From THE DEPARTMENT OF MEDICINE
Karolinska Institutet, Stockholm, Sweden

GENETIC STUDIES OF NON-HLA LOCI IN RHEUMATOID ARTHRITIS: EXPRESSION AND INTERACTION OF CANDIDATE GENES

Marcus Ronninger

Karolinska Institutet

Stockholm 2012

Published by Karolinska Institutet. Printed by Larserics Digital Print AB, Sweden

ISBN 978-91-7457-745-7

To my family

SUMMARY

Genetic predisposition is an important contributor in development of human complex diseases, like rheumatoid arthritis (RA). In this thesis work, we present advances for involvement of non-HLA genetic risk factors for RA. In the same study, by using methods of genetic epidemiology and molecular genetics, we demonstrate how even moderate contribution from candidate genes could be found, interpreted and how this may affect important biological functions.

The majority of the study has been performed in a large population based cohort of Swedish RA patients with matched controls and with additional cohorts from Norway, UK and the Netherlands. Data has been generated with both TaqMan allelic discrimination and DNA array-based genotyping. A subset of the cohort has been used for studying mRNA expression with quantitative PCR. Three risk loci have been investigated in this thesis: the MHC class II Transactivator (CIITA), the Dendritic Cell Immunoreceptor (DCIR) and Protein Tyrosine Phosphatase Non receptor 22 (PTPN22). For CIITA we aimed to produce further evidence for association with disease by replication and fine mapping of the locus. For DCIR and PTPN22 our aim was to examine the gene expression for finding potential regulatory differences.

We present data that CIITA is a valid risk factor for RA and that this risk seems to be population specific. The risk for disease was higher in the subgroup defined by shared epitope (SE) positivity. We extensively analyzed a possible interaction effect for the risk of developing disease in four independent populations. However, no significant interaction between the CIITA and the HLA-DRB1 locus was found. When measuring expression of promoter isoforms of CIITA in cells from peripheral blood, we found that both CIITA_pIII and CIITA_pIV expression are associated with genetic variation in the locus.

For DCIR we could establish that five splice forms were present in blood mononuclear cells, including a novel variant, which were down regulated upon immunostimulation. Transcript DCIR_v4 was associated with genetic variation in the locus. This correlation was similar for both RA patients and controls.

Finally, we present a novel finding that the expression of PTPN22 splice forms is different for RA patients and healthy controls with more of the full-length, putatively more active, splice forms for patients and less of the alternative variant. This mixed effect was replicated in three independent cohorts.

In conclusion, we present a framework for delineating genetic risk association signals by fine-mapping loci and combining with expression analysis of existing splice forms. More specifically, we give further insights for three genetic risk factors for RA that may lead to less expression of HLA class II (CIITA) and stronger inhibition of immune cell signaling (DCIR and PTPN22).

A combined orchestrated effect of all this risk variants together with other risk factors known for RA may be what predisposes certain individuals for rheumatoid arthritis.

LIST OF PUBLICATIONS

I. ***CIITA* gene variants are associated with rheumatoid arthritis in Scandinavian populations**
Morten C. Eike*, Beate Skinningsrud*, Marcus Ronninger*, Alice Stormyr, Tore K. Kvien, Geir Joner, Pål R. Njølstad, Øystein Førre, Berit Flatø, Lars Alfredsson, Leonid Padyukov, Dag E. Undlien and Benedicte A. Lie.
Genes Immun, 2012, Advance online publication, Apr 19; doi:10.1038/gene.2012.11

II. **Interaction analysis between *HLA-DRB1* shared epitope alleles and MHC class II transactivator *CIITA* gene with regard to risk of rheumatoid arthritis**
Marcus Ronninger, Maria Seddighzadeh, Morten Christoph Eike, Darren Plant, Nina A. Daha, Beate Skinningsrud, Jane Worthington, Tore K. Kvien, Rene E. M. Toes, Benedicte A. Lie, Lars Alfredsson, Leonid Padyukov
PLoS One, 2012;7(3):e32861

III. **Genetic control of isoform expression of human MHC class II transactivator**
Marcus Ronninger, Maria Seddighzadeh, Leonid Padyukov
Manuscript

IV. **Differential expression of transcripts for the autoimmunity-related human dendritic cell immunoreceptor**
Marcus Ronninger, Carina Eklöw, Johnny C Lorentzen, Lars Klareskog and Leonid Padyukov
Genes Immun, 2008, Jul;9(5): 412-418

V. **The balance of expression of *PTPN22* splice forms is significantly different in rheumatoid arthritis patients compared with controls**
Marcus Ronninger*, Yongjing Guo*, Klementy Shchetynsky, Andrew Hill, Mohsen Khademi, Tomas Olsson, Padmalatha S Reddy, Maria Seddighzadeh, James D Clark, Lih-Ling Lin, Margot O'Toole and Leonid Padyukov.
Genome Med, 2012, Jan 20;4(1):2

Additional publications

Complexity of a complex disease; understanding genes, environment and immunity in rheumatoid arthritis development
Marcus Ronninger*, Henrik Källberg*, Emeli Lundström*, Annmarie Lindahl*, Lars Klareskog, Lars Alfredsson, Leonid Padyukov
Future Rheumatol, 2007, Oct; 2(5): 485-492

*These authors contributed equally

CONTENTS

LIST OF ABBREVIATIONS

ACPA	Anti citrullinated protein/peptide antibodies
ACR	American college of rheumatology
Anti-ccp	Anti cyclic citrullinated peptides
AP	Attributable proportion
bp	Base pairs, nucleotides
cDNA	Complimentary DNA
CI	Confidence interval
CIITA	MHC class II transactivator
CNV	Copy number variation
CSK	c-SRC tyrosine kinase
DC	Dendritic cell
DCIR	Dendritic cell immunoreceptor
DNA	Deoxyribonucleic acid
EIRA	Epidemiological investigation of rheumatoid arthritis
GWAS	Genome wide association study
HLA	Human leukocyte antigen
IFN	Interferon
kb	Kilo base pairs = 1,000 bp
LCK	Lymphocyte specific protein tyrosine kinase
LD	Linkage disequilibrium
MAF	Minor allele frequency
Mb	Mega base pairs = 1,000 Kb
MHC	Major histocompatibility complex
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis
nt	Nucleotides
OR	Odds ratio
PCR	Polymerase chain reaction
PTPN22	Protein tyrosine phosphatase, non-receptor type 22
qPCR	Quantitative polymerase chain reaction
R	Arginine
RA	Rheumatoid arthritis
RF	Rheumatoid factor
SE	Shared epitope
SNP	Single nucleotide polymorphism
T1D	Type 1 diabetes
TCR	T cell receptor
UTR	Un-translated region
W	Tryptophan

1 INTRODUCTION

This introduction aims at giving the reader a background for the included papers and also lift up the importance that research is carried out in this field. It is also aimed at giving the proper relevance for the aim of the thesis. It covers a brief overview of the disease at topic, rheumatoid arthritis (RA), risk factors for developing disease, the means of studying the disease with genetic tools and how this is done in patient cohorts.

1.1 ABOUT RHEUMATOID ARTHRITIS

Rheumatoid arthritis is a multifactorial disease with established genetic [1] and environmental components [2-5] that has a dramatic effect on life quality and confer great costs for the society. This chapter intends to give a broad introduction of what is known for the disease etiology and the associated risk factors.

1.1.1 Disease characteristics

On average one percent, depending on population, suffer from RA, which is typically a life long diagnosis [6]. The disease is more common for women than men with about a 3:1 ratio. The prevalence of RA is population dependent: it is more common for individuals with European ancestry than Asian [7, 8] with extraordinary high disease frequency for some native American populations (5% for Pima and 7% for Chippewa populations)[9-11]. RA is classified as an autoimmune disease, meaning that the disease, although it may be triggered by something else, is sustained by the individual's own immune system with a clear role of the adaptive immunity. About 3-5% from a general population suffer from an autoimmune disease and for most of these there is a heritable component [6, 12] Evidence for autoimmunity are antibodies directed at self-antigens, so-called autoantibodies that may be detected in early RA and sometimes even before disease onset [13]. It is also said to be a complex disease where possibly multiple genetic risk variants in the genome acts to increase the risk for disease together with environmental exposures. The study of these combinatorial effects of risks is even further complicated by the fact the disease is diagnosed by several criteria. These are described by the American College of Rheumatism 1987 revised criteria for the classification of RA, Table 1. [14]

Table 1. The 1987 revised criteria for the classification of rheumatoid arthritis (ACR-87).

Criterion		Definition
1.	Morning stiffness	Morning stiffness in and around the joints, lasting at least 1 hour before maximal improvement
2.	Arthritis of 3 or more joint areas	At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee ankle and MTP joints
3.	Arthritis of hand joints	At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint
4.	Symmetric arthritis	Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry)
5.	Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxtaarticular regions, observed by a physician
6.	Serum rheumatoid factor	Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive in <5% of normal control subjects
7.	Radiographic changes	Radiographic changes typical of rheumatoid arthritis on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do no qualify)

For classification purposes, a patient shall be said to have rheumatoid arthritis if he/she has satisfied at least 4 of these 7 criteria. Criteria 1 through 4 must have been present for at least 6 weeks. Patients with 2 clinical diagnoses are not excluded. Designation as classic, definite, or probable rheumatoid arthritis is not to be made. PIPs = proximal interphalangeal joints, MCPs = metacarpophalangeal joints, MTPs = metaarsophalangeal joints. Table modified from Arnett et al., 1988 [14].

It is important to point out that RA is a systemic disease, involving the immune system for mediating effects to sites of inflammation, which is evident from the usual symmetry of the affected joints. The cardiovascular and the lymphatic systems are the transport systems mediating the major immune responses by circulating cells and effector molecules.

Criteria number 6, presence of rheumatoid factor (RF), is of special interest since it clearly divides the patients in two distinct groups. RF describes antibodies against the Fc part of IgG giving rise to immune complexes and was first described by Erik Waaler in the year 1940[15]. This factor is not specific for RA and is found in several autoimmune diseases [16-18] and is also found in normal population. In a follow up of RF positive healthy individuals it was found that about a fifth had developed RA [19]. Recently, autoantibodies targeting citrullinated peptides has been found to be much more specific for rheumatoid arthritis than RF where many individuals are seropositive for both [20, 21]. These anti-citrullinated peptide antibodies (ACPA) are present in 50-70% of all RA patients and are found in less than 2% of the normal population [22]. ACPA status is usually determined by anti cyclic citrullinated peptide (anti-ccp)

ELISAs that use a mix of citrullinated peptides [22]. In 2010 a new classification criteria set for RA was presented that included the presence of ACPA as alternative to RF [23].

1.1.2 Autoantibodies, fine specificity

Citrullination, or deamination as it is also called, is a post-translational modification with a conversion of the amino acid arginine (R) to citrulline by peptidiylarginine deiminases (PADs). This modification alters the charge and possibly structure of the protein, giving rise to possible new epitopes foreign for the immune system. Citrullinated proteins have been shown to be abundant in the synovial fluid of arthritis patients where it is missing in osteoarthritis patients [24]. It seems, however, that not all citrullinated proteins trigger antibody production in RA patients. Specifically, four proteins have been described this far: fibrinogen, vimentin, collagen type II and alpha-enolase. Interestingly, the commonly used anti-cyclic citrullinated peptides (anti-ccp) assay does not completely overlap positive detection for antibodies towards these citrullinated proteins. For instance, about 4-7% of RA patients are positive for citrullinated alpha-enolase antibodies but negative for anti-ccp [25].

1.1.3 Cost for society

Due to that RA is a chronic disease that, if left untreated, leads to severe and irreversible damage to the cartilage of joints it leads to loss of physical functions and ultimately to handicap. This reduced quality of life is not only a major consequence for the affected individuals but also a dramatic cost for society. For instance, for the Swedish population it was estimated that the increased amount of sick leave due to RA amounts to a cost of 215 million euro during the year 2007 [26]. This is for a relatively small country and the estimate does not include the cost for care and medicines, which has increased with new therapies for autoimmune disease, hence, there is a lot to gain by learning more about the disease etiology that can lead to more efficient medication and more effective patient diagnosis.

1.1.4 Genetic component

The heritability of RA has been estimated in twin studies of Finnish and British population and is around 60% (Heritability 53-65% in UK resp Finnish population) and concordance rate for monozygotic twins is reported be 15-21% [1, 27, 28]. This establishes that there is a genetic component in the etiology of RA that is partly responsible for development of the disease. Before large genomic screens became available for RA, there were only a handful of genetic risk factors known with variations in the *HLA-DRB1* gene being the absolute strongest.

HLA-DRB1

HLA-DRB1 is an antigen presenting protein and the corresponding gene is encoded in the major histocompatibility complex (MHC) class II locus. This association to arthritis was discovered over 30 years ago [29] and has been replicated numerous times. A more detailed relationship was presented in the shared epitope hypothesis by Gregersen et al. (1987) that described a set of *HLA-DRB1* alleles with a common protein motif to be responsible for the increased risk. These alleles were collectively termed the shared epitope (SE) because of the shared aminoacid seuquence at position 67-74 which is part of the third hypervariable loop, exposed in the binding cleft of *HLA-DRB1* [30, 31]. In further studies it has become clear that the SE effect was only observed for sero-positive individuals, and more specifically presence of ACPA and that there is a dose-response relation [32, 33]. This gives further evidence that the presence of these antibodies distinctly divides patients in two subgroups that have, at least partly, different disease etiologies and where genetic factors predispose to one or the other. Other variants in the *HLA-DRB1* locus have been associated with protection or a milder form of disease [34-39].

PTPN22

Outside the HLA complex, variants in *PTPN22* are the most undisputable genetic risk factor for developing RA. The first report of association with disease came in 2004 for type 1 diabetes (T1D) from a candidate gene study [40]. This was quickly replicated in other cohorts and also for other diseases such as RA [41-44], SLE [42, 43, 45] and Graves' disease [46, 47] and in summary it has been associated with more than ten autoimmune diseases to date. It is also an incredible reproducible risk factor across populations and the associated risk variant exists mainly in individuals with European ancestry with the highest reported allele frequency in Finnish population (15.5% in controls)[48] and lowest for Italian (2%)[40]. The risk variant is almost absent in African individuals (MAF~0.005) and has a low frequency in Asian population (MAF 0.01-0.02)(www.hapmap.org).

The variant associated to disease, 620W (rs2046601), confers a substitution of arginine (R) to tryptophan (W) in the non-catalytic part of LYP, the protein of *PTPN22*, where a SRC homology 3 (SH3) domain is predicted to be affected.

It is predicted from mice studies that this domain has a high affinity with c-Src tyrosine kinase, CSK [49, 50]. This interaction may affect LYPs ability to dephosphorylate lymphocyte-specific protein tyrosine kinase, LCK, which is part of the signal pathway of the T cell receptor (TCR). A reduction of affinity in the SH3 domain by the 620W mutation may then lead to reduced formation of LYP/CSK complexes with a result in altered signaling through TCR [49].

However, studies in T cell lines and primary human cells have shown that the 620W variant is associated with an increase in dephoshporylation of downstream targets that would be a gain of phosphatase activity for LYP leading to an attenuation of TCR signaling [51], which is further supported by a study showing reduced T cell responsiveness to antigen stimulation for individuals homozygous for 620W [52].

