GENETIC STUDIES OF COMPLEX AUTOIMMUNE DISEASE

Alexandra Gyllenberg

Stockholm 2014
To My Beloved Ones
ABSTRACT

In complex autoimmune diseases, there are both genetic and environmental factors that influence our immune system and contribute to the development of disease. The pathways, interactions and mode of inheritance are difficult to unravel, and many discoveries are yet to be done. Here, I have studied three major autoimmune diseases, Type 1 Diabetes (T1D), Multiple Sclerosis (MS) and Rheumatoid Arthritis (RA). They all have an inflammatory component, and share their genetic predisposition in the Human Leucocyte Antigen (HLA) genes. In T1D, the β-cells in the pancreatic Islets of Langerhans are very specifically destroyed in an autoimmune attack, lead by the CD8+ T-cells. This leads to the inability to produce insulin, and a lifelong treatment with daily injections is necessary. The elevated blood glucose levels leads to long-term damages of microvascular circulations, and co-morbidities like neuropathy and cardiovascular disease. In MS, auto-reactive lymphocytes enter the central nervous system (CNS) and causes demyelination and neuronal damage. It results in a periodical occurrence of sclerotic plaques in the brain and spinal cord, and leads to increasing neurological disability. RA is a systemic inflammatory disease induced by activation of auto-reactive T-cells, where the release of antibodies and formation of immune complexes contributes to the severity in anti-citrullinated protein antibody (ACPA) positive but not ACPA negative disease. The target tissue is the synovial joints, but other organs like heart and kidney will also be affected. The inflammation breaks down cartilage and bone and can lead to severe pain and disability.

In this thesis I present my investigations of two candidate genes, CIITA and VAV1, in these diseases. CIITA is the major control factor for transcription of the HLA class II genes, and associated to all three diseases. In paper I we demonstrate genotype variation for markers in the CIITA gene, depending on age among healthy controls. This finding is also replicated in an independent cohort. The consequence of this can be faulty conclusions in association studies, and hence age should be corrected for in genetic case-control studies. We find that association to T1D remains after controlling for age for rs11074932 (p=0.004) and rs3087456 (p=0.001), two markers in the promoter area that also are found to associate to RA but for the opposite allele (paper III). In paper II we replicate the previously reported association between CIITA rs4774 and MS in cases carrying the HLA-DRB1*15 allele (p=0.01, OR: 1.21) but also report association to MS for the same marker when stratifying for the MS protective HLA allele A*02 (p=0.01, OR: 1.33). Interaction between rs4774 and both MS associated HLA alleles is demonstrated. Finally, in paper III we show that the markers found to associate to T1D, MS and RA control the expression of CIITA and MHC class II genes with minor allele homozygotes leading to lower levels of mRNA of the transcripts. In paper IV we investigate the VAV1 gene, important in regulating signals downstream the T-cell receptor. VAV1 has been shown to associate to MS and lead to increased levels of inflammatory cytokines in the CNS. Here we report that the same rs2546133-rs2617822 C-A haplotype is associated only to the ACPA negative subgroup of Rheumatoid Arthritis (p=0.004, OR: 1.28). We also demonstrate that a SNP in the Vav1 gene in rat affects disease severity in pristane- induced arthritis, but not collagen II- induced arthritis, such that the disease in PIA is less severe. Taken together we suggest that these results for VAV1 reflect the heterogeneity between subgroups in human RA disease. In conclusion I have demonstrated genetic susceptibility factors and pathways that are shared between different autoimmune diseases but also that susceptibility genes can be of different importance in subgroups of patients for one disease.
LIST OF PUBLICATIONS

I. **Age-dependent variation of genotypes in MHC II transactivator gene (CIITA) in controls and association to type 1 diabetes**
A Gyllenberg, S Asad, F Piehl, M Swanberg, L Padyukov, B Van Yserloo, EA Rutledge, B McNeney, J Graham, M Orho-Melander, E Lindholm, C Graff, C Forsell, K Åkesson, M Landin-Olsson, A Carlsson, G Forsander, SA Ivarsson, H Larsson, B Lindblad, J Ludvigsson, C Marcus, Å Lernmark, L Alfredsson, K Åkesson, T Olsson, I Kockum, the Swedish Childhood Diabetes Study Group, the Diabetes Incidence in Sweden Study Group, and the Better Diabetes Diagnosis Study group
*Genes and Immunity* (2012) 13, 632–640

II. **Variability in the CIITA gene interacts with HLA in multiple sclerosis**
*Genes and Immunity* (2014), 1–6

III. **Genetic control of isoform expression of human MHC class II transactivator**
M Ronninger, A Gyllenberg, M Lindén, M Seddighzadeh, T James, S K Bedri, D Gomez-Cabrero, F Piehl, T Olsson, I Kockum, L Padyukov
*Manuscript*

IV. **VAV1 shows association to anti-citrullinated protein antibody (ACPA) negative Rheumatoid Arthritis in a Swedish patient cohort**
Ulrika Norin, André Ortlieb Guerreiro-Cacais, Alexandra Gyllenberg, Rasmus Berglund, Amennai Daniel Beyeen, Rikard Holmdahl, Elisabeth Petit-Teixeira, François Cornélio, Abdelhadi Saoudi, Gilbert J. Fournié, Leonid Padyukov, Ingrid Kockum, Maja Jagodic and Tomas Olsson
*Manuscript*

Additional publication

**HTR1A a novel type 1 diabetes susceptibility gene on chromosome 5p13-q13**
*Plos ONE, 2012. Volume 7, Issue 5, e35439*
# CONTENTS

1 Introduction ............................................................................................................. 1
   1.1 The immune System ......................................................................................... 1
      1.1.1 Innate immunity ....................................................................................... 1
      1.1.2 Adaptive immunity .................................................................................... 1
      1.1.3 Cytokines .................................................................................................. 3
      1.1.4 The immune system in autoimmune disease ............................................ 4
      1.1.5 Definition on autoimmunity ....................................................................... 6
   1.2 Autoimmune diseases in this thesis ............................................................... 7
      1.2.1 Type 1 Diabetes OMIM 125853 .............................................................. 8
      1.2.2 Multiple Sclerosis OMIM 126200 .......................................................... 10
      1.2.3 Rheumatoid Arthritis OMIM 180300 ...................................................... 12
   1.3 Risk factors for autoimmune disease .............................................................. 14
      1.3.1 Genetic risk factors .................................................................................... 16
      1.3.2 Environmental risk factors ....................................................................... 18
      1.3.3 The missing heritability problem ............................................................. 19
   1.4 Genetics ............................................................................................................. 19
      1.4.1 General genetics ....................................................................................... 20
      1.4.2 Susceptibility genes and research approach in complex disease studies ....................................................................................... 23
   1.5 Materials & methods ....................................................................................... 26
      1.5.1 Cohorts ..................................................................................................... 26
      1.5.2 PCR .......................................................................................................... 30
      1.5.3 ELISA ....................................................................................................... 30
      1.5.4 Genotyping ............................................................................................... 31
      1.5.5 HLA typing and imputation of HLA types ................................................. 32
      1.5.6 Expression studies .................................................................................... 32
      1.5.7 Animal models .......................................................................................... 34
      1.5.8 Statistical analyses ................................................................................... 36
      1.5.9 Genes studied in this thesis ..................................................................... 39

2 Study aims .............................................................................................................. 41

3 Results and Discussion .......................................................................................... 42
   3.1 The role of CIITA in autoimmune disease (Papers I-III) .............................. 42
      3.1.1 Results from paper I ............................................................................... 42
      3.1.2 Results from paper II ................................................................................ 43
      3.1.3 Results from paper III ............................................................................. 44
      3.1.4 Discussion on papers I-III ....................................................................... 45
   3.2 VAV1 in Rheumatoid Arthritis (Paper IV) ......................................................... 51
      3.2.1 Results from paper IV ............................................................................. 51
      3.2.2 Discussion on paper IV ............................................................................ 54

4 Concluding remarks ............................................................................................... 56

5 Future Perspectives ............................................................................................... 58

6 Acknowledgements ............................................................................................... 59

7 References .............................................................................................................. 62
LIST OF ABBREVIATIONS

AD  Alzheimer’s disease
AP  Attributable proportion
APC  Antigen presenting cell
CD  Cluster of differentiation
CI  Confidence interval
CNS  Central nervous system
CTL  Cytotoxic T lymphocyte
GWAS  Genome wide association study
HWE  Hardy-Weinberg equilibrium
HLA  Human leukocyte antigen
Ig  Immunoglobulin (antibody)
LADA  Latent autoimmune diabetes in adult
LD  Linkage disequilibrium
LPS  Lipopolysaccharids
MAF  Minor allele frequency
MHC  Major histocompatibility complex
MI  Myocardial Infarction
MODY  Maturity onset diabetes of the young
MS  Multiple sclerosis
OR  Odds ratio
PCR  Polymerase chain reaction
qPCR  Quantitative polymerase chain reaction
RA  Rheumatoid arthritis
SNP  Single nucleotide polymorphism
T1D  Type 1 diabetes
T2D  Type 2 diabetes
1 INTRODUCTION

1.1 THE IMMUNE SYSTEM

The immune system is very complex. It is spread throughout the body, from the defensive outer skin layers and mucosa, to highly specialized cells and tissues inside. Without a functioning immune system, we would rapidly succumb to all sorts of infections of bacteria, viruses and parasites. But when something goes wrong in this important system, and the processes starts to turn to our own tissues and cells, that is when we develop an auto-immune disease. The processes of the immune system are extremely intricate and versatile, and will only briefly be described here.

1.1.1 Innate immunity

The first line of defense is the innate immune system. It constitutes of mechanical barriers like skin and mucosa, and antimicrobial substances on epithelial surfaces towards the outside of the body. In the blood, the complement system helps to opsonize (mark out) and lyses invading bacteria. There are also cells in the innate system that recognizes common patterns in different pathogen groups, and respond by consuming them (phagocytic cells like macrophages and neutrophils) or attacking and lysing them (natural killer-NK- cells). Cytokines – cell-signaling molecules-also have an important role of regulating the different cells and pathways.

We are born with the innate system, it is there before we have even encountered any antigen, and it has evolved with us for millions of years, fighting pathogenic invaders. The difference between the innate and the adaptive system is that the innate cells react in mostly the same way to repeated infections, while the adaptive recognizes different invaders and respond stronger and faster to every exposure of that specific pathogen, it has a memory.

1.1.2 Adaptive immunity

The adaptive immune system evolved about 500 million years ago, and has developed to be highly specific in recognizing and clearing those invaders that resists the innate system. The main component is the lymphocytes and their secreted products. The adaptive system is divided into humoral and cell-mediated immunity. The humoral system is mediated by antibodies (immunoglobulins), produced by B lymphocytes (B-
cells, so called plasma cells). The antibodies are highly specific and bind to its antigen with great affinity. They can neutralize microbes and toxins, and activate different effector mechanisms that will lead to elimination of the invader. The cell-mediated immune system is primarily directed toward intracellular pathogens like viruses and some intracellular bacteria (tuberculosis for example) and is constituted of the T lymphocytes; T-helper cells and cytotoxic T cells (CTL). There are also regulatory T-cells which main role is to inhibit immune response and maintain tolerance. Activated B-and T-cells are considered to be among the effector cells that execute the eliminating response of the adaptive immune system. Many other cell types have important functions as well, for example dendritic cells (DCs). DCs are professional antigen presenting cells (APC) that are situated in the tissues of our bodies. They catch foreign microbes and display them to the lymphocytes in the peripheral lymphoid organs. To do this, the APC digest the microbe and present the peptides on its surface in specialized display molecules; the major histocompatibility complex (MHC) molecule. Generally, MHC class I is present on all nucleated cells and presents intracellular peptides, whether they are self-derived or from infecting virus. MHC class II are present mainly on APCs, but can be induced in other cell types. They present extracellular antigens internalized and processed by the cell. When the T-cells encounter such MHC-peptide combination together with other co-stimulatory molecules on the APC, it gets activated. The lymphocytes can be distinguished through surface molecules called CD (cluster of differentiation) molecules. T-cells that have CD4+ molecules will recognize MHC class II molecules, and these are the helper cells. When they are activated they stimulate B-cells to produce antibodies, and phagocytizing cells (macrophages) to kill ingested microbes. T-cells with CD8+ molecules on its surface will recognize class I MHC molecules, and become CTLs. Their action is to attack and lyse other infected cells that present the antigen. Some intracellular pathogens like viruses and even tumor cells down regulates MHC class I molecules to avoid immune response from CTLs. Here the Natural Killer cell from the innate system plays an important role while it recognizes and kills these cells without further activation needed. NK cells are complex lymphocytes with the ability to act in both the innate and adaptive immune response. B-cells recognize antigens through membrane bound antibodies, and cross-binding of these will activate the B-cell to become a plasma-cell, secreting antibodies. For
example, lipopolysaccharides (LPS) from gram-negative bacteria can activate the B-cell directly, but for protein antigens the “help” of CD4+ T-cells co-stimulatory molecules are needed. IgM is the first antibody type to be produced, but with the stimulation from CD4+ T-cells the plasma cell can produce antibodies of different classes, functionally different but with the same specificity as for the antigen first encountered. The classes, or isotypes, are IgG, IgE and IgA, specialized for different pathogens and different tissues. This is called heavy-chain class-switching. The T-helper cell will also stimulate the plasma cell to produce antibodies with an ever higher affinity to the antigen, so called affinity maturation\(^1\).

1.1.3 **Cytokines**

Cytokines secreted by the lymphocytes are important stimuli for the cells to proliferate and differentiate, but also activates effector cells to execute their response to antigens and promotes class-switching in antibodies\(^4\). The cytokines induces different types of immune-response depending on the infectious agent. In innate immunity, Tumor Necrosis Factor-α (TNF-α) is produced mainly by macrophages and is the principal mediator of acute responses towards microbes. LPS from bacteria stimulate the production of TNF-α, which recruits and stimulates innate effector cells (neutrophils and monocytes) to eliminate microbes, but also stimulate endothelial cells to express adhesion molecules for lymphocytes\(^1\). One of the most important early responses in adaptive immune is the production of Interleukin-2 (IL-2), mainly by CD4+ helper T-cells which stimulate the clonal expansion of the lymphocytes that recognizes the antigen. Interferon-γ (INF-γ) is important both in adaptive and innate immune response, and is secreted by T-cells and NK cells. INF-γ activates macrophages and promotes an inflammatory environment\(^1\), induce isotype switching in B-cells and increase antigen processing and presentation by upregulating MHC molecules and co-stimulators on APCs.

