LOOKING FOR ARTHRITIS REGULATING GENES ON MOUSE CHROMOSOME 6 & 14

Marjan Popovic

Stockholm 2008
ABSTRACT

Rheumatoid arthritis (RA) is a chronic autoimmune disease of the joints with a complex aetiology affected by largely unknown genetic and environmental factors. Since ~60% of susceptibility to RA is genetically inherited, one way to progress towards understanding of the disease is to identify the disease regulating genes.

Collagen-induced arthritis (CIA) is the most commonly used model of RA in mice. After immunisation by a subcutaneous injection of collagen emulsified in adjuvant, mice develop a T and B-cell dependent inflammation of the peripheral joints that mimics many characteristics of RA, including production of autoantibodies. Genetic inheritance of CIA has been extensively studied and by now 44 quantitative trait loci (QTL) controlling CIA have been identified. The RA susceptibility genes identified so far point to a central role of T-cell activation along with B cell activation. The central role of the TCR in these processes makes it an appealing candidate and interestingly the previously described Cia6 locus on mouse chromosome 6 is linked to the region encoding the TCR-β chain.

We have mapped seven new CIA loci on chromosome 6 and 14 using congenic strains made between the relatively resistant B10.Q and the highly susceptible DBA/1 strain. The congenics were intercrossed in a partial advanced intercross (PAI) strategy and analysed by linkage analysis, followed by confirmation of found loci in a collection of heterozygous congenic strains with partly overlapping fragments. Using this strategy, we identified seven new arthritis regulating loci including two loci corresponding to the two TCR-chains. Among these loci we found evidence of both epistatic effects and sex-specificity.

Also, we demonstrated the usefulness of the recent advancement in mapping such as genetically heterogeneous stock cross (HS) in the fine mapping of the known Cia36 locus using different lymphocyte sub-phenotypes, that were all associated with the markers within the Slc38a1 gene, as was previously shown for arthritis.

In conclusion, this thesis illustrates the capacity and importance of different mapping strategies in order to identify underlying complex set of loci in an animal model.
LIST OF PUBLICATIONS

I. Identification of New Loci Controlling Collagen-Induced Arthritis in Mouse using a Partial Advanced Intercross and Congenic Strains


II. Phenotyping and Induction of Collagen-Induced Arthritis in a Heterogeneous Stock Inbred-Outbred Cross

Emma Ahlqvist, Therese Lindvall, Malin Hultqvist, Michael Foerster, Kutty Selva Nandakumar, Marjan Popovic, Ivanka Teneva, Dorota Kolodziej and Rikard Holmdahl. Manuscript
TABLE OF CONTENTS

1 Introduction.........................................................................................................................1
2 Rheumatoid Arthritis...........................................................................................................2
  2.1 Aetiology of RA...............................................................................................................2
  2.2 Environmental Risk Factors.........................................................................................3
  2.3 Genetics of RA................................................................................................................4
    2.3.1 Mapping the Susceptibility Genes...........................................................................5
    2.3.2 Genetic Linkage Analysis in RA..............................................................................5
    2.3.3 Genetic Association Studies in RA..........................................................................7
3 Statistics in QTL Mapping.................................................................................................9
  3.1 Single Marker Analysis..................................................................................................9
  3.2 Interval Mapping...........................................................................................................10
    3.2.1 Significance Level....................................................................................................11
4 Animal strains as a mapping tool....................................................................................12
5 Collagen Induced Arthritis-a Model of RA.....................................................................16
  5.1 Looking for Arthritis Regulating Genes on Mouse Chromosome 6 & 14.................17
6 Present investigations.......................................................................................................20
7 Acknowledgements.........................................................................................................22
8 References.......................................................................................................................24
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ab</td>
<td>Antibody</td>
</tr>
<tr>
<td>Ag</td>
<td>Antigen</td>
</tr>
<tr>
<td>AIL</td>
<td>Advanced intercross line</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen presenting cell</td>
</tr>
<tr>
<td>AS</td>
<td>Association study</td>
</tr>
<tr>
<td>BCR</td>
<td>B cell receptor</td>
</tr>
<tr>
<td>Bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>B10.Q</td>
<td>C57BL10 strain with H2(^a) haplotype from DBA/1 at the MHC II locus</td>
</tr>
<tr>
<td>CCP</td>
<td>Cyclic citrullinated peptide</td>
</tr>
<tr>
<td>CFA</td>
<td>Complete Freund’s adjuvant</td>
</tr>
<tr>
<td>CIA</td>
<td>Collagen induced arthritis</td>
</tr>
<tr>
<td>CII</td>
<td>Collagen type II</td>
</tr>
<tr>
<td>cM</td>
<td>Centimorgan</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>Cytotoxic T lymphocyte-associated Antigen 4</td>
</tr>
<tr>
<td>HLA</td>
<td>Human leukocyte antigen</td>
</tr>
<tr>
<td>HS</td>
<td>Heterogeneous stock</td>
</tr>
<tr>
<td>IFA</td>
<td>Incomplete Freund’s adjuvant</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>LA</td>
<td>Linkage analysis</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>LOD</td>
<td>Logarithm of odds</td>
</tr>
<tr>
<td>Mb</td>
<td>Mega base pairs</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>Ncf1</td>
<td>Neutrophil cytosolic factor 1</td>
</tr>
<tr>
<td>PAI</td>
<td>Partial advanced intercross</td>
</tr>
<tr>
<td>PTPN22</td>
<td>Protein tyrosine phosphatase non-receptor type 22</td>
</tr>
<tr>
<td>QTL</td>
<td>Quantitative trait locus</td>
</tr>
<tr>
<td>RA</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>RF</td>
<td>Rheumatoid factor/Recombination frequency</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>RI</td>
<td>Recombinant inbred</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen species</td>
</tr>
<tr>
<td>SE</td>
<td>Shared epitope</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>Th</td>
<td>T helper cell</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll like receptor</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>Treg</td>
<td>Regulatory T cell</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Our immune system is a remarkable defence system evolved to protect us from enormous variety of foreign invaders as well as from different pathological changes occurring in our own cells. Normally, the immune system is in the state of self-tolerance in which it does not attack the normal self-structures of the body. However, a breakdown or failure of the mechanisms of self-tolerance leads to autoimmunity, characterized by overactive immune response and uncontrolled inflammation.

Mostly with unknown aetiology and effective treatment, common autoimmune diseases such as rheumatoid arthritis, multiple sclerosis, diabetes mellitus type I, systemic lupus erythematosus, Grave’s disease, and others affect approximately 5% of the population [1, 2]. As a contrast to monogenic diseases which exhibit classic Mendelian inheritance and are attributable to a single gene with a strong phenotypic effect, those diseases are called complex since they are result of a complex interplay between multiple genetic, environmental and stochastic components[3]. Troubled by genetic and phenotypic heterogeneity, epistatic interactions and environmental influence, in spite of considerable effort to reveal disease-regulating genes, only very few have been identified[4, 5].