Similar studies of B cells from individuals carrying the 620W allele are also pointing at an impaired response of the receptor and an overall reduction of phosphorylated key signaling proteins, which was reversed on inhibition of LYP [53]. How these findings are connected to a mutation not positioned in the active catalytic domain is not clear and does not follow the expected effect of altered affinity of the SH3 domain. In another experiment with Jurkat cells, co-transfecting 620W with CSK significantly increased phosphorylated ERK compared to 620R allele [54]. This would be interpreted as a loss of function with an increased TCR signaling as result.
Thus, it remains to find out what the role of *PTPN22*, and more specifically the 620W allele, has for autoimmune disease.
Besides these two well-established risk factors for RA, this thesis involves two other genes – the MHC class II transactivator (*CIITA*) and the dendritic cell immune receptor (*DCIR*).

CIITA

Initially *CIITA* was discovered in a locus responsible for expression differences of MHC class II molecules between rat strains in a nerve injury trauma model. In a human candidate gene approach, it was found that a variation, -A168G (rs3087456), was associated with RA, multiple sclerosis (MS) and myocardial infarction. It could also be shown that -168G was correlated with a lower expression of *CIITA* and HLA-DRA [55]. The association of *CIITA* with RA has been replicated with varying results [43, 56-59] and a meta-analysis in 2008 concluded that there was no overall association with RA [60]. The association with MS seems however to replicate in several studies [59, 61-63] and association with celiac disease [64], ulcerative colitis [65] and Addison's disease [66] have been reported.
The biological role of *CIITA* was discovered when a cell line derived from a patient with hereditary MHC class II deficiency (or bare lymphocyte syndrome) was complemented with a vector expressing *CIITA* that restored surface class II antigen expression [67]. It was subsequently found that expression of *CIITA* is required for both MHC class II expression and antigen presentation [68]. *CIITA* is part of the transcription complex used by the MHC class II promoters but does not directly bind to DNA itself, instead mediating its essential effect through the other transcription factors involved. Also, *CIITA* exhibits cell specific expression that parallels that of HLA class II [69, 70] and is regulated by multiple promoters leading to four different promoter isoforms, denoted *CIITA*_p1-p4 [71]. These isoforms are expressed in cell specific manner where *CIITA*_p1 is described as dendritic cell and macrophage specific [71, 72], *CIITA*_p2 has been detected in a melanoma cell line [73], *CIITA*_p3 in monocytes, dendritic cells and B cells [74] and *CIITA*_pIV is mainly inducible by IFN-gamma in many antigen presenting cells [71].
This complexity of several promoters of *CIITA* and the fact that it seems to tightly regulate expression of HLA class II suggest that the fine-tuning of antigen presenting molecules is of uttermost importance for the balance of immunity and self-tolerance

and makes it an attractive target to study given the clear involvement of HLA in autoimmune disease.

DCIR

Dendritic cell immunoreceptor (*DCIR*) was mapped in our lab in a quantitative trait locus in a congenic rat strain protected from oil-induced arthritis. This locus contained several antigen-presenting lectin-like receptor genes and five corresponding genes in human were genotyped. Variants in *DCIR* were found to be associated with sero-negative disease with an OR of 1.37 (95% CI 1.12-1.67) [75].

DCIR is also known as *CLEC4A* (official HGNC name), *LLIR, DDB27, CLECSF6* and *HDCGC13P*. *DCIR* is a member of the C-type lectin superfamily where the lectin domain is associated to functions such cell adhesion, cell signaling and other roles in immune response to pathogens and apoptosis [76, 77] with requirement of calcium for binding. *DCIR* also has an ITIM domain, which suggest an inhibitory function. With the assistance of Src-kinases, ITIM can interact with SHP-1 and SHP-2, which are phosphotyrosine phosphatases with potential to dephosphorylate molecules, which can result in a decrease of activating signals [78].

The number of genes (closest to associated genetic variants) associated to RA is now over 30 [79], which gives an insight of the complexity of the genetic background for the disease.

1.1.5 Environmental component

The idea the RA is triggered by an environmental exposure has been thoroughly studied. The resemblance of the characteristics of borreliosis (Lyme disease), an infectious disease caused by bacteria transferred by tics, with chronic rheumatoid arthritis gives the idea that bacterial or viral infections may trigger the disease [80]. There are reports of antibodies toward citrullinated viral peptides in RA patients that may play a role for the disease etiology [81, 82]. One of the hypotheses behind infections as a trigger for RA and loss of tolerance to self is molecular mimicry. An emerging candidate for this is the *Porphyromonas gingivalis* bacterium that causes periodontitis. It has been shown that periodontitis is more common in RA patients than normal populations [83-85] and together with the fact that *P. gingivalis* is the only bacteria known to express the PAD enzyme that converts arginine to citrulline, also shown to convert human proteins in vitro, it is an interesting etiology hypothesis [86].

The most well-known environmental risk factor that has been repeatedly connected to RA is smoking [2-5, 87-90] with a clear dose effect on the risk and the effect exists several years after cessation [2, 87, 88]. The exact mechanism behind this is still not clear but recent discoveries may hint at the process. Firstly, the risk of disease from smoking was greatly increased for individuals carrying the SE alleles in a dose dependent manner, i.e. homozygous individuals have the highest risk [33]. Then studies of smoking and ACPA revealed that smoking was exclusively a risk for the ACPA positive individuals and that this was linked to these specific autoantibodies rather than

RF [32]. This together with information about citrullinated proteins in lungs from smokers lead to a hypothetical disease etiology presented by Klareskog et al, 2006, suggesting that smoking may cause self-reactivity towards citrullinated proteins in a certain genetic background (SE) the eventually leads to RA (Figure 1) [32]. The question, however, how this targets the focus of the disease - joints, is still to be answered.

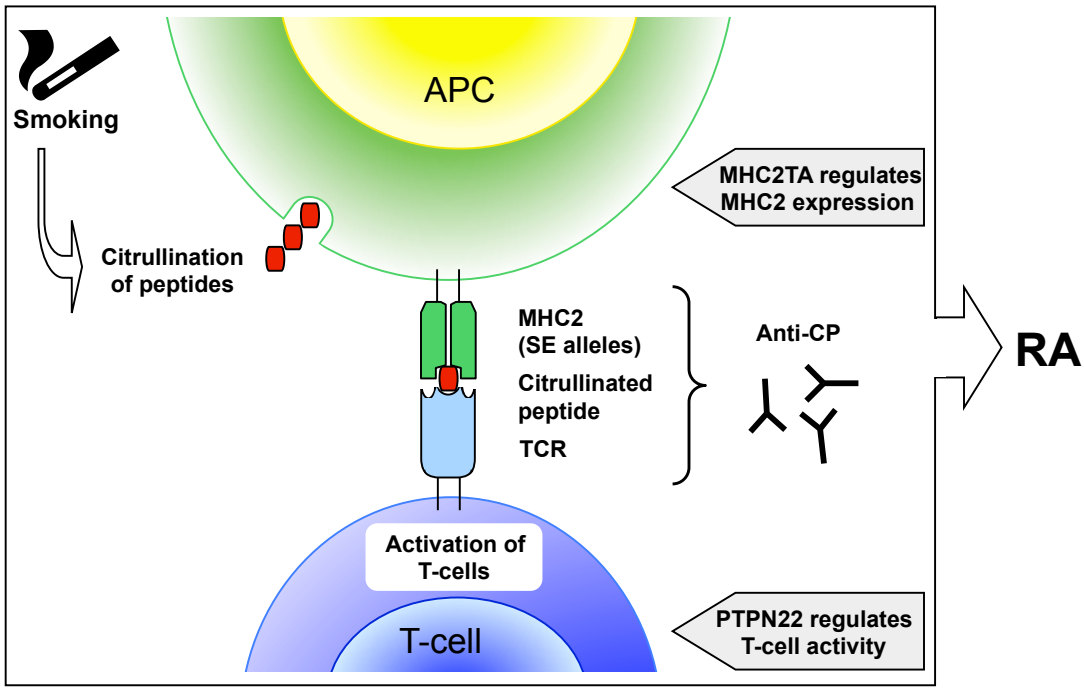

Figure 1. The figure illustrates the proposed etiology by Klareskog et al. by combing the different risk factors found for RA.

There is also support for other environmental factors that are associated with RA patients. Studies have shown that consumption of alcoholic beverages is less frequent for individuals developing RA [91, 92] implying a possible immunosuppressing function, which is supported by other observations [93]. Silica dust [94-97], mineral oil [98] and dietary effects of fruit, red meat and Mediterranean food [99-101] have also been associated with RA, but to a less extent.

1.2 GENETICS

The human genome is full of variations. These variations allow for adaption and survival under different environmental conditions and are key to the evolutionary process. Lately, however, they have become the focus of endemic autoimmune diseases, which may be a result of an effective immune system adapted and evolved to fight of infections.

The common variations in the genome consist of: Single nucleotide polymorphism (SNP), is a mutation of one base to another; Insertions and deletions (indels) which are inserted or lack of nucleotides in the sequence; Copy number variations (CNVs), are

stretches of copied sequences ranging from 1000 bases (1kb) up to several megabases (Mb)[102, 103]; Microsatellites or short tandem repeats are short repetitive elements of about 1-4 bases dispersed mainly in the non-coding part of the genome [104] with an exceptionally high mutation rate [105, 106]; other structural variations such as inversion, translocation and segmental uniparental disomy [107].
With the event of the complete sequence of the human genome we have gained further insight of the genetic diversity. The Human Genome Project (HGP) and Celera published the first draft of the human genome in 2001[108, 109]. The complete sequence (Build 35) was presented by the HGP in 2004 and consisted of 2.85 billion nucleotides (10^9 nt) with an estimate of 20000-25000 protein-coding genes, which occupies only about 1.5% of the genome [108, 110]. This landmark in biology has lead to many insights about the genome structure. It has showed that even though SNPs clearly are in majority of the variant events, other structural variations (CNVs) represent the largest genetic variance by involving 74% of all variant nucleotides. On average, this work estimates that humans are 99.5% similar in the genome where, roughly, SNPs are responsible for 0.1% difference [111]. The increasing number of sequenced genomes, however, results in new variants with every individual hinting that we still have not understood the full complexity of genetic variability. Also, the complexity that the diploid status of the genome should not be forgotten, which results in that a human can have either one or two copies of these affected alleles (heterozygous or homozygous), which may or may not be active.
These described differences, forms the key for genetic studies of susceptibility for disease. During the recent years, SNPs have become increasingly important for conducting large genetic studies due to their distribution and ease of detection. They may even be used for inferring or imputing other types of genetic variability. A deeper description of these variations is given in the next section.

1.2.1 Single nucleotide polymorphisms

The rapid discoveries of SNPs in the human genome, with the HapMap initiative as one of the driving forces [112], have facilitated the performance of large genetic studies. Without this foundation none of the GWASs studies would have been possible. Due to the fact that SNPs may have a small effect they can be found even in coding parts of the genome, though it is less frequent where amino acids are affected and tends to be higher at 5'-UTR regions [113]. This dispersion throughout the genome allows for a better resolution and coverage when comparing individuals by genotype, which is one of the major reasons why it is so commonly used.
Depending on the position of a SNP, either intergenic, 5' or 3' untranslated regions (UTRs), intronic or exonic, different properties are expected. Historically, mutations in the intergenic region of the genome (the so-called "gene desert") were considered to have very little effect on cell biology but discoveries of novel RNA-genes, epigenetic factors affecting histone composition etc., have changed this. This is very likely an

underestimated region where further research is warranted. SNPs in exonic regions can either be synonymous, meaning that the change of nucleotide does not alter the resulting amino acid, or they can be non-synonymous. The latter can further be described as missense, when an amino acid is changed or a nonsense mutation leading to a premature end of the translation. Non-synonymous mutations are often behind rare Mendelian traits due to the big impact they may infer on protein function. Synonymous mutations may still affect the protein by changing the translation and splicing efficiency. This has been studied on a genome wide basis where it was shown that a synonymous SNPs affecting translational efficiency were under negative selection for, amongst others, regulatory genes [114]. SNPs in the close vicinity of coding sequence, i.e. 5'-UTR, intronic or 3'-UTR, may affect very important regulatory features (described in a later section) such as transcriptional regulation by affecting transcription factor binding sites (TFBS) activity, changing of splicing pattern or efficiency by disrupting splice site, exonic splicing enhancers or silencers. The 3' UTR is also a region of gained interested with discoveries of regulation of protein translation by microRNA (miRNA) that often targets mRNA transcripts in this region [115]. An estimate of the amount of genes that may be targeted by miRNA varies (30-60%) but the mechanism seems to be of importance for a large set of genes [116, 117].

1.2.2 Genetic linkage

During the meiosis the chromosome pairs are crossed into new combinations, a process called homologous recombination. This procedure results in an increased diversity and is beneficial for adaption of organisms to environment. The recombination breaks the chromosomes at seemingly random positions and then joins the chromosome pair in cross over fashion. The amount of recombination, *i.e*. the recombination frequency, between two loci is a measure of genetic linkage and is a rough estimate of their distance. The recombination frequency of 1% is termed a centimorgan (cM) and is often used for describing distance between two loci. However, certain regions of the genome are less recombined then expected by a random procedure, which results in that the combination of alleles are inherited intact. This is called linkage disequilibrium (LD) and the measure of LD is called D and is calculated according to Figure 2.
The measure of D is in reality seldom used since it varies with allele frequencies. Instead it is often normalized with the theoretical maximum of observed allele frequencies [118, 119] or the square root of the product of allele frequencies to receive the correlation coefficient *r*.[120]

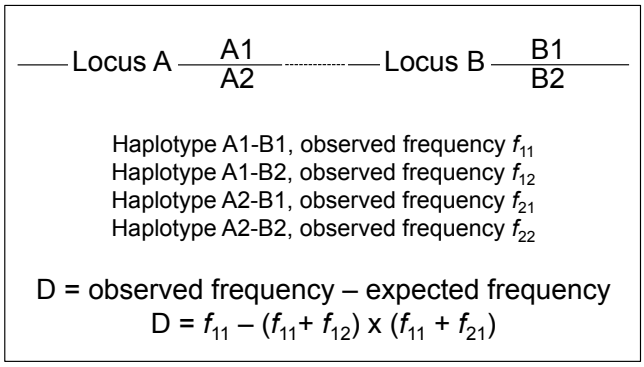

Figure 2. The box explains the LD measure between loci, A and B, with two alleles each. The measure D is calculated from the deviation of observed frequency of a haplotype from the expected in a population of individuals [121].

1.2.3 Transcription

The region directly upstream of the coding sequence is denoted the promoter and contains DNA motifs essential for the transcription initiation, see Figure 3 for schematic representation. The creation of mRNA from genes follows an orchestrated sequence of events the ultimately leads to the correct position of the RNA polymerase for transcription. The core promoter consists of necessary and well-conserved elements for RNA polymerase binding that stretches from about -40bp from the transcription initiation site to roughly +30 bp into the coding sequence. A typical promoter consists of three elements: the BRE motif that binds TFIIB; the TATA motif that binds TBP and the initiator (Inr), which may or may not be present. The most well described element of these is the TATA-motif, named after the nucleotides of the motif (TATAAA) positioned at ~-30 bp [122]. The TATA-motif is conserved throughout evolution and is present in about a third of human promoters [123]. The initiator motif is usually a cytosine at the −1 position and adenosine at the transcription start site and pyrimidines around these nucleotides [124]. These elements determine the direction and efficiency of the transcription of RNA polymerase II.