After the infection has been cleared, memory cells of B-and T-type will survive for many years, ready for a rapid response if the antigen is re-introduced. For some antigens, the memory seems to be lifelong. Then we say that we are immune to that antigen.
1.1.4 The immune system in autoimmune disease

What is it that can make your very own immune cells, there to protect you against pathogenic invaders, to see your own tissues as foreign and attack it? There are many control mechanisms during the development of mature lymphocytes, which function to delete or reprogram self-reactive lymphocytes. Loss of self-tolerance can develop from failure in any of these processes or due to abnormalities in the presentation of self-antigens to the immune cells.

Development of auto-reactive lymphocytes:
The lymphocytes develop from stem cells in the bone marrow, and the B-cells partly mature in the bone marrow, while the T-cells migrate to the thymus to mature, before they both enter the circulation system and populate peripheral lymphoid organs. The antigen receptor of both B-cells and T-cells are produced by genes that undergo somatic rearrangement, to be able to produce a broad diversity of antigen specificity. When the T-cells mature in the thymus, they go through selective processes, where they must both be able to recognize the MHC-peptide complex (positive selection), but not recognize the self-peptides with too high affinity (negative selection). Failure of either one of the checkpoints will lead to elimination of the cell, in the first case by lack of further stimuli, in the latter by apoptosis (programmed cell-death). A variety of self-peptides are presented in the thymus by dendritic cells and thymic epithelial cells, to be able to eliminate self-reactive T-cells.

B-cells go through similar pathways, they have to be able to produce functional membrane bound Ig-molecules to be allowed to survive. If they recognize self-antigens with high affinity they will undergo a process called receptor editing where they produce a new B-cell receptor, hopefully not self-reactive anymore. If this fails they will be eliminated by apoptosis. If any of these “checkpoints” fails, or if the self-antigen is not properly presented in thymus, auto-reactive immune cells will be able to enter the circulation. Since the MHC class II molecule which is a major susceptibility gene in many autoimmune diseases has a crucial role in the presentation of peptides to CD4+ T-cells, and these cells are regulating both the humoral and cellular response to protein antigens, failure in T-cell tolerance is considered to be a main mechanism in autoimmune disease.
Autoreactive T-cells have also been isolated from affected tissues or blood from patients suffering from autoimmune disease.

Autoimmune disease comes in many forms, and they can be classified depending on what hypersensitivity reaction is involved (Table 1) (described by Gell & Coombs 19635). For type I acute allergy reaction—there is no counterpart within autoimmune diseases.

<table>
<thead>
<tr>
<th>Type of hypersensitivity</th>
<th>Description</th>
<th>Mediator</th>
<th>Typical disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>Allergy (immediate)</td>
<td>IgE, IgG</td>
<td>Asthma</td>
</tr>
<tr>
<td>Type II*</td>
<td>Antibody mediated</td>
<td>IgM, IgG</td>
<td>Goodpasture’s syndrome</td>
</tr>
<tr>
<td>Type III</td>
<td>Immune complex</td>
<td>IgG</td>
<td>SLE, RA</td>
</tr>
<tr>
<td>Type IV</td>
<td>T-cell mediated</td>
<td>T-cells</td>
<td>MS, T1D</td>
</tr>
</tbody>
</table>

Table 1; *receptor mediated diseases such as Graves’ disease and Myasthenia Gravis are sometimes classified as Type V.

Autoimmune disease can be organ or tissue specific, as in Autoimmune Thyroiditis, or it can be systemic, as in Systemic Lupus Erythematosus (SLE). The disease can be mediated by antibodies, either contributing to an autoimmune attack towards specific target, or by forming immune complexes with circulating antigen which will deposit in organs and vessels. In both cases, inflammation is induced, leading to tissue injury. Antibodies can also bind directly to receptors and stimulate or inhibit response to hormones and neurotransmitters. This is the case in Graves’ disease and Myasthenia Gravis respectively. The autoimmune disease can also be T-cell mediated, either by inducing delayed-type-hypersensitivity (DTH) reactions or by CD8+ T-cells directly killing target cells. In the first case, T-cells, primary CD4+, recognizes a self-antigen and secretes cytokines that activate macrophages and induce inflammation, such as interferon gamma (INF-γ) and tumor necrosis factor (TNF). They in turn produce cytokines and growth factors contributing to a chronic inflammatory state. It is not common with diseases that are caused solely by CD8+ T-cells (CTL), one example is myocarditis where MHC class I restricted CTLs attack myocardial cells after infection with coxsackie B virus1.
1.1.5 Definition on autoimmunity

A common mistake is to define a disease as autoimmune because it involves immune reactions that cause tissue damage or malfunction. Often it is difficult to establish a self-antigen in these diseases. Presence of association to MHC genes and auto-antibodies also has to be considered. It has been proposed to regard immune diseases along an “immunological disease continuum” with classical autoimmune diseases (for example T1D) in one end and diseases with involvement of the innate immune system, causing site-specific inflammation that is independent of adaptive immune responses (as in hereditary periodic fevers, HPFs) in the other end\(^6\). Diseases could then be classified as purely autoimmune or auto-inflammatory, or a combination of both.

Witebsky’s postulates, formulated in 1957 by Ernst Witebsky and colleagues\(^7\) based on Koch’s postulates and later modified in 1994 by Rose et al\(^8\) are used to classify autoimmune disease. The criteria that are being used include:

- **Direct evidence from transfer of disease by pathogenic antibody or pathogenic T cells.**
  
  This can be seen in neonatal myasthenia gravis and Graves’ disease where IgG auto-antibodies pass the placenta from mother-to-child, causing a transient condition. Also, T1D has been transferred between individuals in bone-marrow transplantations.

- **Indirect evidence based on reproduction of the autoimmune disease in experimental animals.**
  
  In several experimental models, the pathogenic role of T-cells and antibodies can be demonstrated by transfer. In one animal model for T1D, the NOD-mouse, transfer of T-cells from diseased animals into naïve animals induces disease. In the experimental model of MS, experimental Autoimmune Encephalomyelitis (EAE), immunization of animals (rat, mouse) with myelin compounds in adjuvant leads to an inflammatory, demyelinating disease.

- **Circumstantial evidence from clinical clues—for example infiltration of lymphocytes in the affected organ, a genetic association to HLA genes, co-association with other autoimmune diseases or positive response to immune suppressing treatment.**
These findings indicate suspected autoimmune disease but are not enough evidence on their own for a final definition.

1.2 AUTOIMMUNE DISEASES IN THIS THESIS

The papers in this thesis are focused on three of the major autoimmune diseases; type 1 diabetes (T1D), which is mostly seen in children and during adolescence, multiple sclerosis (MS) which affects young adults, and rheumatoid arthritis (RA), most common among middle-aged and elderly. Although these diseases are quite different in symptoms and development of the diseases, as well as in affected organs, they also have some remarkable similarities which could indicate a common autoimmune etiology. Foremost, they share the genetic association to the HLA genes, central in the adaptive immune system, but also other immune related genes are in common, as well as other risk factors. Generally, the autoimmune model with failure in selective processes in thymus leading to release of potentially auto reactive T-cells (and B-cells) specific for certain auto-antigens, is thought to be a major event in these diseases.

There is a big variation in incidence depending on population in these diseases; according to the World Health Organisation (WHO) MS\textsuperscript{9} is generally increasingly common with distance north or south of the equator, but is also prevalent in Australia and New Zealand. RA is more common in the northern hemisphere and the Nordic countries Sweden and Finland has together with Sardinia the highest incidence of T1D\textsuperscript{10} in the world. Even though research in these diseases has been performed for a very long time, even hundreds of years, a lot is unknown regarding which genes, pathways and processes are involved. For every new piece of the puzzle we can discover, there is a potential for understanding the pathogenesis which helps in finding biomarkers and develop protective agents or treatments.
1.2.1 Type 1 Diabetes OMIM 125853

Diabetes is one of the most common non-contagious diseases in the world. Often it is divided into two major types of diabetes; type 1 and type 2, earlier referred to as “childhood diabetes” and “adult onset diabetes” or "insulin dependent diabetes" and "non insulin dependent diabetes". Type 1 diabetes is also often referred to as autoimmune diabetes. In 2013, more than 382 million people worldwide lived with some form of diabetes, and this is predicted to rise to 592 million by 2035. About 85-90% of the cases are type 2 diabetes. Every year about 78 000 children develop type 1 diabetes and its more common among boys than girls. In Sweden, about 50 000 individuals have T1D, and 7-8000 of them are children. The costs for the society are enormous, and the individual sufferings can be severe. There are also other forms of diabetes like gestational diabetes, neonatal diabetes, latent autoimmune diabetes in adults (LADA) and maturity onset diabetes of the young (MODY). All types of diabetes have in common the problem with inability to convert blood glucose into fuel for the cells. This is due to either lack of production of insulin and/or decreased sensitivity to insulin. T1D is considered to be an autoimmune disease, where the insulin-producing beta (\(\beta\)) cells in the pancreas are destroyed by the immune cells of the body. About 80% of the \(\beta\)-cells are destroyed before the insulin production is impaired enough to give symptoms. When the blood glucose levels rises (hyperglycemia), symptoms like excessive thirst, fatigue and frequent urination arises. If the condition remains further undiagnosed, weight loss, dehydration and ketoacidosis occurs from the breakdown of protein and fatty acids. This condition can lead to acute complications and will, if untreated, be fatal. T1D is treated with insulin that has to be administrated directly to the blood-stream. There is today no cure, although transplantation of pancreas or treatment with stemcells has shown successful results. Monitoring blood-glucose levels can be difficult, and there are grave complications related to long-term hyperglycemia. Damages to microvascular circulation leads to impact on nerves and organs such as renal failure, retinopathy, neuropathy, and cardiovascular disease. An increase occurrence of other autoimmune disease, particularly thyroid dysfunction and celiac disease is seen among T1D patients.

According to World Health Organization (WHO 1998, reviewed 2006) an individual with fasting plasma glucose \(\geq 7.0\text{mmol/L}\) or 2--h plasma glucose \(\geq 11.1\text{mmol/L}\) is diagnosed with diabetes. There are also other measurements like “impaired fasting
glucose (IFG)” and “impaired glucose tolerance (IGT)” where the levels for fasting glucose or 2 hour oral glucose tolerance test are elevated but do not reach the diabetes threshold value (6.1-6.9 mmol/L and 7.8 to 11.0 mmol/L, respectively). This indicates a pre-diabetic state with risk of developing diabetes. Measuring C-peptide levels determines the level of insulin secretion and remaining functioning β-cells, generally low in T1D patients but normal or even elevated due to the initial rise in insulin production in T2D patients. Measuring auto-antibodies (described below), present in individuals with T1D but not T2D is also used in differential diagnosis of diabetes. In T2D, there is no autoimmune attack on the β-cells and the symptoms are often more vague before diagnosis. Typically, the insulin sensitivity in the tissue is impaired due to overweight and initially the treatment is focused on dietary changes and exercise, but many patients will also need medical treatment that either increases insulin production or insulin sensitivity.

Pathogenesis of T1D

The autoimmune attack on β-cells is believed to occur due to some event (for example an infection) that breaks the self-tolerance in susceptible individuals, followed by an joint expansion of auto-reactive B-cells, CD4+ and CD8+ T-cells and activation of the innate immune system. There is an initial infiltration of macrophages and dendritic cells in the pancreas, shortly followed by surrounding of the pancreas with T-cells (peri-insulitis). Damage to the islets releases more autoantigens, leading to epitope spreading and further infiltration of lymphocytes in the pancreas. T-cells specific for islet-antigen have been identified in NOD mice and in the peripheral blood of T1D patients. Studies of pancreas from deceased newly-onset T1D patients show infiltrates of mainly CD8+ T-cells, and the effector mechanism of autoreactive CD8+ T-cells is also regarded as the major mechanism of final β-cell destruction. Antibodies like islet cell autoantibodies (ICA) which targets a variety of islet proteins can be detected many years before the onset of disease. There are also auto antibodies towards β-cell specific auto antigens; Insulin autoantibodies (IAA), Glutamic acid decarboxylase 65 autoantibodies (GAD 65), protein tyrosine phosphatase-like molecule (IA-2) autoantibodies, and zinc transporter autoantibodies (ZnT8). The role of auto-antibodies is debated, indicating an autoimmune event prior to clinical
diagnosis, but they are not believed to contribute largely to $\beta$-cell destruction. The antigen specific B-cell however is also efficient in antigen presentation to T-cells, and together with innate cells produces cytokines that enforces the inflammatory environment (ex. TNF-$\alpha$, INF-$\gamma$). It is hypothesized that this will also downregulate the immunosuppressive function of T-regulatory cells.$^{18}$

1.2.2 Multiple Sclerosis OMIM 126200

MS is a neurological disease characterized by demyelination of the nerve axons in the brain and spinal cord, considered as a result of periodical infiltration of auto reactive immune cells. This leads to a progressive accumulation of sclerotic plaques where nerve axons are damaged, in turn leading to increasing neurological disability.$^9$ About 2.5 million people worldwide suffer from MS, in Sweden there are about 17,000 MS cases.$^{21}$ Most patients are diagnosed between the age of 20-40, and the disease is twice as common among women compared to men.$^9$

The disease progress can take different forms (Fig.1) either occurring in isolated attacks (relapsing forms), which occurs in nearly 80% of the cases, or building up over time, worsening the symptoms progressively (progressive forms). About 30-50% of patients with the relapsing-remitting (RRMS) form will after many years enter a secondary progressive form (SPMS). Primary progressive MS (PPMS) is rarer, and it is also generally associated with worse outcome. Some individuals will experience a
single bout of MS symptoms; this is referred to as CIS - clinically isolated syndrome\textsuperscript{22}. As a consequence of the nerve damages, the symptoms can vary between fatigue, blurred vision, motor and sensory problems, muscle spasms, muscle weakness and chronic pain. Emotional problems such as depression or unstable mood are also common\textsuperscript{9}. As the disease proceeds, the damages will occasionally lead to more chronic disabilities. About 10\% of the patients will have a severe disability (wheelchair) in 5-10 years.

MS is diagnosed by a neurologist, based on a set of criteria developed by National Multiple Sclerosis Society (NMSS) of America, the McDonald criteria\textsuperscript{23, 24}. Clinical data on previous attacks, analysis of magnet resonance images (MRI) and testing of cerebrospinal fluid is used in diagnosis. For a final diagnosis more than one lesion in MRI analysis, or evidence (MRI or clinical) of more than one attack over time is needed. Other events can be strengthening like chronic inflammation in the central nervous system and the presence of antibodies, so called oligoclonal bands (OCB).

