started in May 1996 and is still on going. Information regarding life style, environmental exposures, education etc. was collected by using an extensive questionnaire and by taking blood samples from incident cases and controls. Cases were identified at public and privately run clinics and asked by the corresponding personnel if they wanted to participate. Cases were diagnosed according to the 1987 ACR (American College of Rheumatology) criteria. Each case is asked to contribute with blood samples at the corresponding clinic. For each case we randomly select one control from the study base that is matched to the case, in terms of age, sex and living area. We selected controls from the national population register. This register has complete coverage over the Swedish population and is continuously updated. Each control was asked to contribute with a blood sample taken by local medical wards. If a control is not traceable, refused to participate or report having RA a new control is selected from the register. Both cases and controls were supposed to be sufficiently fluent in Swedish to be able to answer the questionnaire.
Blood samples from both cases and controls were posted to the Rheumatology Research laboratory at Karolinska Institutet for further handling and analysis.
The projects in this thesis are based on two generations of data from the EIRA study. The earlier data generation was collected during May 1996 – June 2001 and consists of 930 cases and 1126 controls. In the later generation information from additional 489 cases and 548 controls collected until the end of December 2003 has been added. Information on environmental exposures, length, weight, life events, occupational-, disease- and injury-history from cases and controls has been retrieved by sending all participants identical questionnaires. Questions without answers in the returned questionnaire have been completed thru telephone interviews performed by interview trained personnel. The questions we used for smoking and alcohol consumption in EIRA, used for analyses in the present thesis, are displayed in appendix 1. For more information regarding study design and impact of non-participation I highly recommend another thesis based on the EIRA study by Camilla Bengtsson [64]

3.2 THE NORTH AMERICAN RHEUMATIOD ARTHRITIS CONSORTIUM (NARAC)

This collection of prevalent cases with rheumatoid arthritis recruited from the whole United States is based on families in which at least two siblings were affected by RA according to the 1987 ACR criteria. Cases were investigated by trained NARAC personnel at the 12 nationwide NARAC centres. The family based study design is a

common study design to use in the search of allelic forms of genes involved in the disease process. More information regarding NARAC is given elsewhere [47, 65]. We only used unrelated cases from the NARAC study since the obvious correlation between siblings may cause problems in terms of inflated odds ratios if not considered in the models. These RA cases were matched on age, sex and ethnicity to and healthy controls from the New York cancer project (NYCP). The NYCP is an ongoing cohort study that is based on information from enrolees aged 30 years or more, who live in the New York Tri-State area and have literacy sufficient to complete a mailed questionnaire containing simple questions on topics such as substance use and demographic information among others. More details regarding NYCP are provided elsewhere [66].

3.3 THE LEIDEN EARLY ARTHRITIS CLINIC (LEIDEN EAC)

This cohort consists of cases with RA in the semi-rural area of Leiden in the Netherlands. Leiden EAC was started in 1993 in order to detect and treat inflammatory disorders (especially RA) early. General practicing MDs was informed to pass on patients with suspected arthritis to a rheumatologist for evaluation. If arthritis was confirmed by rheumatologists and the symptoms of arthritis had not been present for more than two years the patients was invited to participate in the cohort. Blood samples, as well as other information such as medical history, smoking etc were collected repeatedly. More detailed information regarding Leiden EAC is given elsewhere [67]. To these RA cases we used hospital based controls with and without deep venous thrombosis as controls. More details regarding the original study in which the controls were recruited is given elsewhere [68].

3.4 THE DANISH CACORA STUDY

The CACORA study is a Danish case-controls study that is based on information from RA patients identified in internal medicine and rheumatology departments throughout Denmark. All patients that were included were between 18 and 65 years of age and fulfilled the 1987 ACR criteria regarding RA. The cases were not supposed to have had their RA diagnose for more than five years before inclusion. Controls were chosen from the Danish civil registry. Controls alive during the recruitment period were matched to each case on sex and birth year. Eligible controls were then randomly selected from a list of eligible individuals born between 1921 and 1985. A detailed description over the CACORA study is given by Pedersen elsewhere [69].

3.5 GENOTYPING AND BIOLOGICAL PARAMETERS

Blood samples from study participants were genotyped for HLA-DRB1 and PTPN22 genes. Among the HLA-DRB1 alleles, DRB1*01, DRB1*04 and DRB1*10 was defined as SE genes [51]. In the case of the PTPN22 gene, the R620W (CC → CT or TT) form was defined as the allelic version of the gene.
In EIRA (paper I-IV) we used the sequence specific primer polymerase chain reaction (SSP-PCR) method [70] to perform HLA-DRB1 specific DNA analysis. For practical reasons we only used low resolution analysis regarding HLA-DRB1 SE alleles. In the beginning of the study a sub sample of 81 cases were sub typed for specific HLA-DRB1*01 and 04 SE alleles. 89 percent of the HLA-DRB1*01 was HLA-DRB1*0101 and 98 percent of the HLA-DRB1*04 was *0401, *0404, *0405 or *0408.

The same method as used in EIRA regarding HLA-DRB1 genotyping was also used for NARAC. In Leiden EAC (paper II) PCR was also used but with biotin-labeled oligonucleotides (PCR-biotin-SSO) [71]. In the CACORA study SSP-PCR was also used according to a fourth method [72].
All genotyping for EIRA, NARAC and Leiden EAC regarding PTPN22 (paper II) was performed by use of kinetic, allele specific PCR [73]. All these analyses were performed in the USA with an average genotyping completeness of 97.8 percent.
For simplicity of evaluation, we assumed a dominant A-allele model for R620W PTPN22 polymorphism and a dominant HLA-DRB1 SE allele model. Subjects with HLA-DRB1 SE alleles were classified as having single or double SE alleles for a separate evaluation of a potential gene dosage effect.

Analysis of ACPA for samples from EIRA (paper I-IV) as well as samples from the Leiden EAC (paper II) and CACORA (paper III) was made with the Immunoscan-RA Mark2 enzyme-linked immunosorbent assay (ELISA) (Euro-Diagnostica, Malmö, Sweden). Cases with serum antibody levels above 25 U/ml were regarded as positive (anti-CCP+ or ACPA+). Rheumatoid factor status for cases was determined using standard procedures.
Antibodies towards CCP in NARAC (paper II) cases were measured using a second-generation commercial anti-CCP ELISA (Inova Diagnostics, San Diego, CA, USA). Cases with antibody levels over 20 U/ml were regarded as anti-CCP positive.

3.6 ENVIRONMENTAL EXPOSURES AND RHEUMATOID ARTHRITIS

In this thesis we have focused on environmental exposures associated with life style factors such as smoking and alcohol consumption.

3.6.1 Smoking

In EIRA each case was given an index year, which was defined as the year when RA symptoms occurred. The cases corresponding index year was also given to the matched control. If a participant started smoking after the index year the participant was considered to be unexposed to smoking. Regarding calculation of smoking doses and duration, only the time of smoking exposure before the index year were considered as the time for exposure. We categorized smoking into ever and never smokers in paper I-IV. In paper I and IV we also estimated the cumulative dose of smoking through number of pack years smoked, classified into three groups (in paper I: (0-5, >5-10 and >10 pack years), in paper IV (0-9.99, 10-19.99 and >19.99 pack years). One pack year was equivalent to smoking 20 cigarettes per day for one year.
In paper III, information regarding smoking exposure from the CACORA study was categorized into ever and never smokers by using pseudo-year which is consistent with the procedure for constructing index year as used in EIRA and described earlier.

3.6.2 Alcohol

In EIRA alcohol consumption was estimated through questions regarding consumption of different alcoholic beverages (in centilitres (cl)) during the week prior to filling in the questionnaire and if the participants had consumed alcohol during the latest 12 months. For each beverage we calculated unit values per week based on following

procedures (all amounts are in cl): amount beer/33 (equals one bottle of beer), amount of wine/20 (equals one glass), amount of strong wine/5 (one small glass), and amount of liquor/5 (one small glass). All these units were the added to estimate weekly consumption for each person. Based on alcohol consumption distribution among the controls we formed four categories: non-drinkers (eg. not having consumed alcohol during the last 12 months), low consumption > 0-2.9 units or drinks per week (0- 50 % up to the median), moderate consumption 2.9-4.9 units or drinks per week (~50- 75 % the third quartile) and high consumption > 4.9 units or drinks per week (~75 %-100 % the fourth quartile). These categories were then used to estimate the odds ratio regarding alcohol consumption and risk of developing RA. We also asked one question regarding if the consumption was representative for a usual week or if it was more or less than usual.
In CACORA the questions regarding alcohol consumption 10 years prior to inclusion in the study was used to form categories in the same way as for EIRA (eg. consumption among the controls in CACORA). In CACORA the units in the questions was glasses and not in centilitres.

3.7 CONFOUNDERS

We also investigated if our analysis were influenced by potential confounding from social economic status; body mass index, marital status, parity and oral contraceptive use (Paper I). These adjustments had minor influence on the estimates and were not retained in the final analysis. In paper III we adjusted for confounding from smoking by using a variable that was classified into the following categories: ever and never smoker. The social class variable was based on the last occupation held. In this thesis we also adjusted for potential confounding from use of oral contraceptives (ever, never), parity (ever, never), silica exposure (ever, never), body mass index (BMI) over 25 (yes, no) and having a degree from a university based education (yes, no). The results from analyses with concomitant adjustment for all potential confounders are not presented in general, since the reported estimates in paper I-IV were practically the same with or without such adjustment.

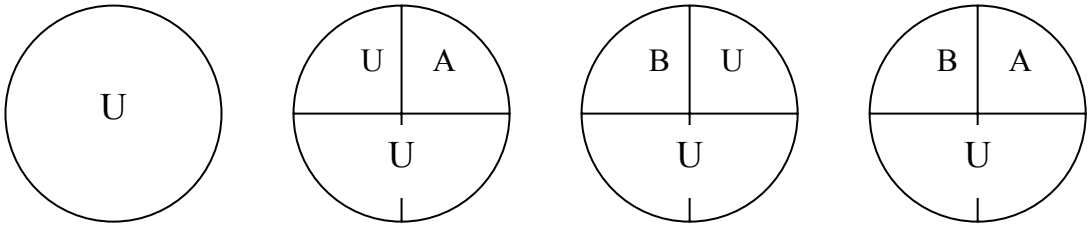
3.8 STATISTICAL ANALYSIS

We primarily used logistic regression models [74-75] to estimate odds ratios together with 95% confidence interval by using the proc logistic procedure in the 9.1 edition of SAS (statistical analysis system) in paper II-IV and the SAS edition 8.2 in paper I. All of the analyses were adjusted for the matching variables (age, gender and living area) by adding these variables to the logistic model. All results displayed are based on unconditional regression analysis. We did however, calculate the matched odds ratios by using conditional logistic regression, but decided to only present the estimates based on unconditional logistic regression. In general, there were no differences between ORs based on unconditional as compared to conditional logistic regression, besides slightly lower estimates (in some analyses) and wider confidence intervals when analysis was based on conditional likelihoods, especially seen when different RA subgroups were analysed. In paper I we present relative risks even though we have calculated odds ratios. Odds ratios can be interpreted as estimates of relative risks, in terms of incidence rate ratio, when using incident cases from a defined study base and selecting controls randomly continuously from the same study base [74].

3.8.1 Interaction and the pie model

Interaction between genetic differences (between alleles, haplotypes) and environmental exposures were evaluated primarily by using Rothmans definition of interaction as a deviation from additivity of effects (biological interaction) [76] (paper I-IV) In order to quantify the amount of biological interaction, we calculated the attributable proportion due to interaction (AP) together with a 95% confidence interval [77]. Attributable proportion is the proportion of the incidence among persons exposed to two interacting factors that is attributable to the interaction per se (i.e. reflecting their joint effect beyond the sum of their independent effects). Additive interaction is displayed in the context of the 'pie model'. The pie model was originally described by Kenneth Rothman in the late 70's. The model is based on the assumption that each disease has one or many causes. If one cause is enough to induce disease then this cause is a sufficient cause. This is illustrated by one pie being filled with only one cause component or factor (this could be one example of a monogenetic disease, where the sufficient cause is the susceptibility allele). A complex disease is made up by many different sufficient causes or pies, where several contributory causes act together with each other in different constellations. In order to describe interaction between two factors, say A and B, with regard to a specific disease, let us divide all sufficient causes (i.e pies) for that disease into four classes. First there is a set of sufficient causes that neither contain A or B, but other factors denoted by U. Second, a set of sufficient causes that contain A, but not B (together with other complimentary causes still denoted by U), and third, the set of sufficient causes that contain B but not A, and finally a set of sufficient causes that concomitantly contain A and B (figure M1). Now, the incidence among those unexposed to both A and B corresponds to the occurrence of the first class of pies. The incidence among those exposed to A but not B to the occurrence of the first two sets of pies. The incidence among those exposed to B but not A to the occurrence of the first and third set and finally the incidence among those exposed to both A and B depends on the occurrence of all four subsets of sufficient causes. The attributable proportion due to interaction (AP) estimates the relative contribution of the fourth set of pies, i.e. the set of pies where the concomitant presence of both A and B are required to cause disease [76].