This thesis illustrates dissection and identification of several new quantitative trait loci (QTL), which regulate a complex disease, arthritis. In the current project, studies have been done to further analyze and identify arthritis modulating QTLs using collagen induced arthritis (CIA), an animal model for rheumatoid arthritis (RA), specifically emphasizing on mouse chromosome 6 and 14. Future use of recent mapping strategies, such as advanced intercross lines (AIL), recombinant inbred lines (RIL) and heterogeneous stock mice (HS) will hopefully lead to identification of CIA- and subsequently RA-prone low-to-moderate risk genes.
2 RHEUMATOID ARTHRITIS

Human rheumatoid arthritis (RA) is a chronic inflammatory disease of peripheral joints, characterized by hyperplasia of synovial membrane and structural damage of the cartilage and bones. It can lead to a long-term joint damage, resulting in chronic pain, loss of function and disability. Clinical manifestations may vary from mild to very severe symptoms, and since it is a systemic disease, meaning it can affect other organs in the body, RA patients may display different extraarticular manifestations, some of which are: subcutaneous rheumatoid nodules, vasculitis, fibrosis of the lungs, pericarditis, kidney disease, dry eyes, Felty’s syndrome, weakness, low-grade fever, depression, fatigue, loss of appetite, and anemia [6]. The effects of the disease can vary from person to person, and there is a growing belief that RA is not a single disease but several different diseases sharing common features. The American Rheumatism Association revised criteria for the classification of RA in the form of seven new disease criteria[7]:

1) Morning stiffness in and around joints lasting at least 1 hour before maximal improvement
2) Soft tissue swelling (arthritis) of at least 3 or more joint areas observed by a physician
3) Arthritis of hand joints: swelling (arthritis) of the proximal interphalangeal, metacarpophalangeal, or wrist joints
4) Symmetric swelling (arthritis): simultaneous involvement of the same joint areas on both sides of the body
5) Rheumatoid nodules (subcutaneous)
6) The presence of serum rheumatoid factor in abnormal amounts and
7) Radiographic erosions and/or periarticular osteopenia in hand and/or wrist joints, which must include erosions or bony decalcification localized to or most adjacent to the involved joints.

Thus, RA is defined by the presence of 4 or more criteria, where criteria 1 through 4 must have been present for at least 6 weeks[7]. It can start in any joint, but it most commonly begins in the smaller joints of the fingers, hands and wrists. The first stage of RA is swelling of the synovial lining, causing pain, warmth, stiffness, redness and swelling around the joint, due to infiltration of lymphocytes and granulocytes. Followed by rapid proliferation and growth of synovial fibroblasts and macrophages together with other cells leading to formation of pannus that causes synovium to thicken[8, 9]. In the late stage, the inflamed cells release enzymes that may digest bone and cartilage, often causing the involved joint to lose its shape and alignment, more pain, and loss of movement. Therefore early diagnosis and treatment of RA is critical to maintain a productive lifestyle and can limit joint damage.

2.1 AETIOLOGY OF RA

What is the exact cause of RA currently is unknown. Although there is a considerable variation of the disease occurrence among different populations, the incidence of RA in the worldwide population is approximately 1% and it is higher in women than in men
with a ratio of 3:1 [10, 11]. South European countries had a lower incidence of RA compared to north European and north American countries, and interestingly RA was not found in some rural African populations [10, 12]. It can affect anyone, including children (Juvenile RA), but the onset usually occurs between 30 and 50 years of age, with a peak in the fifth decade of life [11]. Genetic factors have a substantial contribution to RA in the population, accounting for ~60% [13]. The concordance rate is ~15% for monozygotic twins and ~4% for dizygotic twins, clearly indicating importance of the genetic predisposition [14]. The higher occurrence of RA in females compared to males suggests a possible role of hormones in disease susceptibility. Women arthritic patients go into remission during pregnancy, but disease symptoms can be more severe postpartum. This might be due to high levels of estrogens that shift the pro-inflammatory Th1/Th2 balance towards anti-inflammatory Th2-mediated immunologic responses [15-17].

For many years rheumatoid factor (RF) has been commonly used as a biomarker in laboratory diagnosis of RA patients. It is one of the most known RA autoantibodies that reacts with an epitope on the Fc region of IgG, and can be of various isotypes (IgG-RF, IgA-RF, IgM-RF). RF has been shown to have high sensitivity but lower specificity for RA as a diagnostic tool, when compared to anti-cyclic citrullinated peptide (CCP) antibodies [18]. IgM-RF, the most commonly measured RF isotype in diagnostic laboratories, showed higher sensitivity (66.4%) compared to CCP (64.4%), but the highest specificity was achieved by CCP (97.1%), compared to IgM-RF (82.1%). Citrullination is a post-translational modification of arginine into citrulline in a protein, during cell-death and tissue inflammation. Some of the citrullinated proteins described so far in RA synovium are fibrin and vimentin [19, 20]. It is important to note that not all patients with RA have elevated RF, and not all the people with elevated RF have RA. Instead, the concomitant presence of RF and anti-citrullinated protein antibodies (ACPA) at disease onset was shown to be the best association to predict RA diagnosis [21].

### 2.2 ENVIRONMENTAL RISK FACTORS

Among different environmental factors implicated in RA, infectious agents are suggested to play an important role in triggering or increasing susceptibility to RA. Several infectious agents have been reported to be risk factors for RA, including human parvovirus B19, Epstein–Barr virus, retroviruses, alphaviruses, hepatitis B virus, Mycobacterium tuberculosis, Escherichia coli, Proteus mirabilis and Mycoplasma [22-24]. If the tolerance to self-antigens is broken through molecular mimicry (because of the structural similarity of self-antigens to microbial antigens) it is possible that microorganisms, which may originally be responsible for the immune response might be cleared from the body completely, leaving an autoimmune response where newly targeted self-antigens are in constant supply, thus keeping on the formation of immune complexes indefinitely [25]. Interestingly, no conclusive evidence for a specific viral or bacterial infection is found so far, not excluding that combination of different microorganisms could be responsible for initiation of RA in patients with genetic predisposition.
Smoking appears to be an environmental factor of major importance as a risk factor for RA. It has been shown that disease risk of RF-seropositive RA, associated with one of the classic genetic risk factors for immune-mediated diseases (shared epitope of HLA–DR), is strongly influenced by smoking [26]. Even stronger interaction was observed between smoking and double SE genes (combination of any two HLA–DRB1 genes: DRB1*01, DRB1*04, and DRB1*10) [26], however no interaction was seen between smoking and PTPN22 R620W[27].

Another illustration of the potential role of lifestyle in determining the risk of developing RA is alcohol consumption with a recent demonstration of a preventive effect of alcohol[28].