There is today no cure for MS; however there is much research done on improving treatment. Corticosteroids are used to fight the acute inflammation in an attack, and disease-modifying treatments like interferon-beta protects the myelin by reducing the numbers of autoimmune attacks. There are other treatments affecting the immune system; Natalizumab is a drug that have proven to be efficient, but due to issues with severe side effects it is only used in patients who do not respond to other treatments or in more severe disease cases. It also has to be delivered via injections. This monoclonal antibody blocks cell adhesion molecule α4-integrin, necessary for the lymphocytes to attach and pass the blood-brain barrier. This leads to reduced relapsing and disability progression, but will also give room for opportunistic viral infections like JC virus that can cause a progressive and often fatal demyelinating disease named Progressive Multifocal Leukoencephalopathy (PML)\textsuperscript{25}

Another recent immunomodulating drug is GILENYA (Fingolimod), which is lowering the number of circulating lymphocytes in the blood. It is easy to manage since it is orally taken, but risk of heart problems and severe infections are drawbacks.

It has been debated whether MS should be classified as an autoimmune disease or not. It has been argued that MS could primary be a neurodegenerative disorder and the process of myelin breakdown set of an inflammatory response and autoimmune reactions. The arguments against the autoimmune model points to the lack of an
established specific autoantigen and that autoinflammatory treatment may reduce relapses but have no or low effect on progressive forms or disability outcome\textsuperscript{26}. In the animal model for MS, EAE, disease has been transferred with auto-reactive T-cells which indicate autoimmunity. The animal model has many similarities but the etiological role of the myelin protein used (MOG, MBP) is unclear\textsuperscript{8}, and it is also necessary to use a strong adjuvant for inducing disease. In EAE, the CD4+-T-cell has been proven to mediate disease and induce inflammatory response\textsuperscript{1,26}. Treatment against CD4+ T-cells and TNF-\(\alpha\) is also effective in EAE, but show no effect or even worsening the disease in MS\textsuperscript{27}. However, many evidence is also in favor of MS as primary an autoimmune disease; foremost the strong genetic association to HLA genes that is shared with other known autoimmune disease. Also, almost all other genes found to associate with MS are found to be part of immunological pathways rather than neurodegeneration\textsuperscript{28}. There is also association with other autoimmune diseases in the same individual or the same family, further strengthening this theory\textsuperscript{29}.

\textit{Pathogenesis of MS}

According to the currently most widely accepted autoimmune model in MS, auto-reactive T-cells specific for myelin proteins enters circulation. They get activated in the periphery by dendritic cells that presents self-proteins which leads to upregulation of adhesion molecules and the T-cells attach to receptors on endothelial cells, and then proceed to pass across the blood-brain barrier. When they encounter the antigen in the central nervous system (CNS) they release cytokines that augments the inflammation and activates macrophages. When the myelin is broken down, the process is propagated by epitope spreading and the inflammation, and T-cells with various auto-specificity are activated by APCs. This can also explain the lack of a single antigen driving the disease\textsuperscript{30}.

\textbf{1.2.3 Rheumatoid Arthritis OMIM 180300}

Rheumatoid arthritis is a systemic inflammation of the synovial joints and other organs throughout the body. It can lead to severe disabilities and mortality if not treated. The inflammation degrades cartilage and bone in the joint and as a consequence, excess synovial fluid is gathered. There is also development of fibrous tissue in the synovium and ankylosis (fusion) of the joints. Several organs are affected
as a consequence of chronic inflammation, for example the heart and blood vessels, lung, eyes and kidneys. Symptoms include pain, stiffness and swelling of joints, primarily in smaller joints like fingers and hands. As the disease develops larger joints like the shoulder and knee will also be involved, which can lead to substantial loss of function and mobility.

The exact mechanisms behind disease development are not known, but it is considered to be an autoimmune disease. Autoantibodies against IgG Fc, known as rheumatoid factors (RF), and anti-citrullinated peptide antibodies (ACPAs) are often present, and both B- and T cells seem to play important roles in disease etiology. The disease is occurring worldwide, and about 0.5-1% of the adult population is affected, but prevalence is varying between populations. Its more common (3:1) among women than men, and the usual age of onset is 40-50 years. There is also a juvenile form- Juvenile Idiopathic Arthritis (JIA)-which is a subset of arthritis disorders affecting children under 16 years of age, classified in seven different categories based on specific inclusion and exclusion criteria.

The diagnosis of RA is based on classification criteria, published by the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) 1987 and revised in 2010. It is a detailed point-system used for classifying disease cases in research. In clinical practice, the following criteria are used given that there is no other diagnosis better explaining the synovitis:

- two or more swollen joints
- morning stiffness lasting more than one hour for at least six weeks
- the detection of rheumatoid factors or autoantibodies against ACPA. A negative autoantibody result does not exclude a diagnosis of RA.

The treatment goal is to minimize symptoms such as pain and swelling, keep joint functionality and prevent bone deformity. Disease-modifying anti-rheumatic drugs (DMARD) like methotrexate (immunosuppressive) are the primary treatment for RA, together with nonsteroidal anti-inflammatory drugs (NSAIDS) and cortisone for pain relief and anti-inflammatory therapy. If these first-line treatments are not effective enough, biological agents can be used, for example monoclonal antibodies towards B-cells (rituximab) and tumor necrosis factor alfa (TNF-α) blockers (infliximab).
Pathogenesis of RA

Although the etiology is not fully known, research has show that there is a difference in both genetic and environmental risk factors and severity of disease depending on presence or absence of ACPA, and this points to two different subsets of disease, possibly with somewhat different etiology\textsuperscript{35-37}. ACPA positive disease is the most common, occur in up to 80\% of the cases\textsuperscript{31} and is also associated with higher risk for joint destruction and comorbidities like cardiovascular disease. Citrullination is a post-translation modification where the positively charged aminoacid arginine is converted to the neutral citrulline by the enzyme peptidyl arginine deiminase (PAD). The changed charge and possibly structure of the protein seems to be recognized by the immune system. One theory\textsuperscript{38} is that events triggering immune responses in the lungs, like smoking or infections, in combination with RA associated risk-MHC-molecules on macrophages and dendritic cells presents these citrullinated peptides to CD4+ T cells and B-cells, which get activated and initiates production of anti-citrullin-antibodies. It has been debated whether the inflammation process then is driven by the formation of immune complexes with ACPAs and possibly RF (type III hypersensitivity), or by T-cells triggering release of proinflammatory cytokines such as TNF-\(\alpha\), interleukin-1, and interleukin-6 (type IV hypersensitivity) which in the end also leads to cartilage and bone destruction\textsuperscript{38}. Exactly how the activation in the lung leads to synovial inflammation is not clear, but citrullinated proteins are found in abundance in synovial fluid of RA patients\textsuperscript{39}, and ACPAs can be detected years before disease onset. It is likely that the process is similar in ACPA negative RA, but with different triggers of disease onset, and without the formation of ACPAs and subsequently immune complexes. The milder course of this subset of RA could be seen as an indication of the importance of the B-cell in RA.

1.3 Risk Factors for Autoimmune Disease

To answer the question -who will develop an autoimmune disease- is today not possible, but we can conclude that it is a genetic suscectible individual exposed to enough triggering environmental factors. In complex diseases there are many different factors that on their own are not sufficient for causing disease, but will be part of the total pathogenesis of the disease. For different individuals there can be
different, or overlapping, genes and environmental factors that together will lead to the same disease. Often the disease prevalence in siblings and dizygotic twins, who both share 50% of their genes, are compared to the general population. If risk for disease among relatives is considerably higher than the risk in the general population this indicates a genetic component. The ratio of risk for relatives to that in the general population is called λ, and is referred to as the familial aggregation. If the concordance rate for monozygotic twins, who has almost identical DNA, is not 100%, it shows that other factors than genetics also play a role in disease development.

The heritability (h²) is often based on twin studies and compares the degree of concordance in monozygotic (MZ) twins to dizygotic (DZ) twins. This gives an estimate of how much of the disease risk is attributable to genetic factors. It has been shown to be as high as 66-72% in T1D, 64% in MS and around 60% in RA. The concordance rate in MZ twins displayed in table 2 below demonstrates that the autoimmune diseases studied in this thesis all have a strong genetic component.

<table>
<thead>
<tr>
<th>Concordance rates</th>
<th>MZ twins</th>
<th>DZ twins</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1D</td>
<td>30-50%</td>
<td>16%</td>
</tr>
<tr>
<td>MS</td>
<td>30%</td>
<td>7%</td>
</tr>
<tr>
<td>RA</td>
<td>15%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

Table 2; Approximate values for twins in the different diseases.
1.3.1 Genetic risk factors

1.3.1.1 HLA genes

The diseases studied here have one main genetic risk factor in common, the Human Leukocyte Antigen (HLA) genes on chromosome six (Fig. 2).

Many other autoimmune diseases also show association to HLA for example celiac disease\(^{44}\), SLE \(^{45}\), and narcolepsy\(^{46}\). These genes codes for the MHC molecules that are presenting peptides to the immune cells. Generally, MHC class I (A, B, and C) present peptides from inside the cell (including viral peptides if present), and MHC class II (DR, DQ, and DP) present extracellular antigens, but there is also indications that cross-presentations occur\(^{47}\), which can have impact in auto-immune disease. Also other immune related genes, for example genes for internal processing of antigens, complement genes and tRNA-genes are situated in this region\(^{48}\). Even though the region is very polymorphic, there is also extensive linkage disequilibrium between alleles in the genes, and haplotypes formed by the combination of the genes are of different importance in different diseases.

Both class I and class II molecules consists of two peptide chains. In class I molecules the invariant \(\beta_2\)-microglobulin “light” chain is paired with the “heavy” polymorphic chain, encoded by HLA-A, HLA-B and HLA-C. In class II molecules, both \(\alpha\) and \(\beta\) chains are polymorphic, even if the \(\beta\)-chain is by far the most variable. The \(\alpha\)-chain is encoded by HLA-DRA, HLA-DQA and HLA-DPA, and the \(\beta\)-chain is encoded by HLA-DRB, HLA-DQB and HLA-DPB. The combination of chains creates a peptide binding cleft or groove, and it is here the polymorphic regions are situated, where the peptide to be presented is loaded. This gives the ability for MHC-molecules to load and present a large variety of antigens\(^1\).

In T1D, about 50% of the genetic risk is considered to be contributed by the HLA
genes, mainly the DRB1 and DQB1 genes in the HLA class II region. There are two main risk haplotypes; DRB1*03-DQA1*05:01-DQB1*02:01 and DRB1*04-DQA1*03:01-DQB1*03:02 and one protective haplotype, DRB1*15:01-DQA1*01:02-DQB1*06:02 \(^{49}\) (shortly referred to as DRB1*03, DRB1*04 and DRB1*15). Individuals that carries both risk alleles (DRB1*03/DRB1*04) have the highest risk of developing T1D, and this is seen in 30-40% of diabetes cases compared to 2.4% of the general population\(^ {51}\). It has also been shown that the class I alleles HLA-A and HLA-B are the next in line large HLA association in T1D, and accounts for almost all of the remaining risk mapping to the HLA region after the class II genes. This correlates well with the known importance of CD8+ T-cells in β-cell destruction\(^ {52}\).

In MS, the main HLA risk haplotype is DRB1*15 (DQB1*06:02-DQA1*01:02-DRB1*15:01-DRB5*01:01\(^ {53, 54}\)) which surprisingly is the protective haplotype for T1D. There is also a protective haplotype in MS in the HLA class I allele A*02:01\(^ {54, 55}\). Other HLA-alleles have also been identified as associated in MS, most prominent DRB1*13:03 and DRB1*03:01\(^ {56}\). Due to the very strong LD in the region it has earlier been difficult to distinguish the association signal in the DRB1*15-haplotype in MS, but recent research show that the association primary is explained by DRB1*15:01\(^ {57}\).

In RA; a set of alleles in HLA-DRB1 with a common aminoacid sequence has been described to be responsible for the HLA association in RA. These alleles are collectively termed the shared epitope (SE) due to the shared protein motif which is exposed in the hypervariable region of the binding cleft\(^ {58}\). It has been demonstrated that this association is mainly valid for ACPA positive patient\(^ {59}\), and these factors also associates to the main environmental risk factor; smoking. However, recently it has been shown that aminoacids in three positions in HLA-DRB1 and two positions in HLA-B and HLA-DP almost completely explain the HLA association\(^ {60}\).

These associations of both HLA class I and class II alleles in T1D, MS and RA indicates the involvement of both CD4+ and CD8+ T-cells.

Exactly how the HLA genes biologically confer risk is not know with certainty, but many facts support the hypothesis that associated variants have influences on the structural function of these molecules\(^ {57, 58}\), perhaps affecting the very binding site in its efficiency of presenting self-peptides which will in turn have effects on the T-cell repertoire both in thymus and in the periphery.
1.3.1.2 non-HLA-genes

Genome–wide studies have revealed numerous associated genes in autoimmune disease, and many are also overlapping between the diseases. In RA\textsuperscript{31} and T1D\textsuperscript{61}, the Protein tyrosine phosphatase, non-receptor type 22 (PTPN22) is the largest contributor after the HLA genes, and it is also found to be associated to SLE\textsuperscript{62}, but not to MS\textsuperscript{63}. The Insulin gene\textsuperscript{64} is otherwise the strongest non-HLA gene in T1D, other T1D examples are Cytotoxic T lymphocyte antigen 4 (CTLA-4), also associated in Grave’s disease\textsuperscript{65} and Addisons disease\textsuperscript{66}, while Interleukin-2 receptor alpha chain (IL2RA) is associated to T1D\textsuperscript{67}, MS\textsuperscript{68} and RA\textsuperscript{67}. C-type lectin domain family 16 (CLEC16A) shows association in both T1D\textsuperscript{67} and MS\textsuperscript{69}, and the interleukin-7 receptor (IL7R) is also associated to MS\textsuperscript{70}. All together, there are now 110 established MS susceptibility loci outside the HLA\textsuperscript{63}, 101 RA loci\textsuperscript{71} and 40 T1D loci\textsuperscript{72}. Many of the identified genes are immune-related. However, for many of these genes the biological functional in disease remains unsolved. It is also worth remembering that for many of these association signals it is still unclear which gene is responsible for the association.