Figure M1



$$AP = \frac{RR_{AB} - RR_A - RR_B + 1}{RR_{AB}}$$

Where RR_A is the relative risk associated with exposure to A in the absence of B, RR_B the relative risk associated with exposure to B in the absence of A, and RR_{AB} the relative risk associated with the concomitant exposure to both A and B.
AP > 0 indicates that there is evidence for additive interaction. Confidence intervals for AP were estimated by calculating a symmetrical confidence interval by using the formulas developed by Hosmer and Lemeshow [77]. We primarily used a modified version of a SAS program [78] to calculate interaction effects. But, we have also used and developed other programs and applications to calculate measures of interaction [79-80]

Paper II
In the second paper (Gene-gene and gene-environment interactions involving HLA-DR, PTPN22 and smoking in two subsets of rheumatoid arthritis) we calculated interaction between SE genes and the R620W PTPN22 defined in three different ways. The first definition corresponds to the previously described definition (see under the subscript interaction). The two other definitions were: deviation from multiplicity [74] and deviation from independency of penetrance [81]. Deviation from multiplicity means that interaction is present if the odds ratio of having both the alleles is greater than the product of the odds ratios of the two separate alleles (figure M2).
Deviation from independency of penetrance means that interaction would be present if the penetrance of having both alleles is greater than the product of the separate allelic penetrance. Penetrance means that the gene is associated with a certain phenotype among cases (figure M3). In figure M3 the probability $P(A_1 \cap A_2)$ is the probability of allele A_1 and A_2 being expressed through a certain trait (phenotype).

Figure.M2. Multiplicative model:

$$\log it(P(Y = y, A, B, AB)) = \alpha + \beta_A \times A + \beta_B \times B + \beta_{AB} \times A \times B + \varepsilon$$

In this thesis we consider multiplicative interaction to be present if the following condition is fulfilled:

$OR_{AB} > OR_A \times OR_B$ or if $\beta_{AB} > 0$ in the multiplicative logistic model above (given OR_A and $OR_B > 1$).

Figure.M3. *Independence of penetrance:*

Gene 1	**Gene 2**		
	Allele A_2	**Allele G_2**	**Marginal probability**
Allele A_1	$P(A_1 \cap A_2)$	$P(A_1 \cap G_2)$	$P(A_1)$
Allele G_1	$P(G_1 \cap A_2)$	$P(G_1 \cap G_2)$	$P(G_1)$
Marginal probability	$P(A_2)$	$P(G_2)$	1

Deviation from independency (if the following condition holds): $P(A_1 \cap A_2) \neq P(A_1) \times P(A_2)$

3.8.2 Additional measures and statistical methods

Paper IV
In the fourth paper we used excess fraction percent (EF%) to estimate the contribution of smoking to RA in the population [82]. The formula for excess fraction percent is given by:

$$EF\% = P_E * \frac{(RR_E - 1)}{RR_E}$$

Where P_E is the proportion of cases being exposed and RR_E is the relative risk associated with the exposure.
In the third paper we also estimated the trend for alcohol consumption and risk of developing RA as proposed by Armitage [75]. Basically, this means that we constructed a variable with integers from zero to three, where zero represented no consumption, one equals more than 0 but less than or equal to 2.9 drinks per week, two equals more than 2.9 up to 4.9 drinks per week and finally, three represented consuming more than 4.9 drinks per week. This variable was then used in a logistic regression model to estimate a trend associated coefficient. This procedure for estimating trend was also used in the fourth paper.

4 RESULTS

The results are primarily based on results from EIRA, but we have used additional materials in all studies except in paper IV, which was entirely based on EIRA. Regarding EIRA, the participation proportion was 96 percent for cases and 82 percent for the controls. A total of 1419 cases and 1674 controls have answered the questionnaire from May 1996 to December 2003. Regarding blood samples, the blood contributing proportion was 92 and 63 percent for cases and controls respectively. In paper I we used information from a shorter recruitment period for EIRA than in paper II-IV. Some basic characteristics for the major studies used in this thesis are displayed in table R0.

Table R0. Some basic characteristics for the major studies included in the thesis.

Paper	Study	# cases/controls	ACPA + cases (%)	HLA-DRB1 SE (%) cases/controls	Sex (%)	Age (Std)
Paper I	EIRA	913/631	61	73/54	72	51 (12.5)
Paper II	EIRA	1183/793	61	74/52	73	51.5 (12)
Paper II	NARAC	430/793	81	86/43	82	58 (12)
Paper II	Leiden EAC	364/881	57	70/44	60	50 (15)
Paper III, IV	EIRA	1204/871	61	74/53	73	51.5 (12)
Paper III	CACORA	444/533	69	75/53	65	49.5 (11.5)

4.1 PAPER I

Only the main results from paper I are presented in this section. Additional information is located in the complete paper under the paper I section in this thesis.

We observed a dose-response relationship between proportion of cases with ACPA+ RA and amount of cigarette pack-years smoked. The proportion of ACPA+ cases was also higher depending on number of HLA-DRB1 SE alleles.
We also found citrullinated proteins in BAL cells from smokers as compared to non-smokers.

Presence of HLA-DRB1 SE alleles and smoking was associated with high risks for developing ACPA+ RA. Having two copies of HLA-DRB1 SE alleles and a history of smoking conferred an odds ratio of 21 with a corresponding 95 % confidence interval of: 11.0 - 40.2. The interaction between HLA-DRB1 SE and smoking was observed for men and women separately. Interaction as measured through AP was found for smoking and one or two copies of HLA-DRB1 SE alleles in a dose-dependent way regarding ACPA+ RA. As regards ACPA- RA, we did not find any increased risk, neither for smoking nor for HLA-DRB1 SE or an interaction between them.. These results are displayed in table R1.

Table R1. Relative risk (RR) with 95% confidence interval (95% CI) of developing RA for subjects (men and women) exposed to different combinations of smoking and SE alleles compared with never smokers without SE alleles. Attributable Proportion due to interaction (AP) with 95% CI for the interaction between smoking and SE alleles (single, double).

Anti-ccp Positive (ACPA+) RA**									
	No SE			**Single SE**			**Double SE**		
	ca/co*	**RR**	**95% CI**	**ca/co***	**RR**	**95% CI**	**ca/co***	**RR**	**95% CI**
Never smokers	20/87	1.0	-	72/104	3.3	1.8 – 5.9	36/31	5.4	2.7 – 10.8
Ever smokers	58/184	1.5	0.8 – 2.6	192/146	6.5	3.8 – 11.4	126/31	21.0	11.0 – 40.2
AP					0.4	0.2-0.7		0.7	0.5-0.9
Anti-ccp Negative (ACPA-) RA**									
	No SE			**Single SE**			**Double SE**		
	ca/co*	**RR**	**95% CI**	**ca/co***	**RR**	**95% CI**	**ca/co***	**RR**	**95% CI**
Never smokers	65/87	1.0	-	64/104	0.8	0.5 – 1.3	18/31	0.7	0.4 – 1.5
Ever smokers	84/184	0.6	0.4 – 1.0	76/146	0.8	0.5 – 1.2	18/31	0.8	0.4 – 1.7

* ca/co = number of exposed cases/number of exposed controls

** RR adjusted for age (10 strata), gender and residential area

4.2 PAPER II

In this paper we investigated if there are interaction between the two most established genetic risk factors for RA, HLA-DRB1 SE and R620W PTPN22 alleles in relation to different phenotypes of RA. Three different populations (one Swedish (EIRA), one North American (NARAC) and one Dutch (Leiden EAC)) were used to estimate interaction between the alleles. This paper also included estimation of interaction based on three different definitions of interaction (additive, multiplicative and deviation from independency of penetrance). Smoking was also considered in combination with the two susceptibility alleles in the Swedish EIRA study.

Interaction, defined as additive interaction measured through attributable proportion due to interaction, between HLA-DRB1 SE and R620W PTPN22 alleles was found in all three populations. This interaction was restricted to ACPA+ RA. We also found evidence for interaction, defined as deviation from linkage disequilibrium, in EIRA and NARAC separately and multiplicative interaction when data from all three studies were combined in one analysis. These analyses are displayed in table R2.1.

When investigating the presence of potential interaction between smoking and the R620W PTPN22 allele in EIRA, no evidence for interaction was found either for the ACPA+ or for the ACPA- RA subset. These results are presented in table R2.2.

Further analysis including smoking HLA-DRB1 SE and R620W PTPN22 alleles indicated strong interaction between HLA-DRB1 SE and R620W PTPN22 alleles as well as between HLA-DRB1 SE and smoking regarding ACPA+ RA. No interaction was found in the ACPA- subset. Odds ratios for ACPA+ and ACPA- are displayed in figure 2.1a and 2.1b, respectively.

Table R2.1 . Odds ratios for developing ACPA positive RA according to presence or absence of minor R620W PTPN22 and HLA-DRB1 SE alleles, respectively. Tests for interaction between these susceptible alleles according to three different methods. Results displayed for the EIRA, NARAC and the Leiden EAC studies separately, as well as pooled.

EIRA				
R620W PTPN22 Susceptible allele	**HLA-DRB1 SE allele**	**Number of cases/controls**	**OR [a]**	**95 % CI [b]**
None	None	75/284	1.0	……..
None	Any	416/330	5.0	3.7 – 6.7
Any	None	31/95	1.2	0.8 – 2.0
Any	Any	199/84	9.9	6.8 – 14.3
Deviation from additivity			p < 0.001	
AP [c] together with 95 % CI [b]			0.5 (0.3 – 0.7)	
Deviation from multiplicity			P = 0.06	
Deviation from independency of penetrance			p = 0.022	
NARAC				
R620W PTPN22 Susceptible allele	**HLA-DRB1 SE allele**	**Number of cases/controls**	**OR [d]**	**95 % CI [b]**
None	None	28/348	1.0	……..
None	Any	206/267	9.3	6.0 – 14.3
Any	None	8/69	1.5	0.6 – 3.4
Any	Any	105/47	30.2	17.6– 51.9
Deviation from additivity			p < 0.001	
AP [c] together with 95 % CI [b]			0.7 (0.5 – 0.9)	
Deviation from multiplicity			p = 0.05	
Deviation from independency of penetrance			p = 0.035	
Leiden EAC				
R620W PTPN22 Susceptible allele	**HLA-DRB1 SE allele**	**Number of cases/controls**	**OR [d]**	**95 % CI [b]**
None	None	28/413	1.0	……..
None	Any	122/316	5.7	3.7 – 8.9
Any	None	9/83	1.5	0.7 – 3.3
Any	Any	48/69	11.0	6.4 – 19.0
Deviation from additivity			p = 0.0016	
AP [c] together with 95 % CI [b]			0.4 (0.1 – 0.7)	
Deviation from multiplicity			p = 0.29	
Deviation from independency of penetrance			p = 0.76	
EIRA + NARAC+ Leiden EAC				
R620W PTPN22 Susceptible allele	**HLA-DRB1 SE allele**	**Number of cases/controls**	**OR [e]**	**95 % CI [b]**
None	None	131/1045	1.0	……..
None	Any	744/913	6.1	4.9 – 7.5
Any	None	48/247	1.4	1.0 – 2.1
Any	Any	352/200	13.2	10.2 – 17.2
Deviation from additivity			p < 0.001	
AP [c] together with 95 % CI [b]			0.5 (0.4 – 0.6)	
Deviation from multiplicity			p = 0.025	
Deviation from independency of penetrance			p = 0.027	

[a] Odds ratio adjusted for age, sex and living area (only EIRA)

[b] 95 percent confidence interval

[c] Attributable proportion due to interaction (AP)

[d] Odds ratio adjusted for age, and sex

[e] Odds ratio adjusted for age, sex and study

Table R2.2 Odds ratios for smoking and the R620W PTPN22 allele regarding risk of developing ACPA+/ACPA- RA.

	No R620W PTPN22		**Any R620W PTPN22**	
Smoking status	**No. of exposed cases/controls**	**OR* (95 % CI)**	**No. of exposed cases/controls**	**OR* (95 % CI)**
	ACPA+			
Never smokers	135/251	Referent	74/64	2.2 (1.5 – 3.4)
Ever smokers	385/368	2.1 (1.6 – 2.7)	165/115	2.9 (2.1 – 4.1)
	ACPA-			
Never smokers	160/251	Referent	42/64	1.0 (0.6 – 1.5)
Ever smokers	199/368	0.9 (0.7 – 1.1)	76/115	1.1 (0.8 – 1.6)

* Odds ratio adjusted for age, sex and living area.