Other motifs further upstream of the initiation site (typically ~300 bp) assist in fine-tuning the expression. More distal elements may also affect the transcription with extremes up to 85 kb from coding region [125]. These upstream sequences may be either activating or repressing with various mechanisms of action. Repressing the transcription may be by interfering with activating motifs or modifying chromatin structures [126].

It is estimated that the human genome harbors more than 2600 proteins with DNA binding domains, assumed to be transcription factors [127]. This family of proteins constitutes of about 8% of all encoded human proteins, which highlights the importance of the genetic regulatory mechanisms.

The combination of all these factors is a powerful mechanism that allows for specific regulation of all genes during different environments for different cells. It is not hard to believe that even mutation far out from the coding regions may have great

consequences for the regulation of genes, which may not be detected without studying a specific cell under the right circumstances.

1.2.4 Splicing

Before the mRNA can be translated into amino acids the transcribed molecule needs to be modified to a mature mRNA. The perhaps most advanced step in this procedure is the removal of intervening sequences, so called introns. This phenomenon is called splicing and is carefully guided by the splice machinery and conserved patterns in the sequence. The organelle where splicing takes place is called the splicesome and contains the necessary proteins, with the small nuclear ribonucleoproteins (snRNPs) being the most crucial elements for splicing.
The patterns that determine where to splice are called splice sites, which usually consist of the nucleotides GU at 5' end of the intron and AG at 3' of the intron. About 18 to 40 nucleotides from the 3' end of the intron is a motif called the branch point [128]. Briefly, splicing starts with the snRNP U1 binding to the 5' end of intron. This complex is then looped to form a structure called lariat, by annealing of 5' end of the intron to the branch point. Additional snRNPs then exactly positions the 5' end of the intron to the 3' and the ends are ligated through transesterification and the lariat structure is released together with the snRNPs. Additionally, splicing may be aided by exonic splice enhancers (more common with long introns), which helps to recruit the splice molecules to the correct position, or it can be repressed by silencer elements [129]. Although the splicing of mRNA is exactly regulated by these splice patterns, quite often pre-mRNA is spliced in different constitution of the coding exons. This is a product of alternative splicing, which is commonly occurring and many genes have several variants and some thousands. The mRNA molecule can be differently spliced in different tissues; most likely depending on availability of splice factors. In a gene that has alternative spliced transcripts, most of the exons are constitutive but some exons tend to be excluded, these are called cassette exons and combinations of these may result in a wealth of variants. Exons can also be longer or shorter with alternative splice sites for both 5' and 3' ends. Different promoter structure can result in alternate 5' start exons as well alternate 3' ends. Intron retention is also resulting in alternate transcripts. Splicing can virtually change all functions for a protein, e.g. skipping of membrane domain can result in a soluble protein, changing of ligand binding or enzymatic activity [129].
Alternative splicing greatly increase the protein diversity encoded by the human genome. Estimates from 60% of all human genes [130] to 95% of all multiexonic genes [131] have differently spliced transcripts. Compared to simple organisms, such as bacteria, humans have a moderately increase in number of genes and proteins. Alternative splicing is most likely one explanation for the complexity of higher organisms where the combinatorial approach increases the proteome.

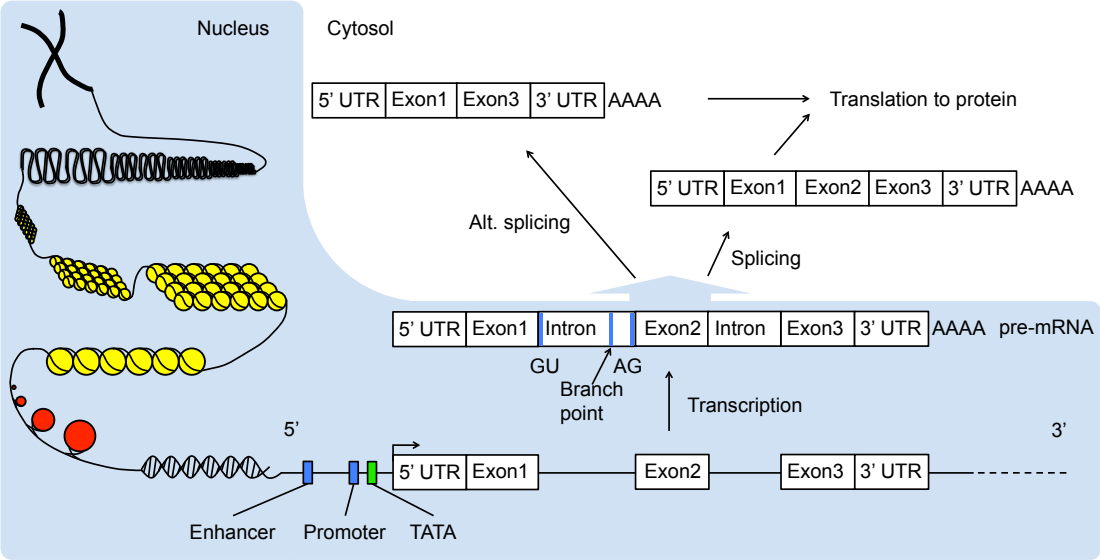

Figure 3. A schematic picture of transcription and splicing of DNA to mRNA.

1.2.5 Study design

To find the responsible genetic changes behind disease traits we need to design an experiment where genetic markers are compared with a phenotype of the trait. This could be done in either experimental animal models of disease or with human patient cohorts. The following description will focus on the latter. This includes linkage, cohort and case-control studies.

Linkage studies

Traditionally, family linkage studies have been used to detect responsible mutations behind disease. The idea is to find genetic markers that show correlated segregation with the trait. To do this it is necessary to follow several generations, which is problematic for obvious reasons. Another approach is to study affected siblings with the goal of finding a shared genetic inheritance.

Family studies have historically worked very well with typical Mendelian diseases, where a single mutation is often causal for the phenotype. But for complex disease with low penetrance the study design is not optimal. Also, in linkage study the resolution is very low meaning that a detected locus typically covers several megabases.

Cohort studies

The cohort study is a very elegant experiment with careful monitoring of exposures for the included individuals and is therefore very suited for studying effects of diets, lifestyle, medical intervention, work conditions, smoking etc. The prospective cohort includes individuals that fulfill certain inclusion criteria regarding exposure and then monitors them over time. The exposure and the expected outcome of this cohort are recorded until a certain pre-determined number of events has occurred, e.g. disease incidence, followed by an evaluation of the exposure and outcome relation (Rothman 2002). In order to get a sufficient number of individuals with the outcome there is often

a need of very large cohorts, specifically if the outcome has a low incidence. This is often the case for endemic diseases.
The clinical trial experiment is a special variant of a prospective cohort where the exposure instead is assigned randomly to half of the cohort and the other is the non-exposure group (the placebo group). This kind of study has a high evidence burden and is typically used for measuring drug efficacy.
A cohort study can also be performed in retrospective, which is in principal the same as the prospective but the information of exposure and outcome already exits for some reason, usually in different registries. This is the most cost-efficient alternative but is limited to data that already exists [132].

Case-control
Genomic components of the more complex common autoimmune diseases have shown to be much harder to assess than for Mendelian disease. Huge progress has been made with the case-control design that allows for studying very large number variables for many individuals with an increase in both power and resolution. The basic design of a case-control experiment is comparing individuals with a certain trait (cases) with individuals without this trait but otherwise of similar genetic background (*i.e.* the same population). If carefully designed and performed, this design offers several advantages over both the family based and cohort based studies. A case-control study is cost-efficient since it can be performed with shorter duration than other experimental designs and with a smaller study population size than a prospective cohort. The case-control studies are very suitable for diseases with a fairly late onset that makes collection of ancestral material more complicated. It is also easy to include large amounts of controls with an increased power as effect at a low cost. Also, when studying genetic variance, compared with a family linkage study the amount of recombination is much more in a case-control material, resulting in an increased resolution of the associated loci.
A major drawback of case-control is the assessment of the exposure that often is subjective. This phenomenon is termed recall bias where affected individuals more often tend to exaggerate the exposure then unaffected. This may result in non-specific conclusions and is one of the reasons why other cohort studies are deemed with a higher evidence burden. Also, the acquisition of controls may also infer errors in the study if not properly selected [132].

1.2.6 GWAS – a new paradigm

Large genetic screens, so called genome wide association studies have introduced a leap in discoveries of disease-associated variations in the human genome. In short time it has become possible to screen all common genetic SNPs of the human genome for large case-control studies. Since the introduction of genome scans over 2000 loci have been discovered for complex traits [79]. Even with this massive increase of new risk

variants, still only a part of the expected genetic contribution for most diseases has been discovered. Reason for this may be other un-observed genetic variations such as structural variations, epistasis effects or genetic interaction. It may also be a result of the high significance threshold, in order to avoid false positives, commonly used for these studies.
A good example of the advancement made possible through GWAS studies is for the disease multiple sclerosis. Many of the MS-associated genes are involved in immune related pathways and may be subject to therapeutic intervention. Also interesting is the two genes in the vitamin D pathways that have been discovered, CYP27B1 and CYP24A1. This finding may eventually explain the role of vitamin D in MS and perhaps also the increased incidence of disease the farther north the population. Another gene, IL2RA, is already a target for therapy, indicating the validity of the findings [133].
For RA similar finding have been made with GWAS. Two successful therapies for RA, TNF-inhibitors (e.g. infliximab) and T cell co-stimulation inhibitors (e.g. abatacept) are targeted by genetic findings from GWAS (TNFAIP3, CLTA-4 (known prior to GWAS) and CD28). However, the function of these variations in relation to disease is still unknown for the majority of the associated genes. Many of the genes are pointing to the NfKB signaling pathway, which controls transcription and plays a central role in regulating the immune system [79].
A major limitation of GWAS studies is the underlying design assumption of LD with causal polymorphisms. This is derived from the hypothesis – Common disease – common variants. This design will miss rare variations since they will most likely not be sufficiently linked with genotyped SNPs and to be detected they need to explain a big part of the risk. It is currently speculated whether this explains a large part of the missing heritability.

1.2.7 Interaction

For the cell to survive and adapt to changes it need to interact constantly. Most of these interactions are carried out with proteins binding and affecting other proteins; in specific patterns we call pathways. These pathways are essential for us to understand to be able to draw conclusions of what happens to the cell when one protein is altered or even removed.
In studying of disease causes we sometimes also use the word interaction for a less physically defined phenomenon, to instead describe a statistical dependency of two factors that together increases the disease risk/protection. This increased risk is compared with the expected effect of both risk factors and if it exceeds this expectation this additional risk is attributed to interaction of the risk factors. The two common ways of quantifying this is departure from additive and multiplicative effects of interaction. In the first model, risk that is additional to the expected sum of the risk from each factor separately is considered due to interaction. The percentage of this additional risk of the

total risk is used as a measure for interaction effect (attributable proportion, see Figure 4). In the corresponding regression model, this is represented exclusively by the group of individuals having both risk variants versus the group of individuals having either of the risk variants [134].

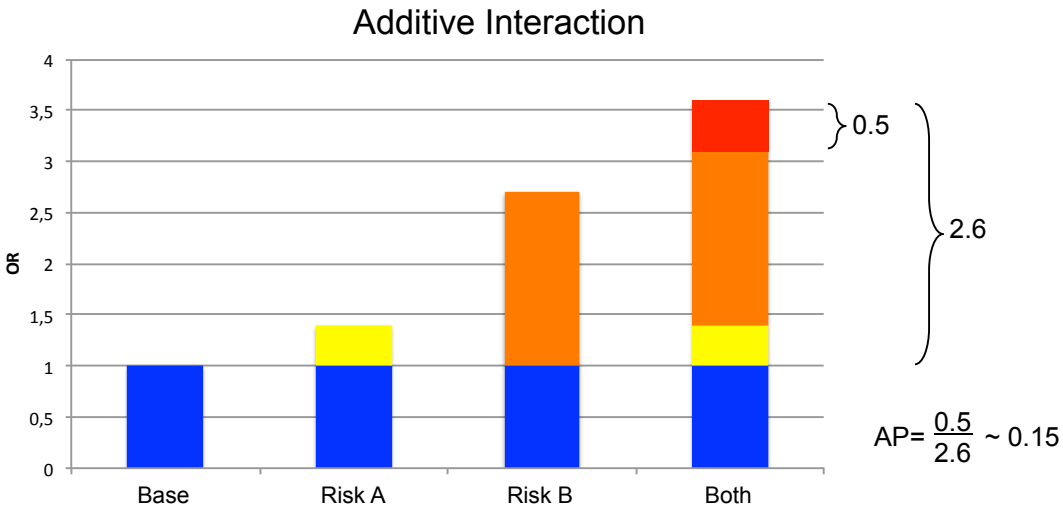

Figure 4. A schematic representation of increased risk due to additive interaction. The attributable proportion (AP) is the percentage extra risk, not explained by the sum of separate risks, of the total risk (from A and B in the picture).

The multiplicative interaction model expects that the overall risk is multiplied for individuals having both risk variants. This is calculated in the regression model by inclusion of a multiplicative variable derived from both risk variants. This analysis can be used for any dichotomous exposure, such as smoking, phenotypic markers or genetic variants (if they are from unlinked loci) [135].
None of the models is better than the other, just different ways of modeling the potential interaction.
Interaction between two genetic loci (genetic interaction) is sometimes referred to as epistasis, which can be misleading in the way that this implies some underlying biological mechanism of interaction. True, this may be so but still it is only a model to analyze if there may be a statistical dependence for two risk factors. Any biologically relevant interaction is still to be proved if a statistical interaction is found [135].

1.3 METHODS

The polymerase chain reaction (PCR) completely changed the field of biotechnology when it was discovered and developed in the mid-1980s by Kary Mullis, a discovery for which he eventually was awarded the Nobel Prize in chemistry (As described in [136]). In fact, it is still the corner stone of most tools for studying genetic diversity because of its extraordinary versatility. The vast majority of sequencing techniques relies on PCR steps and, for instance, one of the recent years largest genetic projects - the human genome project, could not have been carried out on a reasonable time line

without this discovery. The list of techniques involving a PCR step can be made very long and the importance for molecular biology cannot be understated. The following three core steps describes a basic PCR:
Denaturation: The reaction is heated in order to separate DNA strands (~95°C for ~ 30s)
Annealing: The temperature is lowered to allow hybridization of oligonucleotides (primer, 15-30 nt long) specific for the sequence under investigation
Elongation: The reaction is adjusted to optimal temperature for the polymerase, which is 72°C for the commonly used Taq polymerase. The enzyme synthesizes a complimentary a DNA strand with deoxynucleotide triphosphates (dNTPs, nucleotides) that is incorporated in 5' to 3' direction from the annealed primer.
These three steps is cycled until a sufficient amount of copied DNA sequence is received, usually from 20 to 40 cycles.

1.3.1 Genotyping

Genotyping is the procedure where genetic variation is detected for individuals (see chapter on genetics for examples of genetic variation). Many of the genotyping assays are depending on the PCR reaction for reaching detectable levels of DNA. The allelic discrimination method, commonly used for SNP genotyping, uses PCR of the region around the genetic variation and then measures different amount of hybridized probe oligonucleotide specific for either allele. The technique is similar for DNA-arrays (or SNP-arrays) but the reactions are immobilized on a surface making it possible to carry out multiple parallel experiments.