1.3.2 Environmental risk factors

As with the genes, there are environmental factors that are shared, and those that are unique for the different diseases. In RA, smoking is the strongest environmental risk factor\textsuperscript{36, 59}, other suggested risk factors have been mineral oils\textsuperscript{73} and silica dusts\textsuperscript{74}, as well as dietary factors however with weak results\textsuperscript{31}. In MS, viral infection seem to be the strongest environmental risk factor, mainly Epstein-Barr virus\textsuperscript{75}. Smoking\textsuperscript{76}, lack of sunlight/vitamin D deficiency\textsuperscript{77} and high BMI before 20 years of age\textsuperscript{78} has also been shown to participate in MS susceptibility. In T1D, the hygiene hypothesis is the most referred, stating that too good hygiene in development countries leads to few infections in childhood and less “trained” immune system, which in turn leads to more severe infections later in life that could trigger autoimmunity. Enterovirus, rotaviruses and rubella are suggested, as well as intestinal microbial balance. Dietary (cow’s milk, wheat protein) and other factor like vitamin D have also been demonstrated, but the effects are not large\textsuperscript{17} A general hypothesis for autoimmune diseases is molecular mimicry were a pathogen-component resembles self-proteins and trigger an faulty immune response. This have been suggested for P.gingivalis in RA\textsuperscript{79}, coxsackievirus and GAD in T1D\textsuperscript{17} and EBV and Myelin basic protein for MS\textsuperscript{80}. 
1.3.3 The missing heritability problem

Even if hundreds of susceptibility variants have been identified in complex diseases, most of them confer relatively small increase in risk, and can explain only a small proportion of the familial clustering or heritability. This lack of explanation for a large portion of the genetic risk expected from population studies is called the missing heritability. Many explanations for this missing heritability have been suggested:

- Many more variants of smaller effect are yet to be found and contribute to the heritability.
- Rare variants has a larger influence than expected, and these have been poorly detected by available genotyping arrays that focus on variants present in >5% of the population.
- Identified variants of modest effect are not the causal ones, and the same gene might harbor rare variants with large effect.
- Gene-gene interaction and gene-environment interaction is not accounted for, therefore the estimates of risk associated with a susceptibility variant may be too low and hence the heritability explained is estimated too low. It is also possible that we have failed to identify genetic associations because we have not allowed for interaction with other genes or environmental risk factors.
- Epigenetic modifications might affect the phenotype and contribute to heritability.

1.4 GENETICS

The Austrian monk Gregor Mendel (1822-1884) is regarded as the father of genetics, with his discovery of dominant and recessive inheritance patterns in pea plants. Monogenic diseases (example cystic fibrosis) follow the autosomal or X/Y-linked dominant or recessive patterns, and are often easy to follow in a family pedigree. There are also genetic diseases involving whole chromosomes, like trisomi 21 (Down’s syndrome), and mitochondrial diseases which involve the genes in the mitochondria and therefore only are inherited from the mother (mitochondria are generally not passed over from the sperm). However, the most common diseases with a genetic component are the complex, or multifactorial diseases, where several genes, environmental factors and interaction between genes and between genes and environment together cause the disease. Examples are cardiovascular disease, Type 2
diabetes and many autoimmune diseases like the ones discussed in this thesis.

1.4.1 General genetics
When the discovery of the DNA double helix was published in 1953 by Francis Crick and James D. Watson, it was an elegant solution to the question of how genetic information is held inside organisms and how it is passed from generation to generation. It has been the base of later discoveries and research like deciphering the genetic code and the Human Genome project (HUGO) which aim was to map all genes of the human genome.

DNA structure
The DNA (deoxyribonucleic acid) helix is built by paired nitrogenous bases with hydrogen bounds on a sugar-and-phosphate backbone. There are 4 bases which always pairs the same; adenine (A) pairs with thymine (T) and cytosine (C) with guanine (G). The unit of one base + sugar + phosphate is termed nucleotide and is the basic repeat unit of a DNA strand.

The genetic information is encoded by the linear sequence of bases in one strand, the primary structure. Two anti-parallel strands with complementary bases form the helix shape. When the cell is dividing, the DNA is replicated by DNA polymerase enzyme in a semi-conservative way using one strand as a template for a complementary strand, which results in two new molecules identical to the parental one (Fig.3).

The central dogma of molecular biology; the flow of information
The expression of a gene is considered to be a one-way system where the flow is DNA → RNA → protein. RNA –ribonucleic acid-is similar to DNA but is generally single stranded, based on ribose instead of deoxyribose, and the base thymine is substituted by uracil. Shortly, a gene is first transcribed with the aid of the enzyme RNA polymerase into single-stranded messenger-RNA or mRNA.
The mRNA is transported to the ribosomes, and here the genetic code is translated for the building of a peptide, which after post-translational processing will form part of a protein. Each set of three nucleotides on the mRNA is termed a codon and codes for one amino acid (Fig. 4). The region upstream of the gene to be transcribed contains the promoter; initiating site of transcription and binding motifs for the polymerase, as well as other motifs with regulating effects. The transcription can both be repressed (silenced) and enhanced by upstream sequences. After the gene is transcribed, the mRNA is modified by splicing. This takes away intervening parts of the code called introns, and put together the exons into the mature mRNA ready to be translated into protein. Different promoters or splice variants for a gene can be used in different tissues or cell types. The splicing mechanisms results in several variants of the gene and greatly increases the protein diversity.

Variations in the genome

There are many kinds of sequence variations in the DNA. Historically, this has a role in evolution though it contributes to population heterogeneity which improves the possibility for the species to survive environmental changes. Many variations are also used in research to identify susceptibility genes for complex diseases. These variations include insertions and deletions of segments (INDELs), copy number variations (CNVs) and variation in repeats. The number of repeats of a certain segment within a stretch of DNA is different between individuals, and can be used to distinguish individuals (as in crime scene investigations or paternity tests) but also in identifying disease-associated genes. Repeats are generally classified as microsatellites, or short tandem repeats, minisatellites, or variable number of tandem repeats (VNTR) and megasatellites depending on the length. While microsatellites are short repeats of 1-6 bases, minisatellites can be up to 90 bases and megasatellites are several Kb. The smallest variation is called a Single Nucleotide Polymorphism (SNP) and is the exchange of a single base to another in a specific position in the DNA sequence. Larger
variations can of course lead to dysfunction and damages like in Huntington’s disease where repeats grow longer for each generation, resulting in a dysfunctional protein. Small changes like SNP are often silent or synonymous; they do not change the aminoacid sequence. Some SNPs do change the code leading to a new amino acid (non-synonymous) at that position, or even a stop codon halting the translation process and resulting in a truncated protein. They can also more indirect affect transcription levels or splice variants of the gene. SNPs have proven to be very useful in genetic research owing to their distribution all over the genome and to that they are easy to measure with modern methods. The less common allele in a population is called the minor allele compared to the major allele, and the minor allele frequency (MAF) is measured in %.

**Linkage Disequilibrium**

When two markers are inherited together more often than expected, they are said to be in linkage disequilibrium (LD) with each other. It means that there is no random crossing-over or recombination in the meiosis where the gametes are made. The recombination event is a process for exchanging genetic material between chromosomes and increasing diversity. It may result in a new combination of alleles which can be evolutionary advantageous. Generally, markers close to each other are in higher LD than those further apart, but there are great variations in LD patterns and many things can influence the LD. For example, genetically conserved regions of high importance through evolution are generally in high LD, there are few crossing over events here due to the risk of losing important genes. This can be seen in the HLA area on chromosome six. Selection and mutations among other factors can affect LD patterns. Two markers inherited together are said to be on a haplotype. The distance for recombination between two markers can be measured in centi-Morgan (cM), where 1 cM is a 1% probability of crossing-over event between these two loci from one generation to the next. Physical distances on the chromosome are generally measured in base pairs (bp). The LD between two markers A and B is measured by calculating $D$, where $D = P_{AB} - p_1 q_1$. ($P_{AB}$ is the frequency of the AB-haplotype, and $p_1$ and $q_1$ is the allele frequencies respectively.) $D=0$ means there is no LD. Often, a derived form of $D$ is used, $D'$, which is normalized on the theoretically maximum of $D$, or $r^2$.
which is a correlation coefficient between alleles. For $r^2$, the range is 0-1 where 1 means complete LD. All these measurements are based on allele frequencies.

_Hardy-Weinberg equilibrium_

The Hardy-Weinberg equilibrium (HWE) principle states that allele and genotype frequencies in a population will remain constant from generation to generation. This assumption is only true when there are no evolutionary influences like for example non-random mating, mutations, selection or genetic drift. Of course this cannot be true for a real population where all these events occurs, but HWE can be used to test observed numbers of each genotype compared to expected numbers for markers in study samples. If there is a significant deviation, it can depend on some kind of sample bias for example genotyping errors, or population stratification. This in turn can lead to a faulty conclusion in that study, and therefore markers deviating from HWE are often excluded. The expected frequencies are calculated from the equation: $p^2 + 2pq + q^2 = 1$ where $p^2$= major allele homozygotes (AA), $q^2$=minor allele homozygotes (aa) and $2pq$=heterozygotes (Aa) and compared to the observed ones in a $\chi^2$ test.

1.4.2  Susceptibility genes and research approach in complex disease studies

It is important to remember that in complex diseases there are not any specific disease genes, just “normal” variation in genes that might confer an increased risk and in combination with other factors lead to disease. A lot of people have the same gene variant without getting the disease, and also individuals get the disease without that specific gene variant. Many gene variants might even have been beneficial during evolution, perhaps through a generally increased immune activation which may have been an advantage for surviving certain infections. Research is often directed at identifying susceptibility genes; genes with variants statistically associated with an increased occurrence of the studied disease. Different approaches can be used for this, either there is a pre-decided candidate gene to investigate or the search is hypothesis-free. Information for the choice of a candidate gene may come from animal models or it can be genes involved in immune functions thought to be part of the pathogenesis.
Linkage studies

In linkage studies genetic markers co-segregating with the trait are identified in families with several affected individuals. In parametric linkage analysis a model of inheritance of the disease is used and recombination between the trait loci and genetic marker is estimated, when the recombination fraction is low linkage is declared. A problem for autoimmune diseases is that there is no clear pattern of inheritance therefore non-parametric linkage is often used. In non-parametric linkage allele sharing between affected individuals is compared to that expected for markers not linked to the disease, linkage is declared when increased allele sharing is observed. Both microsatellites and SNPs can be used for linkage studies. One problem with the method is the low resolution; the area detected is often very large and can contain hundreds of genes. Also, while the method works well in monogenic diseases the low penetrance of complex diseases makes it hard to follow a trait. For autoimmune diseases linkage has successfully been identified in the HLA region, but few linkage signals have been identified for other susceptibility loci.

Cohort studies

In a prospective cohort study a group of people are followed during a set time-period and exposures are measured as well as outcome, normally how many people develop a certain disease. It is a very useful method, both when studying the effect of environmental factors like work conditions, smoking etc. and for genetic studies, but the drawback is that the cohort has to be unreasonable large (and hence expensive) in order to be able to study low-incidence diseases like autoimmune diseases. Also, retrospective cohort studies can be done based on already registered information on exposure and outcome. This might however be very limited or inconclusive data.

Case-control studies

In a case-control study the frequency of an allele or genotype among affected individuals is compared to the frequency in controls, or healthy individuals. If the allele is more common in the affected group, the marker is said to be associated to the disease (or phenotype). Information about exposures to environmental factors can also be added. However, it cannot be said with certainty that the specific marker is the
pathogenic one due to LD between markers. This study design has an advantage in 
time and costs efficiency, but care must be taken to control for certain factors;

1) Population stratification - due to heterogeneity between populations for 
allele frequencies, it is important that cases and controls are from the same 
population in order to avoid false association results. Principal component 
analysis (PCA) is a method to reduce the dimensionality of a dataset by 
analyzing the co-variance between variables, and is often used to remove 
population outliers.

2) Other confounding factors that might affect the results must be considered, 
for example age, sex or treatment. It can often be corrected for in a logistic 
regression model with the factor as a covariate.

3) The definition of the affected group or the trait studied is also important, 
some diseases might have subgroups and if the association of a gene variant 
is different in these subgroups, the total association outcome could be falsely 
negative or missed due to dilution (loss of power). Also, studying self-
estimated parameters (for example smoking the latest 10 years) can be 
affected by recall bias between cases and controls which could result in 
declaring a false association.

**GWAS; genome-wide association studies**

Technical developments the last years have introduced the possibility to perform large 
genetic scans throughout the genome in very large case-control cohorts. The method 
is based on LD patterns, and tag-SNPs are chosen to represent LD blocks covering the 
genome. When it was first introduce in mid-2000\textsuperscript{67}, there was high hopes to find many 
associated variants for complex diseases and that it should be possible to pinpoint the 
causal genes. Even though there have been many findings, only a part of the 
heritability in these diseases can be explained by the identified genetic variants.

Criticism has also been raised to the fact that the majority of found variants have no 
established biological relevance to disease or possible implication in prognosis or 
treatment. A limitation to the model is the “common-disease-common variant” 
theory; only variations with a minor allele frequency (MAF) of more than 5% are 
cluded on the genotyping chips, leading to that rare variations may be missed\textsuperscript{28}. It is 
now discussed whether the magnitude of rare variants is larger than first estimated.
Also the significance level for association in GWAS is quite stringent \((P=5\times10^{-8})\) to avoid false positive findings when so many statistical tests are performed, which will exclude many findings of less strength.

*Family-based association studies*

While the case-control study is population based, association studies can also be performed in families. Here the transmission frequencies of alleles or haplotypes associated with disease from parents to affected children are evaluated. This study design has the advantage that there is no problem with population stratification.

*Animal models*

Often experimental models of the human disease are studied in animals. It has the advantage that experiments not possible to perform in human can be made, for example inducing disease, remove organs to study cell infiltrations, etc. Also, in genetic studies inbred strains of rat or mice which are homozygous at all position can be used to pinpoint causal genetic variations. Different inbred strains with different susceptibility for disease can be crossed in order to identify risk loci through linkage analysis. Another method is the use of congenic animals with an identical genetic setup except for the position \((\text{locus})\) to be studied. The shortcoming is of course that mice are not humans, and even if some genes are corresponding in the two species, results from animal studies cannot be directly translated into humans. Also, the autoimmune diseases studied are similar and well documented in many models, but still there are differences that can be of profound effect to the human variant. Animal models have proven very useful to identify candidate genes for further human studies.

1.5 MATERIALS & METHODS

1.5.1 Cohorts

All included cohorts, patient material and analyses in our studies were approved by Regional Ethical Review Boards in respective city or country. Informed consent from all study participants or their parents were obtained. Investigations were carried out according to the guidelines from the Declaration of Helsinki. All individuals of known
non-Scandinavian or western European origin were excluded from the studies. Many
different cohorts have been used in the studies in this thesis, below follows a short
overview. More details on each cohort are described in the corresponding paper.

**T1D patients and controls (paper I and II)**

*Diabetes incidence study in Sweden 1 (DISS1)*
The DISS1 cohort consists of DNA samples from 839 incident T1D patients in the
Diabetes Incidence Study in Sweden (DISS) registry\(^87\), diagnosed between 1987 and
1989 at the age of 15–36 years, as well as 625 sex, age and residence- matched
controls.