Fig 2.1A. Anti-CCP positive RA

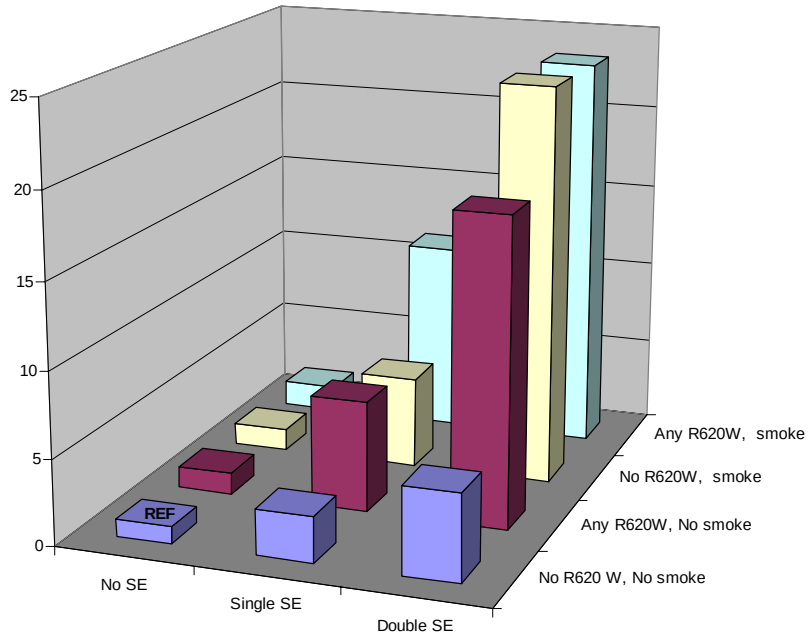


Fig 2.1B. Anti-CCP negative RA

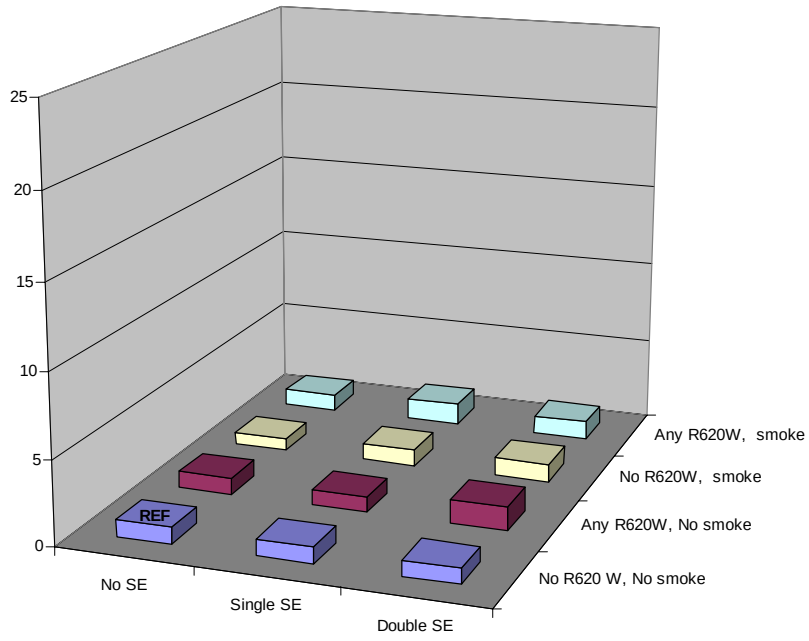


Figure 2.1. Histograms on odds ratios for developing anti-CCP positive RA (Fig 1A) and anti-CCP negative RA (Fig 1B) for different combinations of absence or presence of R620W PTPN22, single or double HLA-DRB1 SE alleles and smoking (based on the EIRA study).

4.3 PAPER III

In paper III data from two studies (the Swedish EIRA and the Danish CACORA) were used to investigate the association between alcohol consumption and risk of developing RA. The alcohol consumption differed between the controls in EIRA and CACORA, the Danish controls drank more than Swedish controls (table R3.1). The categories displayed in table R3.1 were used to classify cases and controls in the corresponding studies for further analysis.

Table R3.1. Alcohol consumption categories among controls in EIRA and CACORA

Alcohol consumption quartile distribution	**Alcohol consumption categories**	**EIRA (units/week)**	**CACORA (units/week)**
None	Non-drinkers	…….	……
<= median	Low	0-2.9	0-5
Median<cons.<=third quartile	Moderate	>2.9 - 4.9	>5-12
> third quartile	High	>4.9	>12

In order to investigate changes in alcohol consumption patterns due to disease duration, we investigated the proportions of cases with low, moderate and high alcohol consumption based on duration of symptoms and found that the alcohol consumption were similar and the largest group was low consumers irrespectively of symptoms duration (table R3.2).

Table R3.2. Alcohol consumption among EIRA cases by symptom duration

	Alcohol consumption among RA cases				
Duration of symptoms (Months)	**# cases (%)None**	**Low**	**Moderate**	**High**	**Total**
<= 1	3 (30)	5 (50)	1 (10)	1 (10)	10 (100)
1-2	8 (17)	22 (48)	8 (17)	8 (17)	46 (100)
2-3	18 (17)	55 (51)	18 (17)	17 (16)	108 (100)
3-4	19 (16)	70 (59)	13 (11)	16 (14)	118 (100)
4-5	14 (11)	69 (55)	25 (20)	17 (14)	125 (100)
5-6	14 (13)	62 (60)	20 (19)	8 (8)	104 (100)
6-7	16 (15)	55 (53)	14 (13)	19 (18)	104 (100)
7-8	14 (13)	60 (56)	14 (13)	20 (19)	108 (100)
8-9	15 (16)	52 (56)	12 (13)	14 (15)	93 (100)
9-10	5 (8)	46 (70)	6 (9)	9 (14)	66 (100)
10-11	14 (23)	33 (55)	8 (13)	5 (8)	60 (100)
11-12	12 (19)	36 (56)	8 (13)	8 (13)	64 (100)
>=12	30 (15)	109 (55)	24 (12)	35 (18)	198 (100)
Totalt	182 (15)	674 (56)	171 (14)	177 (15)	1204 (100)

We found a dose-dependent association regarding increased alcohol consumption and decreased odds ratios for RA in both EIRA and CACORA. The effect was more pronounced in the ACPA+ RA group.

In EIRA there were similar patterns for ACPA+ and ACPA- for the odds ratios except in the non-drinking group, in which the non-drinkers had higher odds ratio for ACPA+ RA than ACPA- RA. Odds ratios for alcohol consumption and different phenotypes of RA are displayed in table R3.3.

A significant interaction as measured through AP was found between smoking and no alcohol consumption regarding risk of developing ACPA+ RA in both EIRA (AP: 0.6 (95% CI: 0.5 - 0.7)) and CACORA (AP: 0.4 (95% CI: 0.3 - 0.5)). In addition to the observed interaction between smoking and alcohol, we also found interaction between homozygous HLA-DRB1 SE alleles and no alcohol consumption in both EIRA and CACORA. The odds ratios for different combinations of all factors (HLA-DRB1 SE, smoking and alcohol) are displayed in figure 3.1

Table R3.3. Odds ratio (OR) with 95% confidence interval (95% CI) of rheumatoid arthritis for subjects in different alcohol consumption categories, by ACPA status and study (EIRA and CACORA)

		EIRA					CACORA				
	Alcohol Consumption	**exp ca/co***	**OR†**	**95 % CI**	**OR‡**	**95 % CI**	**exp ca/co***	**OR†**	**95 % CI**	**OR‡**	**95 % CI**
RA overall											
	Non-drinkers	182/109	1.0	0.8 - 1.3	1.1	0.8 - 1.4	80/54	1.7	1.1 - 2.5	1.7	1.1 - 2.5
	Low	674/396	1.0 §	……	1.0 §	……	186/215	1.0 §	……	1.0 §	……
	Moderate	171/175	0.6	0.4 - 0.7	0.6	0.4 - 0.7	111/134	1.0	0.7 - 1.3	1.0	0.7 - 1.3
	High	177/190	0.5	0.4 - 0.6	0.5	0.4 - 0.6	67/130	0.7	0.4 - 1.0	0.6	0.4 - 0.9
	Trend		p<0.0001		p<0.0001			p=0.0006		p=0.0003	
ACPA-positive RA											
	Non-drinkers	120/109	1.2	0.9 - 1.6	1.3	1.0 - 1.8	62/54	2.0	1.3 - 3.0	2.0	1.3 - 3.1
	Low	410/396	1.0 §	……	1.0 §	…….	125/215	1.0 §	……	1.0 §	…….
	Moderate	97/175	0.5	0.4 - 0.7	0.5	0.4 - 0.7	77/134	1.0	0.7 - 1.4	1.0	0.7 - 1.4
	High	108/190	0.5	0.4 - 0.7	0.5	0.3 - 0.6	44/130	0.6	0.4 - 0.9	0.6	0.4 - 0.9
	Trend		p<0.0001		p<0.0001			p<0.0001		p<0.0001	
ACPA-negative RA											
	Non-drinkers	62/109	0.9	0.6 - 1.3	0.9	0.6 - 1.3	18/54	1.1	0.6 - 2.1	1.1	0.6 - 2.1
	Low	264/396	1.0 §	……	1.0 §	……	61/215	1.0 §	……	1.0 §	……
	Moderate	74/175	0.7	0.5 - 0.9	0.7	0.5 - 0.9	34/134	0.9	0.6 - 1.5	0.9	0.6 - 1.5
	High	69/190	0.5	0.4 - 0.7	0.5	0.4 - 0.7	23/130	0.9	0.5 - 1.5	0.9	0.5 - 1.5
	Trend		p=0.0004		p=0.0005			p=0.46		p=0.43	

* Number of exposed (exp) cases (ca) and controls (co), † Odds ratio adjusted for sex, age and residential area (EIRA only),‡ Odds ratio adjusted for sex, age, smoking (never/ever) and residential area (EIRA only), § Reference category.

In order to investigate if OR for developing RA was different depending on NSAID or Methotrexate treatment or not, we performed stratified analysis based on these treatments among the cases. There were no major differences regarding ORs associated with alcohol consumption based on treatment with NSAID or Methotrexate in cases (table R3.4)

Table R3.4 Odds ratios for alcohol regarding risk of developing ACPA positive and ACPRA negative RA according to different treatments among the cases (Only EIRA)

Alcohol consumption, with NSAID, Methotrexate treatment				
Alcohol consumption	**ACPA status**	**Exp ca/co***	**OR¶**	**95 % CI**
None	Positive	66/109	1.0	0.7 - 1.3
Low	Positive	246/396	Ref	……
Moderate	Positive	53/175	0.6	0.4 - 0.7
High	Positive	67/190	0.6	0.4 - 0.8
None	Negative	24/109	0.6	0.4 - 1.0
Low	Negative	145/396	Ref	……
Moderate	Negative	38/175	0.7	0.5 - 1.0
High	Negative	32/190	0.5	0.3 - 0.8
Alcohol consumption, without NSAID, Methotrexate treatment				
Alcohol consumption	**ACPA status**	**Exp ca/co***	**OR¶**	**95 % CI**
None	Positive	54/109	1.3	1.0 - 1.9
Low	Positive	164/396	Ref	……
Moderate	Positive	44/175	0.7	0.5 - 1.0
High	Positive	41/190	0.5	0.3 - 0.6
None	Negative	38/109	1.1	0.7 - 1.6
Low	Negative	119/396	Ref	……
Moderate	Negative	36/175	0.6	0.4 - 0.9
High	Negative	37/190	0.5	0.4 - 0.8

* Number of exposed cases (ca) and controls (co)

¶ Odds ratios adjusted for age, sex and living area.

Figure 3.1a

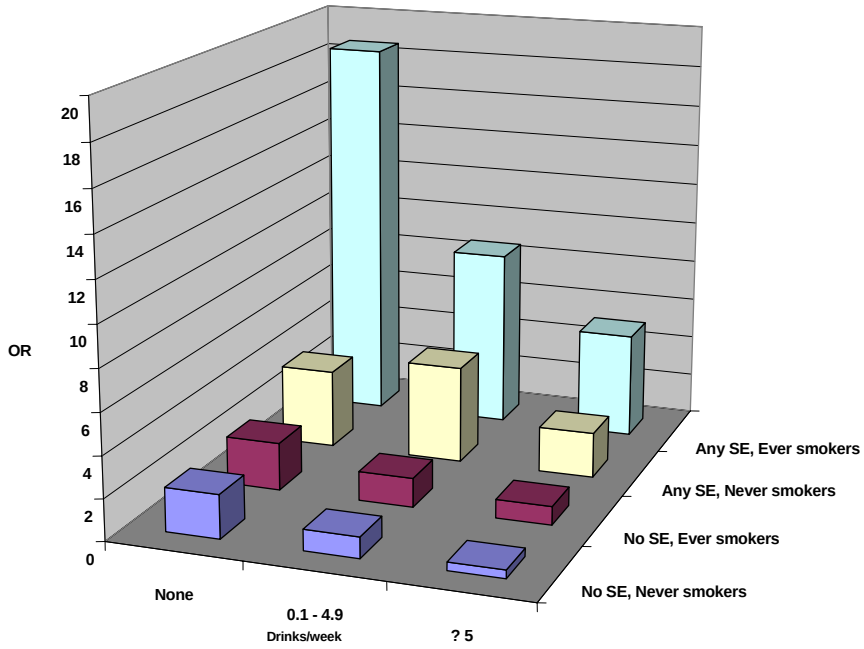


Figure 3.1b

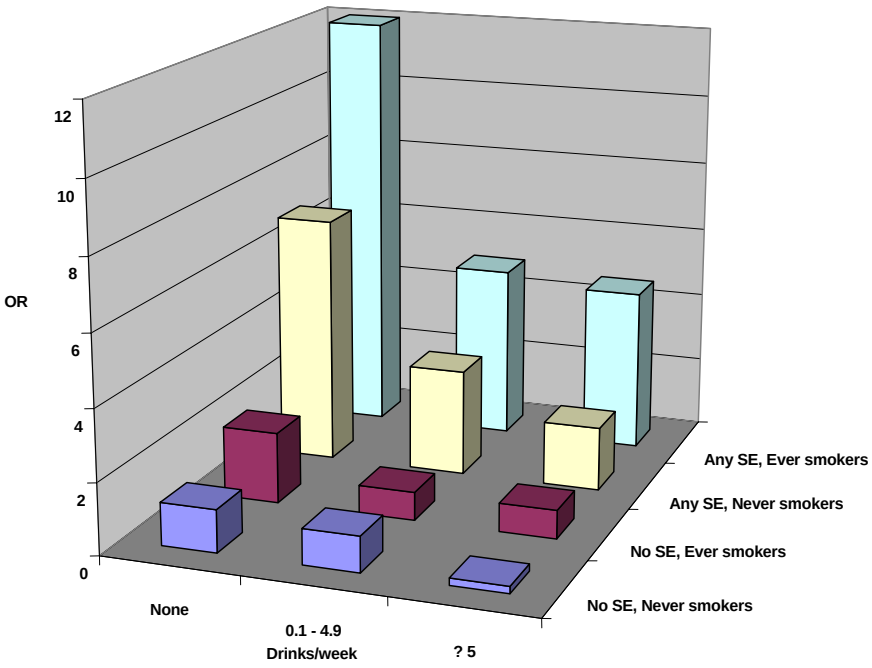


Figure 3.1: Histograms showing odds ratios (OR) of ACPA-positive RA for different combinations of alcohol consumption (three categories), smoking (never/ever) and presence or absence of HLA-DRB1 SE alleles compared with never smokers with low alcohol consumption (0.1-4.9 drinks/week) and without SE alleles, by study (EIRA fig 3.1a and CACORA fig 3.1b).