Figure 1. A very simplified illustration of a complex interplay between multiple genetic and environmental components required for development of a complex disease.

### 2.3 GENETICS OF RA

A genomic region involved in regulation of a continuously variable trait (i.e. complex disease) is called a quantitative trait locus (QTL). For complex diseases such as RA no single gene is sufficient for disease development, instead multiple genes with weak phenotypic effects and incomplete penetrance. Different combinations of susceptibility genes, together with environmental factors and gene interactions will result in different magnitudes of disease, once a certain threshold is reached (Figure 1).

The first identified and most consistently characterised genetic risk factor contributing to RA across all populations was within human leukocyte antigen (HLA) gene region on chromosome 6 [29]. Different HLA-DRB1 alleles, such as DR1 (DRB1*0101), DR4 (DRB1*0401, DRB1*0404) and DR10 (DRB1*1001), which are involved in antigen (Ag) presentation, are all associated with both disease susceptibility and disease
severity[30-32]. All of these alleles share the same conserved five amino acid sequence in the third hypervariable region of class II DRβ1 chain, so called shared epitope (SE). Since this sequence occurs in the Ag binding site it may influence the efficiency of Ag presentation. However, in total HLA region provides almost 40% of the genetic contribution to RA[33]. An interesting observation is that HLA–DRB1 alleles containing the SE are strongly associated with susceptibility to RA in African Americans, where the finding of a higher degree of European ancestry among African Americans with SE alleles suggests that a genetic risk factor for RA was introduced into the African American population through admixture [34].

In addition to HLA, as the single most important genetic factor, there is a number of other susceptibility genes identified to contribute to RA development, which are described bellow.

2.3.1 Mapping the Susceptibility Genes

Identification of genetic factors contributing to RA is of great importance since it would give invaluable clues to the largely unknown pathogenic mechanisms of RA, as well as for the improvement of diagnostic and prognostic markers and development of new therapeutic targets. Basically genetic mapping aims to discover how often two loci are separated by meiotic recombination, where the recombination frequency (RF, θ) is a measure of the distance between the two loci. Thus the two loci in between where there is 1% chance of being split by a recombination are said to be 1 centimorgan (cM) apart on a genetic (linkage) map. Determining the position of loci relative to one another RF defines genetic distance, which is not the same as physical, and the goal is to convert a genetic map into a physical map of precise chromosomal position measured in base pairs (bp). Although non-intuitive, RF never exceeds 0.5 (50% recombinants) however far apart the loci are [35].

It is possible to localize genome regions in which the loci lie and to estimate how much of the total variation is accounted for by QTL variation in each region. This analysis is done by intercrossing or backcrossing two strains that clearly differ in the quantitative trait and differ in alleles at known genetic markers. A marker that is unlinked to a QTL will have the same mean value of the quantitative trait phenotype for all its genotypes, whereas the one that is closely linked to a QTL will differ in its mean phenotype from one of its genotypes to another. How different the mean phenotype is between the marker genotypes depends both on strength of the QTL effect and how linked the marker and the locus are[36].

There are two main strategies to identify genetic factors contributing to a complex trait such as RA: linkage analysis and association studies (Figure 2).

2.3.2 Genetic Linkage Analysis in RA

Linkage analysis (LA) aims to identify genomic regions containing RA susceptibility genes by observing coinheritance of the disease and genetic markers within families. It is expected that affected relatives (for example, affected sibling pairs) will share haplotypes identical by descent (IBD) in chromosomal regions containing disease predisposing genes[37]. Contrary to Mendelian disorders, for complex diseases
nonparametric (model free) linkage analysis has been developed, due to absence of precise genetic model, mode of inheritance, gene frequencies and penetrance of each genotype[35]. The closer the marker is to a putative polymorphism, the lower the chances that they will be split due to recombinations during meiosis. That means the higher chances that it will be co-inherited with the disease susceptibility gene, and thus be shared among affected siblings. In this case two linked alleles which are inherited together in the same haplotype more often than expected are said to be in linkage disequilibrium (LD)[38]. In a linkage analysis genome can be genotyped partially or completely (whole genome scan) using microsatellite (di-, tri- or tetrancleotide repeats) or single nucleotide polymorphism (SNP) (biallelic) genetic markers[39]. A genetic marker is any polymorphic DNA sequence that allows us keeping track of a specific genomic region. Nowadays there is a possibility to perform linkage studies in large crosses of inbred animal strains, thereby mimicking large families in humans. Unfortunately, the small number of recombinations per single meioses, i.e. poor resolution of the crosses, is usually insufficient to separate closely positioned loci in a linkage analysis[40].

Figure 2. A schematic view of the two main strategies for mapping QTLs.
Several RA genome-wide linkage studies have been carried out, but apart from the well-established HLA susceptibility locus (6p21), very few of the reported significant regions of linkage have been replicated (1p13, 1q43, 6q21, 16p, 18q21)[31]. Among these only one candidate gene mapping within 1p13 region was identified, the PTPN22 gene. Therefore, it is still possible to combine linkage analyses from different studies in meta-analysis that provide greater power to identify regions showing only weak evidence for linkage in individual studies[41].

2.3.3 Genetic Association Studies in RA

Association studies (AS) compare the distribution of certain genetic markers between a well-defined, unrelated group of affected individuals (cases) and a healthy group (controls) from the same population. Marker alleles with a higher frequency in the group of patients will be identified as disease associated. Association differ from linkage in that the same allele (or alleles) is associated with the trait across the whole population, while linkage allows different alleles to be associated with the trait in different families[42]. Genetic AS arise because human populations share common ancestry and it can be seen as just a special form of LA in which the extended family is the wider population. Taking advantage of numerous recombinations collected in the population, association studies have much higher power than linkage analysis to fine map regions displaying small effect size, but requires genotyping of many more markers over the region of interest. However, to avoid false positive associations and population stratification a very careful study design is required[42].

Association studies are used to test candidate genes in so-called candidate gene association studies where SNPs are analysed in candidate genes or regions, which can be selected based on their implications in disease pathogenic mechanisms or based on their location in the regions that have previously displayed linkage to disease[43]. In a candidate gene association study interesting SNPs found within or in a close proximity of a candidate gene are tested for association. Large areas of the genome were suggested to be divided into haplotype blocks characterized by small LD between blocks and limited haplotype diversity within blocks[44]. The so-called tag SNPs ensure that most of variation of particular genomic region will be captured to uniquely identify those haplotypes, and the effects of unassayed SNPs are detected through linkage disequilibrium with tag SNPs[39, 45]. The obvious candidate genetic polymorphisms are the ones found in animal models, or related human diseases, which are altering protein structure and function, but also non-coding polymorphisms in regulatory regions of genes.