*Diabetes incidence study in Sweden 2 (DISS2)*
The DISS2 cohort consists of DNA samples from 778 incident diabetes patients aged
15–36 years and from the DISS registry during 1992 and 1993, and 836 sex- and age-
matched controls\(^88\). In all, 586 of the patients were classified with T1D by the treating
physician at yearly follow-up and these subjects are included in this study.

*Swedish childhood study (SV2)*
A total of 497 cases of children between 0–14 years with newly diagnosed T1D were
collected from the Swedish Childhood registry 1986-87\(^89\). Controls were
geographically, gender and age matched to all cases over 7 years of age. For patients
under 7 years, a control was selected among patients being treated at the hospital for
reasons other than T1D (n=53) due to restrictions in the ethical permit.

*Better diabetes diagnosis (BDD) study*
A total of 2700 incident diabetes patients under 18 years at diagnosis were collected
between 2005 and 2009 from 40 pediatric clinics in Sweden for the BDD study\(^90\).

*Diabetes registry (DR) in Southern Sweden*
A total of 804 T1D patients, 436 men and 368 women, with onset age between 1 and
75 years of age, from the Diabetes Registry in Southern Sweden\(^91\), were all enlisted at
the Department of Endocrinology at Malmo University Hospital, Sweden, and
collected between 1996 and 2005. Additionally, 2312 healthy controls, 1695 men and 617 women between 45 and 75 years of age were added to the cohort. Because of risk for overlap among patients in the DISS2 and DR cohorts, those individuals that could possibly occur in both cohorts were identified and removed from the DR cohort (n=73).

**MS patients and controls (paper I, II and III)**

All MS cases have been diagnosed either according to McDonald’s criteria\(^\text{23}\) or Poser’s criteria\(^\text{92}\).

**The Epidemiological Investigation of Multiple Sclerosis (EIMS)**

A population based nation-wide case–control cohort with incident cases of MS that has been described previously\(^\text{76}\). The controls were randomly selected from the national population register and matched to the case’s sex, age and residential area; 625 cases and 663 controls from this group were used in the current studies.

**The Immunomodulation and Multiple Sclerosis Epidemiology cohort (IMSE)**

The cohort consists exclusively of cases (n=318) with relapsing-remitting MS from clinics throughout Sweden who are treated with natalizumab\(^\text{93}\).

**The Stockholm Multiple Sclerosis cohort (STOP MS).**

The patients were recruited by neurologists at the Karolinska University Hospital Huddinge and Solna sites in Stockholm, Sweden. The patients in the cohort were between 22 and 91 years of age and the controls were matched for ethnicity and constitutes of blood donors between 21 and 76 years of age. A total of 1060 cases and 1215 controls were used from this group\(^\text{94}\).

Samples from 29 MS patients, 4 CIS cases and 23 patients with other neurological diseases are included in paper III from this cohort.
RA patients and controls

The Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) (paper I, II, III, IV)

The Swedish RA cohort (EIRA) is composed of 3026 incident RA cases (1943 ACPA positive, 1079 ACPA negative and 4 unknown) and 2202 population-based controls, matched by age, sex and residential area. The cases were collected from clinics in the middle and southern parts of Sweden and diagnosed according to the 1987 American College of Rheumatology criteria for RA by rheumatologists at the different clinics. The controls were randomly selected from The Swedish National Population registry and matched to the cases on age, sex and residential area as described.

The French RA cohort (paper IV) constitutes of one hundred unrelated trio families (one RA patient and both parents) 78 were ACPA positive, 21 were ACPA negative and 1 was of unknown ACPA status. The RA families were recruited through a national media campaign followed by selection of individuals who fulfilled the 1987 American College of Rheumatology criteria for RA according to the physicians in charge of the patients.

Additional controls from the studies of:

Myocardial infarction—SCARF (paper I and II)

From the SCARF study, the control group consists of 387 sex- and age-matched healthy persons between 40 and 60 years of age, and recruited from the general population of the same county as the cases with MI, of self-reported Caucasian origin.

Alzheimer’s disease—SNAC (paper I)

A total of 424 healthy controls of 60–73 years of age, from a longitudinal study of AD (the Swedish National Study on Aging and Care in Kungsholmen - SNACK, in Stockholm, Sweden). Also, originally added to this cohort are 39 individuals, which are autopsy cases from the Karolinska Brain Bank who died from cardiovascular or malignant diseases, between 56 and 91 years of age and without a medical history of dementia.
Population-based control cohorts from osteoporosis study (PEAK-25, OPRA) (paper I)

The PEAK-25 cohort consists of 1005 healthy women, all 25 years old and randomly selected from the Malmö city files between 1999 and 2003. The second cohort consists of 1010 healthy women from the Malmö OPRA study, all aged 75 years and randomly selected from the Malmö city files between 1995 and 1999. Individuals in both groups are all of Swedish or Northeuropean ancestry.

1.5.2 PCR

The polymerase chain reaction (PCR) is the base for many developed techniques in molecular biology, like genotyping methods and qPCR. Kary Mullis was awarded the Nobel Prize in Chemistry in 1993 for this invention. The template (DNA), primers for the target, a mix including free dNTP’s (nucleotides) and heat-stable DNA polymerase is mixed in a test tube or plate. Generally, the protocol includes three steps (for short sequences the elongation step can be excluded);

1) Denaturation – heating to 95°C separated the DNA strands
2) Annealing – the primers binds to the DNA sequence, temperature depends on the primers melting point, but around 50-60°C. The polymerase initiates.
3) Elongation – around 72°C the polymerase activity incorporates the dNTP’s and synthesizes a complimentary DNA strand to the target.

Around 40 cycles of the process is repeated to generate a sufficient amount of the target for further study, for example allelic discrimination of a SNP.

1.5.3 ELISA

Enzyme-linked Immunosorbent assay (ELISA) paper IV

ELISA is a method that uses antibodies and color change to identify presence or absence of a specific substance. The sample investigated is coated onto plates, and a specific antibody is added that will bind to the substance if present. This antibody is linked to an enzyme and in the final step the enzyme’s substrate is added. The subsequent reaction produces a detectable signal, most commonly a color change.

The measurement of ACPAs for samples included in the EIRA study was performed with Immunoscan-RA Mark2 anti-CCP2 ELISA (Euro-Diagnostica, Arnhem, The Netherlands) and samples with antibody levels above 25 U/ml were regarded as positive according to instructions provided by the manufacturer.
1.5.4 Genotyping
Genotyping is the process of identifying the variation (for example a SNP) at a specific locus in DNA from an individual. Most genotyped SNPs in this thesis can be found on the NCBI dbSNP homepage; [http://www.ncbi.nlm.nih.gov/snp/](http://www.ncbi.nlm.nih.gov/snp/). Different methods of genotyping have been used for different cohorts as described in the papers:

**Taqman ABI 7900 (paper I-IV)**
The majority of SNPs have been genotyped with the allelic discrimination method for Taqman ABI 7900 HT machine (Applied Biosystems, Foster City, California, USA). Assays where ordered from Applied Biosystems and genotyping was performed according to the manufacturers protocol as described. Shortly, the assay consists of two probes, specific for each allele of the SNP, with a fluorophore each (VIC and FAM) coupled to a quencher. The quencher will be cleaved of by Taq polymerase once the probe is bound to the DNA, and the fluorescent signal can be detected using the SDS 2.2.1 sequence detection software.

**DASH- dynamic allele-specific hybridization (paper I and II)**
Genotyping was performed according to the manufacturers protocol. Shortly, a biotinylated DNA strand is bound to a streptavidin-coated plate, and an allele specific probe is added as well as SYBR green dye which binds to double-stranded DNA. As the temperature is increased, the probe dissociates and a melting curve can be detected. The SNP is distinguished on how well the probe binds to the DNA which will lead to a lower melt-point for a SNP.

**MassArray chip-based matrix-assisted laser desorption/ionization time-of flight mass spectrometer (paper I and II)**
MALDI-TOF or iPLEX (Sequenom Inc., San Diego, CA, USA) using the Homogeneous MassEXTEND chemistry as described. The method uses the different mass of the DNA strands to distinguish between the alleles, and many SNPs can be analyzed in the same pool.
The Immunochip project is collaboration for the study of susceptibility genes for autoimmune diseases. Around 200,000 markers in 250 genes where genotyped with the Illumina Immunochip, alleles were called using Opticall. QC of markers was carried out as described\textsuperscript{63}. Genotypes for 70 SNP markers between position 10878975 and 10930150 (Chr16_pos build 36.3) were included in paper III.

1.5.5 HLA typing and imputation of HLA types
Since HLA has such a central role in the susceptibility to autoimmune disease, the individual haplotype or genotype for a specific variant is often included in our studies also when investigating the effect of other factors. There are several methods to genotype the HLA area, and different methods have been used for the many cohorts in this thesis;

\textit{DISS1 and SV2:} Restriction Fragment Length Polymorphism (RFLP) was used for DR typing, and genotyping for DQB1, DQA1 and DRB1 was performed with PCR amplification followed by dot blot hybridizations\textsuperscript{104}.

\textit{DISS2:} HLA genotyping for DQB1, DQA1 and DRB1 was performed with PCR amplification followed by dot blot hybridizations and by RFLP as previously described\textsuperscript{63} except that allele-specific PCR amplification (PCR-SSP) of DRB1 alleles was also used\textsuperscript{105}.

\textit{BDD:} A method based on an asymmetrical PCR and a subsequent hybridization of allele-specific probes was used, as described previously\textsuperscript{106}. Established haplotypes in the European population were used to determine DR genotypes in the BDD cohort where only DQA1 and DQB1 were genotyped.

\textit{MS and RA cohorts:} HLA typing of DRB1*15 in MS and RA cohorts and A*02 in the MS cohorts was performed by allele-specific amplification as described earlier\textsuperscript{105}.
For individuals also included in the MS Immunochip study\textsuperscript{63} we used imputed genotypes for HLA where we lacked this information. The imputation is based on LD structures and is used to reduce the need for genotyping. Reference panels of known haplotypes are utilized (HapMap, 1000 genomes). In \textit{paper III} it was performed using HLA*IMP:02\textsuperscript{107} based on SNPs genotyped on the Immunochip custom array\textsuperscript{108}.

1.5.6 Expression studies
To study the expression of genes the mRNA is collected from cells and complementary DNA (cDNA) is synthesized with the aid of reverse-transcriptase (RT) enzyme, the
process is called RT-PCR. This enables the amplification of the product with a standard PCR reaction with primers for the gene of interest. To measure the corresponding original amount of mRNA, the amount of PCR product is measured after each reaction cycle. The measurement is done by adding a dye (ex. SYBR green) that binds to dsDNA, or a specific probe with a fluorescent reporter that will bind to the PCR product and the fluorescence can be detected by a laser. This process is termed real-time PCR, or quantitative-PCR (qPCR)\textsuperscript{109}. The measured values are compared to an internal control, often a “house-keeping” gene that has a stable expression to be able to compare increase or decrease in expression.

When more than a few genes are studied, a more efficient method is required. In RNA sequencing all the mRNA from a sample is instead sequenced using next generation sequencing and the numbers of reads mapping to the specific targets are used as a measure of expression of that target.

\textit{Cell cultures and RNA extraction (paper III)}

The cell culture procedures are described in paper III and only shortly described here;

RA samples;
Peripheral blood mononuclear cells (PBMCs) were collected in Sodium citrate vacuum tubes (Vacutainer, Becton-Dickinson, Franklin Lakes, NJ, USA), cleaned, viable cell counted and cultivated with stimuli (5*10\textsuperscript{5} cells, IFN-gamma) for 6 hours at 37°C. After incubation cells were lysed with RLT buffer from the mini RNA prep kit and subsequently RNA extracted (RNeasy Mini, Qiagen, Hilden, Germany). mRNA was stored at -80°C prior to cDNA synthesis, which was performed with iScript cDNA synthesis kit following the manufacturer’s protocol (BioRad, Hercules, California USA).

MS samples;
PBMCs were isolated immediately after taking whole blood from the individuals with BD Vacutainer™ CPT™ Tube (Becton-Dickinson, Franklin Lakes, NJ, USA) and stored in liquid nitrogen, diluted in freezing medium containing Fetal Calf Serum (FCS) and 20% DMSO. For the experiment, 50,000 cells /well were stimulated for 21h with 1μg/mL Concanavalin A (ConA) and unstimulated cells for the same period of time was used as controls. RNA extraction and DNA synthesis was performed as described for the RA samples above.
Quantitative reverse transcriptase polymerase chain reaction (qPCR) (paper III)

Individual CIITA transcripts were assessed with transcript-specific qRT-PCR primers (positioned over exon boundaries, see details in paper III), which were used with SYBR Green dye (iTaq SYBR Green Supermix, Bio-Rad, Hercules, California) in triplicate and measured using a BioRad CFX384. For CD74, a BioRad iQ5TM iCycler Detection System (Bio-Rad Laboratories, Ltd) was used. GAPDH and β-glucuronidase were used as endogenous controls.

RNA sequencing (paper III)

Total RNA was isolated from PBMC as described above for MS samples, but immediately after isolation it was pelleted and kept in -80 degrees. cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, CA, USA), as dictated by the TruSeq protocol. Sequencing was carried out on Illumina Hiseq machine (Illumina, CA, USA) at Science for Life Laboratories (Stockholm, Sweden) using a paired end design at an average depth of 24x10^6 reads per sample. Sequence reads were mapped to human reference sequence build 19 using Tophat. The trimmed mean of M-values (TMM) normalization method is used to report the gene expression and TMM values of the target genes were used for correlation to CIITA genotypes.

1.5.7 Animal models

Congenic rat strains were used for studying the effect of the VAV1 locus using two different methods for inducing arthritis in paper IV.

Generation of congenic animals

The DA.BN-Eae4 congenic strain was developed using a speed congenic strategy as described previously by crossing the RA susceptible strain Dark Agouti (DA) with the RA resistant strain Brown Norway (BN). The resulting DA.BN-Eae4 (also named DA.BN-R25) strain has thus a DA background containing a ~2 centimorgan region from the BN strain in chromosome 9.
Induction and evaluation arthritis

Aged matched DA (Hannover) and DA.BN-R25 congenic male rats were immunized by injection, at the base of the tail, of 100µl Pristane (2, 6, 10, 14-tetramethylpentadecane; (ACROS Organics) for Pristane-induced arthritis or by injection of 100µg of pepsin-digested rat collagen type II (as described in [112]) emulsified in 100µl incomplete Freund’s adjuvant (Difco) for collagen-induced arthritis. Immunizations were performed under Isoflurane anesthesia. DA and DA.BN-R25 rats were mixed in cages and each animal was scored in a blinded manner every second or third day from day 10 to day 27. Arthritis development was monitored using a macroscopic scoring system as previously described[113].