4.4 PAPER IV

In this paper interaction between smoking and HLA-DRB1 SE alleles regarding increased risk of developing ACPA+ RA was investigated further by investigating different amounts of smoking as measured by pack-years. All analysis presented here are restricted to the ACPA+ RA group.

The increased risk of ACPA+ RA due to smoking was observed to decrease along with the duration of time since cessation. Among women the trend was statistically significant (*p* trend .002), but not among men (*p* trend 0.14). Among women, twenty years after smoking cessation, there was no longer a significantly elevated risk whereas an increased risk was maintained among men (RR 2.1, 95% CI 1.1–4.0)

A strong interaction between smoking and HLA-DRB1 SE alleles was observed. A clear dose-dependent association between smoking and risk of developing ACPA+ RA was observed among all HLA-DRB1 SE genotypes, i.e. both among those without SE alleles, as well as those being heterozygous or homozygous for HLA-DRB1 SE alleles. The highest OR was seen in smokers having smoked more than 19.99 pack years and being homozygous for HLA-DRB1 SE alleles (OR: 37.6 (95% CI: 18.3-77.4). Different amounts of pack years smoked without considering HLA-DRB1 SE alleles status was associated with ACPA positive RA in a dose response manner, with the highest OR for ever smokers who had smoked more than 19.99 pack years. Smoking more than 19.99 pack years among those without the HLA-DRB1 SE alleles was associated with an increased OR of 1.9 (95% CI:1.1 - 3.5). This indicates that heavy smoking seems to be a risk factor for ACPA+ RA even without the presence of HLA-DRB1 SE alleles.(table 4.1)

The hypothesised proportion of ACPA+ RA cases that could be attributable to smoking was estimated to 35 percent indicating that smoking prevention would reduce the incidence of ACPA+ RA significantly if hindered. This corresponds to a reduction of 20 percent of all RA cases. Furthermore, in the group homozygous for HLA-DRB1 SE alleles smoking was estimated to be attributable for 55 percent of the ACPA+ RA cases. We did not find the same pattern in the ACPA- subset of RA.

Table R4.1. Odds ratios and attributable proportions due to interaction for different doses of smoking and HLA-DRB1 SE alleles, regarding the risk to develop ACPA positive RA.

		No SE	**Heterozygotic SE (SSE)**		**Homozygotic SE (DSE)**	
Smoking dose	**No. case/controls**	**OR* (95 % CI)**	**No. case/controls**	**OR* (95 % CI)**	**No. case/controls**	**OR* (95 % CI)**
No Smoke	38/154	1.0 (REF)	107/150	3.2 (2.0 - 4.9)	62/43	6.3 (3.7 - 10.9)
≤9.99 pack years	25/104	1.0 (0.6 - 1.8)	80/96	3.4 (2.1 - 5.6)	54/19	12.0 (6.2 - 23.0)
AP**	……	……	……..	0.09 (-0.32-0.49)	……	0.47 (0.12 - 0.83)
9.99>pack-years≤ 19.99	18/68	1.2 (0.6 - 2.2)	83/56	7.3 (4.3 - 12.4)	43/10	24.6 (10.9 - 55.8)
AP**	……	……	……..	0.53 (0.30 - 0.76)	……	0.73 (0.50 - 0.95)
>19.99 pack years	30/87	1.9 (1.1 - 3.5)	112/71	8.7 (5.3 - 14.4)	83/13	37.6 (18.3 - 77.4)
AP**	……	……	……..	0.51 (0.31 - 0.72)	……	0.80 (0.67 - 0.94)
p-value for trend (OR)		P = 0.11		p < 0.0001		p < 0.0001

*** Adjusted for gender, age and living area**

**** Attributable proportion due to interaction with 95 % confidence interval**

5. DISCUSSION

Our results points out that there are gene-environment and gene-gene interaction, in terms of developing serological defined types of RA. This type of reasoning is probably true for many complex diseases with multi factorial aetiology. Interestingly, our results regarding an interaction between smoking and HLA-DRB1 SE for developing ACPA+ RA has been confirmed in a Dutch study, a Danish study, and to some extent in a French study [32, 60, 33]. The Dutch and the French studies are case-only studies which might dilute the associations. In the French study there is a higher prevalence of HLA-DRB1 SE alleles in the ACPA- group (68 %) as compared to the EIRA study (57 %) and Leiden EAC (53 %). The combination of SE alleles, smoking and ACPA+ RA has also been associated with increased hazard ratios for CVD related death [83], which might indicate that there is an association between SE, smoking and ACPA, regarding amount of inflammation.

Our results suggest that smoking is involved in the process of citrullination of self peptides and that SE genes is associated with increased affinity to citrullinated self peptides, eventually leading to autoimmune reactions and disease.

In the paper IV we also observed a strong dose response relationship within different categories of SE alleles and that smoking a lot may be a risk factor for ACPA+ RA without the presence of SE alleles. We also concluded that smoking is a major preventable factor for RA, which could, if hindered, decrease the proportion of new ACPA+ RA cases by approximately 35 percent.

The gene-gene interaction between HLA-DRB1 SE and R620W PTPN22 regarding risk of developing anti-CCP+ RA, based on the EIRA material was also confirmed by the results we obtained when we analysed material from NARAC and Leiden EAC. The interaction between SE and R620W PTPN22 alleles suggests that disease onset is likely to happen be some sort of MHC II dependent T-cell activation, where the R620W PTPN22 allele is associated with the threshold for T-cell activation in the RA subset characterised by presence of ACPA. In addition to our results one large UK study has reported gene-gene interaction between HLA class II and PTPN22, regarding diabetes type I, which is also considered to be an autoimmune disease [84]

Our results regarding alcohol consumption and the attenuation of the risk factors smoking and HLA-DRB1 SE alleles for RA has not yet been confirmed in any study. Notable is though that a Korean study found alcohol consumption to be associated with less extra articular manifestations in RA cases [85]. Additionally, the evidence for alcohol being an immunological attenuator is convincing. [36]

There are however some question marks regarding study design, recruitment, data management, data-, and biological-, analysis that needs to be discussed before we proceed and draw firm conclusions.

5.1 METHODOLOGICAL AND STUDY DESIGN CONSIDERATIONS

One potential source of bias is related to missing cases that should be included in the study. RA cases that are: not referred to rheumatologists, have other diseases that are considered to be more important than RA, have died or not meet a doctor at all, are all potential threats to the generalizibility or external validity of the study. Hopefully, these problems have a minor impact on our study but it is hard to estimate the effect. Bengtsson et al. have investigated characteristics of participating, non-participating cases and controls, respectively, with regard to demographic and socioeconomic variables. They found that non participating cases were in general older than participating cases. Among controls, the non participators were more likely to be male, being slightly younger, unmarried and living in urban areas as compared to participators. While, having a low income, not being born in Sweden and being less educated was associated with non-participation in both cases and controls. The observed differences were considered to only marginally bias the estimated odds ratios though [64].
The replications of our findings in other materials could also be considered to strengthen the generalizbility of the results. For some genetic questions it will probably be impossible to replicate findings due to genetic difference in the populations. In this thesis we have focused on the two most established genetic risk factors for RA that have been replicated in different populations (HLA-DRB1 SE alleles and the R620W PTPN22 allele).
By using the 1987 ACR criteria for inclusion in EIRA we might miss to include cases that will develop RA in the near future , since the 1987 ACR RA criteria might be to strict for inclusion of cases with milder symptoms. This inclusion problem is probably more problematic regarding inclusion of cases with less severe RA, or cases having ACPA- RA. The main results in this thesis are focused on risk factors associated with ACPA+ RA, which should decrease the potential impact of this inclusion problem regarding inferences made.

There is a possibility of misclassification of disease due to ACPA status, since it was only measured once. This could potentially lead to a dilution of associations between exposures and risk of RA assuming the misclassification is uncorrelated to the exposure. ACPA is however, associated with higher specificity and positive predictive value for RA than RF [7]. There are also indications on ACPA levels being a stable phenotype during follow-up in early RA cases [13].
In EIRA we used a population based design with very high participation proportion for cases and controls (96 % for the cases and 82 % for the controls). The high participation proportion and population based design should minimize the possibility of a large selection bias. Even if the total difference between cases and controls (14 %) regarding participation proportion constituted of smoking controls we would obtain significantly increased odds ratios between smoking and ACPA+ RA (pseudo estimated odds ratio of 1.4 (95 % CI: 1.2 - 1.7) as compared to the actual odds ratio of 1.8 (95 % CI: 1.5 - 2.1)). It is not likely though, that all controls not participating are smokers. The question regarding selection bias is also a concern regarding differences between proportions of cases and controls contributing with blood samples. In the case of EIRA, this is not a likely source of selection bias leading to over estimation, since we found slightly higher odds ratios for smoking and risk of developing ACPA+ RA

when we used information from cases and controls without HLA-DRB1 SE allele information (the OR for smoking and risk of RA was 2.5 (95% CI: 1.3 - 4.9) as compared to the results based on cases and controls with genetic information regarding HLA-DRB1 SE allele (OR = 1.8 95% CI: 1.4 - 2.2)). In order to check for possible selection bias regarding smoking and genetic information we investigated if there were any association between smoking and leaving a blood sample among the controls (OR = 1.0 95% CI (0.8 - 1.2)). This indicates that there is no selection bias regarding smoking and genetic information among the controls.

Similar analysis regarding alcohol consumption did not result in any increased OR for any of the categories used (Non-, Low-, Moderate-, or High-consumers).

When using the case-control design to study genetic associations, it is possible that observed differences are a result of underlying structure in the population and not a result of a disease associated locus [86]. This sort of bias is referred to as population stratification. By using information regarding SNPs unlinked to RA from a wide genome analysis in EIRA the potential impact from population stratification as measured through a modified version of the Armitage trend test [87] was estimated to λ_{GC} =1.03 for ACPA+ RA after removing the MHC region. This value is well below 1.1, which is the threshold that indicates presence of population stratification. The genetic distribution among cases and controls in EIRA and NARAC are in Hardy Weinberg equilibrium indicating that alleles are randomly distributed in cases and controls, respectively [55]. In addition to this control, it would be strange if the findings regarding HLA-DRB1 SE alleles and PTPN22 alleles in three different populations originally from countries with heterogeneous and mixed populations would be a result of population stratification.

Retrospective cases-control studies, such as EIRA, are often criticised for being associated with recall bias. Recall bias appear when cases and controls differ in the way they remember events or exposures. In this thesis we used information from questions regarding smoking habits and alcohol consumption. It is hard to estimate the potential effect of recall bias without performing a specific study with the aim to investigate the possibility of a differential reproducibility between cases and controls. In EIRA we used cases with a recent onset of RA in order to decrease the impact of potential recall bias. The average time between symptoms onset and diagnosis was 10 months. This is a quite short time which should minimize the risk of cases altering their exposures or remembering exposures differently than controls. It has been observed that patients that change their living habits afters they have received their diagnosis tend to extrapolate their new habit backwards in time [88-89]

Misclassification due to intervention may be an issue in the case of measuring alcohol consumption. Cases may have changed their alcohol consumption behaviour due to advice given from the doctor responsible for medication. We did not find any differences regarding ORs for alcohol and risk of RA when performing separate analysis based on Methotrexate, NSAID treatment [table R3.4]. In addition to the absence of treatment differences, the alcohol consumption was quite similar for cases with different symptoms durations [table R3.2]. This could be considered to be an indication of alcohol consumption being quite stable for cases with recent onset of RA and that the cases don't change their drinking behaviour initially, which otherwise could have changed during the disease course and potentially explained our findings.

Besides smoking there are few established environmental agents that could be considered as confounders in our analysis. One exposure that has been associated with

increased risk for RA is exposure to silica or stone dust [34]. Silica or stone dust exposure is a quit rare exposure though, making it an unlikely to confound the observed associations.

Social class indicators such as not having a degree from higher education has been found to be associated with increased risk for developing RA. Bengtsson et al. reported that having a university degree was associated with a decreased risk for RF+ RA [90]. There are some potential confounders that we have considered in this thesis (table D1) such as having a university degree, BMI, silica exposure, mineral oil exposure, use of oral contraceptives and parity. These potential confounders did not change our estimates in general and could not explain our findings.

Table D1. Adjusted odds ratio* for alcohol consumption, PTPN22 and smoking dose associated with risk of developing ACPA+ RA

	Ordinary	**No SE**	**Any SE**
Exposure	**OR* (95 % CI)**	**OR* (95 % CI)**	**OR* (95 % CI)**
None alcohol	1.4 (1.0 - 1.9)	……	…….
Low alcohol	Ref.	……	……
Moderate alcohol	0.5 (0.4 - 0.7)	…….	……
High alcohol	0.4 (0.3 - 0.6)	…….	……
None alcohol	…….	4.1 (1.5 - 11.0)	11.2 (4.7 - 26.6)
0.1 - 4.9 drinks/week	…….	1.7 (0.7 - 3.8)	9.3 (4.2 - 20.5)
> 4.9 drinks/week	…….	Ref.	3.9 (1.6 - 9.3)
No PTPN22	…….	Ref.	5.0 (3.5 - 7.1)
Any PTPN22	…….	1.2 (0.7 - 2.1)	8.4 (5.4 - 13.0)
No Smoking	Ref.	Ref.	7.8 (5.5 - 11.0)
0 - 9.99 pack years smoking	1.2 (0.9 - 1.5)	2.2 (1.2 - 3.8)	9.8 (6.7 - 14.4)
10 - 19.99 pack years smoking	1.9 (1.4 - 2.5)	2.0 (1.0 - 4.0)	15.9 (10.3 - 24.6)
> 19.99 pack years smoking	2.8 (2.1 - 3.8)	4.0 (2.2 - 7.3)	21.7 (14.3 - 33.0)

* All odds ratios are adjusted for having a university degree, BMI, silica exposure, mineral oil exposure, use of oral contraceptives and parity.