There have been several important candidate gene association studies in RA outside the HLA region. An example of a candidate gene first associated with one disease-type I diabetes, and later with multiple autoimmune diseases, is protein tyrosine phosphatase non-receptor 22 gene (PTPN22), the second largest genetic risk to development of RA after HLA-DRB1 gene [46, 47]. PTPN22 gene encodes intracellular PTP lymphoid-specific phosphatase (LYP), which inhibits T cell activation through suppression of T cell receptor signalling. A non-synonymous SNP (R620W), arginine substituted by
tryptophan, is the causing polymorphism also associated with RA and SLE[48, 49]. This mutation, located in the motif which otherwise enables physical binding of LYP with the negative regulatory kinase Csk, reduces its binding affinity, and probably the amount of LYP-Csk complexes, thus allowing hyper-reactive T cells to take part in a destructive autoimmune response against auto-Ag after viral infection or other immune stimulation[46].

Another negative regulator of T-cell activation, cytotoxic T lymphocyte-associated antigen-4 (CTLA-4), which was also associated with common autoimmune diseases, such as Grave’s disease, autoimmune hypothyroidism, TID[50], has association with RA in number of studies and is found to be stronger in RA patients who are seropositive for anti-citrulline antibodies [51].

Autoantibodies to cyclic citrullinated peptides (CCP) are highly specific and present in the majority of RA patients[52], therefore peptidyl arginine deiminase type IV (PADI4), which is involved in posttranslational modification of arginine into citrulline, is an excellent candidate gene for RA pathogenesis, and has been replicated in different population based association studies[51, 53].

A recent study investigating a chromosome 2q region under the peak of linkage identified association for both RA and SLE with a SNP within signal transducer and activator of transcription 4 gene (STAT4), suggesting a shared pathway for these illnesses[54]. The association between the STAT4 polymorphism and RA has been confirmed in different populations[55].

In addition, several other genes have been proposed as candidates for RA, such as macrophage migration inhibitory factor (MIF) associated with high levels of radiological joint damage[56], solute carrier family 22, member 4 gene (SLC22A4) and Fc receptor-like 3 gene (FCRL3) replicated in Japanese population[57, 58], and transcription factor RUNX1 gene[59]. However, associations have been difficult to reproduce not only in ethnically different populations, but also within the same ones[53].

Advances in technology enabling hundreds of thousands of SNPs to be affordably genotyped in large samples, allowing for good coverage of common variation in the human genome, has made hypothesis-free genome wide association studies (GWAS) a powerful approach to the susceptibility gene identification. A recent genome wide study identified a common genetic variant in the TNF receptor associated factor 1 and complement component 5 locus (TRAF1-C5) to be associated with increased risk of anti-CCP-positive RA[60]. A huge GWAS performed by The Welcome Trust Case Control Consortium revealed, beside known HLA-DRB1 and PTPN22, an additional strong association signal for RA on chromosome 7, with a pronounced sex difference[61].
3 STATISTICS IN QTL MAPPING

The aim of any method for QTL mapping is to identify genomic regions for which there is an association between variations of a phenotype on the one hand, and one or more genetic markers on the other. It is important here to understand some basic statistical principles used in the mapping methods, and to distinguish between: pointwise p value, genome-wide p value, and logarithm of the odds (LOD) score. Pointwise p value is the probability of exceeding the observed LOD score value at the specified position in the genome assuming the null hypothesis of no linkage, whereas genome-wide p value is the probability that the observed LOD score value will be exceeded anywhere in the genome if there is no linkage[35]. The ratio of the likelihood of QTL existence at the specific genome position and the likelihood of no QTL gives the odds of the linkage, and the logarithm of the odds is the LOD score. For example, a LOD score of 5 means that it is $10^5$ times more likely that there is a true linkage than the null hypothesis of no linkage in the given data, whereas p value of $10^{-5}$ means that the stated LOD score will be exceeded only once in $10^5$ times if there is no true linkage.

Linkage is usually reported as a LOD score, which can be presented as a function of recombination frequency ($\theta$) between marker and disease locus or map position of the disease locus with respect to a fixed map of markers[37]. The objective of the analysis is to estimate $\theta$ between individual markers and the disease locus (two-point) or position of the disease locus relative to a fixed map of markers where the location of each marker is known (multipoint). The null hypothesis represents no linkage between disease and marker locus ($\theta=0.5$), whereas the alternative hypothesis assumes linkage exists ($\theta<0.5$), thus the LOD score function can be defined as:

$$LOD(\theta) = \log_{10}\left(\frac{\text{Like}(\theta)}{\text{Like}(\theta = 0.5)}\right)$$

The LOD score function is maximized with respect to $\theta$ in two-point analysis (a single marker and disease locus), or map position in multipoint analysis (disease locus and at least two markers at fixed relative position). The value of $\theta$ which gives the maximum LOD score is the maximum likelihood estimate of $\theta$.

3.1 SINGLE MARKER ANALYSIS

The simplest method for QTL mapping is a single marker analysis using t-test (in a backcross for two possible genotypes at a marker), or analysis of variance (ANOVA, sometimes called “marker regression”)(in other types of crosses such as intercross, where there are more than two possible genotypes) at a marker. At each genotyped marker, individuals are split into groups according to their genotypes at the marker, and the phenotype mean values and distributions of the groups are compared[62]. If they differ there is an indication that the marker is linked to a QTL. The main advantage of ANOVA at the marker loci is that genetic map of the markers is not required, and it can easily be extended to account for multiple loci. However, it has some important weaknesses: QTL location is indicated only by looking at which marker gives the greatest phenotype difference between genotype groups, all individuals with missing genotypes at the marker must be discarded, and if the markers are widely spaced the QTL may be quite far from the marker thus decreasing the power for QTL detection.
3.2 INTERVAL MAPPING

Interval mapping (IM) overcomes problems caused by limited informativeness of markers. Some meioses might be informative for one marker, but other meioses not for that one, instead for some other nearby marker. In that case only linkage analysis of the disease with both markers extracts the full information. The marker map is taken as given and the aim is to locate the disease gene in one of the intervals of the framework of markers[35]. Like ANOVA IM assumes the presence of a single QTL. Taking into account genotypes at the two nearest flanking markers method calculates the probability of a position between them to be of a certain genotype. This method uses genetic map of the typed markers and each location in the genome is postulated, one at a time, as the true location of the QTL, by calculating its LOD score. As a result from all LOD scores, which are the strength of evidence for the presence of a QTL, LOD curve over the marker map is obtained. Confidence intervals which indicates the most plausible interval location for the QTL are hard to deduce analytically but can be calculated[63], and a widely accepted support interval extends to recombination fractions at which the LOD score is 1 unit below the peak value (the lod-1 rule)[35, 64], or even LOD drop of 1.5 [62], or 2 units[65].

IM has several advantages as it estimates QTL position in between markers, thus providing more precise localization, and overcomes the incomplete marker genotype data, so if an individual is missing the marker genotype for a flanking marker, it moves to the next flanking marker for which the genotype is known. This method was used in the first paper of the thesis for all the missing genotypes.