Serum levels of anti-collagen antibody

MaxiSorb plates (Nunc) were coated over night at 4°C with 10µg/ml of pepsin digested rat collagen type II followed by blocking for 1 hour at room temperature with 2% milk. After extensive washing, sera (1:1000 dilution) from the day the experiments were terminated were added and plates were incubated at room temperature for 2 hours and washed. A secondary biotinylated antibody for total IgG (Goat-anti rat IgG H+L, Jackson ImmunoResearch) was added and plates were incubated at room temperature for 1 hour and washed. Streptavidin-Europium diluted in ASSAY buffer (PerkinElmer) was added and plates were incubated for 30 minutes and then washed. Enhancement solution was added and fluorescence was measured in a Victor/1420 Multilabel counter (both from Perkin Elmer).

Serum levels of rheumatoid factor

MaxiSorb plates (Nunc, Roskilde Denmark) were coated over night at 4°C with 10µg/ml Rabbit IgG (Jackson ImmunoResearch) followed by blocking for 1 hour at room temperature with 2% milk. After extensive washing, sera (1:50 dilution) from the day the experiments were terminated were added and plates were incubated at room temperature for 2 hours and washed. Biotinylated secondary antibodies for total IgG (Goat-anti rat IgG H+L, JacksonImmunoResearch) and IgM (mouse-anti rat IgM, BD Biosciences) were added and plates were incubated at room temperature for 1 hour and washed. Streptavidin-Europium diluted in ASSAY buffer (PerkinElmer) was added
and plates incubated for 30 minutes and then washed. Enhancement solution (PerkinElmer) was added and fluorescence was measured in a Victor/1420 Multilabel counter (both from Perkin Elmer).

1.5.8 Statistical analyses

*Tag-SNPs and haplotype blocks (paper I)*

Genotypes from 28 SNPs in the *CIITA* gene in a selected cohort of 373 controls were used to analyze LD block structure and to identify haplotype-tagging SNPs. The results from this analysis revealed five LD blocks, and common haplotypes in each block could be resolved by typing 1–3 haplotype tagging SNPs. LD blocks and haplotype-tagging SNPs were accomplished using the HapBlock analysis program\(^\text{114-116}\). The block partitioning algorithm was set as the dynamic programming algorithm\(^\text{114}\), the common haplotype method was used for block partitioning and the method for identifying haplotype-tagging SNPs was capable of identifying all common (45%) haplotypes.

**Association**

Univariate association was tested using Pearson’s \(\chi^2\) test for differences in genotype and allele frequencies between cases and controls.

Logistic regression analysis using generalized linear modeling between cases and controls was used in multivariate analysis to correct for the effect of age and HLA. HLA coding was generally defined as presence or absence of DRB1*15 and DRB1*03 and DRB1*04 alleles. For DRB1*04, only individuals with the DRB1*04-DQB1*03:02 haplotype were considered positive for DRB1*04.

Association between genotypes and expression levels in the RNAseq cohort was carried out with linear regression model taking into account significant confounders (batch in cDNA library prep for *HLA-DRA*, *DEXI* and *CIITA*, and age for *CIITA*) using the statistical software PLINK\(^\text{117}\).

For comparison of transcripts between cases and control groups in *paper III*, the non-parametric Kruskal-Wallis test was used. For analysis of genotype effect on transcript levels for more than 2 groups Man-Whitney test was use.
**Odds ratio (OR)**

OR is often presented with its 95 % confidence interval, and this describes the effect of the associated variable. Often it is described as the *risk*, even if it is not really applicable in a case-control cohort. However, in a rare disease the risk ratio (RR), or relative risk is ≈odds ratio

Example:

<table>
<thead>
<tr>
<th></th>
<th>exposed</th>
<th>non-exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>cases</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>controls</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Odds ratios are calculated by the odds among cases / odds among controls:

\[
\frac{a}{b} = \frac{ad}{cb}
\]

For a significant association, OR>1 indicates increased risk, while OR<1 indicates decreased risk, or protection.

**Age-stratification (paper I)**

The controls were divided into 15 5-year intervals with respect to age at sampling (0–4, 5–9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69 and >70 years) (for the rs3087456 SNP there was one more agegroup; >75). The \( \chi^2 \) test for trend in proportions was used to detect the overall trend in variation of genotype over age.

**Meta-analysis (paper I and II)**

An age-stratified meta-analysis test of association in *CIIA* with T1D in the combined cohort of five case–control studies was performed (paper I) for the first eight age groups (0–39 years) using fixed effect Mantel–Haenszel analysis and Woolf’s test for heterogeneity in R using the meta.MH command in the rmeta package (R package version 2.16. [http://CRAN.R-project.org/package=rmeta](http://CRAN.R-project.org/package=rmeta)). In rs3087456_GG, age group 1 (0–4 years) was removed from the analysis due to heterogeneity between groups (\( P=0.04 \)). In paper II, we conducted a meta-analysis in the same way on the material presented by us and that of Bronson et al.\(^{118} \) with stratification of cases for the presence of DRB1*15. No heterogeneity between groups was discovered here.
**Interaction**

Statistical interaction between two risk factors is said to occur when the risk for individuals with double exposure is higher than expected from the risk for each exposure separately. The departure from the expected can be measured on different scales e.g. Multiplicative and additive. Interaction on the multiplicative scale was tested using logistic regression including both variables (a, b) investigated, as well as the interaction variable (axb) and confounders (age). In this model it is assumes that the risk for individuals with double exposures is the product of the risk for individual exposures, when there is a significant departure from this it is assumes that the two risk factors interact. In additive interaction, departure from additivity was estimated by calculating attributable proportion (AP) due to interaction which measures the proportion of disease among individuals exposed to both risk factors that is attributable to the interaction. The additive model assumes that the risk for individuals exposed to two risk factors equals the sum of the risk for the individual risk factors. Deviations from this, which is detected as AP ≠ 0, indicated that there is interaction between the risk factors. According to the sufficient cause model proposed by Rothman\textsuperscript{119} such an interaction indicates that the two factors are involved in the same sufficient cause for the disease. (The analysis was performed as described previously\textsuperscript{120} using the generalized linear modeling (glm) in R and the vcov command to get the covariance matrix.) Both measurements concerns statistical dependence and does not tell us about biological interaction on a protein or cellular level.

**Statistical software programs**

Association analyses were performed in the statistical computer program R version 2.6.2\textsuperscript{121} and in Unphased 3.1\textsuperscript{122} using the cocaphase command. Visualization and calculations of transcript and genotype data for paper III was done using GraphPad Prism v5 software program. Linkage Disequilibrium structures was analyzed and illustrated using the program Haploview 4.2\textsuperscript{123}
1.5.9 Genes studied in this thesis

In this thesis, I have focused on two candidate genes that I have studied during my PhD education. Both genes had a history of findings in our group, based in animal models. The aim was to further investigate these genes in the settings of the initial or other autoimmune diseases in humans, and to some extent explain some of the relation to biological function in disease.

1.5.9.1 CIITA

The official name for the CIITA gene (also called C2TA, MHC2TA) is class II major histocompatibility complex transactivator. It is situated on chromosome 16p13. Four different promoter variants have been described in human, three of these seem to be functionally expressed\(^\text{124}\). The different promoters (PI-IV) have been shown to be used to different extent in different celltypes. PI mainly controls CIITA expression in myeloid dendritic cells and macrophages, whereas PIII is active in B cells, activated T cells and plasmacytoid dendritic cells, and the PIV promoter regulates IFNγ-inducible CIITA expression in cells of non-hematopoietic origin and in thymic epithelium. The function of the PII promoter in humans has not yet been fully characterized\(^\text{125}\). The encoded CIITA protein acts as a positive regulator of class II major histocompatibility complex (MHC) gene transcription. Dysfunction of the CIITA gene results in type II bare lymphocyte syndrome\(^\text{126}\) where all cellular expression of the MHC class II molecule is missing, leading to severe immunodeficiency. The CIITA gene is also found to be conserved in chimpanzee, rhesus monkey, dog, cow, mouse, rat, and zebrafish further indicating its importance\(^\text{127}\).

Functionally, the CIITA protein is not a DNA-binding transcription factor itself, but rather acts as a co-activator in that it binds to the enhanceosome of transcription factors in the promoter region of MHC class II and initiates transcription\(^\text{128}\). The transcription factors are constitutively expressed and the regulation of expression of MCH class II genes is dependent on CIITA, which also has been defined as the master control factor for expression of the MHC class II genes due to its crucial role\(^\text{129}\).
1.5.9.2 VAV1

The VAV1 gene is part of the family of VAV (VAV1-3) genes. VAV1 is located on chromosome 19p13.2 and was initially identified as a proto-oncogene. Generally, the VAV proteins are cytoplasmic guanine nucleotide exchange factors (GEFs) for Rho family GTPases that transduce extracellular signaling by activating pathways leading to actin- and cytoskeletal remodeling, cell migration and gene expression. While VAV2 and VA3 are more ubiquitous expressed, VAV1 is expressed only in hematopoietic cells and activated in response to T-cell antigen receptor (TCR) stimulation. It has importance for both positive and negative selection of T-cells in the thymus, as well as in T-cell activation. It is also a gene highly conserved between species (chimpanzee, dog, cow, mouse, rat, fruit fly, and mosquito). Several isoforms has been observed, depending in splice variants of the gene transcript.
2 STUDY AIMS

The overall aim in this thesis is to study the association of 2 candidate genes, CIITA and VAV1, to autoimmune disease, and investigate the possible biological background for the identified association.

In paper I, we aimed to further investigate a genetic area on chromosome 16 where our group earlier had found linkage to T1D, and CIITA is situated. We wanted to evaluate CIITA as a candidate gene for T1D and study its interaction with HLA. We also aimed to analyze genotype variation within the CIITA gene depending on age in a large control cohort.

In paper II we aimed to reproduce the association reported to MS for CIITA and to study its interaction with HLA. The goal was to find further support that variability in the CIITA gene affects autoimmune disease risk.

In paper III we wanted to functionally evaluate the role of SNPs in CIITA as expression quantitative trait loci (eQTLs) affecting CIITA transcript expression, but also neighboring genes and HLA genes. We studied control of CIITA expression in both MS and RA, which show association to different parts of the CIITA gene and we hoped to better understand the connection of the genetic association to a functional variation in these diseases.

In paper IV, we aimed to investigate a suspected shared risk gene between MS and RA. The VAV1 gene, earlier found to confer risk in MS, was analyzed in Swedish and French RA cohorts. By using two animal models, theoretically different in pathogenesis of arthritis, we try to explain to some extent the role of VAV1 in ACPA subgroups of RA patients.
3 RESULTS AND DISCUSSION

3.1 THE ROLE OF CIITA IN AUTOIMMUNE DISEASE (PAPERS I-III)

The MHC class II transactivator gene – *CIITA*, has been of interest as a susceptibility gene for autoimmune diseases for a long time. This is due to the fact that *CIITA* codes for a transcription factor that acts as an assembler of all transcription factors necessary to express the MHC class II genes\(^{135}\). These genes are by far the most strongly associated genes in many autoimmune diseases. In paper I-III we have studied the association of *CIITA* to T1D and MS, we have studied influencing and confounding factors as well as interaction with HLA, and genetic regulation of expression depending on markers in *CIITA* in MS and RA.

3.1.1 Results from paper I

Our group have previously found linkage in DRB1*03/DRB1*04-positive T1D patients to the region of chromosome 16 where *CIITA* is situated\(^{136}\). In this paper we performed a thorough investigation of *CIITA* association in T1D. Initially, we genotyped 41 markers evenly dispersed over the gene in a cohort of 373 controls. Successful markers were then used to analyze LD structures in the gene and based on that chose Tag-SNPs representing the common haplotypes in LD blocks. The selected markers were then genotyped in a T1D case-control cohort (DISS2) and here we found association to SNPs in an area between promoter I and promoter III. We therefore extended the analysis to include markers upstream the promoter I to define the associated area. After quality control and correction for multiple testing, significant association (p<0.05) to one SNP remained; rs11074932 situated upstream promoter III but in high LD with markers in the promoter region.

We were aware that there was inconclusive evidence for association to MS and RA\(^ {137-140}\) for markers in this region, and we wanted to investigate if confounding factors could affect the varying outcomes of association studies to *CIITA*. Age was one such factor since different control groups varying in age had been used for the different studies. We performed a test- for- trend analysis for correlation between age at sampling and allele frequency in a collection of controls from studies of T1D, MS, RA, Myocardial Infarction (MI) and Alzheimers Disease (AD), between 3300 up to 7300
samples for three markers in the \textit{CIITA} gene. The results showed a significant variation of genotype frequency depending on age for two markers in the investigated samples, such that the minor allele homozygote was more rarely occurring in older age-groups compared to younger (rs11074932, \(P=4\times10^{-5}\); rs3087456, \(P=0.05\)) (paper I, Fig. 1).

This finding was replicated in an independent cohort of 1000 samples from 25-year old women and 1000 samples from 75-year old woman, all from a study of osteoporosis. Here we found significant variation for all three markers (rs11074932, \(P=0.006\); rs3087456, \(P=0.007\); rs774, \(P=0.03\)) (paper I, Fig. 2).

This finding could explain some of the varying results in \textit{CIITA}-association studies. We then performed age-corrected association analysis for the investigated \textit{CIITA} SNPs in a combination of T1D cohorts with over 3400 cases and 3800 controls. Association remained for the two markers in the promoter III area (rs11074932, \(P=0.004\) and rs3087456, \(P=0.001\)).

In an age-stratified analysis for the different genotypes of the markers we observed that for both the associated markers it was the major allele homozygote that was associated with disease (rs11074932-TT, OR: 1.24, 95\%CI: 1.07-1.44; rs3087456-AA, OR: 1.27, 95\%CI: 1.11-1.45).

We also showed that this association was independent of the nearby gene \textit{CLEC16A}, established as a T1D-risk gene with genome-wide significance.\cite{141}

With the role of \textit{CIITA} as the key protein in the control of expression of MHC class II alleles and the previous observation of linkage in the \textit{CIITA} region being affected by HLA genotype in mind, we also performed interaction analysis between \textit{CIITA} and T1D-associated HLA-haplotypes (\textit{DRB1*03}, \textit{DRB1*04} and \textit{DRB1*15}). No multiplicative interaction was detected, but for both markers additive interaction with \textit{DRB1*15} was detected, such that lack of the protective \textit{DRB1*15} haplotype increased the “risk” associated with the major allele homozygotes of the markers (rs11074932, AP: \(-0.32\), 95\%CI: \((-0.50)\)\textendash\((-0.09)\); rs3087456, AP: \(-0.36\), 95\%CI: \((-0.59)\)\textendash\((-0.13)\) (paper I, Fig.5).