We have not considered confounding from any dietary factors, though. But it is unlikely that these difficult to study factors could explain our strong findings, which are only observed in a defined sub group of RA, were ACPA is present.

We have used unconditional logistic regression in all studies despite the design with 1:1 matching in EIRA. Using unconditional logistic regression when conditional logistic regression should be used, can bias the estimates. We did however also perform conditional logistic regression and found no large differences as compared to unconditional analysis, other than slightly lower estimates for some estimates and wider confidence intervals for the ORs when analysis was based on conditional analysis. There largest difference between conditional and unconditional analysis regarding AP for the different papers in this thesis was found in the HLA-DRB1 SE-R620W PTPN22 interaction analysis (paper II) where the AP was estimated to 0.4 (95% CI.: 0.1 - 0.7)

based on the conditional analysis as compared to 0.5 (95% CI.: 0.3 - 0.7) based on the unconditional analysis.

In a recent study regarding when to break the matching in case-control studies, unconditional logistic regression performed better than conditional logistic regression in terms of method bias (differences regarding estimates for full data model and reduced data model) when data is missing [91]. This is comforting results since we would loose power if we only considered conditional regression analysis based on individually matched cases and controls with complete information.

In all papers we have used attributable proportion due to interaction (AP) as an indicator of gene-gene-, gene-environment- and environment-environment- interaction. The confidence interval for AP was estimated by using a formula proposed by Hosmer and Lemeshow. This method has been targeted for discussions regarding poor performance [92-93]. When using the MOVER method to estimate confidence intervals (method of variance estimates recovery [92]), the results were almost identical for the gene-gene interaction in paper II, and the limits for the 95 percent confidence intervals only differed on the second decimal. In paper II we also compared different methods for investigating interaction effects; additive, multiplicative and deviation from independency of penetrance of two unlinked loci (referred to as LD statistic). We showed that when we combined data from all studies in the same analysis we obtained significant results for all interaction methods. Unfortunately, the LD statistic did not offer the possibility to adjust for additional confounding factors. Multiplicative interaction was associated with lower sensitivity but higher specificity for finding additative interaction as compared to using attributable proportion due to interaction.

5.2 BIOLOGICAL CONSIDERATIONS

The overall interpretation of the results in this thesis relay on autoimmunity being induced by the process of citrullination. Citrullination is a post translational process which occurs when the amino acid arginine is converted to citruline [94]. In paper I we found a higher proportion of citrullinated peptides in bronchoalveolar lavage cells in smokers as compared to non-smokers. Interestingly, the proportion of citrullinated BAL cells was highest in smokers with pulmonary inflammation. In an extended study regarding smoking and citrullinated peptides in BAL cells, a higher expression of the enzyme peptidylarginine deiminase 2 (PAD2) in smokers was considered to be a potential cause of the higher proportion of citrullinated cells [95].

In paper I this citrullination was suggested to induce the production of anti-bodies toward citrullinated peptides. ACPA as measured in the present thesis is in fact, associated to a collection of anti-bodies toward different citrullinated auto antigens [96]. This sub group diversity might be helpful in the search for a more specific disease mechanism. The most common antibody was IgG antibodies toward citrullinated-fibrinogen and thereafter was IgG antibodies toward citrullinated alpha-enolase and citrullinated C1-epitope of collagen II. The sub groups of anti bodies toward citrullinated peptides were primarily associated with the HLA-DRB1*04 genotype.

Genetic studies on RA have repeatedly shown regions on chromosome one and six to be associated with RA [56, 97-99]. So far, the most important genetic factors belong to regions on chromosome six, in particular on the HLA-DRB1 chain where the SE alleles are located. The SE alleles have repeatedly been associated with ACPA+ RA in different populations [28, 56, 60-61, 63]. In a recent study SE was considered to be

important in African Americans [100]. This could however be a consequence of admixture between Europeans and African Americans. Despite the issue of admixture, SE is a major risk factor for ACPA+ RA only. In all our studies (paper I-IV) we use a low resolution analysis of HLA-DRB1 and we define SE alleles as being present if a persons have at least one copy of HLA-DRB1*04, HLA-DRB1*01 or HLA-DRB1*10. In a Dutch study based on the Leiden EAC a high resolution analysis on HLA-DRB1 indicated that primarily DR*10, DR*1 and smoking was associated with interaction with regard to risk of ACPA+ RA and that DR*0401 conferred the highest risk for ACPA+ RA [61]. It is not clear though if the authors would have found interaction between smoking and HLA-DRB1*0401 if they had considered additive or biological interaction instead of multiplicative interaction. A rough estimate of AP for interaction between smoking and HLA-DRB1 *04 based on figures from the Leiden EAC article resulted in an AP corresponding to 0.19 with a 95 % CI corresponding to: -0.37 - 0.76 (based on covariance terms equal to zero). This indicates that there is a lack of power to detect additative interaction in Leiden EAC.

In EIRA also DR*0401 and smoking was found to interact regarding development of ACPA+ RA and the strongest interaction (highest AP) was observed for smoking and DR*0401 homozygousity [101]. The amino acid sequences for the DR*0401 allele is QKRAA which is quite similar to the sequences QRRAA and RRRAA, which are associated to the other SE alleles. Despite the similar amino acid sequences, the results indicate that MHC class II molecules coded for by DR*0401 alleles might have higher affinity for certain citrullinated peptides than other MHC class II molecules. In DRB1*0401 transgenic mice T cell immunity to citrullinated peptides may contribute to arthritis [102-103].

The second most "important" genetic risk factor for RA is the R620W PTPN22 allele. PTPN22 is short for protein tyrosine phosphatase 22. This enzyme is important in the regulation of the process of suppressing T-cell activation. The functional polymorphism at position 1858 constitutes of C being substituted by a T. The PTPN22 gene is located on another chromosome than the HLA-DRB1 gene, which indicates that interaction effect is due to physiological mechanisms and not a consequence of linkage disequelibrium between the genes coding for HLA-DRB1 and PTPN22 molecules.

In a sub study based on the Nurse's Health Study, the authors report, in contrast to our result (table R2.1) interaction between R620W PTPN22 and smoking regarding RF+ RA [104]. Furthermore, they could not replicate our findings in paper II, where we found gene-gene interaction between SE alleles and R620W PTPN22. Their results did however indicate presence of interaction effects and the power for drawing conclusions regarding absence of interaction was low. The authors, however, were not able to include information on ACPA status or to consider the potential impact of SE alleles regarding gene-environment interaction with smoking in that particular manuscript. A further analysis regarding the potential interaction between HLA-DRB1 SE and PTPN22 alleles stratified by smoking, in order to investigate effect modification as well as confounding from smoking should be of interest in this sub sample of the Nurse's Health Study.

In paper II the higher odds ratios observed for NARAC, might indicate presence of a true relationship regarding the gene-gene interaction between HLA-DRB1 SE and R620W PTPN22 alleles. Especially since the cases in NARAC were selected to have established and severe RA. The proportion of ACPA positive cases in NARAC was 81

percent, which is a much higher proportion than in EIRA and the presence of ACPA is also associated with more severe RA.
Additional genetic factors that have been associated with RA are TRAF1-C5 and STAT4 [97, 105]. Interestingly, TRAF1 is considered to bind intracellular proteins such as the nuclear factor-κβ inhibitory protein TNFAIP3 [106]. The activity of nuclear factor-κβ in monocytes has been observed to be down regulated by alcohol [37]. Alcohol was also shown to decrease TNF-α production in monocytes. Also this result is interesting since blocking of TNF-α is one treatment that is being used for RA. Previous observations make alcohol an interesting environmental factor, potentially influencing the risk to develop RA. In paper III there was no association between alcohol consumption and decreased risks for ACPA- RA in CACORA. In CACORA the proportion of ACPA+ RA cases was higher than in EIRA. This might be due to the inclusion of prevalent cases in CACORA than EIRA that included incident cases. Alcohol is considered to have an effect on both innate and adaptive immunity [36]. Recently, similar findings as ours with regard to alcohol consumption and risk of RA, was presented from a Swedish research group using a prospective cohort study conducted in south of Sweden. [107]. Previous studies on RA and alcohol consumption gives a mixed picture on a potential protective effect. In the Iowa study the participants mainly consisted of older women which may be a problem due to the amount of exposure to alcoholic beverages [38]. In fact, only a few of the women in the Iowa study drank more than one glass of wine during a week, which is a potential explanation to why no effect was observed.
In an experimental study on mice, increased levels of testosterone due to alcohol consumption was considered to be one explanation to the damped immune response. Unfortunately, only male mice were used in this experiment. But, alcohol consumption has been observed to be associated with increased levels of testosterone also among women [108-109].
Other interesting tracks regarding potential important biological mechanisms include exposures associated to increased levels of IL-6. IL-6 is a pro inflammatory cytokine that is secreted by T-cells, macrophages and from contracted muscle cells [110-111]. Prolonged increased levels of IL-6 caused by inflammation, could act as a catalyst in predisposed individuals eventually causing RA onset. In the context of muscle induced secretion of IL-6, physical workload might be one additional environmental risk factor for onset of RA [112].

5.3 A HYPOTHETICAL MODEL

Based on the results in the present thesis a hypothetical model on how genetic and environmental factors coexist and interact is presented in figure D1 which is an expanded version of the figure presented in Future rheumatology [113]. In this simplified hypothetical model the IL-6 and TNF-α has been omitted despite their importance.

If SE alleles and PTPN22 coexist the risk of developing ACPA positive RA was elevated more than one could expect from the sum of the separate effect of SE and PTPN22 alleles. This interaction indicates that the presence of major histocompability complex 2 (MHC-2) associated to SE alleles on the antigen presenting cell (APC) could initiate the T-cell activation, which is maintained in the presence of the PTPN22 allele, eventually leading to autoimmunity (fig D1, paper II). Again, this interaction was only observed for ACPA positive RA. There was no evidence for interaction between PTPN22 and smoking [table 2 in the result section, 114], in contrast to the observed interaction between SE alleles and smoking. This gene-environmental interaction has also been then reproduced in other studies as mentioned before [32.-33, 62]. In article I we also found that citrullinated peptides were more common in smokers than non smokers. This result together with the interaction between SE alleles and smoking regarding ACPA positive RA leads to the hypothesis that smoking induces citrullination of peptides in a dose dependent way (paper IV), which makes APCs with certain type of MHC-class II molecules (SE alleles) to present these peptides efficiently and eventually trigger autoimmunity (fig D1). This process may be slowed down by introduction of alcohol through alcoholic beverages, which primarily decrease the innate immune response and the overall immune system activity (paper III).

The strong interaction effects between smoking, PTPN22 and SE alleles described above are restricted to the subgroup of RA where ACPA antibodies are present, which pinpoint the importance of defining the subgroups of RA. The sub-grouping of the disease is probably equally important for finding successful treatments for RA, due to the obvious differences regarding disease mechanisms as described earlier. Due to the different features of ACPA+ and ACPA- RA, it may be appropriate to change the diagnosis criteria regarding RA will be changed in a near future, in order to optimize the treatments for the patients.