1.3.2 Sequencing

The Sanger sequencing have been instrumental for modern genetic research and is, just as PCR, depending on the polymerase enzyme [137, 138]. Basically it is a normal polymerase reaction but with the addition of dideoxynucleotides (ddNTP), which when incorporated will terminate the elongation process. Originally a sequencing experiment was carried out in four reactions with a specific ddNTP that will only terminate the sequence for that nucleotide but at random positions. The ddNTP were radioactively labeled and the result could be read on a polyacrylamide gel. This procedure was greatly enhanced by the use of fluorescently labeled ddNTP, each with a different emission. This allows for running the sequence in one reaction and then reading the result with capillary electrophoresis.
The problem with Sanger sequencing is that it is difficult to scale up the technique for massive sequencing. The development of another technique, pyro sequencing, has however opened the field for array based parallel sequencing [139]. This method detects the release of pyrophosphate release with every incorporated nucleotide without

the need of a chain termination, as for the Sanger method, and thereby can be used for sequencing-by-synthesis.

1.3.3 Transcript expression

The traditional way of measuring mRNA is by reverse transcriptase PCR (RT-PCR). Reverse transcriptase is an enzyme that synthesizes DNA sequence from a RNA template, that is the reverse of the transcription and hence the name. With this enzyme it is possible to create a complimentary DNA (cDNA) strand from the mRNA extracted from cells and then apply a normal PCR to amplify the product. To be able to exactly measure the template of a PCR the amount of product after every cycle of the reaction can be monitored by fluorescence. By using a specific emission threshold, the number of PCR cycles for the targets to reach this threshold is used to calculate the initial amount. This threshold is set so that the read off is in the exponential phase of the PCR, thus making it possible to compare different reactions. This procedure is called real-time PCR, or quantitative PCR (qPCR) and is a very precise way of measuring the amount of mRNA. There are two major techniques in use for qPCR: the double stranded DNA fluorescent dye and a reporter probe method. The first method uses, in addition to the normal PCR reagents, fluorescence dye that emits light at a specific wavelength upon binding to double-stranded DNA. By this it is possible to follow the increase of the double stranded product after every cycle of the PCR reaction. A common dye for this is SYBR green. The other method uses a probe specifically designed to hybridize with the PCR product. To the probe a fluorescent reporter and a quencher is attached at separate ends and when the polymerase reaches the hybridized probe the quencher and reporter are separated by breakdown of the probe. This allows for emission of light upon excitation and the fluorescence of reporters are proportional to the amount of produced PCR product. [140]
The rapid development of DNA sequencing has made it feasible to instead sequence all the mRNA from a sample and use the number of sequence reads as proxy for the amount. This is a very promising technique that has the advantage of covering all targets at once and is rapidly exchanging the array-based methods [141, 142].

1.4 CONCLUSION OF INTRODUCTION

With this survey I believe I have made the following statements valid as a background for my study:
1. Rheumatoid arthritis is genetically heritable disease that indicates that at least some patients must have a more or less common genetic component(s).
2. The human genome is highly diverse which allows for genetic studies.
3. There is a huge regulatory potential outside the coding regions, which may be very important for genetic contribution to disease.

4. The regulatory features in these regions have effects on expression of genes and also on the composition of transcripts (alternative splicing).

2 STUDY POPULATIONS

All the included papers in this thesis are using data from the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA). This is a population based case-control cohort that includes the major part of incident cases of RA from the middle region of Sweden starting from the year 1996. For each included patient in the EIRA study, 1-2 controls are selected based on age, gender and residence area. All included individuals are asked to fill in an extensive questionnaire and leave blood samples for DNA extraction. [32, 97]. From a small subset of these individuals we have also collected blood for mRNA extraction.

This thesis also includes several other cohorts that are described in the included papers.

3 AIMS

The aim of this thesis is to explore the RA risk associated genes *CIITA*, *PTPN22* and *DCIR* to get further knowledge about the role for development of disease.
The first part of this thesis is devoted to studies of *CIITA*. In papers I and II we aimed to replicate and to study in detail the association of *CIITA* with RA. The first paper includes an updated meta-analysis with evidence of association for a novel marker in the locus and a relationship with SE. In the second paper this relationship is further investigated in additional cohorts.
In paper III we aim to explain the biological background for the previous found association of *CIITA* expression with the associated risk variant by specifically targeting the different promoter isoforms of the gene.
This is followed by similar studies of expression of mRNA isoforms for *DCIR* and *PTPN22* with the hypothesis that the risk variants may have different effect on the specific transcripts, stemming from alternative splicing the genes, which may be specifically affected and it may suggest mechanisms for disease development.

4 MAIN RESULTS

This section is a summary of the results presented in papers I-V in the end of this thesis.

4.1 PAPER I: *CIITA* GENE VARIANTS ARE ASSOCIATED WITH RHEUMATOID ARTHRITIS IN SCANDINAVIAN POPULATIONS

Ever since a variant in the *CIITA* promoter, SNP rs3087456 (-168A/G), was discovered to be in association with RA and MS [55] there have been conflicting reports whether it is an actual risk factor for the disease RA or not. A meta-analysis from 2008 [60] concluded that the overall evidence was negative, but findings for other disease than RA showed association, specifically for MS [43, 61, 62] and also extending this association for other markers in locus. Given this, and the observation that the effect is moderately strong in some population (Swedish: OR 1.19, 95% CI 1.02-1.38 [55]; OR 1.46, 95% CI 0.80–2.66 [43]; Japanese: OR 1.47, 95% CI 1.16-1.87 [58]) where it is abolished in other [56, 57, 143] we believed a thorough investigation was warranted. Therefore we extended the data for rs3087456 in the Swedish EIRA cohort (partly used in [55]) and added a Norwegian cohort with RA (n=819), a juvenile idiopathic arthritis (JIA; n=524), a type 1 diabetes (T1D; n=1211) with controls (n=2149). We also tested a novel risk variant, rs8048002, which has indicated a strong association with risk for Addison's disease [66], suggesting it may be a better marker for disease.

CIITA rs3087456 was significantly associated with RA for the Norwegian cohort (GG vs GA+AA, OR 1.50, 95% CI 1.10-2.05) and also in the combined material of Swedish and Norwegian cohorts (OR 1.37, 95% CI 1.11-1.69), Table 2.

Also rs8048002 was significantly associated with RA, but none of the markers exhibited association with JIA and T1D (Table 1 and Table 2 in paper I).

We also performed an updated meta-analysis for rs3087456 with all available studies published, which supported association of *CIITA* with RA (p=0.02). The effect was, as anticipated, stronger in the Scandinavian populations (including three independent cohorts), which was evident from meta-analysis of subgroups (OR 1.39 (95% CI 1.16-1.66), $p=4x10^{-4}$, Figure 5).

Table 2: Association analyses of rs3087456 in Scandinavian patients and controls

Rs3087456		Genotype count (%)			MAF %	GG vs. GA+AA	
Material	**n**	**GG**	**GA**	**AA**	**G**	**OR (95% CI)**	***P*-value**
Norway							
RA	799	68 (8.5)	271 (33.9)	460 (57.6)	25.5	1.50 (1.10-2.05)	0.0093
JIA	501	28 (5.6)	171 (34.1)	302 (60.3)	22.7	0.96 (0.63-1.46)	0.84
T1D	1190	84 (7.1)	443 (37.2)	663 (55.7)	25.7	1.23 (0.92-1.64)	0.16
Control set #1	932	51 (5.5)	354 (38.0)	527 (56.5)	24.5		
Control set #2	1112	68 (6.1)	424 (38.1)	620 (55.8)	25.2		
Controls all	2044	119 (5.8)	778 (38.1)	1147 (56.1)	24.9		
*Sweden new**							
RA	1212	91 (7.5)	427 (35.2)	694 (57.3)	25.1	1.32 (0.90-1.93)	0.16
Controls	706	41 (5.8)	254 (36.0)	411 (58.2)	23.8		
*Sweden all**							
RA	2479	174 (7.0)	876 (35.3)	1429 (57.6)	24.7	1.28 (0.97-1.70)	0.081
Controls	1332	74 (5.6)	448 (33.6)	810 (60.8)	22.4		
*Combined new**							
RA new**	2011	159 (7.9)	698 (34.7)	1154 (57.4)	25.3	1.42 (1.12-1.81)	0.0043; corrected: 0.017
Controls new	2750	160 (5.8)	1032 (37.5)	1558 (56.7)	24.6		
*Combined all**							
RA all***	3278	242 (7.4)	1174 (35.0)	1889 (57.6)	24.9	1.37 (1.11-1.69)	0.0030; corrected: 0.012
Controls all	3376	193 (5.7)	1226 (36.3)	1957 (58.0)	23.9		

n refers to the number of successfully genotyped individuals. *P*-values are uncorrected and are for Pearson χ^2 test. CI, confidence interval; JIA, juvenile idiopathic arthritis; MAF, minor allele frequency; OR, odds ratio; RA, rheumatoid arthritis; T1D, type 1 diabetes. *"New" refers to previously uncharacterised Swedish samples; "All" includes samples reported in [55]. **Meta-analysis, fixed effects model; heterogeneity: Chi²=0.28, df=1 (*P*=0.59); I²=0 %. ***Meta-analysis, fixed effects model; heterogeneity: Chi²=0.56, df=1 (*P*=0.45); I²= %. This table is modified from Table 1 in paper I where the full table including additional analysis can be found.

When stratifying the Norwegian and Swedish cohorts by SE status we detected an increased combined risk for disease for rs3087456 and rs8048002 in the SE-positive and SE-negative groups, respectively (Table 3 in paper I). This interesting relationship with SE alleles, the strongest genetic predisposition marker for RA, was analysed in detail in paper II.

Conditional regression analysis could not distinguish which of the variants, rs3087456 or rs8048002, was the better marker for disease (supplementary Table 2 in paper I).

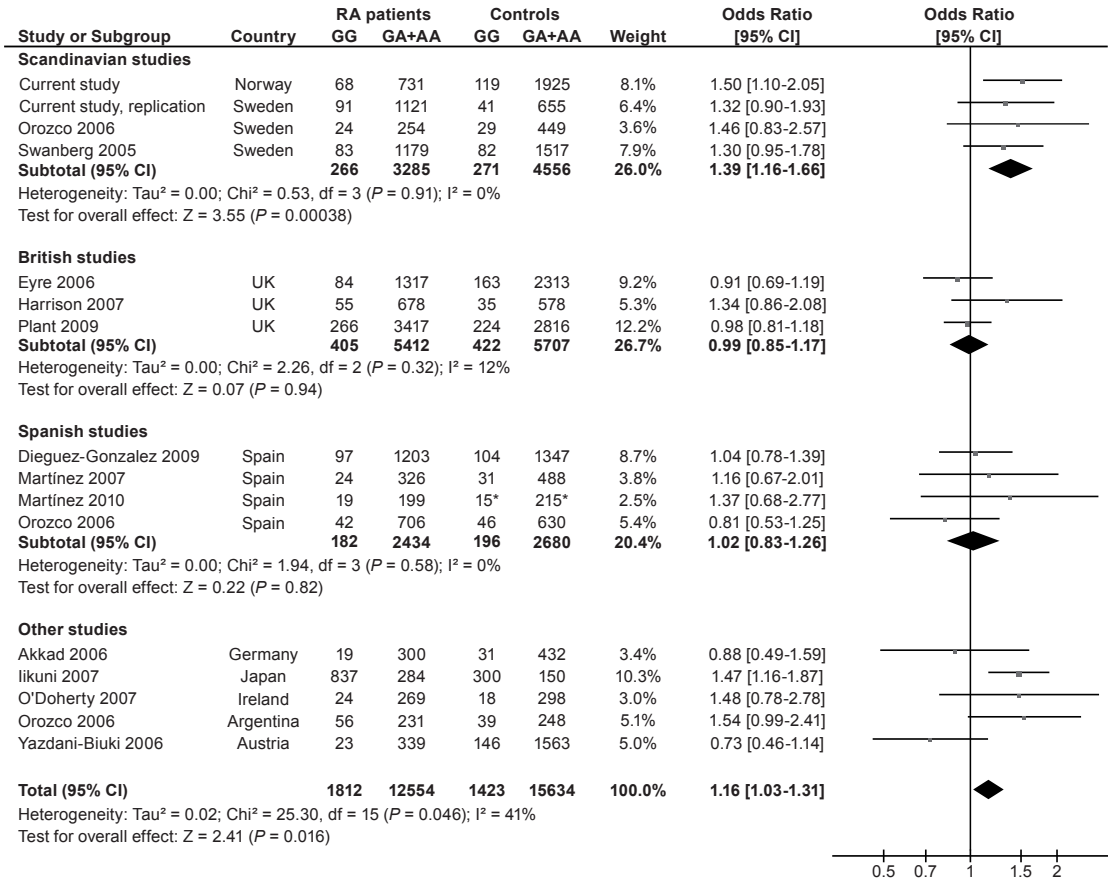

Figure 5. An updated meta-analysis of *CIITA* -168GG as a risk factor for RA. The analysis is performed in subgroups based on origin of population combined with a complete analysis of all available populations. Combined associations were tested with a random effects model. RA, rheumatoid arthritis; CI, confidence interval.

4.2 PAPER II: INTERACTION ANALYSIS BETWEEN *HLA-DRB1* SHARED EPITOPE ALLELES AND MHC CLASS II TRANSACTIVATOR *CIITA* GENE WITH REGARD TO RISK OF RHEUMATOID ARTHRITIS

The role of *CIITA*, a key regulator for the human leukocyte antigen (HLA) class II proteins, for disease risks is interesting due to the strong effect of HLA on autoimmune disease. In paper I, we re-evaluated the association of *CIITA* with RA and showing a possible population dependent mode of action. We also detect an increased level of risk for developing RA with *CIITA* risk variants in the subgroups defined by the presence of SE alleles. The risk conferred by rs3087456 was stronger in the SE-positive group. To further study this effect and evaluate whether this increase is the result of an interaction of the risk variants we performed a detailed analysis of interaction for *HLA-DRB1* (SE) alleles and *CIITA* (rs3087456) in four independent cohorts with a total number of 6649 RA patients and 5118 controls.

We could not establish a consistent significant interaction (additive or multiplicative) between the risk factors rs3087456 and SE alleles in the cohorts studied (Table 4).

The result was similar when individuals were stratified by presence of ACPA. The specific alleles for SE were also scrutinized where it seems that a small effect was detected for the *DRB1**04 allele in the Swedish cohort but it did not replicate in the Norwegian cohort (Table 3).

The variant rs8048002 did not give reliable results since the low minor allele frequency resulted in very small data groups.

A screen of 22 SNPs in the *CIITA* locus did not reveal any other interacting variants.

Table 3. Summary data of the interaction analysis for *HLA-DRB1* SE allelic groups and SNP rs3087456 for the Swedish cohort

Group	rs3087456 and:	AP	CI 95 low	CI 95 high	P value
All	SE (yes/no)	0.2	-0.2	0.5	0.5
	*DRB1**01	-0.3	-1.2	0.6	0.5
	*DRB1**04	0.2	-0.2	0.6	0.3
	*DRB1**10	0.05	-1.5	1.6	0.9
ACPA+	SE (yes/no)	0.3	-0.05	0.6	0.1
	*DRB1**01	-0.2	-1.1	0.7	0.7
	*DRB1**04	0.3	-0.1	0.6	0.2
	*DRB1**10	-0.1	-2.0	1.8	0.9

Additive interaction is presented as attributable proportion (AP) with 95% confidence interval (CI). For additional analysis see supplementary table S5. SE = shared epitope; ACPA+ = anti citrullinated protein antibody positive RA patients.