The chance of detecting a QTL is called the “power”. So, the power to detect modest-risk alleles depends on the expected size of the effect of the QTL which can be defined as the percent of the total phenotypic variance in the final generation that would be observed if there were no environmental variation, measurement error, or other QTL [62, 66]. Large relative risks can lead to large differences in allele frequency between cases and controls, whereas small relative risks (for modest-risk alleles) produce smaller differences [39]. Further, statistical power of genetic linkage analysis is strongly determined by: sample size, marker density, type of cross, genetic, phenotypic and environmental heterogeneity of the populations, stringency of the chosen LOD threshold, genotype errors and the missing data. [37, 62, 67].

IM can be used to identify multiple QTLs, and mapping multiple QTLs simultaneously has several advantages: greater power to detect QTLs, greater ability to separate linked QTLs, and the ability to identify interactions between QTLs[62]. And intra- and inter-chromosomal genetic interactions have been shown to confer significant contribution to the variation of a complex disease such as CIA[68]. In an additive interaction the final phenotypic effect is the sum of effects of at least two loci where the effect of QTL1 is the same, irrespective of the genotype at QTL2, but if there is a significant deviation from this additive effect when compared to full effect it can be attributed to epistatic interactions[62, 69]. Although today more widely used, originally epistasis was defined as an interaction where one locus completely masks the effect of another. The simplest
method for multiple QTL mapping is multiple regression, which is an extension of ANOVA.

3.2.1 Significance Level

Large number of significance tests are difficult to interpret because doing tests for long enough is more likely to give some significant associations than not. Association studies, and in particular GWAS which are using high number of genetic markers, have as a major concern setting of an adequate significance level, since there is a positive correlation between the number of tests performed (markers tested in the study) and false positive associations[70]. Therefore, because each test performed carries an independent risk of a false positive association, methods that are correcting for multiple testing should be applied, and one commonly used is Bonferroni correction: $\alpha_{corrected} = \alpha / n$ where $\alpha$ is conventional significance level of 0.05 and n is a number of independent tests (potential associations) performed[71]. However, closely linked markers in linkage disequilibrium with each other may not be independent and Bonferroni correction is likely to be too conservative in the current dense SNP GWAS which consider even 500,000 or 1 million SNPs, therefore suggesting for other approaches[70, 72].

Extensive genome scan simulations have been performed to establish criteria for significant linkage in the whole genome scans[73]. Assuming a very dense map of fully informative markers mathematical theory was used to establish threshold of significance required for a LOD score to achieve genome-wide significance of 5%[74]. Thus linkage is suggestive if it occurs one time by chance in a genome scan, and it’s significant if it occurs 0.05 times in a genome scan (meaning that 5% of the time significant linkages will turn out to be false positives, i.e. once in 20 genome scans). Therefore, suggested LOD score thresholds for mapping complex trait loci in an F2 intercross are 1.9 for suggestive and 3.3 for significant linkage, but if both additive and dominance components are estimated in an intercross rather stringent ones of 2.8 for suggestive and 4.3 for significant linkage are proposed[74]. However, the use of very stringent significance levels leads to a loss of power and increases the rate of false negatives[72].

There is an empirical method, based on permutation test, for determining an appropriate threshold value for significance that is specific for the given experimental data set. This empirical method estimates how often a certain p value (evidence of association/linkage) would arise by chance in the given data set if there were no true associations or linkages. The data is randomly shuffled i.e. each trait value is reassigned to a new individual by computing permutations, thus destroying any association between the trait values and the analysis points (genotyped markers). The resulting maximum LOD score at each analysis point is recorded in each simulated dataset and the entire procedure (shuffling and analysing) is repeated N times[75]. However, replication of the results from genetic association study in an independent sample is always the preferred confirmation of the true association[43].
4 ANIMAL STRAINS AS A MAPPING TOOL

Sequencing and comparative analysis of mouse and human genomes indicate that approximately 80% of mouse genes have a direct 1:1 ortholog in the human genome[76, 77]. The availability of commonly used laboratory inbred strains of mice has provided a powerful tool to map QTLs that underlie heritable phenotypic variation, where this large level of human/mouse conservation is very likely to give important insights of the corresponding human susceptibility genes and which subgroups of patients are relevant for further investigation. Inbred animal strains, in which relevant models for human diseases can be studied, dramatically raise the power to identify susceptibility genes for several reasons[67]:

I. Mouse and rat inbred strains have unique, well defined, fixed genotype that greatly reduces genetic heterogeneity, and genetic segregation can be followed making linkage analysis (LA) informative.

II. Environmental influence can be controlled.

III. Genes can be isolated by selective breeding, positional cloning and transgenic technology.

IV. Sample size is “unlimited” and sampling is greatly simplified.

V. Gene interactions can be studied through crossbreeding of different congenic strains.

VI. Genome can be modified, making it possible to test the function of specific gene variants (e.g. knock-in, knock-out).

VII. Animal studies of complex diseases enable mapping of small effect genes that are hardly detectable in human studies.

Inbred strains are animals that are nearly genetically identical to each other due to long inbreeding. Mating of brother-sister pairs for at least 20 generations will result in lines that are roughly 98% genetically identical, which is usually sufficient to be considered as an inbred strain (compared to identical twins or clones which are 100% genetically identical). Up to date a total of 478 inbred strains of mice are described (www.informatics.jax.org). Inbred strains are widely used in laboratories for experiments, where for reproducibility of conclusions, all the test animals should be as similar as possible. In addition, inbred strains can be used in different breeding schemes to obtain genetically very useful populations of animals that can be utilized for mapping of genes (Figure 3).

To get the first inkling where the genes are on the chromosomes, it is important to establish a genome-wide evidence for linkage or association usually performing an F2 intercross or N2 backcross[78]. In this primary genome screens the size of QTL is usually from 10 to 30 cM containing a few hundreds of genes, which is too many candidates to start functional evaluation of each gene. Therefore, the next step would be fine-mapping of the region aiming to reduce the size of QTL as much as possible, preferably to less than 1 cM[79]. And this can be done with the use of different high-resolution crosses and congenic strains.
A congenic strain is an inbred strain where a desired chromosomal segment is introduced to a given inbred genetic background by repetitive backcrossing of donor region to the recipient strain for at least 10 generations, followed by intercross[80]. As this appears to be a long way to remove most of the contaminating donor genome outside the region of interest, it is suggested that use of “speed congenic” strategy can decrease this time to half. Availability of numerous polymorphic DNA markers allows for marker-assisted breeding schemes based, beside selection of animals that contain region of interest, on additional selection of animals with the lowest percent of contaminating donor genome. The optimal breeding strategy appears to be a 10cM genome scan of 50 or fewer progeny per backcross generation and congenic strain can be produced as early as the fifth backcross generation[80]. Congenic strains made from relatively resistant C57Bl/10Q (B10.Q) (prof. Jan Klein, Tubingen University, Germany) and highly susceptible DBA/1 strains (Jackson laboratory) have been used in the first paper of this thesis.