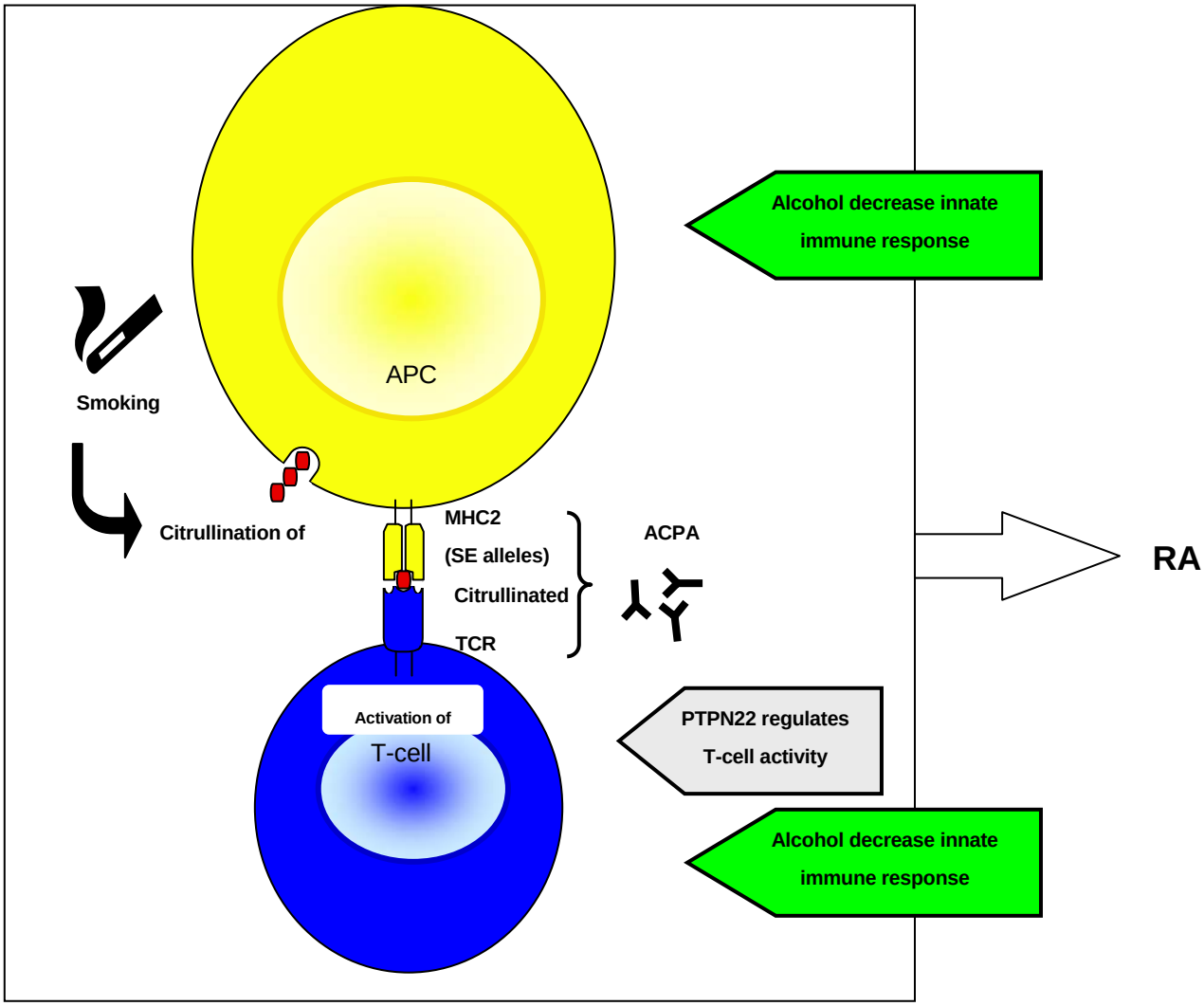


Figure D1. Hypothetical model for development of ACPA+ RA in the presence of smoking, alcohol, R620W PTPN22 and HLA-DRB1 SE alleles.

5.4 FUTURE THOUGHTS

Due to the nature of a complex disease such as RA, it is not efficient with a reductionist approach to search for major aetiological factors any more. The future spells: cooperation new methods (both biological and statistical methods) and most of all a holistic approach where genes are not considered to function without the impact of environmental factors or that environmental factor are not considered to interact with our genes. This approach demand that we use methods to investigate how genes are expressed, how gene function is modified from epigenetic changes, and what effect the proteins have and how they interact with other proteins in an environmental setting. Figure D2 show how the paradigm of systems biology is expanded to include the environment and show that the importance of the environment as a disease cause is increasing as we move further away from the genomics towards metabolomics.

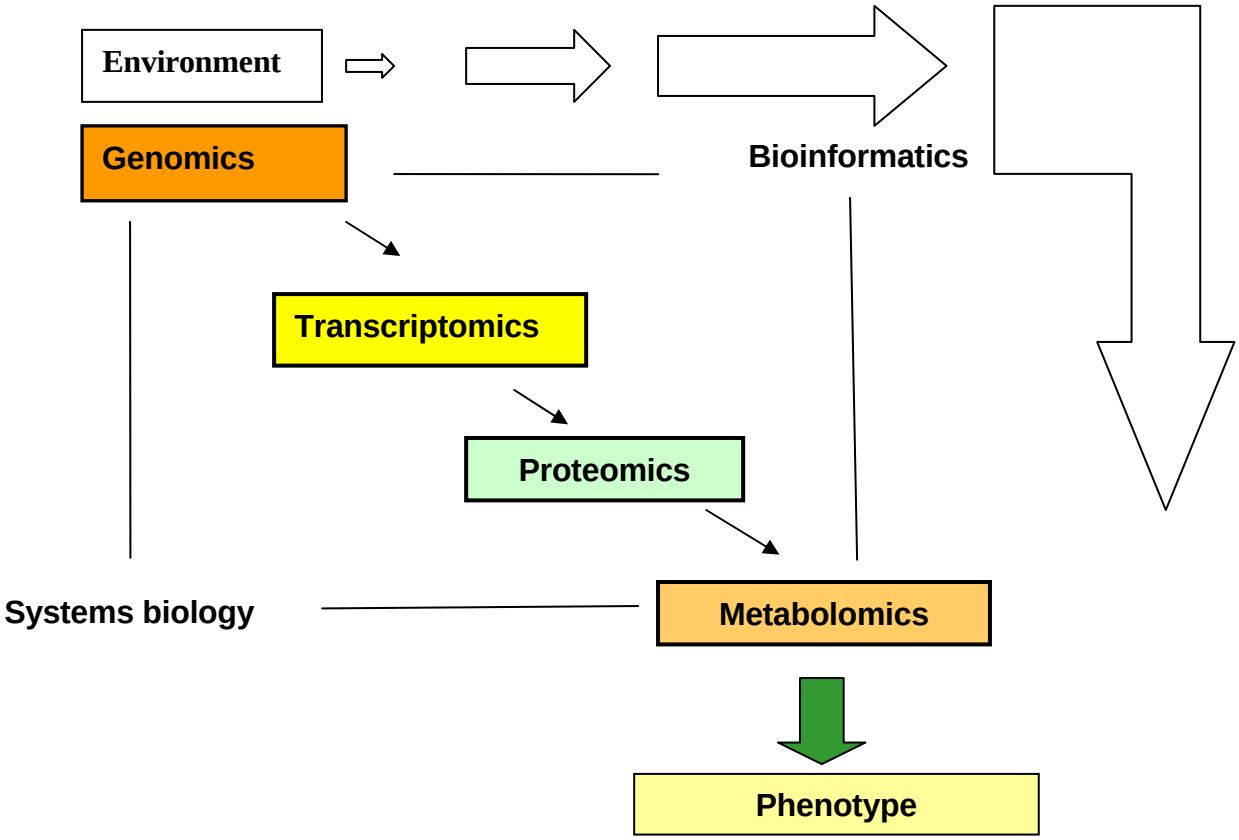


Figure D2. Schematic picture over the different parts in systems biology and the omics cascade in an environmental context.

This implies that we need to develop new, as well as use present biological methods for investigating physiological activities in the body. We also need to develop new statistical methods as well as use existing models for handling complex data material and to test complex ideas. Regarding statistical methods, I think we need to move away from the traditional frequentistic ideas aiming to estimate p-values in favour of the Bayesian way of thinking, where the results are considered to be the actual results and not a hypothetical results that would not occur by chance if we were to repeat our study many times in the exact way using the same participants etc. Bayesian networks offer a flexible yet stringent way to model causal relationships and have been proved to be

very useful in laboratory based research as presented by Eric Schadt and colleagues [115]. Additionally and maybe most important of all: we need to use well designed studies and really consider different sources of biases. Study designs have a central role in the field of epidemiology and therefore should epidemiological "thinking" have a central role in planning and analysis of complex studies on complex diseases.
In the future we will probably use computers even more than now, as they will be a necessary tool for handling these issues regarding complexity and we will probably start using, as well as develop complex computer algorithms for handling huge amount of complex information. We probably have to glance on other scientific disciplines, in order to find new ways to increase our knowledge. In the case of RA, we have already taken small steps towards integration of environmental and genetic data in an epidemiological setting. But, we have just started and the steps are so far, only the thriving steps of a baby learning to walk, holding the hand of a parent called science. Hopefully, environmental data bases will be created with the aim to study and perhaps simulate disease mechanisms from an environmental perspective, in a more complex way than genetic and pathway data bases such as NCBI and KEGG are used today [116-117].
Finally, it is likely that the results from this thesis will be rejected because of new and more refined findings, in a way Karl Popper once described. But who knows, maybe Popper was wrong.
What symbol could better than yin yang, display a holistic viewpoint, as exemplified by the interplay of genes and environment, as well as the importance of genes in the environmental context and the other way around?

6 CONCLUSIONS

- Smoking and presence of HLA-DRB1 SE alleles were observed to strongly interact regarding risk of developing Rheumatoid Arthritis characterized by presence of anti-bodies toward citrullinated peptides. Furthermore, citrullinated peptides were more common in bronco alveolar cells from smokers than in cells from non smokers. These results provided the basis for an aetiology model for development of Rheumatoid Arthritis.

- Having both the HLA-DBR1 SE and R620W PTPN22 alleles was associated with gene-gene interaction effects for developing ACPA positive Rheumatoid Arthritis. This gene-gene interaction was found in three different materials (one Dutch, one North American and one Swedish study) by using interaction defined as departure from additivity of effects. Interaction defined as departure from multiplicativity of effects was found when all three different materials were combined. No gene-environmental interaction was found between smoking and the R620W PTPN22 allele in the Swedish study.

- Alcohol consumption was associated with a dose dependent decrease in risk of developing RA, irrespective of ACPA status, in two studies from two different countries (Denmark and Sweden). The most pronounced risk decrease of developing ACPA positive RA was observed for high alcohol consumers being ever smokers with HLA-DRB1 SE alleles. There was also an indication of interaction between smoking and absence of alcohol consumption regarding risk of developing ACPA positive RA.

- Finally, smoking was associated with strong dose dependent effects regarding risk of ACPA positive RA. A dose-dependent effect was seen in each category of HLA-DRB1 genotype. The interaction between smoking and HLA-DRB1 SE alleles seemed to be dose dependent as well. Smokers with double SE had almost 40 times higher risk to develop ACPA+ RA compared with never smokers without SE; the corresponding attributable proportion due to interaction was 0.8 (0.7 - 0.9). In addition, the public health impact from smoking as a major preventable factor for ACPA positive RA was considered. It was found that the incidence of ACPA positive RA cases could in theory be decreased by 35 percent by means of smoking prevention. This corresponds to a reduction of 20 percent of all RA cases.

7 SAMMANFATTNING PÅ SVENSKA

Reumatoid artrit [RA] är en kronisk inflammatorisk sjukdom som framförallt drabbar kvinnor. Sjukdomens specifika karaktäristik kännetecknas av ledinflammation och progressiv destruktion av brosk, ben och ligament i leder. RA betecknas ofta som en autoimmun sjukdom vilket innebär att kroppens immunförsvar angriper den egna kroppen istället för främmande ämnen. Kunskapen kring orsaksfaktorer för uppkomst av RA är fortfarande knapp men både genetiska-, och miljömässiga-faktorer antas ha betydelse. Hittills har allela former av generna HLA-DRB1, PTPN22 samt den miljömässiga faktorn rökning associerats med ökad risk för RA.

Målet med denna avhandling är att studera betydelsen av genetiska och miljömässiga faktorer samt dessas potentiella interaktion med avseende på risken att insjukna i serologiskt definierade subgrupper av reumatoid artrit. Specifikt avses rökning och alkohol konsumtion och allela former av generna HLA-DRB1 och PTPN22 studeras. Med serologiskt definierade grupper av RA menas framförallt fall med eller utan antikroppar mot citrullinerade peptider (ACPA).

Avhandlingen baseras i huvudsak på material från EIRA (Epidemiological Investigation of Rheumatoid Arthritis). EIRA är en populationsbaserad fall-kontroll studie där 1419 fall som diagnostiserats med RA enligt ACR kriterier från 1987 och 1674 kontroller inkluderats mellan 1996 och 2004. Fall och kontroller fick på frivillig basis fylla i en omfattande enkät över bland annat levnadsvanor och yrkesliv, dessutom fick fall och kontroller frivilligt lämna blodprov för genetisk och serologisk analys. Förutom EIRA studien användes avidentifierade data från USA (delarbete 2), Holland (delarbete 2) och Danmark (delarbete 3) för att undersöka gen-gen interaktion i delarbete 2 samt alkohol konsumtion och risk att utveckla RA i delarbete 3.

- Rökning uppvisade ett dos-responssamband beträffande risken att utveckla RA med förekomst av antikroppar mot citrullinerade peptider. Rökning och HLA-DRB1 SE alleler var vidare associerade med interaktionseffekter for risken att drabbas av ACPA positiv RA. Bland friska kontroller var proportionen citrullinerade lungceller högre bland rökare jämfört med icke rökare. Dessa resultat har resulterat i följande potentiella förklaringsmodell till att rökning och HLA-DRB1 SE alleler kan ge upphov till RA: rökningen orsakar citrullinering av peptider som i sin tur fäster lättare på MHC klass II molekyler som modifieras av HLA-DRB1 SE alleler vilket leder till aktivering av T-celler som initierar vidare immunologiska reaktioner.

- Två av de mest etablerade genetiska risk faktorerna (R620W PTPN22 och HLA-DRB1 SE alleler) för RA är förenade med interaktionseffekter beträffande risken att drabbas av ACPA positiv RA. Additativ interaktion mellan HLA-DRB1 SE och PTPN22 observerades i tre olika populationer (en amerikansk, en holländsk och i den svenska EIRA studien). En signifikant multiplikativ interaktion observerades då alla tre studierna analyserades tillsammans. Mellan rökning och R620W observerades ingen interaktion beträffande risk för ACPA positiv RA.

- I studie III observerades att alkohol konsumtion är associerad med minskade risker för att drabbas av RA på ett dosberoende sätt. Dessutom observerades interaktionseffekter mellan rökning, HLA-DRB1 SE alleler och alkoholkonsumtion för risken att insjukna i ACPA positiv RA. Dessa resultat bygger på analyser i två skandinaviska studier (en dansk (CACORA) och en svensk studie (EIRA)).

- Rökning och HLA-DRB1 SE alleler är associerade med starka interaktionseffekter beträffande risk att drabbas av ACPA+ RA. Dessutom är rökning med hänsyn tagen till HLA-DRB1 SE alleler kopplade till dosberoende riskökningar att drabbas av ACPA+ RA. Rökning beräknades vara orsaken till 35 procent av alla ACPA positiva RA fall och 20 procent av alla RA fall.