### 3.1.2 Results from paper II

Previous reports\cite{118,137} had found association to MS for markers in \textit{CIITA}, and also stronger association in \textit{DRB1*15} stratified MS patients. In this paper we reproduce some of these findings, and also show that the effect is modulated by MS-associated
HLA haplotypes. Our data supports the hypothesis that variability in the *CIITA* gene has influence on MS risk.

Initially, association between rs4774 and MS was detected in a combined cohort of 2000 MS cases and up to 6900 controls (P=0.01, OR: 1.17, 95%CI: 1.03–1.31). When we stratified for MS associated HLA DRB1*15 and A*02 alleles, we observe an increased effect (P=0.03, OR: 1.23, 95%CI: 1.02–1.49) and (P=0.01, OR: 1.33, 95%CI: 1.07–1.64) respectively. These associations are shown to be independent of the adjacent confirmed MS susceptibility gene *CLEC16A*.

In interaction analysis we discover interaction between rs4774 and HLA; additive interaction between rs4774 and DRB1*15:01 such that individuals carrying the risk allele for rs4774 and DRB1*15:01 have a higher than expected risk for MS (attributable proportion: 0.17, 95%CI: 0.01–0.33, P=0.04) (paper II, Fig.1), and also multiplicative interaction for rs4774 and A*02 (P=0.03, OR: 1.40, 95%CI: 1.02–1.90).

### 3.1.3 Results from paper III

To follow up on our genetic association findings in T1D and MS we were interested in performing functional studies on *CIITA* gene expression in the settings of autoimmune disease. Here we collaborate with a research group that studied *CIITA* in RA, and expand the paper with our results from MS cohorts. The previous findings showed that there was a difference in expression of the two main *CIITA* transcripts in RA patients (n=44) compared to controls (n=48). There was also evidence of genetic regulation of expression of the two most abundant transcript isoforms of *CIITA*, *CIITA*\_pIII and *CIITA*\_pIV, in peripheral blood mononucleated cells (PBMCs) from RA patients depending on SNPs in *CIITA* promoter area.

We performed analyses of the same *CIITA* transcripts, but also CD74, the MHC class II invariant chain, in PBMC from a cohort of 33 patients with MS and 23 patients with other non-inflammatory neurological disease (OND) such as depression or headache. No difference in expression levels of *CIITA*\_pIII, *CIITA*\_pIV or CD74 between MS and OND cases was discovered.

We chose 4 markers based on earlier association to MS (rs4774, rs3087456) and the LD blocks defined in the RA cohort (rs6416647, rs4781009) and investigated their influence on the mRNA levels of the mentioned genes. No effect on either of the *CIITA* transcripts where discovered, however we observed that both rs4774 and rs6416647
had a significant effect on CD74 transcript levels, (P=0.03 and P=0.02, respectively) and rs3087456 had a suggestive effect (P=0.06).

We also had access to genotypes for 132 samples from MS and OND patients from the Immunochip project with genome-wide expression data generated by sequencing RNA. 70 markers in and close to CIITA was analyzed for eQTL-effects on CIITA, neighboring genes DEXI and CLEC16A (an established MS susceptibility gene) and MHC class II molecule transcripts; CD74, HLA-DRA and HLA-DQA1. Here we found several markers that affected HLA-DRA transcription, mainly one SNP that remained significant after Bonferroni correction (rs58497481).

When we took a closer look at the expression data from the RA cohort we observed that for all markers with a genetic association to RA, it was the same allele associated to RA that also lead to lower expression of CIITA transcripts, altogether 8 markers (6 where nominally significant, 2 suggestive association) situated in the promoter region.

3.1.4 Discussion on papers I-III

Age variation, association, interaction and expression

Because of the CIITA gene’s intricate role as transcription factor assembler for the transcription of MHC class II genes, it has been studied in many autoimmune diseases. Since the HLA, and mainly class II genes, are the main genetic determinant for many of these diseases, and no transcription of them can occur without the aid of the CIITA protein, surely there must be some effect of this gene?

Many affirmative studies have also been conducted, and association has been observed to T1D, MS, RA, Myocardial Infarction (MI) Addison’s disease, Celiac disease, Systemic Lupus Erythematosus (SLE) among others. But there have also been negative findings, sometimes for the very same markers found associated in another study. Neither have any of the associations except for celiac disease reach genome wide significant levels (P<10^-8), even when large, well-powered studies have been made.

Some of the explanation can lie in population stratification. Different populations have different allele frequency for SNPs, and genes can also be of different importance in these populations, the involvement of the gene can even be depending on interaction with local environmental factors. Also, if there is a mixture of populations within a study cohort this might influence the results. This fact is well known and care is often
taken to create homogenous study cohorts in that sense. But there can be other factors, and confounders that affect the association. Swanberg et al reported association for rs3087456 to MS, RA and MI, and it became a target for many replication studies with contradictive results. There was a difference in association in Swanbergs’ study depending if they used healthy controls matched to MS or RA patients which seemed a little strange. In study I in this thesis we investigated whether age among the controls could have an effect. The RA controls are generally older than MS controls if they are matched on age. What we found was quite interesting; there is a variation in genotype/allele frequency for markers in the CIITA gene depending on age among the controls (Fig. 5). It manifests as higher frequency of the major allele homozygote genotype among older individuals compared to younger. This was seen as a trend from younger to older age groups, and was confirmed in 2000 women of 25 and 75 years of age. Age is often matched for in case-control studies, but not always. It is also common to use blood-donors as controls, and they are known to be “extra healthy” compared to the general population. It is probable that an individual who is often sick will be less dedicated to participate as a control in a research study. When we investigated a part of our control cohort constituted of blood donors, we observed allele frequencies similar to the older groups, diverging from younger groups.

![Fig.5; Allele frequencies over age groups for rs11074932, total n=3747, p-value: 1.4e-05](image)

If age among cases versus controls is not considered, the results of association studies can be affected due to the different allele frequencies in different age groups.
Take rs3087456 for example, if the control group is older than the patient group, they will have a lower frequency of minor allele than the cases and thus create a false association of the marker to disease. This is probably what happened in the Swanberg article when using controls matched to RA cases when testing association to MS. When we corrected for age with logistic regression in our T1D cohort, we could still observe significant association for the two investigated markers (rs3087456 and rs11074932) to T1D.

What could be the reason for the variation of allele frequency over age that we observe in study I? Surely, a strong selection pressure on certain genes affecting survival for a certain infection would create a genetic effect like this. For this to occur, the infection in question must have conferred a high rate of mortality in young individuals, a scenario similar to what occurred due the Spanish flu epidemic. Also, gene variants affecting longevity would also increase in older age groups compared to younger. Both these effects are quite extreme, and our study is not designed to investigate such scenarios. Our hypothesis is rather that we observe a selection bias for healthy controls. If a genotype is associated with being healthier, for example affecting the severity or incidence of recurrent common infections, it might be more likely that individuals with this genotype are included as healthy controls for a medical study. It is reasonable to think that such conduct would be of larger effect in older individuals, resulting in a skewing of genotypes throughout the cohort.

There can also be other confounders that we have not considered. It is for example known that infectious agents, such as cytomegalovirus (CMV), can down-regulate CIITA for immune evasion\(^ {148}\), and we have observed a variation in genotypes among controls from MS cohorts, depending on CMV status (unpublished). Variables like this could have an effect on the heterogeneity of the cohort if there is some selection bias depending on the CMV infection.

Many of the SNPs in the promoter area (PI-IV, Fig.6) are in quite high LD with each other, and it will be difficult to sort out which marker is the true pathogenic one. There can be variations in LD pattern between populations that will effect which SNP will identified as associated when a stretch of markers are investigated.
Fig. 6; Linkage disequilibrium plot of the CIITA to CLEC16A gene region in the DISS2 cohort; darker grey indicates higher $r^2$ between markers. (HaploView 4.2).

Both in the T1D study and in the MS study we took care to correct for variations in age, and also to investigate the interaction between CIITA and disease-associated HLA alleles. In T1D, where the HLA-DRB1*15 acts protective from disease, interaction was found between lack of DRB1*15 and the disease associated allele in CIITA, such that the risk (or OR) increased more than expected when both occurred. In MS it was instead the joint effect of the associated CIITA allele and presence of the DRB1*15 allele, which in MS is conferring risk for disease, that increased risk. Also, the MS associated rs4774 had no significance in HLA-DRB1*15 negative individuals, but OR was strengthened when we stratified for DRB1*15 presence compared to the whole cohort. Clearly, the effect of CIITA in T1D and MS is partly through its impact on HLA genes. In summary, markers in the promoter region are associated in T1D and RA, and also to celiac disease\textsuperscript{146}, while for MS this region cannot be confirmed as associated but rather we observe association further downstream the gene, outside the strong LD blocks of the promoter region (Fig. 6).

Exploring the variable association and interaction data in paper I and II regarding CIITA, and considering the known function of CIITA as a transcription regulator for
MHC class II genes, we were eager to investigate any functional impact that these associated markers could have. Possibly, it could clarify underlying biological function that might affect autoimmune disease onset. In paper III we investigate SNPs in and close to CIITA as eQTLs, affecting transcription of self and nearby genes, as well as HLA genes. Our collaborators from an RA research group had previously observed a variation in CIITA transcript from promoter III (CIITA_pIII) and promoter IV (CIITA_pIV) depending on SNPs in the promoter region. These SNPs correlated with earlier reported association to RA, and when we further investigated the data we observed that the associated allele always was corresponding to lower level of expression of the CIITA transcript. We could not repeat this in our MS cohort, but it is plausible that it was due to power issue; at least for the rs3087456 SNP we see the same tendency as in RA with lower expression for the minor homozygote genotype. However, in the MS cohort we did observe genetic regulation of transcription of MHC class II genes; CD74 and DR-A, both in our own cell-based experiments as well as in our analysis of genome-wide expression data from RNA sequencing and genotypes from the Immunochip project. This has also been shown earlier for the rs3087456 SNP in RA patients.

Fig. 7: Correlation between expression of CIITA_pIV and rs3087456 genotype in INF-γ stimulated cells from RA patients, p=0.007 (a), and correlation between expression of CD74 and rs4774 genotype in ConA stimulated cells from MS and OND patients (p=0.03).

It is striking that for all associated SNPs in CIITA there was a corresponding lower expression level of CIITA and/or MHC class II genes, depending on the same allele as
the association. This was valid both in RA and MS cohorts. Generally, the markers in
the promoter region affected CIITA transcripts and MHC class II transcript, possibly
through the downstream regulation of transcription via CIITA. For rs4774 and
rs58497481, situated downstream the promoter area, we saw an effect only on the
MHC class II transcripts (Fig. 7). It is likely that the effect we observe due to these
markers outside the promoter area is more depending on the function of the CIITA
protein, and not on the level of CIITA mRNA. Analyzing the protein substitution
(glycine to alanine) caused by the rs4774 SNP in computer programs (PolyPhen2, SIFT,
PredictProtein) does not reveal any remarkable effects, and this part of the protein
seems to be hidden and not in any binding-active site. However, exactly how it will
affect the function in real life is hard to say, and it is not unlikely that rs4774 is not the
true causing SNP due to LD patterns. The PredictProtein analysis also showed that
substitution of many of the surrounding amino acids close to the rs4774 SNP would
have a strong effect on protein function.

It has been shown that CIITA promoter pIV is essential for positive selection of CD4+
thymocytes because it drives CIITA and MHC class II expression in thymic epithelial
cells (TECs) in mice. Promoter III on the other hand is necessary for expression in
lymphoid cells, mainly B-cells. If the lower expression of CIITA transcripts and /or
MHC class II transcripts we observe due to SNPs in the gene also correlates to lower
levels of CIITA and subsequently MHC class II molecules in thymus, it could in
extension lead to less efficient clearance of auto reactive T-cells, and hence increased
risk for autoimmune disease.

This theory correlates with our hypothesis from paper I that individuals considered to
be healthier than average has a lower frequency of the minor allele, the same allele
here found to lead to lower expression of CIITA and MHC class II. Could that be due to
less efficient antigen presentation in the periphery for those carrying that allele?

Many of the markers associated in RA are also found to be associated in T1D (paper I),
but it is the opposite alleles that are associated with increased risk of disease. T1D is a
disease with strong known islet auto-antigens like insulin. Possibly it is not beneficial
to have a normal level of MHC class II molecules and antigen presentation under these
circumstances, and that’s why we see a protective effect in T1D exerted by the minor
allele of these SNPs. The key functions of the immune system, to both establish
central tolerance in the thymus as well as maintain self-tolerance in the periphery and to be able to mount adaptive immune responses, are diverse and often opposing. It is also important to remember that CITA also is involved in expression of other immune genes as well as HLA class I\textsuperscript{135,149}, and any effects of these genes have not been taken into account here.

Taken together, we have shown in these papers that CITA has a prominent role for three different autoimmune diseases. SNPs in CITA are genetically associated to these diseases, and functionally this might have effect through inadequate clearance of auto-reactive immune cells, and possibly also through presentation of auto-antigens.

### 3.2 VAV1 IN RHEUMATOID ARTHRITIS (PAPER IV)

Our group have previously reported a role for Vav1 in EAE, the animal model for MS, and also found genetic association to the VAV1 gene in seven human MS case-control cohorts of more than 12,000 individuals. The MS-associated rs2546133-rs2617822 C-A haplotype also lead to increased expression of VAV1 in peripheral blood and cerebrospinal fluid cells in MS patients\textsuperscript{111}. In this paper we show that the same haplotype is associated in ACPA-negative but not ACPA-positive RA, and further demonstrating difference in severity in arthritis induced by either pristane or collagen type II in a congenic rat strains harboring a SNP in the coding region of Vav1.

#### 3.2.1 Results from paper IV

12 SNP markers in the VAV1 gene, selected from the previous MS study discussed above were analyzed in a Swedish RA cohort of 3026 incident RA cases and 2202 population-based controls. Of the 3026 RA cases, 1943 were ACPA positive, 1079 were ACPA negative and 4 were of unknown ACPA status. We observed association in the ACPA negative group for the MS associated rs2546133-rs2617822 C-A haplotype (p=0.004, OR: 1.28, CI: 1.09-1.51) but no association could be detected in the ACPA positive group. The haplotype was analyzed in a French family (trio) cohort with 100 RA (78 ACPA positive, 21 ACPA negative, 1 unknown) cases with parents. Here we could not detect any association in the whole cohort or in stratified groups.
We used the congenic rat strain described earlier\textsuperscript{111}, here called DA.BN-R25, with a RA susceptible Dark Agouti (DA) background containing a region of approximately ~2 centimorgan on chromosome 9 from the RA resistant strain Brown Norway (BN). Both DA and DA.BN-R25 animals were immunized with pristane and rat collagen type II in incomplete Freund’s adjuvant with the purpose to induce pristane-induced arthritis (PIA) and collagen-induced arthritis (CIA), respectively. We observed a significant lower arthritis score in the PIA model for the DA.BN-R25 strain which seems to depend on severity score rather than incidence. This was also correlated to a tendency of decreased weight-loss in DA.BN-R25, which is a measurement for disease severity (\textit{Fig.8}). No difference between strains could be detected in the CIA model and no difference in the measured antibodies could be detected between groups.
Fig. 8; DA.BN-R25 congenic rats develop milder arthritis after PIA induction compared to DA rats.