It has been shown that congenic approach is a good choice to narrow down a QTL to a level of a single or a few genes, as seen in positional identification of Ncf1 in rats[81]. However, different problems may occur in practice: a) as we narrow down a QTL, finding new polymorphic markers may be difficult as well as obtaining new recombination events thus requiring increased breeding population, b) phenotype used to keep track of a QTL may become insufficient to continue further due to e.g. disruption of epistatic interactions with other QTLs during congenization or fine mapping, c) original size of the QTL effect may be too small to be detected, d) initial linkage was actually a false positive.

A number of different experimental crosses are available today allowing for high resolution mapping while avoiding large increase in tested population size. Advanced intercross lines (AIL) are made by random intercross of two inbred strains for many generations, or by selecting the least related pairs for breeding at each generation in order to minimize fixation of loci within families[82, 83]. This approach generates genetically unique offspring where accumulation of recombination events over a number of generations allows for more precise QTL mapping. An AIL variant called partial advanced intercross (PAI) was engaged in the first paper. PAI is made similarly as AIL just starting from one or several congenic strains that are intercrossed in subsequent generations[84]. This strategy is particularly useful when there is a previous knowledge about broader QTL position in the genome (e.g. from an F2 crosses) that are to be dissected further in the fine-mapping or to study possible interactions between different congenics, as in the first paper here. The aim of PAI would be to produce all possible combinations of genotypes at the location of the initial linkage peaks.

An alternative method for fine mapping of small-effect QTL that uses outbred mice of known ancestry has been developed. It uses genetically heterogeneous stock (HS) of mice descended from many inbred progenitor strains, and in the second paper we used Northport HS made from eight inbred strains A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, CBA/J, DBA/2J and LP/J[85, 86]. Because the stock has been on intercross for 55 generations, using the circular breeding design (i.e., a male from family 1 is bred to a female from family 2, and so on) each chromosome is a fine-grained mosaic of the progenitor strains. The average distance between recombinants is small (less than 2
cM) so that the HS provide high resolution mapping of multiple QTLs across the genome[87]. However, high density of recombinations demands high number of markers and adequate analytical methods. One such method has been developed to incorporate information from the flanking markers and the progenitor haplotypes, in order to determine the probability of each founder strain being the ancestor of a given allele in the HS[66].

Beside these kinds of experimental crosses and strains, there is a possibility to use several other as a mapping tool, such as recombinant inbred strains, chromosome substitution strains, and collaborative cross [88, 89].
Figure 3. A scheme of some of the experimental crosses used for mapping of genes.
5 COLLAGEN INDUCED ARTHRITIS-A MODEL OF RA

Animal models of rheumatoid arthritis (RA) provide unique opportunity to investigate similar disease pathways operating during different stages in development of human RA, and although maybe not the same, some might be shared at least in part, and thus can give valuable clues about the targets in human disease. Collagen induced arthritis (CIA) is the most commonly used model of RA in mice, induced by subcutaneous injection of native (triple helical) heterologous or autologous collagen type II (CII) usually in complete Freund’s adjuvant[90].

This model mimics human disease in different clinical, histological and immunological aspects. The major epitopes of CII, which is the predominant protein of articular cartilage, recognized by autoantibodies and T cells are shared between mice and humans[91, 92]. The arthritis associated MHC class II molecules in both mice and humans bind and present to T cells the immunodominant epitope CII260-270 which is glycosylated in the healthy joint cartilage, but less in arthritic one, suggesting an important role in changes of the posttranslational modifications of CII in development of RA and CIA[92]. Both diseases depend on particular susceptibility and protective alleles within MHC class II. The most susceptible strains for CIA are those expressing H-2^q and H-2^r haplotypes[93]. The specific MHC class II gene that is coding for Aq molecule, responsible for susceptibility to CIA in mice, whose peptide-binding pocket is very similar to that of DR4 and DR1 molecules (i.e. shared epitope), has been identified[94]. The relevance of CIA to human RA has been confirmed in experiments where mice that are not susceptible for CIA became susceptible after transgenic introduction of RA susceptibility HLA class II alleles HLA-DR4 (DRB1*0401) and HLA-DR1 (DRB1* 0101)[95, 96].

The autoimmune arthritis is accompanied by strong T and B cell responses that react with both autoantigens, murine CII, and heterologous human CII. Mice lacking αβ T cells are completely resistant to CIA and the observation that early administration of monoclonal antibodies to T-cell receptor-αβ (TCR) can suppress CIA provides further support for the role of T cells in mediating disease [97, 98]. Different studies suggested that defects in CD4+CD25+ regulatory T cells (Treg) function are involved in increased severity of CIA[99, 100]. The lack of IL-17, secreted by the novel T helper subset Th17 cells, has suppressive effect on CIA[101], and it has been shown that IL-17 promotes osteoclastic bone erosion[102]. Histopathology of arthritic joints from DR4 and human CD4 transgenic mice shows severely deformed joints in CIA, inflammatory pannus tissue eroding the bone and cartilage surface, with infiltrating CD4+T cells, activated macrophages and neutrophils, and new cartilage formation[103].These mice develop high levels of Abs to CII, with both IgG1 and IgG2a responses, but with dominating IgG2a, indicating a Th1 type of response.

The crucial role of B cells in pathogenesis of CIA is clearly demonstrated when shown that B-cell deficient mice do not develop CIA[104]. The antibody response to CII is directed towards distinct triple-helical peptides of the major B cell epitopes along the CII molecule[105], and binding and accumulation of anti-CII Abs may initiate articular inflammation through the activation of complement cascade[106]. Activation of
complement, especially C5a recruits neutrophils and macrophages, which activated by FcγR ligation secrete proinflammatory cytokines such as IL-1β, TNF-α, IL-8, IL-6, nitric oxide (NO), and prostaglandins (PGE2) that will recruit and activate different immune cells: NK cells, DCs, T and B cells, as well as activate synovial macrophages and synovial fibroblasts[107]. A new therapeutic concept for the treatment of arthritis has been identified using streptococcal enzyme IdeS, which cleaves IgG2a in vivo and thus protects mice against collagen antibody induced arthritis (CAIA), delays the onset and reduces the severity of CIA[108].

5.1 LOOKING FOR ARTHRITIS REGULATING GENES ON MOUSE CHROMOSOME 6 & 14

Although major susceptibility genes for CIA are associated with MHC II haplotypes, CIA similar to RA is a polygenic disease and genetic studies have shown associations with genes outside the MHC[109]. It has been observed in studies of experimental models of different autoimmune diseases, that many of the genetic linkages to disease development are located in the same genetic regions and potentially could be controlled by the same gene(s). Furthermore, comparison of the mouse/rat genetic regions with regions of association to human inflammatory diseases, also demonstrates some homologous loci shared between species. For example, Ncf1, an arthritis susceptibility gene identified first in rats using congenic approach, causing a lower oxidative burst that is associated with the activation of autoreactive T cells and development of severe arthritis, has shown the same effect in Ncf1 mutated mice[110].