Table 4. Risk of developing RA for combinations of the *HLA-DRB1* SE and rs3087456 alleles in Swedish, British, Dutch and Norwegian cohorts

CIITA 168GG	SE	Ca/Cont	OR	95% C.I.	AP (95% C.I.)	Add	Mult
Sweden all					0.2(-0.2-0.5)	P=0.4	P=0.9
No	None	571/582	1.0	…			
No	Any	1655/630	2.7	2.3-3.1			
Yes	None	41/30	1.4	0.9-2.3			
Yes	Any	117/33	3.6	2.4-5.4			
Sweden ACPA+					0.3(-0.05-0.6)	P=0.1	P=0.4
No	None	222/582	1.0	…			
No	Any	1205/630	5.0	4.2-6.0			
Yes	None	11/30	1.0	0.5-2.0			
Yes	Any	86/33	6.8	4.4-10.5			
Norway all					0.4(0.03-0.7)	P=0.03	P=0.4
No	None	186/682	1.0	…			
No	Any	533/751	3.8	2.1-3.2			
Yes	None	14/44	0.9	0.6-2.2			
Yes	Any	50/43	3.1	2.7-6.6			
Norway ACPA+					0.4(0.05-0.7)	P=0.02	P=0.7
No	None	56/682	1.0	…			
No	Any	363/751	6.2	4.4-7.9			
Yes	None	5/44	0.7	0.5-3.6			
Yes	Any	35/43	5.5	5.9-16.7			
UK. all					-0.2(-0.7-0.3)	P=0.4	P=0.7
No	None	429/638	1.0	…			
No	Any	1354/529	3.0	3.2-4.5			
Yes	None	36/57	0.9	0.6-1.5			
Yes	Any	97/46	2.1	2.2-4.5			
UK ACPA+					-0.1(-0.5-0.3)	P=0.7	P=0.6
No	None	198/638	1.0	…			
No	Any	1023/529	5.9	5.1-7.5			
Yes	None	13/57	1.0	0.4-1.4			
Yes	Any	79/46	4.3	3.7-8.2			
Netherlands all					-0.3(-1.4-0.8)	P=0.6	P=0.8
No	None	136/146	1.0	…			
No	Any	9/11	2.6	2.2-4.1			
Yes	None	321/116	1.2	0.4-2.2			
Yes	Any	20/10	4.3	1.0-4.8			
Netherl. ACPA+					-0.4(-1.7-1.0)	P=0.6	P=0.8
No	None	27/146	1.0	…			
No	Any	2/11	5.9	3.7-9.6			
Yes	None	127/116	1.4	0.2-4.7			
Yes	Any	8/10	9.9	1.6-12.0			

Evidence for additive and multiplicative interaction is displayed as significance (P value) of deviation from expected risk given no interaction. AP = attributable proportion; SE = shared epitope alleles; OR = odds ratio; ACPA = anti citrullinated protein antibody positive; CI = confidence interval.

4.3 PAPER III: GENETIC CONTROL OF ISOFORM EXPRESSION OF HUMAN MHC CLASS II TRANSACTIVATOR

It was previously shown that the *CIITA* risk variant rs3087456 was correlated to the mRNA expression of the gene [55]. Interestingly, *CIITA* has four promoter isoforms that are expressed at various degrees in different cell types [71]. This probably gives an advantage in specific regulation of the HLA class II genes where different promoters are under control of different transcription factors. With this complexity we felt the need of quantifying the isoforms separately since they may be affected by different mechanisms. We characterized the expression of isoforms *CIITA*_pIII and *CIITA*_pIV in samples from the EIRA cohort and an asthma cohort. Isoform *CIITA*_pI was below detection limit in our material and *CIITA*_p2 is rarely expressed (for instance detected in a human melanoma cell line [73]).

For *CIITA*_pIII the expression was similar for RA patients and controls but this was not the case for *CIITA*_pIV, where patients had significantly higher expression (Figure 1a and 1b in paper III). For PBMCs stimulated with IFN-gamma, the expression was higher for both *CIITA*_pIII and *CIITA*_pIV for patients (Figure 1b and 1d in paper III). Also, the expression for stimulated cells increased more for the patients samples compared to controls (Figure 1e and 1f in paper III). In the cohort with asthma patients and controls we found no difference in expression between patients and controls (Figure 2 in paper III).

Perhaps more interestingly, we detected a correlation with genotype and *CIITA* expression. This was the most pronounce for *CIITA*_pIV, but could also be seen for *CIITA*_pIII in stimulated cells. The expression of *CIITA*_pIII and *CIITA*_pIV for controls and for the samples in cohort of asthma patients and controls was not correlated with the SNP rs3087456 however.

For patients, we correlated a further 22 SNPs in the *CIITA* locus with the isoform expression and could detect other SNPs significantly associated with expression, primarily in the promoter region (Table 1 in paper III). This association was coinciding with the allelic association with risk of developing RA (Figure 6). This may be due to extensive LD in the region (Figure 6) but is still a sign of consistence for our finding.

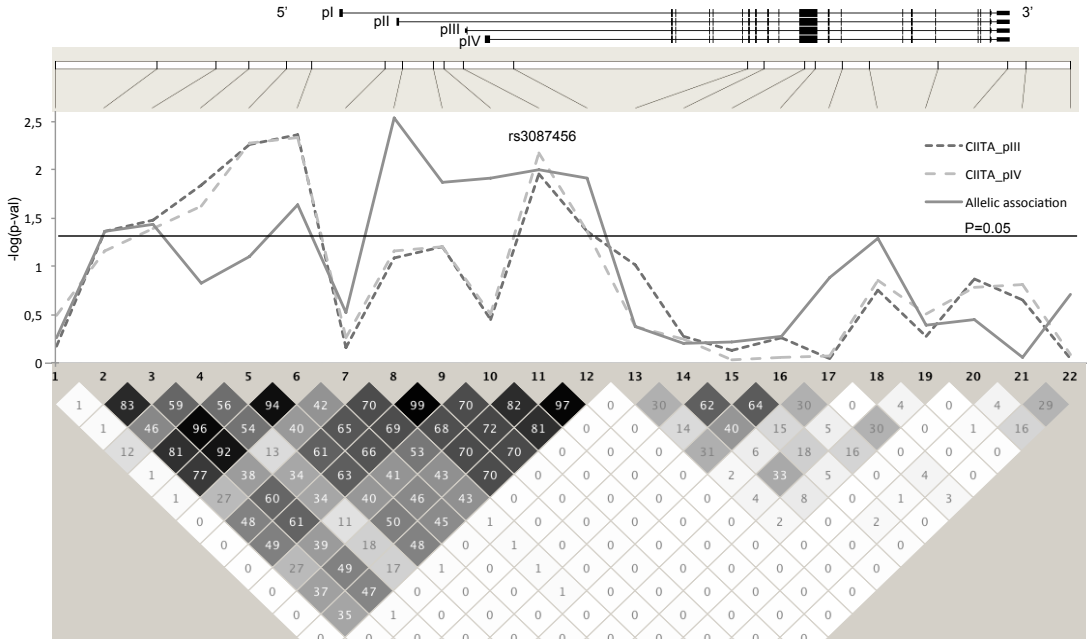

Figure 6. The middle part is a graph representing the association (-log p-val) of SNPs (n = 22) with mRNA expression (dashed line) and with risk for disease (continuous line) for the *CIITA* locus. Under the graph is a LD-plot where numbers correspond to r-square values and above is a schematic illustration of the *CIITA* isoforms exonic-intronic structure.

4.4 PAPER IV: DIFFERENTIAL EXPRESSION OF TRANSCRIPTS FOR THE AUTOIMMUNITY-RELATED HUMAN DENDRITIC CELL IMMUNORECEPTOR

In this paper we present result of an expression study for the gene *DCIR* in peripheral blood mononuclear cells focusing on the four known expressed gene transcripts. The study was conducted in mRNA samples from the EIRA cohort with replication in a cohort of asthma patients.

The mRNA expression was evident for all four splice forms of *DCIR* in PBMC and we could not detect any differences for RA patients versus controls (Figure 1a and figure 2 in paper IV). During the basic investigation of splice forms we could detect a novel fifth variant in 12.5% of all individuals and sequencing revealed that this form was lacking exons three and four (Figure 7).

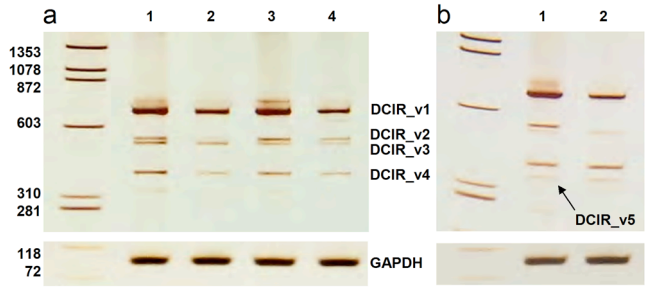

Figure 7. A gel picture of stained PCR product from of all known splice forms of *DCIR*. The splice forms were all amplified by the same primer-pair. In a), the four previously known transcripts are shown and in b) the additional fifth transcript is shown. All PCR products were of the expected size and the structure was confirmed by sequencing.

In line with the assumed function of *DCIR*, as part of an inhibitory cell-signaling pathway, we could show that it was significantly down regulated with IFN-gamma stimulation of cells (Figure 3 in paper IV). This down regulation was similar for all the targeted splice forms and was not affected by disease status.

To assess whether the RA associated variants could have an effect on the mRNA expression we used a set of 19 SNPs for patients and 6 SNPs for controls across the *DCIR* locus. This revealed a clear association of the fourth *DCIR* splice form with variants in the locus (Figure 8), where the SNP rs2024301 was the strongest (patients P=0.0009, controls P= 0.0078 and combined samples P=6.5*10^{-6}), Kruskal–Wallis, Figure 5 in paper IV). We could also replicate this finding in the smaller cohort of asthma cases and controls.

We could not see an effect on the expression when stratifying the patients by ACPA status or any other disease measure.

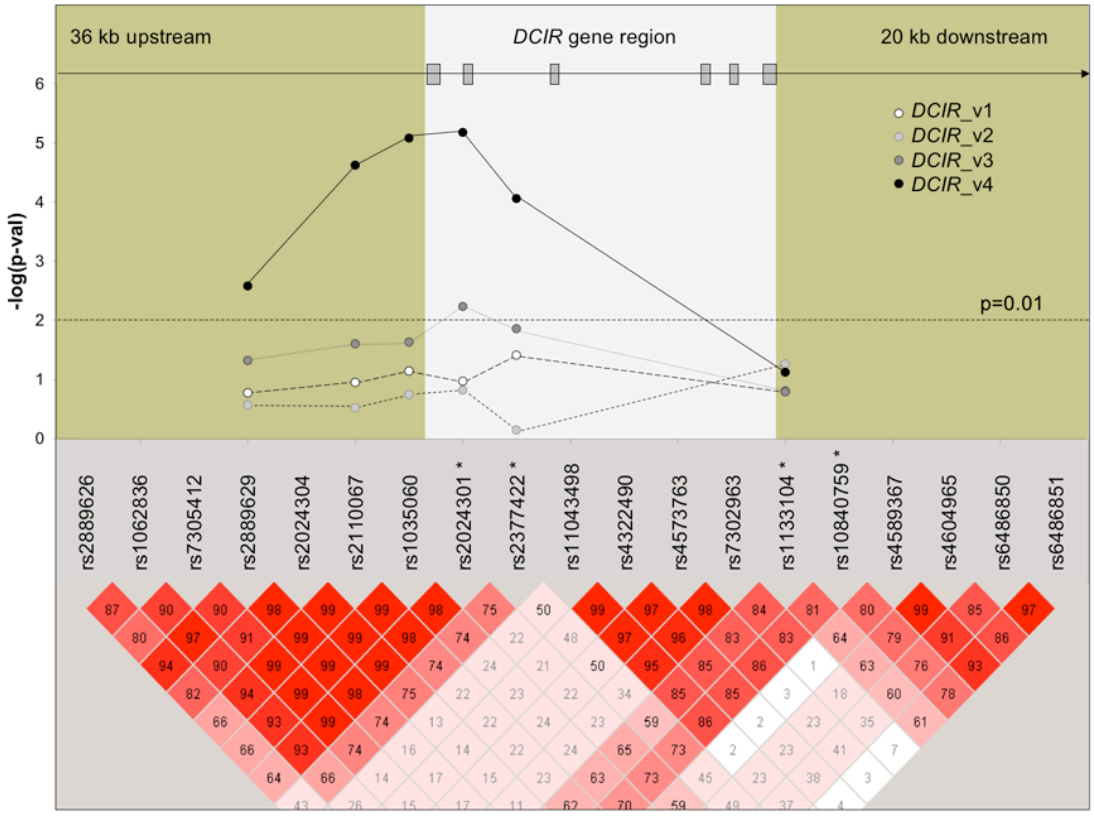

Figure 8. The graph represents association of SNPs with expression of *DCIR* transcripts from analysis of both RA patients and controls. SNPs marked with an asterisk are associated with ACPA negative RA. The upper part of the figure shows the exonic configuration in genomic DNA of *DCIR*_v1.

4.5 PAPER V: THE BALANCE OF EXPRESSION OF *PTPN22* SPLICE FORMS IS SIGNIFICANTLY DIFFERENT IN RHEUMATOID ARTHRITIS PATIENTS COMPARED WITH CONTROLS

Genetic variation in protein tyrosine phosphatase non receptor 22, *PTPN22*, has repeatedly been shown to associate to autoimmune diseases where the risk is detected, mainly, through the SNP rs2476601. This is a non-synonymous polymorphism affecting the aminoacid at position 620 in the protein with a change from arginine to tryptophan (R -> W). This difference in the protein has been associated with an increased function of the phosphatase. This gain of function leads to impaired receptor signaling for T-cells [51] and B-cells [53]. There are, however, other variants in the locus that has been shown to associate with disease independently from rs2476601 [144, 145], which is evidence that there could be other mechanisms for PTPN22 involvement in the risk for disease.

In order to look for other mechanisms by which disease associated polymorphisms can exert their effect we measured the level of mRNA expression of two transcripts from *PTPN22* for which it has been shown there are corresponding protein isoforms, *PTPN22*_v1 (NM_015967.5) and *PTPN22*_v4 (NM_012411.3) [146]. The measurement for the longer of these two, considered normal, transcripts also involved two other transcripts (*PTPN22*_v2 and *PTPN22*_v3 (NM_012411.4 and NM_001193431.1 respectively)), which were discovered recently. An illustration of the *PTPN22* transcripts with functional domains can be seen in Figure 9. This study was conducted with three independent cohorts of RA patients and controls for exclusion of sampling and sample preparation errors and qPCR measurements was conducted at different labs.