8 APPENDIX

Smoking questions:

1. Do you smoke? □ No □ Yes, regularly □ Yes, occasionally

2. If you don't smoke
   Have you ever smoked? □ No □ Yes, regularly □ Yes, occasionally

3. If you have smoked,
   What year did you stop smoking…… Write year when you stopped

4. If you smoke or have smoked,
   When did you start smoking regularly? Write year when you stopped smoking……

5. How much do you smoke,
   Or used to smoke before you stopped? 1. Cigarettes/day
   ……st
   2. Cigarillos/day
   ……st
   3. Cigars/day
   ……st
   4. Pipetobacco/day
   ……gram

Alcohol consumption questions:

1. Have during the last 12 months
   Drinked at least one glas of beer, strong liquor,
   Wine or strong wine? □ No □Yes
   (if No continue to section N)

2. Approximately how much alcoholic
   beverages did you drink during last week?
   1. Strong beer …… cl
   2. Wine ……..cl
   3. Strong wine ……cl
   4. Strong liquor ……cl

3. Have your alcohol consumption
   during last week been:
   □ Much less than usual consumption
   □ less than usual consumption
   □ slight less than usual consumption
   □ as usual consumption
   □ slight more than usual consumption
   □ More than usual consumption
   □ Much more than usual consumption

9 ACKNOWLEDGEMENTS

Lars Alfredsson
Lars Klareskog
Leonid Padyukov
Camilla Bengtsson (Silva)
Anita Berglund
Marie-Louise Serra
Lena Nise
Bo Ding
Patrik Stolt
Eva Jemseby
Emilie Lundström
Marcus Rönninger
Peter Gregersen
Robert Plenge
Tom Huizinga
Anette van der Helm-van Mill,
Berit Sverdrup, Johan Rönnelid, Elisabeth W. Karlson, Katherine Liao, Ingvar Lundberg, Monica Mortimer,
Alan J Silman,
All the EIRA study group members and the patients and controls participating in the study.
Eva Skillgate, NPH, All students that I have meet, Konung Gustav den V:s jubileums fond, The National Institute of Health of the USA.
Anette Wigertz, Mats Halldin, Anna Bennet, Bengt Sjögren, Maria-Pia Hergens, Jenny Selander, Helen Rosenlund, Lena Holm, Ann-Marie Wesley, Tony Bohman
Slobodan Zdravkovic, Anders Ahlbom, Stefan och Maria Lönn, Ann Dsilna, Anna Persson, All BJJ practitioners and trainers (lollo) at Bålsta JJK and Marcel Yogui, Sara Wedren, Yvonne Ehnman, Maria Feychting, Sören Jacobsen, Morten Frisch, Peter Garred, Sten Torhold, Karin Lundberg, Birgitta Nordmark, Marie Gunnarsson, Bruna Gigante, Ulf de Fair, Hiba Mahdi, Maria Seddighzadeh, Max Wikström, Camilla Etzell, Sofie Jacobsen, Camilla Gard, Birgitta Ohlander, Eleonor Fransson, Sofia Carlsson, Karin Leander, Karin Lycke, Magdalena Rosell, Magnus Hellgren, Niklas Noren, Fredrik Andersson, Familjen Leone.
And a bunch of other people (ingen nämnd ingen glömd!)

Johan Bunner med familj, Ray Eriksson med familj, Jan och Lena Bunner

Mamma (Margareta Källberg), Pappa (Johnny Källberg), Syster (Christina Nordlander), Anna, Amanda och Christian, Mormor (Viola Nilsson), Farbror (Kjell Källberg), Birgitta Berg,

Mina älsklingar: Charlotta Bunner, Astor Källberg och Sverre Källberg.

Support RDF

10 REFERENCES

1. Garrod AB. Treatise on nature and treatment of gout and rheumatic gout. London: Walton and Maberly, 1859.

2. Lagier R. Nosology versus pathology, two approaches to rheumatic diseases illustrated by Alfred Baring Garrod and Jean-Marie Charcot. Rheumatology (Oxford) 2001;40;467-71

3. Rotschild BM. Rheumatoid arthritis at a time of passage. J Rheumatol 2001;28:245-50

4. Michaud K, Wolfe F. Comorbidities in rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2007 Oct;21(5):885-906.

5. Rantapää Dahlqvist S, Jacobsson L. Reumatoid artrit/ledgångsreumatism. In: Klareskog L, Saxne T, Enman Y eds. Reumatologi. 1st ed. Lund, Sweden. Studentlitteratur, Lund; 2008;49-74.

6. Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest. 1998 Jan 1;101(1):273-81.

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

8. Sauer, U. et al. (17 April 2007). "Getting Closer to the Whole Picture". Science 316: 550.

9. Moreland LW, Curtis JR.. Systemic Nonarticular Manifestations of Rheumatoid Arthritis: Focus on Inflammatory Mechanisms. Semin Arthritis Rheum. 2008 Nov 18. [Epub ahead of print]

10. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31:315-24.

11. Aho K, Palosuo T, Heliovaara M, Knekt P, Alha P, von Essen R. Antifilaggrin antibodies within "normal" range predict rheumatoid arthritis in a linear fashion. J Rheumatol 2000;27:2743-6.

12. Nielen MM, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MH, Habibuw MR, Vandenbroucke JP, Dijkmans BA. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum. 2004;50:380-6.

13. Rönnelid J, Wick MC, Lampa J, Lindblad S, Nordmark B, Klareskog L, van Vollenhoven RF. (2005) Longitudinal analysis of anti-citrullinated antibodies (anti-CP) during 5 years follow-up in early rheumatoid arthritis: Anti-CP status is a stable phenotype that predicts worse disease activity and greater radiological progression. Annals Rheum Dis. online May 5, [Epub ahead of print]

14. Hellgren L. The Prevalence of Rheumatoid Arthritis in Occupational Groups. Acta Rheumatol Scand 1970; 16: 106 – 13

15. Lundberg I, Alfredsson L, Plato N, Sverdrup B, Klareskog L, Kleinau S. Occupation, Occupational exposure to chemicals and Rheumatological disease. Scand J Rheumatol 1994; 23; 305 – 10.

16. Reckner Å, Skogh T, Wingren G. Occupational determinants for rheumatoid arthritis. Scand J Work Environ Health 2000; 26(3):243 – 49.

17. Olsson AR, Skogh T, Axelson O, Wingren G. Occupations and exposures in the work environment as determinants for rheumatoid arthritis. Occup Environ Med. 2004 Mar;61(3):233-8

18. Vessey MP, Villard-Mackintosh L, Yeates D. Oral contraceptives, cigarette smoking and other factors in relation to arthritis. Contraception. 1987 May;35(5):457-64.

19. Silman AJ, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. Arthritis Rheum. 1996 May;39(5):732-5.

20. Jorgensen C, Picot MC, Bologna C, Sany J. Oral contraception, parity, breast feeding, and severity of rheumatoid arthritis. Ann Rheum Dis. 1996 Feb;55(2):94-8

21. Sverdrup B, Kallberg H, Bengtsson C, Lundberg I, Padyukov L, Alfredsson L, Klareskog L; EIRA Study Group. Association between occupational exposure to mineral oil and rheumatoid arthritis: results from the Swedish EIRA case-control study.Arthritis Res Ther. 2005;7(6):R1296-303.

22. Stolt P, Kallberg H, Lundberg I, Sjogren B, Klareskog L, Alfredsson L. Silica exposure is associated with increased risk of developing rheumatoid arthritis: results from the Swedish EIRA-study. Ann Rheum Dis. 2004 Aug 19 [Epub ahead of print]

23. Oliver JE, Silman AJ. Risk factors for the development of rheumatoid arthritis. Scand J Rheumatol. 2006 May-Jun;35(3):169-74.

24. Voigt LF, Koepsell TD, Nelson JL, Dugowson CE, Daling JR. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. Epidemiology. 1994 Sep;5 (5):525-32.

25. 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 Rheum 1999;42(5):910-7.

26. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, Alfredsson L. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Annals of the Rheumatic Diseases 2003; 62:835 – 841.

27. Reckner Olsson A, Skogh T, Wingren G. Comorbidity and lifestyle, reproductive factors, and environmental exposures associated with rheumatoid arthritis. Ann Rheum Dis. 2001 Oct;60(10):934-9.

28. Klareskog L, Stolt P, Lundberg K, Källberg H, Bengtsson C, Grunewald J, Rönnelid J, Harris HE, Ulfgren AK, Rantapaa-Dahlqvist S, Eklund A, Padyukov L, Alfredsson L.. A new model for an etiology of RA: Smoking may trigger HLA-DR (SE)- restricted immune reactions to autoantigens modified by citrullination. 2006 Arthritis Rheum. 54(1): 38-46

29. Bhatia SS, Majka DS, Kittelson JM, Parrish LA, Ferucci ED, Deane KD, Arend WP, Rewers M, Michael Holers V, Norris JM. Rheumatoid factor seropositivity is inversely associated with oral contraceptive use in women without rheumatoid arthritis. Ann Rheum Dis. 2007 Feb;66(2):267-9. Epub 2006 Jul 25.

30. Costenbader K H, Feskanich D, Mandl L A, Karlsson E W. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. (2006) Am. J. Med. 119; 503-511.

31. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. Arthritis Res Ther 2006;8 (4):R133.

32. Linn-Rasker SP, van der Helm-van Mil AH, Van Gaalen FA, Kloppenburg M, de Vries R, le Cessie S, Breedveld FC, Toes RE, Huizinga TW. (2005) Smoking is a risk factor for anti-CCP antibodies only in RA patients that carry HLA-DRB1 Shared Epitope alleles. Ann Rheum Dis. Jul 13, [Epub ahead of print]

33. Michou L, Teixeira VH, Pierlot C, Lasbleiz S, Bardin T, Dieudé P, Prum B, Cornélis F, Petit-Teixeira E. Associations between genetic factors, tobacco smoking and autoantibodies in familial and sporadic rheumatoid arthritis. Ann Rheum Dis. 2008 Apr;67(4):466-70. Epub 2007 Jul 27.

34. Stolt P. Cigarette smoking and silica exposure as determinants for the development of rheumatoid arthritis (2004). Doctoral thesis; Karolinska Institutet: ISBN- 91-7140-082-6.

35. Hill, A.B. (1965). "The environment and disease: association or causation?". Proceedings of the Royal Society of Medicine 58: 295–300.

36. Waldschmidt TJ, Cook RT, Kovacs EJ. Alcohol and inflammation and immune responses: Summary of the 2005 Alcohol and immunology research interest group (AIRIG) meeting. Alcohol 38(2006):121-125.

37. Mandrekar P, Catalano D, White B, Szabo G. (2006) Moderate alcohol intake in humans attenuates monocyte inflammatory response: inhibition of nuclear regulatory factor kappa B and induction of interleukin 10. Alcohlism: Clinical and experimental research. Vol 30; 135-139.

38. Cerhan JR, Saag K, Criswell LA, Merlino LA, Mikuls TR. Blood Transfusion, Alcohol Use, and Anthropometric Risk Factors for Rheumatoid Arthritis in Older Women. The Journal of Rheumatology 2002;29:246-254

39. Heliövaara M, Aho K, Knekt P, Impivara O, Reunanen A, Aromaa A. Coffee consumption, rheumatoid factor, and the risk of rheumatoid arthritis. Ann Rheum Dis 2000; 59: 631-635

40. Hazes JM, Dijkmans BA, Vandenbroucke JP, de Vries RR, Cats A Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. Ann Rheum Dis. 1990 Dec;49(12):980-2.

41. Jonsson IM, Verdrengh M, Brisslert M, Lindblad S, Bokarewa M, Islander U, Carlsten H, Ohlsson C, Nandakumar KS, Holmdahl R, Tarkowski A.(2007) Ethanol prevents development of destructive arthritis. Proc Natl Acad Sci U S A. 104:258-63. Epub 2006 Dec 21.

42. MaCgregor A, Silman A, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis and rheumatism. vol. 43; January 2000. 30-37.

43. Struan F.A.G, Steinsson K, The inheritance of rheumatoid arthrirtis in Iceland. Arthritis and rheumatism. vol. 44; October 2001. 2247-2254.

44. Criswell L.A, Gregersen P.K. Current understanding of the genetic aetiology of rheumatoid arthritis and likely future developments. Rheumatology. Vol 44(suppl 4); 2005: iv9-iv13

45. Oliver JE, Worthington J, Silman AJ. Genetic epidemiology of rheumatoid arthritis. Curr Opin Rheumatol. 2006 Mar;18(2):141-6.

46. Bowes J, Barton A. Recent advances in the genetics of RA susceptibility. Rheumatology (Oxford). 2008 Apr;47(4):399-402. Epub 2008 Feb 7

47. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, Kern M, Criswell LA, Albani S, Nelson JL, et al. (2001) A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. Am. J. Hum. Genet.68: 927 – 936

48. MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, Barrett J, Lee D, White S, John S, Brown MA, Bell J, Silman A, Ollier W, Wordsworth P, Worthington J. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. Arthritis Rheum. 2002 Mar;46(3):632-9.

49. Yamamoto K, Yamada R. Genome wide single nucleotide polymorphism analyses of rheumatoid arthritis. Journal of autoimmunity. Vol. 25; 2005. 12 – 15.

50. Jawaheer D, Gregersen P.K, et al. Dissecting the genetic complexity of the associations between leukocyte antigens and rheumatoid arthritis. Am. J. Hum. Genet. Vol 71; 2002: 585 – 594.