Studies in SKG mice that have a recessive mutation in a key signalling molecule in T cells encoded by ZAP-70 gene, where defective signal transduction through ZAP-70 results in failure in thymic negative selection of highly self-reactive T cells, points to a central role of TCR signal transduction in arthritis development[111]. Also, the expression of a very restricted TCR-α/β repertoire is associated with the development and susceptibility to CIA in mice, and experiments in HLA-DR-transgenic mice demonstrated that DR1 and DR4 stimulate a highly restricted subset of T cells [112]. T cells from draining lymph nodes of DBA/1 (H-2b) mice immunized with bovine CII showed restricted Vβ, Vα, and Jα gene segments involved in the pathogenesis of CIA [113, 114]. In addition, the adoptive transfer of Vβ3+CD4+ T cells, specific to CII, had a clear arthritogenic role in CIA induction [115]. The central role of TCR in these processes makes it a desirable candidate as a genetic determinant of arthritis susceptibility.

Genetic contribution to CIA has been extensively investigated so far, and together with the seven new loci identified in this thesis in total 44 QTLs controlling CIA in mice have been identified (http://www.informatics.jax.org/). The Cia6 locus that is linked to the region encoding TCR-β chain, was previously identified in the F2 intercross where the alleles from susceptible strain DBA/1 were surprisingly found to confer protection from arthritis [116]. Interestingly, after splitting males and females, the two peaks of the chromosome 6 LOD score curve were found to be related to the gender of the mice, a case that we also noted for Cia50 locus. The BQ.DQ14 region on chromosome 14 that we included in the PAI didn’t include any known arthritis loci, but included the
TCR-α locus. Therefore, our starting hypothesis was that Cia6 locus is caused by a polymorphism in the TCR-β chain region, that both the TCR-α and TCR-β could be arthritis susceptibility loci, and that these might interact to determine arthritis susceptibility as well. Our findings don’t exclude this possibility, however, in the performed experiments we haven’t been able to narrow down the loci considerably, still we have found that these regions have many interesting genes as potential candidates (Figure 4).

**Genes**

- Eosinophil-associated ribonuclease A family
- Prostaglandin D & E receptors
- Suppressor of cytokine signaling 4
- Solute carrier family 35
- Toll-like receptor 11
- Solute carrier family 39
- T-cell receptor alpha/delta VDJJC gene segments
- Solute carrier family 7
- Matrix metalloproteinase 14
- Solute carrier family 22
- Interleukin 25
- Interleukin 17D
- Fibroblast growth factor 9
- C1q and tumor necrosis factor related protein 9
- Tumor necrosis factor receptor superfamily

![Diagram showing TCR-α and TCR-β loci with genes associated with arthritis susceptibility](image-url)

Figure 4. A schematic view of all known arthritis QTLs on mouse chromosome 6 and 14 together with some possible candidate genes around TCR alpha and beta chain regions.

Bbaa: Borrelia burgdorferi-associated arthritis
Pgla: Proteoglycan induced arthritis
(C) Position of loci and order of genes according to http://www.informatics.jax.org/ and http://www.ensembl.org/
The sex specificity or preponderance as we have observed for Cia50 locus in males doesn’t exclude the possibility that another gene other than TCR-α might be the causative one.

On the other hand, both TCR loci have been more precisely defined and are included in two new loci: TCR-α in Cia50, and TCR-β in Cia48. The data is thus in accordance with a possible role of both TCR-α and TCR-β polymorphisms, but further experiments are needed to determine the genes that truly cause the loci. Interestingly, several other arthritis regulating loci have been mapped to the same regions on chromosome 6 and 14: Pgia19 (Proteoglycan induced arthritis) and Cia17 are colocalized with Cia48 and Cia49 respectively, and Bbaa20 (Borrelia burgdorferi–associated arthritis) and Bbaa21 are close to Cia48 and Cia50, respectively, thus suggesting for the shared genetic pathways in different models of arthritis[117-119].

The results from the experiments that we have performed in heterogeneous stock inbred-outbred cross suggest a promising resource for the future fine-mapping of many known CIA loci as well as for identification of the new ones, and hopefully the underlying causative genes.
6 PRESENT INVESTIGATIONS

The initial aim of this thesis was to identify new arthritis susceptibility genes. We haven’t identified new genes, but instead we identified seven new CIA regulating loci. Additional experiments will be required to identify the genes underlying these QTLs.

Paper I: Identification of New Loci Controlling Collagen-Induced Arthritis in Mouse using a Partial Advanced Intercross and Congenic Strains

In the first study we mapped loci on chromosome 6 and 14 in mouse using a partial advanced intercross (PAI) strategy based on three congenic strains, BQ.DQ6-1, BQ.DQ6-2 and BQ.DQ14. The two BQ.DQ6 congenic fragments covered the Cia6 locus, including the TCR-β locus that had previously been identified in an F2 cross between the relatively resistant B10.Q and the highly susceptible DBA/1 strain. Based on the central role of TCR in antigen recognition in arthritis the TCR-β locus was considered as a main candidate for Cia6 locus. The BQ.DQ14 region on chromosome 14 did not include any known arthritis loci but included the TCR-α locus. The congenic strains were intercrossed in a PAI and analysed by linkage analysis, followed by confirmation of found loci in a collection of heterozygous congenic strains with partly overlapping fragments. Using this strategy, we identified seven new arthritis loci including two loci corresponding to the two TCR-chains. The Cia6 locus, previously mapped to a broader region on chromosome 6, was now separated into three different loci: two protective and one promoting. On chromosome 14 we identified four new loci: three promoting and one protective. Among these loci we found evidence of both epistatic effects and sex-specificity. This study demonstrated the usefulness of PAI and congenic approach as a strategy for identification and fine mapping of the loci.