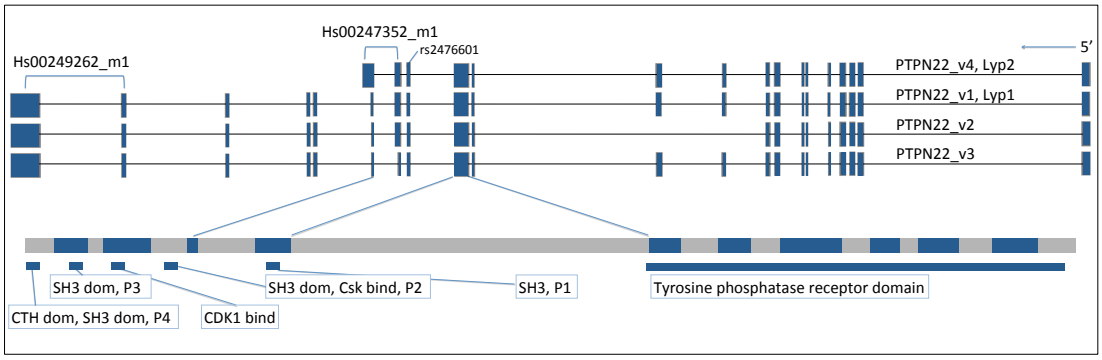

Figure 9. A genomic illustration of *PTPN22* at human chromosome 1p32. Hs00247352_m1 and Hs00249262_m1 are assay numbers (Applied Biosystems). The four SRC Homology 3 (SH3) domains are denoted by P1-P4 and are typically involved in protein-protein interaction. Known binding sites for C-terminal Src kinase (Csk) and CD2-binding protein 1 (CD2BP1) are located in P2 and P4 (C-terminal homology (CTH) domain) respectively. The C-terminus of Lyp1 contains a consensus motif for cyclin-dependent kinase 1 (CDK1). Exonic and basic domain structure is deduced from Ensembl (Ensembl 64: Sep 2011)[147] and UCSC Genome Browser (GRCh37/hg19 assembly)[147, 148] [148] and published articles [146, 149].

We discovered and could replicate a consistent difference in expression of *PTPN22* splice forms for patients and controls in all the three RA cohorts (Table 5). Patients had in general a higher expression of *PTPN22*_v1 and lower expression of *PTPN22*_v4. Using the ratio for each individual of the expression for these two transcripts the difference was even more pronounced.

Table 5. Expression of *PTPN22* transcripts in PBMCs from cohort I, II and III.
Median relative quantity values

		Control group	RA patient group	P-value [a]
Cohort I		*n*=44	*n*=44	
	*PTPN22*_v4	1.05	0.85	0.08
	PTPN22-long	1.03	1.20	0.006
	long/v4	1.01	1.42	$6x10^{-9}$
Cohort II		*n*=19	*n*=47	
	*PTPN22*_v4	0.98	0.96	0.85
	PTPN22-long	0.97	1.07	0.25
	long/v4	0.96	1.15	0.02
Cohort III		*n*=48	*n*=48	
	*PTPN22*_v4	0.96	0.72	$1.2x10^{-4}$
	PTPN22-long	0.97	0.92	0.2
	long/v4	1.04	1.24	0.01

[a] P-value for comparison of patient group and control group, Mann-Whitney test.

This difference was abrogated in IFN-gamma stimulated cells where the healthy controls behaved similarly as the patients (Table 2 in paper V). We could control for effect of medication in cohort 1 since not all individuals were on treatment at the time of sampling. Also, we performed cell experiments with methotrexate treatment of PBMCs from healthy controls and also Jurkat and Daudi cell lines (T lymphocyte and B lymphoblast cell lines respectively) and the results showed no trend for up- or down-regulation of the ratio for *PTPN22* expression by this treatment.
We also measured *PTPN22* transcripts in a multiple sclerosis (MS) cohort (patients n=60, controls n=44) and the difference between MS patients and controls were inversed from RA patients and controls. This marked difference from RA coincides with the fact that *PTPN22* is not a risk factor for MS disease.
Of the studied risk factors, neither the disease associated SNP rs2476601 nor any other variant in the locus (25 SNPs spanning Chr1:114114146-114303491) were in correlation with the expression of the *PTPN22* transcripts or the ratio of those.

5 DISCUSSION

5.1 OVERVIEW

In this thesis work, variants in three seemingly separate genetic risk genes have been studied thoroughly. The aim was to expand and consolidate the evidence for a disease-causing role of the current genes and to explain possible biological mechanisms for previously found associations.

Single nucleotide polymorphisms (SNPs) are central to conducting large well-powered studies of common complex disease. The last five years, genome wide association studies (GWAS) has pushed our knowledge about the underlying genetic contribution for many diseases with several new potential risk loci in our genome to study further [79]. It is, however, in most cases unknown exactly how the associated risk variants acts to increase the risk for disease. Also, it is not always trivial to connect an associated SNP with a particular candidate gene locus, which could be evaluated for a studied effect.

The genes of topic are, at varying degree, established risk factors for RA.

PTPN22

The risk variant rs2476601 in *PTPN22* has been replicated several times for RA and is considered the strongest genetic risk factor outside the major histocompatibility complex for autoimmune disease, showing association to several diseases, such as type 1 diabetes, systemic lupus erythematous, Hashimoto's thyroiditis, Grave's disease, Addison's disease, myasthenia gravis, vitiligo, systemic sclerosis juvenile idiopathic arthritis and psoriatic arthritis (reviewed in [150]).

CIITA

The effect of variants in *CIITA* for RA is still under debate with conflicting results in different populations. However, there are reports of variants in *CIITA* as risk factors for MS [43, 55, 59, 61, 62], SLE [151] and Addison's disease [66]. Shared risk factors, pleiotropy, is a phenomenon that is often seen for autoimmune disease which increase the likelihood that *CIITA* has a true effect for RA.

DCIR

The *DCIR* gene is the least investigated among these three targets. With the background from a rodent model for oil-induced arthritis this locus was associated with sero-negative RA [75]. This is particularly interesting since there are very few genetic candidates for this subgroup. And perhaps even better evidence, mice with a knockout for the *DCIR* gene develops autoimmune like disease [152].

Our intentions have been, for each of these autoimmune associated loci, to enhance the connection to disease with further genetic and gene expression analyses.

5.2 *DCIR* EXPRESSION

For *DCIR* we could show that a transcript variant was significantly correlated with the disease associated polymorphism. This transcript variant is a putative soluble protein, which could have an opposing effect on the normal signaling pathway for *DCIR*.
The lack of replicate studies for *DCIR* is not very surprising. Firstly, the effect found between RA patients and healthy controls was small, (1.27 (95% CI 1.06–1.52)) and to replicate this with 80% power in a candidate gene study, it would need around 1000 patients and 1000 controls at significance level 0.05 [153]. In most cases, a lower significance level is wanted with the extreme in GWAS studies, where $5x10^{-7}$ is considered a reasonable threshold for significance [154], which results in that a larger study population is needed (>3000 patients in this case). Secondly, the association was found for the ACPA negative subpopulation of RA, which in Swedish EIRA cohort represents about 40% of all patients. In other cohorts this is many times smaller or sometimes non-existing. These two facts combined hamper the likelihood of a successful replication of this risk locus.
There are, however, a couple of facts that speak in favor that *DCIR* is involved in the etiology of RA. DCIR protein is abundantly expressed in the synovial fluid and tissue of RA patients; it is not detected in the joint of healthy individuals. This expression is decreased with local glucocorticoid treatment [155]. DCIR protein is expressed on many professional antigen presenting cells, but perhaps the most interesting is expression on both CD4+ and CD8+ T cells. This expression was shown to be restricted to cells from synovial joint compared with cells from blood [155]. Also, IFN-alpha seems to be interesting for regulation of DCIR. INF-alpha induces a reduction (78%) of cell surface DCIR expression and when triggering *DCIR* it down-regulates TLR9-mediated IFN-alpha production by pDCs [156]. IFN-alpha is a pro-inflammatory cytokine, which is highly expressed in the RA joint and is also co-expressed with TLR3/7 [157].
Additionally, we have observed a clear co-expression of DCIR and FOXP3 on T cells, which is further enhanced for cells in synovial fluid. Interestingly, these DCIR+FOXP3+ T-cells were not CD25bright. This data suggests that there may be a unique subset of Tregs distinguished by DCIR+FOXP3+CD25low specifically in the rheumatic joint (unpublished data).
In the paper included in this thesis, we showed that there was no difference of the mRNA expression of *DCIR* in circulation for a sufficiently sized cohort of RA patients and matched controls, which indicates that the effect may be confounded to the actual site of inflammation – the rheumatic joint. This is interesting since the expression of all *DCIR* mRNA isoforms was down regulated by IFN-gamma in our study and previous studies has shown a similar down regulation followed by IFN-alpha [156].
The correlation of rs2024301 with *DCIR* expression is equal for patients and controls gives a hint that this is a biological effect of the SNP and not biased by the disease status. A possible functional effect for the *DCIR* RA associated allele was presented as

significant lower expression of the shortest mRNA splice form. This isoform lacks the membrane-binding domain and could give rise to a potentially soluble gene product. The *DCIR* full length variant is assumed to have an inhibitory function (supported both by the ITIM motif and by the down regulation in cells treated with IFN-gamma) where the soluble form may counteract this. Less of this soluble form may therefore lead to a higher inhibitory cell signaling.

5.3 TRANSCRIPT BALANCE FOR *PTPN22*

PTPN22 R620W is without a reasonable doubt associated to autoimmune disorders, which is not only reproduced in several diseases with a strong antibody component but also across several different populations where this polymorphism exists [41]. This latter fact supports a causal and functional effect of the SNP. There is however no clear mechanism for how this mutation acts to favor disease.

Studies of primary human cells suggest that R620W is a gain of function mutation with an increased de-phosphorylation and inhibitory function as result [51, 52, 158]. In opposite, in Jurkat cells (Lymphoma T cell line), co-expression *PTPN22* W620 and CSK (C-terminal Src kinase) have an increase phosphorylation of the downstream MAPK Erk compared to wild-type R620 [54]. This result implies a loss of function with an increased effect of TCR signaling. With these effects pointing in the opposite direction, when it comes to the effect of R620W, it is not impossible that the mutation may exert both positive and negative regulation in the different pathways in same or different cells. The proposed interaction of LYP and CSK, which stems from results from the homologous mouse protein Pep studies [49, 50], could have different effect on the LYP isoforms or possibly no effect on the shortest.

PTPN22 is expressed exclusively in a majority of the hematopoietic cell [159], with the highest expression in NK cell and neutrophils and lowest in monocytes and CD4+ T cells [41]. This may, however be subject to change with activation of cells.

Interestingly, tissue from thymus has been shown to express all LYP isoforms, which may point to an important function of the transcripts for the positive and negative selection of T cells.

In our work we illustrate that the expression of *PTPN22* is different for RA patients compared with healthy individuals and that this is specific for the different isoforms. If *PTPN22* may indeed have an ambiguous role, with both a gain of function and a loss of function for different pathways, an altered balance between transcripts may increase the direction of autoimmunity. Also, exactly how the R620W mutation affects the enzymatic capability of *PTPN22* is unclear since the position is well outside the catalytic domain (see Figure 9). This mutation may very well have different effect for the different isoforms or just be in LD with a causal variant.

Another interpretation of our findings may be that we see a marker of active inflammation for RA patients where this is not the case for MS patients which show an opposite balance of isoforms compared with controls. If this is true, it is very likely that

this difference in expression is similar for other autoimmune diseases that associate with *PTPN22* R620W. In conclusion, the altered balance between isoforms could be used as a marker for inflammation and even useful for diagnosis of disease.

5.4 *CIITA*

5.4.1 *CIITA* is re-associated to RA

The fact that *CIITA* has a well-defined function in the expression of class II HLA, including *HLA-DRB1*, gives it a promising role as a gene candidate to study for autoimmune disease. When a variant in *CIITA*, -168G (rs3087456), was presented as risk factor for RA, MS and MI (myocardial infarction) [55] it became a target for many replication studies. These studies gave an inconclusive picture and when used in a meta-analysis by Bronson et al., 2008, the combined results was that there was no evidence regarding *CIITA* as a risk locus. There was, however, still evidence for an effect in the Swedish population [43, 55] that needed to be followed up, particularly since evidence for *CIITA* involvement in other diseases is growing [61, 63-66, 151]. When extending the *CIITA* research with data from Norwegian population, the results points in the same direction as for the Swedish population and with an updated meta-analysis, the *CIITA* -168G risk variant is a marginally significant risk factor for RA. In the Scandinavian population, this is much more evident (3551 patients and 4827 controls; OR 1.39 (95% CI 1.16-1.66), $P=3.8x10^{-4}$; no heterogeneity: $I^2=0\%$, P=0.91). It seems, not exclusively, that *CIITA* plays a bigger role as risk factor in northern populations, either because of interaction with unknown environmental factors or due to difference in allelic frequency and higher statistical power to detect association. Mutations that exist in several populations are very seldom risk factors for only a certain subset of these, but even so, this seems to be the case for the risk variant -168G in RA. One of likely explanations for this is that rs3087456 is not the causal SNP for this association; rather it is in linkage disequilibrium (LD) with a variant that has direct or indirect functional effect on disease risk. If this causal variant has a low frequency in the population, it may even be missing in some, even small differences in LD may affect the association to disease [160].
It should also be noted that heterogeneity between different cohorts, not only in the inclusion of cases but also controls, might infer these different conclusions. For instance, we see a clear difference in association for subgroups determined by the SE status; rs3087456 has a stronger effect in the SE-positive subgroup (SE-positive: OR 1.58 (95% CI 1.20-2.10), P=0.0013 vs all RA: rs3087456 OR 1.37 (95% CI 1.11-1.69), P=0.0030). This data is extended in paper II.
We also detected a, for RA, new risk variant, rs8048002, in the Norwegian cohort and replicated it in the Swedish population. We could not show that this variant was independent of rs3087456 and the relation between these risk variants is unclear.

5.4.2 Interaction with MHC class II

The discovery that *CIITA* has stronger effect in the subgroup of SE-positive patients prompted us to perform a detailed investigation whether the *CIITA* locus is statistically interacting with *HLA-DRB1* locus, in other words: do individuals carrying risk variants in both loci have an unexplained increase of risk, not expected by merely adding the separate risks. If they are dependent risk factors, this should be the case. We already know they are biologically interacting and it could be that certain alleles from both loci are driving the disease more strongly.
In this work we were using four cohorts from Swedish, Norwegian, British and Dutch populations. As for the previous study, results were differing between populations but the overall conclusion was that a significant interaction could not be established. Since SE is a synthetic marker for a specific set of *HLA-DRB1* alleles, these were also focus for deeper analysis of interaction with *CIITA* but with no conclusive results.
It seems that the relationship of *CIITA* and SE alleles that was found in paper I is not due to a significantly large interaction. Interestingly, the other variant studied in paper I, rs8048002, had a stronger effect in the ACPA negative subgroup. Since LD between markers indicates a relation ($D'=0.96$ and $r^2=0.20$ in the Swedish cohort) this opposite relation is surprisingly and the effect of SE should therefore be interpreted with caution. However, a relation for *CIITA* and *DRB1* has been observed in another study of a MS cohort [61] where a dependency for *CIITA* rs4774 and *HLA-DRB1**1501 is described. This result indicates that there may be more to discover for these two risk loci and that we need to refine our hypothesis.