51. Gregersen P.K, Silver J, Winchester R.J. The Shared Epitope Hypothesis. Arthritis Rheum 1987.30(11):1205-1213.

52. MacGregor A, Ollier W, Thomson W, Jawaheer D, Silman A. HLA-DRB1*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. J Rheumatol 1995;22:1032-6.

53. Meyer JM, Evans TI, Small RE, Redford TW, Han J, Singh R, et al. HLA-DRB1 genotype influences risk for and severity of rheumatoid arthritis. J Rheumatol. 1999;26:1024-34.

54. Begovich AB, Carlton VEH, Honingberg LA, et al. A missense single nucleotide polymorphism in a gene encoding a protein phosphatase (PTPN22) is associated with rheumatoid arthritis. Am. J. Hum. Genet. Vol 75; 2004: 330-337.

55. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, Wolfe F, Kastner DL, Alfredsson L, Altshuler D, Gregersen PK, Klareskog L, Rioux JD. (2005) 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. Am J Hum Genet. 77(6):1044-60

56. Wellcome Trust Case Control Consortium, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. Nature. 2007 Jun;447(7):661-678.

57. T.W.J. Huizinga, C.I. Amos, A.H.M. van der Helm-van Mil, W. Chen, F.A van Gaalen, D Jawaheer, et al. Refining the Complex Rheumatoid Arthritis Phenotype Based on Specificity of the HLA-DRB1 Shared Epitope for Antibodies to Citrullinated Proteins. Arthritis Rheum.2005 (52): S 3433-3438

58. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH, Toes RE, Huizinga TW, Klareskog L, Alfredsson L; Epidemiological Investigation of Rheumatoid Arthritis study group. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet. 2007 May;80(5):867-75.

59. 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 Rheum. 2004 Oct;50(10):3085-92.

60. Pedersen M., Jacobsen S., Garred P., Madsen H.O., Klarlund M., et al. Strong Combined Gene-Environment Effects in Anti-cyclic citrullinated peptide-positive rheumatoid Arthritis: a nationwide Case-Control study in Denmark. Arthritis Rheum. 2007 May;56(5):1446-53.

61. 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 Rheum. 2007 Feb;56(2):425-32.

62. Criswell LA, Saag KG, Mikuls TR, Cerhan JR, Merlino LA, Lum RF, Pfeiffer KA, Woehl B, Seldin MF. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. Ann Rheum Dis. 2006 Sep;65(9):1163-7.

63. Lee HS, Irigoyen P, Kern M, Lee A, Batliwalla F, Khalili H, Wolfe F, Lum RF, Massarotti E, Weisman M, Bombardier C, Karlson EW, Criswell LA, Vlietinck R, Gregersen PK. Interaction between smoking, the shared epitope, and anti-cyclic citrullinated peptide: a mixed picture in three large North American rheumatoid arthritis cohorts. Arthritis Rheum. 2007 Jun;56(6):1745-53.

64. Bengtsson C. The risk of developing rheumatoid arthritis: epidemiological studies on associations with socioeconomic status, psychosocial work stress and smoking (2008). Doctoral thesis; Karolinska Institutet: ISBN 978-91-7409-038-3.

65. Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. (2004) Clustering of disease features within 512 multicase rheumatoid arthritis families. Arthritis Rheum. 50(3):736-41

66. Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D; New York Cancer Project. The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health. 2004 Jun;81(2):301-10.

67. Van Aken J, van Bilsen JH, Allaart CF, Huizinga TW, Breedveld FC. (2003) The Leiden Early Arthritis Clinic. Clin Exp Rheumatol;21(5 Suppl 31):100-105

68. van der Meer FJ, Koster T, Vandenbroucke JP, Briet E. (1997) The Leiden Thrombophilia Study (LETS). Thromb Haemost. Jul;78(1):631-5

69. [Unpublished]. Merete Pedersen PhD thesis 2006, University of Copenhagen.

70. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue Antigens 1992;39:225-235.

71. Verduyn W, Doxiadis II, Anholts J, Drabbels JJ, Naipal A, D'Amaro J, Persijn GG, Giphart MJ, Schreuder GM. Biotinylated DRB sequence-specific oligonucleotides. Comparison to serologic HLA-DR typing of organ donors in eurotransplant. Hum Immunol. 1993 May;37(1):59-67

72. Kimura A, Sasasuki T. Eleventh International Histocompatobility Workshop Reference protocol for the HLA DNA typing technique. In: Tsuij K, Aizawa M, Sasazuki T, editors HLA 1991, vol 1. Oxford: Oxford University Press; 1992.

73. Germer S, Holland MJ, Higuchi R (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. Genome Res 10:258–266.

74. Rothman K J. Epidemiology an introduction. Oxford University press 2002.

75. Armitage P, Berry G, Matthews J.N.S. Statistical methods in medical research. 4th ed. Massachusetts. Blackwell publishing 2002.

76. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol 1980; 112: 467-470.

77. Hosmer DW, Lemeshow S. (1992) Confidence interval estimation of interaction. Epidemiology. 3: 452-6.

78. Lundberg M, Fredlund P, Hallqvist J, Diderichsen F. A SAS program calculating three measures of interaction with confidence intervals. Epidemiology. 1996 Nov;7(6):655-6.

79. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol. 2005;20(7):575-9.

80. Källberg H, Ahlbom A, Alfredsson L. Calculating measures of biological interaction using R. Eur J Epidemiol. 2006;21 (8):571-3.

81. Zhao J, Jin L, Xiong M. Test for Interaction between Two Unlinked Loci . Am. J. Hum. Genet.2006;79: 831-845.

82. Coughlin SS, Benichou J, Weed DL. (1994) Attributable risk estimation in case-control studies. Epidemiol Rev.16(1):51-64.

83. Farragher TM, Goodson NJ, Naseem H, Silman AJ, Thomson W, Symmons D, Barton A. Association of the HLA-DRB1 gene with premature death, particularly from cardiovascular disease, in patients with rheumatoid arthritis and inflammatory polyarthritis. Arthritis Rheum. 2008 Feb;58(2):359-69.

84. Deborah J. Smyth, Jason D. Cooper, Joanna M. M. Howson, Neil M. Walker, Vincent Plagnol, Helen Stevens, David Clayton, and John A. Todd. *PTPN22* Trp^{620} explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. Diabetes Publish Ahead of Print published online ahead of print February 29, 2008 DOI: 10.2337/db07-1131.

85. Kim SK, Park SH, Shin IH, Choe JY. Anti-cyclic citrullinated peptide antibody, smoking, alcohol consumption, and disease duration as risk factors for extraarticular manifestations in Korean patients with rheumatoid arthritis. J Rheumatol. 2008 Jun;35(6):995-1001. Epub 2008 May 1.

86. Lander, E. S. and Schork, N. J. (1994). Genetic dissection of complex traits, Science 265(5181): 2037–2048.

87. Devlin, B. and Roeder, K. (1999). Genomic control for association studies, Biometrics 55(4): 997–1004.

88. Persson PG, Norell SE. Retrospective versus original information on cigarette smoking. Implications for epidemiologic studies. Am J Epidemiol 1989;130:705-712.

89. Hammar N, Norell S. Retrospective versus original information on diet among case of colorectal cancer and controls. Int J Epidemiol 1991;20: 621-7.

90. Bengtsson C, Nordmark B, Klareskog L, Lundberg I, Alfredsson L; EIRA Study Group. Socioeconomic status and the risk of developing rheumatoid arthritis: results from the Swedish EIRA study. Ann Rheum Dis. 2005 Nov;64(11):1588-94. Epub 2005 Apr 20.

91. Hansson L, Khamis HJ. Matched samples logistic regression in case-control studies with missing values: when to break the matches. Stat Methods Med Res. 2008 Mar 28. [Epub ahead of print]

92. Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond. Am J Epidemiol. 2008 Jul 15;168(2):212-24. Epub 2008 May 28.

93. Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. Epidemiology. 1996 May;7(3):286-90.

94. van Venrooij W J, Pruijn G J M. Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. Arthritis Res. 2000; 2(4): 249–251.

95. Makrygiannakis D, Hermansson M, Ulfgren AK, Nicholas AP, Zendman AJ, Eklund A, Grunewald J, Skold CM, Klareskog L, Catrina AI. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. Ann Rheum Dis. 2008 Oct;67(10):1488-92.

96. Snir O, Widhe M, von Spee C, Lindberg J, Padyukov L, Lundberg K, Engström A, Venables PJ, Lundeberg J, Holmdahl R, Klareskog L, Malmström V. Multiple antibody reactivities to citrullinated antigens in sera from rheumatoid arthritis patients - association with HLA-DRB1 alleles. Ann Rheum Dis. 2008 Jul 17. [Epub ahead of print]

97. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klareskog L, Gregersen PK. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. N Engl J Med. 2007 Sep 20;357(12):1199-209. Epub 2007 Sep 5.

98. Chang M, Rowland CM, Garcia VE, Schrodi SJ, Catanese JJ, van der Helm-van Mil AH, Ardlie KG, Amos CI, Criswell LA, Kastner DL, Gregersen PK, Kurreeman FA, Toes RE, Huizinga TW, Seldin MF, Begovich AB. A large-scale rheumatoid arthritis genetic study identifies association at chromosome 9q33.2. PLoS Genet. 2008 Jun 27;4(6):e1000107.

99. Julià A, Ballina J, Cañete JD, Balsa A, Tornero-Molina J, Naranjo A, Alperi-López M, Erra A, Pascual-Salcedo D, Barceló P, Camps J, Marsal S. Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility. Arthritis Rheum. 2008 Aug;58(8):2275-86

100. Hughes LB, Morrison D, Kelley JM, Padilla MA, Vaughan LK, Westfall AO, Dwivedi H, Mikuls TR, Holers VM, Parrish LA, Alarcón GS, Conn DL, Jonas BL, Callahan LF, Smith EA, Gilkeson GS, Howard G, Moreland LW, Patterson N, Reich D, Bridges SL Jr. The HLA-DRB1 shared epitope is associated with susceptibility to rheumatoid arthritis in African Americans through European genetic admixture. Arthritis Rheum. 2008 Feb;58(2):349-58.

101. E. Lundström, H. Källberg, L. Alfredsson, L. Klareskog, L. Padyukov. Gene-environment interaction involving HLA-DRB1 shared epitope and smoking in ACPA-positive Rheumatoid Arthritis: all alleles are important. Ann Rheum Dis 2008;67(Suppl II):440

102. Hill JA, Southwood S, Sette A, Jevnikar AM, Bell DA, Cairns E. Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. J Immunol 2003; 171: 538-541.

103. Hill JA, Bell DA, Brintnell W, Yue D, Wehrli B, Jevnikar AM, Lee DM, Hueber W, Robinson WH, Cairns E. Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in DR4-IE transgenic mice. J Exp Med. 2008 Apr 14;205(4):967-79.

104. Costenbader KH, Chang SC, De Vivo I, Plenge R, Karlson EW. Genetic polymorphisms in PTPN22, PADI-4, and CTLA-4 and risk for rheumatoid arthritis in two longitudinal cohort studies: evidence of gene-environment interactions with heavy cigarette smoking. Arthritis Res Ther. 2008;10(3):R52. Epub 2008 May 7.

105. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357(10):977-86.

106. Bradley JR, Pober JS. Tumor necrosis factor receptor-associated factors (TRAFs). Oncogene. 2001 Oct 1;20(44):6482-91.

107. Bergstrom U, Jacobsson L, Nilsson J A, Berglund G, Turesson C. Smoking, Low Level of Formal Education and Infrequent Alcohol Consumption Are Independent Predictors of Rheumatoid Arthritis. Presented at ACR 2007 (abstract No 374). Arthritis Rheum. 2007;56(suppl): S192

108. Rinaldi S, Peeters PH, Bezemer ID, Dossus L, Biessy C, Sacerdote C, Berrino F, Panico S, Palli D, Tumino R, et al. (2006) Relationship of alcohol intake and sex steroid concentrations in blood in pre- and post-menopausal women: the European Prospective Investigation into Cancer and Nutrition. Cancer Causes Control. vol 17;1033-43

109. Sarkola T, Fukunaga T, Mäkisalo H, Eriksson P. (2000) Acute effect of alcohol on androgens in premenopausal women. Alcohol & alcoholism. Vol. 35; 84-90

110. Jones SA (2005). "Directing transition from innate to acquired immunity: defining a role for IL-6". J. Immunol. 175 (6): 3463–8

111. Febbraio MA, Pedersen BK (2005). Contraction-induced myokine production and release: is skeletal muscle an endocrine organ?. Exerc Sport Sci Rev 33 (3): 114–9.

112. H.J. Källberg, M. Mortimer, I. Lundberg, L. Klareskog, L. Alfredsson. Physical workload and the risk of developing Rheumatoid Arthritis. Results from the Swedish EIRA study. Ann Rheum Dis 2006;65(Suppl II):132.

113. Rönninger M, Källberg H, Lundström E, Lindahl A, Klareskog L, Alfredsson L, Padyukov L. Complexity of a complex disease; understanding genes, environment and immunity in rheumatoid arthritis. Future Rheumatol. 2007;2(5): 485-492.

114. H. Källberg, L. Padyukov, C. Bengtsson, L. Klareskog, L. Alfredsson. Smoking and R620W PTPN22 are independent risk factors associated with anti-CCP positive RA. Results from the Swedish EIRA study. Ann Rheum Dis 2008;67(Suppl II):303

115. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE. Variations in DNA elucidate molecular networks that cause disease. Nature. 2008 Mar 27;452(7186):429-35. Epub 2008 Mar 16.

116. National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/genbank/

117. Kyoto Encyclopedia of Genes and Genomes: http://www.genome.jp/kegg/