Paper II: Phenotyping and Induction of Collagen-Induced Arthritis in a Heterogeneous Stock Inbred-Outbred Cross

In the second article, we describe the successful induction of arthritis in the HSxBQ mice as well as a phenotyping protocol used to collect phenotypes from 1764 mice for QTL analysis. To map disease modifying loci we used an inbred-outbred cross strategy to introduce arthritis permitting MHC alleles into the HS by crossing them with an inbred strain, C57BL/10.Q that has an arthritis allowing H2q haplotype. The aim was to produce arthritis susceptible mice having high-density recombinations that would be a perfect tool for fine mapping of the CIA loci. In order to maximise the benefit of HS experiments, it was desirable to map in parallel as many phenotypes as possible. Therefore, in addition to arthritis related phenotypes such as arthritis scores, antibody titers, lymphocyte population analysis and oxidative capacity, we collected some un-related phenotypes such as coat colour, weight and length (to calculate BMI). Different organs and parts of the body were collected for the future studies of phenotypes such as mRNA expression level, bone density, etc. The successful mapping of different lymphocyte phenotypes to the known Cia36 locus confirmed the usefulness of the cross for QTL mapping. The data suggests that HSxBQ
Concluding remarks and future perspectives

So far, genetic studies have led to identification of a large number of quantitative trait loci, however, identification of the underlying genes involved in pathogenesis of the complex diseases, such as rheumatoid arthritis, has proven to be much more difficult. Using combination of different genetic strategies such as partial advanced intercross (PAI) and generation of congenic strains, we have identified seven new CIA modulating loci on mouse chromosome 6 and 14. Chromosome 6 region harboured two protective (Cia46 and Cia47) and one promoting (Cia48) loci, whereas chromosome 14, with no previously known CIA loci, harboured three promoting (Cia49, Cia50, Cia51) and one protective (Cia52) loci. The protective Cia6 locus was previously mapped to a broader region on chromosome 6 and the results of this thesis illustrate the importance of breaking up congenic regions into smaller sub-congenic fragments to unmask the complex set of multiple loci acting in opposite directions. The TCRα chain region within Cia50 and TCRβ chain region within Cia48 locus are well within the confidence intervals of these loci, hence still remain as interesting candidates.

However, resolution of the crosses in the present study was insufficient to conclusively pinpoint these genes, because many interesting candidate genes are still present in the vicinity of TCRα and TCRβ genes. To address the question it would be interesting to test the TCR response in different models, by immunizing the congenics with different antigens besides collagen, such as MOG (myelin oligodendrocyte glycoprotein) or ovalbumin, and perform T-cell proliferation assays.

We have also here demonstrated the usefulness of the heterogeneous stock mice as a fine-mapping tool for the already known Cia36 locus, which is also promising to be a future step for many of our congenics.

One way to proceed further towards positional cloning of the genes is, certainly to continue narrowing down the regions by identification of small recombinations within the loci. Furthermore, in order to study interaction(s) between the QTLs, in particular the ones harbouring TCR genes it would be interesting to establish double congenics for these loci and perform PAI experiments.

Identification and further dissection of the gene regions having modulating effect on collagen induced arthritis, will not only reveal the responsible candidate genes, but also give the possibility to define and understand critical molecular pathways leading to arthritis development, and thus give some fresh prospective in development of novel and more effective therapies for RA.
7 ACKNOWLEDGEMENTS

This work was performed at the Medical Inflammation Research at Lund University and Karolinska Institute and supported by EU Marie Curie Early Stage Training.

The past three years have been an exciting period of my life not only because of learning a lot about genetics and research but also for meeting many new colleagues and friends from all over the world. Therefore I would like to express my gratitude to those people as well as to all the people that supported me throughout the years. Especially, I would like to thank:

My supervisor Professor Rikard Holmdahl for the great opportunity to do this work and improve in different ways, for encouraging me and being full of understanding.

Eszter Rockenbauer for the warm and friendly welcome when I came to the new group, and for teaching me a lot in a short period of time.

Emma Ahlqvist, my unofficial supervisor, for teaching me different lab techniques, answering many of my never-ending questions, and for great organization of our HS project.

To all the MIR people I have worked with in the lab, especially to Nan for always giving an instant help, for all the discussions about immunology and life, and for critical reading of this thesis. Dorota K. for being a good friend and for all the funny time we had. Patrick M., the best computer guy, for all the help and kindness. Ingrid L. for making the lab a much nicer place to work, for many interesting talks and for being a true friend. To Jia H. for the Chinese wisdom and all the help, Ia K. for being so nice, Maria L. for sharing the same taste in movies and for the funny conversations, Emma M. for talking in Swedish over the phone instead of me and for all the jokes, Kristin B., prof. Ragnar M., Duojia C., Stefan C., Malin H., Tsvetelina B., Bruno R., Åsa A., Ulrika N., Michael F., Meirav H., Thomas B., Jonatan T., Angela P., Johan B., Therese L., Geng H., Carola R., Alexandra B., Tiina K. for all the help and for making the lab such a unique place to work.

I would also like to thank some previous members of the lab for everything they did to make the work in the lab go smoothly: Malin N., Robert B. for the initial work on my project, my student Marie L., Solveig B., Balik Dz., Ivanka T., Kyra G., Lina O., Martina J., Myassa D., Margareta S., and Jenny K.

To Angel for making the administration work less hard, and for being an initiator of many pleasant events in the lab.

To all the present and past animal caretakers for taking good care of our mice, especially to Sandy L., Isabell B., Carlos P., Rebecka Lj. and Tomek.

Thanks to all my friends outside the lab who made my time in Lund and Malmö joyful, in particular: family Bogojevic for incredible support, many exciting events and for
being my “second” family, family Gajic for all the happy time we spent, and for introducing me into the world of good music, Zana for the Malmö’s night life, Arni and Karolina for being such good friends, Nenad and Milan S. for all the coffees, my nice corridor mates: Markus, Zia, Martin, Megan, Maria D, Anja, Lieke, Ruben, and Meiki. To Nega for all the laughs, for being so positive and magnificent, for unforgettable time in Paris and her outfit, which left Mona Lisa unnoticed…

My amazing friends in Belgrade, Sweden and further…. Milja M. for all the caring talks and all her support, Sasa J. for taking care of me at the “drunk” parties, Twiggy for the coolest places to go out and all her love, Angela for being always there for me and for all the nice memories, Vlada A. for being my voice of consciousness, Kuki for all the care and making things simple, Kale for 16 successfully completed days in Göttingen, Bane V. for crazy parties and irreplaceable humour, Jelena M. for her magical brain, Dragana M. for all her support, Milos T. for funny emails, Magi B. for her irresistible laughter, Zlata N. for nice talks and beautiful time in Holland, Maja T. for being the best book-fair companion, Maja J. for the nice time at Karolinska Institute, Sanja H. for our internal jokes, Bojan V., Emina S., Ana B., for all this and much more - Thank you!

To all other friends and relatives, not mentioned but always on my mind, who make my life better.

Many, many thanks to my sister Maja, for all her love, care and support. I am happy to have you!

At the end, I would like to express my eternal gratitude to my mother and father for everything they did for me, for their support, love, and encouragement. For everything I am, and everything I can be.
8 REFERENCES


30. Wordsworth BP, Lanchbury JS, Sakkas LI, Welsh KI, Panayi GS, Bell JI. HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. Proc Natl Acad Sci U S A 1989;86:10049-53.


61. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.


polygenic and exhibits consistent linkage to multiple loci on chromosome 5 in four independent mouse crosses. Genes Immun 2001;2:388-97.