(A) Mean arthritis score for 27 days of monitoring.

(B) Arthritis incidence in DA and DA.BN-R25 rats after PIA induction.

(C) Arthritis sum score of DA and DA.BN-R25 rats still in experiment at day 27.

(D) Correlation of weight change and arthritis sum score, comparing day 27 to day 10. DA rats are represented by filled circles and DA.BN-R25 by open circles.

(E) Weight loss in percent at day 27 compared to weight at day 10. Each dot represents one animal.

Data in A, C and E are the mean ±SEM from n = 10 DA rats and n=10 DA.BN-R25 congenic rats. *P<0.05; **P<0.01; calculated by Mann-Whitney U test.
3.2.2 Discussion on paper IV

In this paper we studied two methods of inducing arthritis in animal models; pristane and collagen induce arthritis, both with much resemblance of human rheumatoid arthritis. Similarities are the chronic disease course, symmetrical involvement of peripheral joints and the presence of auto-reactive antibodies\textsuperscript{150}. Pristane immunization induces a strong response to the joints that is CD4+ T cell driven, and transfer of the pathogenic CD4+ T-cells also induces arthritis in naïve animals. The arthritis is improved by T cell depletion and the disease is genetically associated to MHC class II. RF antibodies but not ACPAs can be detected in this model\textsuperscript{113}. The collagen-induced arthritis is also genetically associated to MHC class II and includes a CD4+ T cell response in priming the immune response, but differently from PIA transfer of serum from collagen-immunized animals into naïve recipients induces disease. ACPAs and RF antibodies can be found in the serum, suggesting a somewhat larger participation of B cells in the pathology of this model\textsuperscript{151}.

We used both methods to induce arthritis in RA susceptible DA rats and congenic rats with DA background but harboring a SNP in exon1 of \textit{Vav1} leading to an aminoacid change at position 63 (Arg63Trp). It was observed that there was a decreased severity in the rat harboring the SNP than in the “wild-type” animal, but only in the PIA model. Previously it was shown\textsuperscript{111} that the same SNP conferred less severe EAE, and correlated to lower levels of \textit{Vav1} mRNA in CD4+ T-cells and lower levels of INF-\gamma producing cells in the CNS. Since \textit{VAV1} is a protein activated downstream the TCR and is important for development and activation of T-cells\textsuperscript{132}, it was hypothesized that the R63W SNP causes a defect in the \textit{Vav1} protein and subsequently affecting T-cell proliferation and potency, leading to less severe disease. VAV-signaling is also important in B-cells, but here the malfunctioning \textit{Vav1} protein seems to be compensated by Vav2 and Vav3, which are ubiquitously expressed\textsuperscript{132} and the capacity to develop B cell responses remains fairly intact. Lack of a phenotype variation in B cell response is also supported by the absence of any differences in generation of RF antibodies of both IgM and IgG isotypes in both our models and the presence of anti-collagen antibodies in CIA.

This might explain why an effect of the \textit{Vav1} polymorphism is observed in the PIA model, where T cells more directly contribute to the phenotype. However, functioning
CD4+ T-cells are also important for a strong B-cell response and development of plasma cells for generating antibodies.

In the human studies we observed association for VAV1 in ACPA negative RA but not for ACPA positive RA. ACPAs are seen in up to 80% of RA and is also associated with more severe disease and higher risk for comorbidities\textsuperscript{31}. Hypothetically, in ACPA positive RA with its strong auto-antigen in the citrullinated peptides, massive immune response and the effect of the antibodies in activating complement and immune complexes\textsuperscript{38}, a modest variation in VAV1 function would not affect the phenotype. In ACPA negative RA, both the T-and B-cell response are relatively weak in that there is no presence of ACPAs or a strong association to the shared epitope HLA genes\textsuperscript{152}. An increase in T-cell proliferation and upregulation of inflammatory cytokines could in this group have a more profound effect in the final disease pathology. The common rs2546133-rs2617822 C-A haplotype associated to MS and RA does not disrupt the VAV1 protein like the Arg63Trp SNP in rat, but rather allows for a more potent activation of auto-reactive lymphocytes and pro-inflammatory cytokine production. Upregulation of VAV1-mRNA, TNF and INF-\gamma was also observed in individuals with MS compared to controls\textsuperscript{111}. The animal model has in this case shown a good example of its use in identifying immune-regulatory genes.
Complex, or multifactorial diseases are difficult but exciting to study just because they are complex. In this thesis I have focused on three complex autoimmune diseases; Type 1 diabetes, Multiple Sclerosis and Rheumatoid Arthritis. Many contradictory factors contribute to the total picture of these diseases; there is a familiar aggregation indicating heritability. Yet, most affected individuals do not have a family history of disease. There is a strong genetic component in the HLA genes and a substantial part of patients carry HLA risk variants. Still, many healthy persons in the population have the same gene variants without developing disease. Many environmental risk factors have been suggested, but it is complicated to pinpoint the exact effect of these. Smoking has for example emerged as the perhaps largest risk factor, but there are many known high-mortality risks with smoking that are more likely to develop that an autoimmune disease. Foremost, what people in general wants to hear is “eat this and avoid that, then you will never develop this disease”. I believe that might never be a possible scenario. To be able to identify individuals at risk and to be able to prevent disease, for example by vaccination or other preventive treatment is of course very desirable. Next, efficient drugs and treatments are highly prioritized. At least, reliable information on your prognosis is very important for those with an incurable disease. Clinical research today is very focused on finding drug targets and to identify biomarkers for disease, to be used in prognosis of severity or drug treatment for example. Great successes have also been achieved, but more knowledge is needed. Questions like- “Is this disease primarily autoimmune or are the autoimmune reactions we see secondary to another event?” or “How do we classify sub-groups of the disease, do they share etiology or only final symptoms?” are very important when it comes to identify mechanisms that are possible as drug targets. In this thesis I have studied two genes, *CIITA* which have been shown to associate to T1D, MS and RA, and *VAV1*, shown to associated to MS and RA. For *CIITA* “the master regulator of MHC class II transcription”, we demonstrated an interesting age-dependent effect on genotype distribution among healthy individuals; this is perhaps a reflection of *CIITA*’s role in health and disease. We observe association to T1D, MS and RA, and interaction with the important HLA genes. The expression levels of *CIITA* transcripts where also studied, and we find that alleles associated with T1D, MS and
RA also leads to a down-regulation of CIITA transcripts and also of MHC II invariable-chain transcripts, which can depend either on lower levels of CIITA protein or more direct effects possibly on CIITA protein functionality.

For VAV1, important in T-cell function, we demonstrate different effects of the gene in two animal models of arthritis. We also find the gene associated in ACPA negative, but not ACPA positive RA patients. This is further strength to the notion that these two subgroups have profound differences that might even reflect different etiologies. None of the associations demonstrated here reach genome-wide significance ($P=5\times10^{-8}$) and needs to be confirmed in independent materials before stated as susceptibility-genes, but it is reasonable to argue that many genes of low or modest effect contributes to the etiology of complex diseases, and can be of different importance in different groups of patients and even in different individuals. It is also possibly that the SNPs presented here are not the true pathogenic ones, and the association observed is a reflection of a stronger association in unexplored markers due to LD. The functional effect of the SNPs in CIITA that we demonstrate in paper III, and the animal model correlation in VAV1 further strengthens the importance of these genes.

These findings further underscore the notion that the same processes and pathways are involved in several autoimmune diseases, but also that there can be profound heterogeneity within a disease, like in the case with ACPA positive and ACPA negative RA. Considering the central roles of CIITA and MHC class II molecules, as well as the T-cells and their proliferation in autoimmune disease, my hope is that this research will add a little piece to the grand puzzle.
5 FUTURE PERSPECTIVES

There are many possibilities in the continued study of CIITA. I would like to investigate CIITA more in T1D patients. First, to confirm association of CIITA to T1D while controlling for age, since no other publication so far has replicated this association. Also might it be interesting to see if the association to T1D is only found among T1D patients with celiac disease, given that CIITA is associated to celiac disease on a genome-wide level\textsuperscript{146}.

Since RA and T1D turned out to associate to the same alleles, but in different directions, it would be interesting to investigate the expression of CIITA in T1D. I would like to investigate CIITA in sorted cells. Since it is known that CIITA is expressed by different promoters in different celltypes, especially for dendritic cells, it would be interesting to see if a specific celltype is affected by eQTLs in CIITA. Also, more functional studies in larger materials are of interest, since some of the negative results observed in MS might depend on power issues.

Our results in the research of VAV1 indicate that further studies of disease activity, severity and age of onset in RA patients can be of interest. The ACPA negative group within RA disease is not as well explored as the ACPA positive group and further findings to explain the differences between the groups are important. Also, it would be of interest to investigate whether the rs2546133-rs2617822 C-A haplotype leads to the same expression pattern of VAV1 and cytokine environment in RA as it did in MS.
6 ACKNOWLEDGEMENTS

To all of you who have helped me, supported me, shared my time as a PhD-student, a big THANK YOU from the bottom of my heart.

Ingrid Kockum – my main supervisor. For 10 years we have worked together, we have experienced marriages, children, setbacks and successes. You have always been kind and patient and you give so much of yourself. Thank you for having me as your student and for our friendship that will extend beyond this thesis.

Pernilla Nikamo – co-supervisor. Even more than my supervisor you are my friend, and we have known each other since I first started in Ingrid’s lab. Uncountable are the times you have been there for me, showing me everything in the lab as well as explaining statistics. Thank you.

Petter Höglund – co-supervisor. We didn’t get so many co-operations as planned, but even though you are very busy you always took time to meet, read and have a positive input on my ideas, thank you for that.

Professor Tomas Olsson – Thank you Tomas for accepting me as a PhD student in your group

Fredrik Piehl - Without you the CIITA papers wouldn’t exist...Thank you for always being so helpful and spending a lot of time reading manuscripts, coming up with new ideas and promote connections between research groups.

Leonid Padyukov – always so friendly and calm, you have supported our work endlessly with both materials and input on papers and presentations. Thank you.

CMM Co-workers, collaborators and closest friends;

Samina Asad - min kära kollega, hur skulle jag ha klarat mig utan dig! Tack för allt vi delat genom åren. Guds sol ska alltid lysa över dig och din familj.

Elizabeth Ekelund – Min kära vän, tack för allt vi delat under dessa år, och allt kul vi gjort med barnen. Vi måste ses mer!

Malin Almgren – Allt roligt vi har haft, trevliga luncher och familjeskoj, men också våra allvarliga samtal och din kraft att orka också det svåra ska jag aldrig glömma.

59
Marcus Rönninger – Tack Marcus för ditt eviga tålamod i att förklara allt som gäller gen-expression...utan dig hade det inte blivit något paper III!

Cecilia Dominguez – Stort tack Cecilia för all din hjälp med celler och för din vänskap, du är sannerligen en stjärna! Lycka till med nya bebisen 😊


Tojo James – Thank you Tojo for all your help in paper III, and our nice conversations about travelling!

Pernilla Stridh – Tack P för våra samtal, all din pepp! Och livsvisdom. Du är expert på att belysa och framföra det wichtigaste tycker jag, oavsett om det gäller forskning eller dagisplatser.

Emilie Sundqvist – Tack för vår tid tillsammans, du har bidragit till mycket av min kunskap om epidemiologi och statistik.

Biborka Bereczky-Veress - Tack för våra diskussioner, och ditt engagemang i Naturvetarna.

Jenny Link – Vi går lång tid tillbaka faktiskt, tack för allt du hjälpt mig med under dessa år!

Nada Abdelmagid- Thank you Nada, you are a person who really cares about people, and do your best to help them. You must be a great MD!

Present and former members of the Neuroimmunology group, I would like to thank you all, I will just mention a few;

André – Always so patient and friendly, thanks for all your help.

Mohsen – Tack för att du alltid hjälpt mig i stort och smått.

Roham – Tack Roham för att Emelie fick komma och titta på makrofager!

Maria Swanberg – Tack för vårt fina samarbete Maria!

Venus – Tack för ditt engagemang och din hjälpsamhet.

Cynthia – Thanks for all our nice chats and lunches.
Från CMM00; Anna Witasp, Lollo Sjöholm, Philip Melas, Karin Luttrop, Ida Nilsson, Selim Sengul, Sivonne Arvidsson, Jeanette Johansen, Catharina Lavebratt...och många fler! Tack för vår härliga tid på plan 0!

Från CGB; Susanna “SusPot” vår solstråle. Hoppas du hittat din plats i (arbets-)livet och att vi ses snart igen! Helen-Melon; My friend, thank you for enlightening my way! Camilla, Viv, Linda och Mia. Tack för allt vi delade.

Sofi, min vän sen första dagen på SU! Tack för allt vi delat sen dess.

Tack till Miklagårds förskola för era fantastiska “fröknar” som tagit hand om mina barn i vått&törrt under dessa år.

Även; Tack till Riddarhusets Stiftelser och NEURO förbundet för generösa bidrag som gjort denna forskning möjligt.

Min käraste familj:

Mina bästisar sen lågstadiet: Camilla, Pernilla, Linda, Cecilia och Anna. Tack för att ni finns, för alla glada dagar och gråtna tårar; även då vi inte ses finns ni hos mig.

Farmor & Farfar, Mormor & Morfar; Önskar ni kunde varit här idag. Vi ses igen.

Mamma Christine & Pappa Staffan: Tack för att ni alltid stöttat mig, uppmunrat mig och gett er ovillkorliga kärlek till oss.


Syrrorna! Bästa Beatrice o Therese, vad skulle livet varit utan er? Tack för att ni finns, och alltid ha trott på mig. (Tack för barnvaktning också...the nanny!)

Tom – thank you for being a wonderful part of our family, and your great patients with the kids on long journeys through beautiful English countryside 😊

Eric; min älskade tack för ditt eviga stöd, din tålmodighet och din kärlek. Det finns ingen som du.

Finaste; Emelie, Amanda, Sofia; Framför allt finns ni i mitt hjärta för evigt.
7 REFERENCES


78. Hedström AK, Olsson T, Alfredsson L. High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. *Mult Scler* 2012; 18(9): 1334-6.


83. Marian AJ. Elements of 'missing heritability'. *Curr Opin Cardiol* 2012; 27(3): 197-201.