5.4.3 Expression of CIITA isoforms

As seen in the original work describing the association of *CIITA* with disease [55], the risk variant correlates with *CIITA* mRNA expression for IFN-gamma stimulated PBMCs and also downstream levels of class II HLA mRNA. In paper III we investigate if this effect from -168G variant (rs3087456) specifically affects *CIITA* isoforms. We could establish that this variant was correlated to both *CIITA*_p3 and *CIITA*_p4 but with a more pronounced effect for _p4. This could mean that both these promoters have a common transcription factor binding site that is upstream of the transcription initiation site. Or, it could be more several mutations in linkage affecting both promoters in a similar fashion. Due the extensive and strong LD pattern in the promoter region of *CIITA* this is difficult to discern. This LD block is most likely also the reason why we see an extensive overlap between the SNPs correlating with isoform expression and the SNPs associating to disease (Figure 6). The well-conserved LD block also extends a good portion 5' upstream from the first *CIITA* promoter exon indicating that this region may contain important enhancer elements for transcription.
If the SNP -168G (rs3087456) would be a true causal variant, with an effect on *CIITA* expression, it would most likely not effect transcription of both isoforms. More likely,

another variation in linkage with -168G is responsible for both association to disease and correlation to expression. This conclusion also fits with our results presented in paper I.

It is interesting that we only detect the correlation for expression and genotype for patients. It could be argued that a variant should have the same effect regardless of diagnosis. However, the expression of *CIITA* is considerably higher for patients and this up-regulation might be leading to the detectable difference. This highlights the need for studying expression in suitable cohorts or there may exist a substantial possibility for missing these discoveries.

Due the extensive linkage in this locus, further molecular and mechanistic studies are needed to explain *CIITA*s role in the etiology of rheumatoid arthritis.

6 CONCLUDING REMARKS

In this thesis work, we present advances for non-HLA genetic risk factors behind RA, a topic that has gained focus with the recent years development of whole genome techniques. These risk factors are very subtle and hard to distinguish and the effect of these variants is very likely also small. However, even small differences may have large consequences during the life span of a human being.
A common theme for the studies of topic is increase of inhibitory regulation of disease associated variants. Effect of RA associated SNPs could lead to more restricted immune signaling with:
DCIR full length variant is assumed to have an inhibitory function and a soluble form may counteract this. Less of this soluble form, which is seen in paper IV, may therefore lead to a higher inhibitory cell signaling.
CIITA RA re-associated variant rs3087456, as we show in paper I, has a confirmed effect of both *CIITA*_pIII and *CIITA*_pIV with a lower expression for the RA risk allele. The association of *CIITA* with RA seems to be independent of the risk from *HLA-DRB1* (paper II).
PTPN22 620W variant is shown to be a gain of function mutation in many studies. In our paper V we complement this picture by showing that the expression of splice forms is different for RA patients and healthy controls with more of the full-length, putatively more active, splice form. This may increase the restriction for the T-cell receptor and also other cell surface receptors that may include *PTPN22* for their signaling pathway.

A combined orchestrated effect of all this risk variants together with other risk factors known for RA, perhaps in different time-points of disease development and at different cellular locations, may be what predispose certain individuals for rheumatoid arthritis, a complex disease which is heritable but with a low penetrance of genetic factors.

7 FUTURE PERSPECTIVES

To continue to scrutinize the impact of *CIITA* on the risk for developing RA, studies in additional cohorts are warranted. If it is true that the effect is population restricted, further studies in theses populations could both reassure and reveal further risk variants. Also, this should be complemented with sequencing of the locus to find possible rare variants.

As mentioned before, the *CIITA* locus has an extended LD around the promoter regions, which makes it almost impossible to separate the effect from a causal variant from other close variants. To find out if it is a specific effect of the rs3087456 SNP expression should be studied in a cell line with a point mutation at the -168 position.

The difference in expression we have found for *PTPN22* transcripts is from a pool of mononuclear cells from blood. It is possible, or even likely, that the effect is driven by a subset of these cells that will give a more pronounced effect. Therefore we should collect new samples from patients and controls and sort the cells in the major cell subset. Suitable cells would be CD4+ and CD8+ T-cells (separate naïve and active/memory cells), B-cells (naïve and active/memory), monocytes and dendritic cells. A longitudinal study with collection of samples starting from early RA patients could also answer whether this is a secondary pattern that increases with disease duration.

With a collection of such samples, of separated and also possibly stimulated cells with interleukins and interferons, it would be possible to re-address the expression of all the targets in this thesis and also other to elucidate differences between patients and controls. Combining this with data from previously performed GWAS would lead to further powerful analyses.

To cover as much data as possible for these samples, whole transcriptome sequencing would be a good alternative (cDNA sequencing) as a complement for targeting specific transcripts.

8 ACKNOWLEDGEMENTS

Without these people this thesis would not have come about. Thanks to:

Leonid Padyukov, my main supervisor who is there 24-7-365. Life isn't always easy, but never give up! Thank you for all the memorable discussions that showed me the way forward.

Lars Alfredsson, my co-supervisor who helped with epidemiological issues and was a great support for writing papers.

Lars Klareskog, head of the lab, for building up a wonderful work environment. Your visions are shining brighter than most leaders. Thank you for having me.

Ingrid Lundberg, for leading me to science by teaming up with Lars. Thank you for all the great years.

All co-authors, specifically **Morten Christoph Eike, Darren Plant, Nina A. Daha, Beate Skinningsrud, Margot O'Toole** and **James D Clark**. Thanks for your great support.

Eva Jemseby, Gull-Britt Almgren, Gloria Rostvall and **Julia Boström** for processing and keeping track of all our samples.
The EIRA secretariat for managing the EIRA project that is very essential for our research
Gunnel Bemerfeldt for all administrative help and **Susanne Karlfeldt** for keeping order for Lars. You are doing a fantastic work in structuring the chaos.

Maria, Clemens, Mai, Too and **Espen** for creating a motivating group and good meetings.
Micke, our discussion about brewing has enlightened the entire lab! I hope that our adventures will continue.
Lasse and **Vijole**, the most fantastic couple! I look forward to future oyster-dinners.
Carina, for bringing something extra to the group.
Nånnis, for the scientific and medical support.
Gustavo, always with a big smile! Thanks for the great time in the Rheuma lounge.
Vivi, for all the advise on science-related matters and great beer tips!
Henrik, for facilitating the work with statistical interaction.
The **HMGB1 group** for nice lunches, continuous laughter and a fun cruise!
All former and present members at the **Rheumatology** and **Neuroimmunology Unit.**

The CMM-IT, for managing the infrastructure of CMM. Keep up the good work!

To all **my friends** for reminding me that life isn't always work. You are all very dear to me.

Åsa and **Maris**: Thanks for all the great times, trips and experiences.

To my family:

Åsa, min underbara storasyster som alltid är där när det behövs, och **Andreas**, som ser till att min syster orkar lite till! **Tyra** och **Märta**, mina underbara systerdöttrar!

Mamma och **Pappa**! Min nödutgång i livet, när allt känns jobbigt är ni ett enormt stöd vare sig ni vet det eller inte. Tack för all hjälp genom åren.

Emeli: Vad skulle det ha blivit utan dig?

9 REFERENCES

1. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. 2000;43(1):30-7.
2. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am J Med. 2006;119(6):503 e1-9.
3. Heliovaara M, Aho K, Aromaa A, Knekt P, Reunanen A. Smoking and risk of rheumatoid arthritis. J Rheumatol. 1993;20(11):1830-5.
4. 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;39(5):732-5.
5. Uhlig T, Hagen KB, Kvien TK. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. J Rheumatol. 1999;26(1):47-54.
6. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996;85(3):311-8.
7. Abdel-Nasser AM, Rasker JJ, Valkenburg HA. Epidemiological and clinical aspects relating to the variability of rheumatoid arthritis. Seminars in arthritis and rheumatism. 1997;27(2):123-40.
8. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. Autoimmunity reviews. 2005;4(3):130-6.
9. Del Puente A, Knowler WC, Pettitt DJ, Bennett PH. High incidence and prevalence of rheumatoid arthritis in Pima Indians. American journal of epidemiology. 1989;129(6):1170-8.
10. Harvey J, Lotze M, Stevens MB, Lambert G, Jacobson D. Rheumatoid arthritis in a Chippewa Band. I. Pilot screening study of disease prevalence. Arthritis Rheum. 1981;24(5):717-21.
11. Harvey J, Lotze M, Arnett FC, Bias WB, Billingsley LM, Harvey E, et al. Rheumatoid arthritis in a Chippewa band. II. Field study with clinical serologic and HLA-D correlations. J Rheumatol. 1983;10(1):28-32.
12. Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs. Nature medicine. 2001;7(8):899-905.
13. Rantapaa-Dahlqvist S, de Jong BA, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum. 2003;48(10):2741-9.
14. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31(3):315-24.
15. Waaler E. On the occurrence of a factor in human serum activating the specific agglutintion of sheep blood corpuscles. 1939. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica. 2007;115(5):422-38; discussion 39.
16. Dorner T, Egerer K, Feist E, Burmester GR. Rheumatoid factor revisited. Curr Opin Rheumatol. 2004;16(3):246-53.
17. van Jaarsveld CH, ter Borg EJ, Jacobs JW, Schellekens GA, Gmelig-Meyling FH, van Booma-Frankfort C, et al. The prognostic value of the antiperinuclear factor, anti-citrullinated peptide antibodies and rheumatoid factor in early rheumatoid arthritis. Clin Exp Rheumatol. 1999;17(6):689-97.
18. van Zeben D, Hazes JM, Zwinderman AH, Cats A, van der Voort EA, Breedveld FC. Clinical significance of rheumatoid factors in early rheumatoid arthritis: results of a follow up study. Ann Rheum Dis. 1992;51(9):1029-35.
19. Jonsson T, Thorsteinsson J, Kolbeinsson A, Jonasdottir E, Sigfusson N, Valdimarsson H. Population study of the importance of rheumatoid factor isotypes in adults. Ann Rheum Dis. 1992;51(7):863-8.
20. 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. The Journal of clinical investigation. 1998;101(1):273-81.
21. Girbal-Neuhauser E, Durieux JJ, Arnaud M, Dalbon P, Sebbag M, Vincent C, et al. The epitopes targeted by the rheumatoid arthritis-associated antifilaggrin autoantibodies are posttranslationally generated on various sites of (pro)filaggrin by deimination of arginine residues. J Immunol. 1999;162(1):585-94.
22. Schellekens GA, Visser H, de Jong BA, van den Hoogen FH, Hazes JM, Breedveld FC, et al. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis Rheum. 2000;43(1):155-63.
23. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, 3rd, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010;62(9):2569-81.
24. Kinloch A, Lundberg K, Wait R, Wegner N, Lim NH, Zendman AJ, et al. Synovial fluid is a site of citrullination of autoantigens in inflammatory arthritis. Arthritis Rheum. 2008;58(8):2287-95.
25. Mahdi H, Fisher BA, Kallberg H, Plant D, Malmstrom V, Ronnelid J, et al. Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. Nat Genet. 2009;41(12):1319-24.
26. Neovius M, Simard JF, Askling J. How large are the productivity losses in contemporary patients with RA, and how soon in relation to diagnosis do they develop? Ann Rheum Dis. 2011;70(6):1010-5.
27. Bellamy N, Duffy D, Martin N, Mathews J. Rheumatoid arthritis in twins: a study of aetiopathogenesis based on the Australian Twin Registry. Ann Rheum Dis. 1992;51(5):588-93.
28. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, et al. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. British journal of rheumatology. 1993;32(10):903-7.
29. Stastny P, Fink CW. HLA-Dw4 in adult and juvenile rheumatoid arthritis. Transplant Proc. 1977;9(4):1863-6.
30. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987;30(11):1205-13.
31. Morel PA, Erlich HA, Fathman CG. A new look at the shared epitope hypothesis. Am J Med. 1988;85(6A):20-2.
32. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum. 2006;54(1):38-46.
33. 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;50(10):3085-92.
34. Debaz H, Olivo A, Vazquez Garcia MN, de la Rosa G, Hernandez A, Lino L, et al. Relevant residues of DRbeta1 third hypervariable region contributing to the expression and to severity of rheumatoid arthritis (RA) in Mexicans. Hum Immunol. 1998;59(5):287-94.
35. Lundstrom E, Kallberg H, Smolnikova M, Ding B, Ronnelid J, Alfredsson L, et al. Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum. 2009;60(4):924-30.
36. Mattey DL, Hassell AB, Plant MJ, Cheung NT, Dawes PT, Jones PW, et al. The influence of HLA-DRB1 alleles encoding the DERAA amino acid motif on radiological outcome in rheumatoid arthritis. Rheumatology (Oxford). 1999;38(12):1221-7.
37. Taneja V, Mehra NK, Kailash S, Anand C, Malaviya AN. Protective & risk DR phenotypes in Asian Indian patients with rheumatoid arthritis. The Indian journal of medical research. 1992;96:16-23.
38. Toussirot E, Auge B, Tiberghien P, Chabod J, Cedoz JP, Wendling D. HLA-DRB1 alleles and shared amino acid sequences in disease susceptibility and

severity in patients from eastern France with rheumatoid arthritis. J Rheumatol. 1999;26(7):1446-51.
39. van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, Houwing-Duistermaat JJ, et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. Arthritis Rheum. 2010;62(5):1236-45.
40. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004;36(4):337-8.
41. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet. 2004;75(2):330-7.
42. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet. 2005;76(4):561-71.
43. Orozco G, Robledo G, Linga Reddy MV, Garcia A, Pascual-Salcedo D, Balsa A, et al. Study of the role of a functional polymorphism of MHC2TA in rheumatoid arthritis in three ethnically different populations. Rheumatology (Oxford). 2006;45(11):1442-4.
44. Wesoly J, van der Helm-van Mil AH, Toes RE, Chokkalingam AP, Carlton VE, Begovich AB, et al. Association of the PTPN22 C1858T single-nucleotide polymorphism with rheumatoid arthritis phenotypes in an inception cohort. Arthritis Rheum. 2005;52(9):2948-50.
45. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet. 2004;75(3):504-7.
46. Skorka A, Bednarczuk T, Bar-Andziak E, Nauman J, Ploski R. Lymphoid tyrosine phosphatase (PTPN22/LYP) variant and Graves' disease in a Polish population: association and gene dose-dependent correlation with age of onset. Clinical endocrinology. 2005;62(6):679-82.
47. Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, et al. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab. 2004;89(11):5862-5.
48. Seldin MF, Shigeta R, Laiho K, Li H, Saila H, Savolainen A, et al. Finnish case-control and family studies support PTPN22 R620W polymorphism as a risk factor in rheumatoid arthritis, but suggest only minimal or no effect in juvenile idiopathic arthritis. Genes Immun. 2005;6(8):720-2.
49. Cloutier JF, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. The Journal of experimental medicine. 1999;189(1):111-21.
50. Gregorieff A, Cloutier JF, Veillette A. Sequence requirements for association of protein-tyrosine phosphatase PEP with the Src homology 3 domain of inhibitory tyrosine protein kinase p50(csk). J Biol Chem. 1998;273(21):13217-22.
51. Vang T, Congia M, Macis MD, Musumeci L, Orru V, Zavattari P, et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet. 2005;37(12):1317-9.
52. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, Buckner JH. Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. J Immunol. 2007;179(7):4704-10.
53. Arechiga AF, Habib T, He Y, Zhang X, Zhang ZY, Funk A, et al. Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. J Immunol. 2009;182(6):3343-7.
54. Zikherman J, Hermiston M, Steiner D, Hasegawa K, Chan A, Weiss A. PTPN22 deficiency cooperates with the CD45 E613R allele to break tolerance on a non-autoimmune background. J Immunol. 2009;182(7):4093-106.

55. Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, Jagodic M, et al. MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. Nat Genet. 2005;37(5):486-94.
56. Eyre S, Bowes J, Spreckley K, Potter C, Ring S, Strachan D, et al. Investigation of the MHC2TA gene, associated with rheumatoid arthritis in a Swedish population, in a UK rheumatoid arthritis cohort. Arthritis Rheum. 2006;54(11):3417-22.
57. Harrison P, Pointon JJ, Farrar C, Harin A, Wordsworth BP. MHC2TA promoter polymorphism (-168*G/A, rs3087456) is not associated with susceptibility to rheumatoid arthritis in British Caucasian rheumatoid arthritis patients. Rheumatology (Oxford). 2007;46(3):409-11.
58. Iikuni N, Ikari K, Momohara S, Tomatsu T, Hara M, Yamanaka H, et al. MHC2TA is associated with rheumatoid arthritis in Japanese patients. Ann Rheum Dis. 2007;66(2):274-5.
59. Martinez A, Sanchez-Lopez M, Varade J, Mas A, Martin MC, de Las Heras V, et al. Role of the MHC2TA gene in autoimmune diseases. Ann Rheum Dis. 2007;66(3):325-9.
60. Bronson PG, Criswell LA, Barcellos LF. The MHC2TA -168A/G polymorphism and risk for rheumatoid arthritis: a meta-analysis of 6861 patients and 9270 controls reveals no evidence for association. Ann Rheum Dis. 2008;67(7):933-6.
61. Bronson PG, Caillier S, Ramsay PP, McCauley JL, Zuvich RL, De Jager PL, et al. CIITA variation in the presence of HLA-DRB1*1501 increases risk for multiple sclerosis. Hum Mol Genet. 2010;19(11):2331-40.
62. Martinez A, Alvarez-Lafuente R, Mas A, Bartolome M, Garcia-Montojo M, de Las Heras V, et al. Environment-gene interaction in multiple sclerosis: human herpesvirus 6 and MHC2TA. Hum Immunol. 2007;68(8):685-9.
63. Zuvich RL, Bush WS, McCauley JL, Beecham AH, De Jager PL, Ivinson AJ, et al. Interrogating the complex role of chromosome 16p13.13 in multiple sclerosis susceptibility: independent genetic signals in the CIITA-CLEC16A-SOCS1 gene complex. Hum Mol Genet. 2011;20(17):3517-24.
64. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010;42(4):295-302.
65. McGovern DP, Gardet A, Torkvist L, Goyette P, Essers J, Taylor KD, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. Nat Genet. 2010;42(4):332-7.
66. Skinningsrud B, Husebye ES, Pearce SH, McDonald DO, Brandal K, Wolff AB, et al. Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. J Clin Endocrinol Metab. 2008;93(9):3310-7.
67. Steimle V, Otten LA, Zufferey M, Mach B. Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). Cell. 1993;75(1):135-46.
68. Chang CH, Flavell RA. Class II transactivator regulates the expression of multiple genes involved in antigen presentation. The Journal of experimental medicine. 1995;181(2):765-7.
69. Reith W, Mach B. The bare lymphocyte syndrome and the regulation of MHC expression. Annu Rev Immunol. 2001;19:331-73.
70. Sims TN, Halloran PF. MHC class II regulation in vivo in the mouse kidney. Microbes and infection / Institut Pasteur. 1999;1(11):903-12.
71. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. EMBO J. 1997;16(10):2851-60.
72. Reith W, LeibundGut-Landmann S, Waldburger JM. Regulation of MHC class II gene expression by the class II transactivator. Nat Rev Immunol. 2005;5(10):793-806.
73. van der Stoep N, Quinten E, Alblas G, Plancke A, van Eggermond MC, Holling TM, et al. Constitutive and IFNgamma-induced activation of MHC2TA promoter type III in human melanoma cell lines is governed by separate regulatory

elements within the PIII upstream regulatory region. Mol Immunol. 2007;44(8):2036-46.
74. LeibundGut-Landmann S, Waldburger JM, Reis e Sousa C, Acha-Orbea H, Reith W. MHC class II expression is differentially regulated in plasmacytoid and conventional dendritic cells. Nat Immunol. 2004;5(9):899-908.
75. Lorentzen JC, Flornes L, Eklow C, Backdahl L, Ribbhammar U, Guo JP, et al. Association of arthritis with a gene complex encoding C-type lectin-like receptors. Arthritis Rheum. 2007;56(8):2620-32.
76. Cambi A, Figdor C. Necrosis: C-type lectins sense cell death. Current biology : CB. 2009;19(9):R375-8.
77. Drickamer K. C-type lectin-like domains. Current opinion in structural biology. 1999;9(5):585-90.
78. Barrow AD, Trowsdale J. You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling. European journal of immunology. 2006;36(7):1646-53.
79. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012;90(1):7-24.
80. Jouben LM, Steele RJ, Bono JV. Orthopaedic manifestations of Lyme disease. Orthopaedic review. 1994;23(5):395-400.
81. Anzilotti C, Merlini G, Pratesi F, Tommasi C, Chimenti D, Migliorini P. Antibodies to viral citrullinated peptide in rheumatoid arthritis. J Rheumatol. 2006;33(4):647-51.
82. Pratesi F, Tommasi C, Anzilotti C, Chimenti D, Migliorini P. Deiminated Epstein-Barr virus nuclear antigen 1 is a target of anti-citrullinated protein antibodies in rheumatoid arthritis. Arthritis Rheum. 2006;54(3):733-41.
83. Dissick A, Redman RS, Jones M, Rangan BV, Reimold A, Griffiths GR, et al. Association of periodontitis with rheumatoid arthritis: a pilot study. Journal of periodontology. 2010;81(2):223-30.
84. Kasser UR, Gleissner C, Dehne F, Michel A, Willershausen-Zonnchen B, Bolten WW. Risk for periodontal disease in patients with longstanding rheumatoid arthritis. Arthritis Rheum. 1997;40(12):2248-51.
85. Mercado FB, Marshall RI, Bartold PM. Inter-relationships between rheumatoid arthritis and periodontal disease. A review. Journal of clinical periodontology. 2003;30(9):761-72.
86. Wegner N, Wait R, Sroka A, Eick S, Nguyen KA, Lundberg K, et al. Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis. Arthritis Rheum. 2010;62(9):2662-72.
87. 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 research & therapy. 2006;8(4):R133.
88. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis. 2003;62(9):835-41.
89. Vessey MP, Villard-Mackintosh L, Yeates D. Oral contraceptives, cigarette smoking and other factors in relation to arthritis. Contraception. 1987;35(5):457-64.
90. Mikuls TR, Sayles H, Yu F, Levan T, Gould KA, Thiele GM, et al. Associations of cigarette smoking with rheumatoid arthritis in African Americans. Arthritis Rheum. 2010;62(12):3560-8.
91. Maxwell JR, Gowers IR, Moore DJ, Wilson AG. Alcohol consumption is inversely associated with risk and severity of rheumatoid arthritis. Rheumatology (Oxford). 2010;49(11):2140-6.
92. Kallberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, et al. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case-control studies. Ann Rheum Dis. 2009;68(2):222-7.

93. 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. 2006;38(2):121-5.
94. Brown JM, Pfau JC, Pershouse MA, Holian A. Silica, apoptosis, and autoimmunity. Journal of immunotoxicology. 2005;1(3):177-87.
95. Caplan A. Certain unusual radiological appearances in the chest of coal-miners suffering from rheumatoid arthritis. Thorax. 1953;8(1):29-37.
96. Stolt P, Yahya A, Bengtsson C, Kallberg H, Ronnelid J, Lundberg I, et al. Silica exposure among male current smokers is associated with a high risk of developing ACPA-positive rheumatoid arthritis. Ann Rheum Dis. 2010;69(6):1072-6.
97. 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. 2005;64(4):582-6.
98. Sverdrup B, Kallberg H, Bengtsson C, Lundberg I, Padyukov L, Alfredsson L, et al. Association between occupational exposure to mineral oil and rheumatoid arthritis: results from the Swedish EIRA case-control study. Arthritis research & therapy. 2005;7(6):R1296-303.
99. Oliver JE, Worthington J, Silman AJ. Genetic epidemiology of rheumatoid arthritis. Curr Opin Rheumatol. 2006;18(2):141-6.
100. Symmons DP. Epidemiology of rheumatoid arthritis: determinants of onset, persistence and outcome. Best practice & research Clinical rheumatology. 2002;16(5):707-22.
101. Pattison DJ, Symmons DP, Young A. Does diet have a role in the aetiology of rheumatoid arthritis? The Proceedings of the Nutrition Society. 2004;63(1):137-43.
102. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004;305(5683):525-8.
103. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949-51.
104. Tautz D. Notes on the definition and nomenclature of tandemly repetitive DNA sequences. Exs. 1993;67:21-8.
105. Jeffreys AJ, Royle NJ, Wilson V, Wong Z. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature. 1988;332(6161):278-81.
106. Kelly R, Gibbs M, Collick A, Jeffreys AJ. Spontaneous mutation at the hypervariable mouse minisatellite locus Ms6-hm: flanking DNA sequence and analysis of germline and early somatic mutation events. Proceedings Biological sciences / The Royal Society. 1991;245(1314):235-45.
107. Robinson WP. Mechanisms leading to uniparental disomy and their clinical consequences. BioEssays : news and reviews in molecular, cellular and developmental biology. 2000;22(5):452-9.
108. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
109. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
110. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931-45.
111. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS biology. 2007;5(10):e254.
112. The International HapMap Project. Nature. 2003;426(6968):789-96.
113. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. Nat Genet. 2008;40(3):340-5.
114. Waldman YY, Tuller T, Keinan A, Ruppin E. Selection for translation efficiency on synonymous polymorphisms in recent human evolution. Genome biology and evolution. 2011;3:749-61.
115. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nature reviews Genetics. 2004;5(7):522-31.

116. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120(1):15-20.
117. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105.
118. Lewontin RC, Kojima K. The Evolutionary Dynamics of Complex Polymorphisms. Evolution. 1960;14:458-72.
119. Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics. 1964;49(1):49-67.
120. Hill WG, Robertson A. Linkage disequilibrium in finite populations. Theor Appl Genet. 1968(38):226-31.
121. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995;29(2):311-22.
122. Lifton RP, Goldberg ML, Karp RW, Hogness DS. The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harbor symposia on quantitative biology. 1978;42 Pt 2:1047-51.
123. Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, et al. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res. 2001;11(5):677-84.
124. Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P. Promoter sequences of eukaryotic protein-coding genes. Science. 1980;209(4463):1406-14.
125. Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science. 1998;281(5373):60-3.
126. Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. Annual review of genetics. 2000;34:77-137.
127. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Current opinion in structural biology. 2004;14(3):283-91.
128. Konarska MM, Grabowski PJ, Padgett RA, Sharp PA. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature. 1985;313(6003):552-7.
129. Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry. 2003;72:291-336.
130. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet. 2002;30(1):13-9.
131. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008;40(12):1413-5.
132. Rothman KJ. Epidemiology : an introduction. New York, N.Y.: Oxford University Press; 2002. viii, 223 p. p.
133. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
134. Rothman KJ. Epidemiology : an introduction. New York: Oxford University Press; 2002. viii, 3 , 223 s. p.
135. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 2002;11(20):2463-8.
136. Bartlett JM, Stirling D. A short history of the polymerase chain reaction. Methods Mol Biol. 2003;226:3-6.
137. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology. 1975;94(3):441-8.
138. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.
139. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry. 1996;242(1):84-9.
140. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques. 2008;44(5):619-26.

141. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nature reviews Genetics. 2011;12(2):87-98.
142. Qi YX, Liu YB, Rong WH. [RNA-Seq and its applications: a new technology for transcriptomics]. Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji. 2011;33(11):1191-202.
143. Akkad DA, Jagiello P, Szyld P, Goedde R, Wieczorek S, Gross WL, et al. Promoter polymorphism rs3087456 in the MHC class II transactivator gene is not associated with susceptibility for selected autoimmune diseases in German patient groups. Int J Immunogenet. 2006;33(1):59-61.
144. Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, et al. PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. Am J Hum Genet. 2005;77(4):567-81.
145. Rodriguez-Rodriguez L, Taib WR, Topless R, Steer S, Gonzalez-Escribano MF, Balsa A, et al. The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in Caucasian case-control samples. Arthritis Rheum. 2011;63(2):365-72.
146. Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM. Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. Blood. 1999;93(6):2013-24.
147. Ensembl. 2011 [cited 2011]; Available from: http://www.ensembl.org.
148. UCSC Genome Browser. 2011 [cited 2011]; Available from: http://genome.ucsc.edu/.
149. Vang T, Miletic AV, Arimura Y, Tautz L, Rickert RC, Mustelin T. Protein tyrosine phosphatases in autoimmunity. Annu Rev Immunol. 2008;26:29-55.
150. Burn GL, Svensson L, Sanchez-Blanco C, Saini M, Cope AP. Why is PTPN22 a good candidate susceptibility gene for autoimmune disease? FEBS letters. 2011;585(23):3689-98.
151. Bronson PG, Goldstein BA, Ramsay PP, Beckman KB, Noble JA, Lane JA, et al. The rs4774 CIITA missense variant is associated with risk of systemic lupus erythematosus. Genes Immun. 2011;12(8):667-71.
152. Fujikado N, Saijo S, Yonezawa T, Shimamori K, Ishii A, Sugai S, et al. Dcir deficiency causes development of autoimmune diseases in mice due to excess expansion of dendritic cells. Nature medicine. 2008;14(2):176-80.
153. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006;38(2):209-13.
154. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.
155. Eklow C, Makrygiannakis D, Backdahl L, Padyukov L, Ulfgren AK, Lorentzen JC, et al. Cellular distribution of the C-type II lectin dendritic cell immunoreceptor (DCIR) and its expression in the rheumatic joint: identification of a subpopulation of DCIR+ T cells. Ann Rheum Dis. 2008;67(12):1742-9.
156. Meyer-Wentrup F, Cambi A, Joosten B, Looman MW, de Vries IJ, Figdor CG, et al. DCIR is endocytosed into human dendritic cells and inhibits TLR8-mediated cytokine production. Journal of leukocyte biology. 2009;85(3):518-25.
157. Roelofs MF, Wenink MH, Brentano F, Abdollahi-Roodsaz S, Oppers-Walgreen B, Barrera P, et al. Type I interferons might form the link between Toll-like receptor (TLR) 3/7 and TLR4-mediated synovial inflammation in rheumatoid arthritis (RA). Ann Rheum Dis. 2009;68(9):1486-93.
158. Aarnisalo J, Treszl A, Svec P, Marttila J, Oling V, Simell O, et al. Reduced CD4+T cell activation in children with type 1 diabetes carrying the PTPN22/Lyp 620Trp variant. Journal of autoimmunity. 2008;31(1):13-21.
159. Arimura Y, Yagi J. Comprehensive expression profiles of genes for protein tyrosine phosphatases in immune cells. Science signaling. 2010;3(137):rs1.
160. Wray NR. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2005;8(2):87-94.