

From THE DIVISION OF NEUROLOGY  
DEPARTMENT OF CLINICAL NEUROSCIENCE  
Karolinska Institutet, Stockholm, Sweden

**GENETIC ANALYSIS OF *IL7R*  
AND OTHER IMMUNE-  
REGULATORY GENES IN  
MULTIPLE SCLEROSIS**

Frida Lundmark



**Karolinska  
Institutet**

Stockholm 2007

The cover and Figure 1 are painted by Mattias Lundmark.

All previously published papers are reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Frida Lundmark, 2007  
ISBN 978-91-7357-278-1

*To know that we know what we know,  
and to know that we do not know what we do not know,  
that is true knowledge.*

*Copernicus*

*To my family*



## ABSTRACT

Multiple sclerosis is a chronic neurological disease, where both genetic and environmental factors are influencing the susceptibility and pathogenesis. Epidemiological studies have clearly demonstrated the existence of a genetic component by comparing the degree of shared genetic material and the risk of MS, where the degree of shared genetic material clearly correlates with the risk of MS. Until recently only one confirmed genetic risk factor for MS has been identified, HLA-DR\*1501. In this thesis we present evidence for a new genetic risk factor for MS, interleukin 7 receptor alpha chain, (*IL7R*) (*study I* and *IV*). The *IL7R* was initially identified in *study I*, where 66 genes were investigated in up to 672 MS patients and as many controls. The genes investigated were selected based on chromosomal location and biological functions presumed to be of importance in MS. Two genes, the *IL7R* and the lymphocyte activating gene (*LAG3*), were identified to be associated with MS. In addition, two haplotypes in *IL7R* presented significant differences between cases and controls. The *IL7R*, located on chromosome 5p13, is important in the maturation and survival of T-cells in humans. *LAG3*, located on chromosome 12p13, is important in inhibiting activated T-cells.

In *study II* we analysed *LAG3* and *CD4* in two independent populations; a Swedish case/control material used for the initial study and a Nordic case/control material used for confirmation. *CD4* was included due to the location close to *LAG3* and the LD patterns between the genes as well as a prior association of *CD4* with MS. None of the SNPs associated with MS in *LAG3* in *study I* were confirmed in the Nordic material. Initial analysis of nine SNPs in *CD4* revealed three associated SNPs, but none of these survived the confirmation step. From this data we conclude that *CD4* and *LAG3* do not present evidence to influence the genetic susceptibility in these populations.

In *study III* we investigated two polymorphisms located in the promotor region of the myeloperoxidase (*MPO*) gene. A number of studies have been reported for one of the SNPs (-463) and MS, without any conclusive result. The other SNP (-129) has not previously been investigated in MS. Neither of the SNPs presented any evidence of influencing the susceptibility to MS in this study. In addition, we investigated if any of the two SNPs showed any association with disease severity by using Multiple Sclerosis Severity Score, but no association between disease severity and genotype could be detected. We therefore conclude that these two polymorphisms do not contribute to either disease susceptibility or severity in our material.

In *study IV* we confirmed the three associated SNPs and the two haplotype associations in *IL7R* from *study I* in a large independent Nordic case/control material. In addition we fine-mapped the LD block harbouring the *IL7R* in a Swedish case/control material using a tagSNP approach. At this stage, three additional SNPs showed significant associations with MS, where one non-synonymous SNP in exon 6 presented the most significant p-value, and the importance of this SNP was proved by logistic regression analysis. Haplotype analysis presented convincing evidence for a protective effect of the most common haplotype. Analysis of cerebrospinal fluid from MS patients and from patients with non-inflammatory neurological diseases revealed an increased expression of *IL7R* in MS patients adding to the hypothesis of this pathway in MS.

Due to the mounting evidence for an importance of *IL7R* in MS we investigated the ligand, interleukin 7 (*IL7*), in *study V*. Nine SNPs were genotyped and no significant association was identified for any of the markers, thus we conclude that *IL7* does not contribute to the genetic susceptibility in MS. These negative findings strengthen the role of *IL7R* in MS, as the functional regulation of this complex has been suggested to be due to the receptor and not the ligand.

# LIST OF PUBLICATIONS

The thesis is based on the following publications which are referred to by the Roman numerals:

- I. Zhang Z, Duvefelt K, **Svensson F**, Masterman T, Jonasdottir G, Salter H, Emahazion T, Hellgren D, Falk G, Olsson T, Hillert J, Anvret M.  
Two genes encoding immune-regulatory molecules (*LAG3* and *IL7R*) confer risk to multiple sclerosis.  
*Genes and Immunity* 2005;6(2):145-152
- II. **Lundmark F**, Harbo HF, Celius EG, Saarela J, Datta P, Oturai A, Lindgren CM, Masterman T, Salter H, Hillert J.  
Association analysis of the *LAG3* and *CD4* genes in multiple sclerosis in two independent populations.  
*Journal of Neuroimmunology* 2006;180:193-198
- III. **Lundmark F**, Salter H, Hillert J.  
An association study of two functional promotor polymorphisms in the myeloperoxidase (*MPO*) gene in multiple sclerosis  
*Multiple Sclerosis* 2007;13(6):697-700
- IV. **Lundmark F**, Duvefelt K, Iacobaeus E, Kockum I, Wallström E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J.  
Variation in interleukin 7 receptor  $\alpha$  chain (*IL7R*) influences risk of multiple sclerosis.  
*Nature Genetics* 2007;39:1108-1113
- V. **Lundmark F**, Duvefelt K, Hillert J.  
Genetic association analysis of the interleukin 7 gene (*IL7*) in multiple sclerosis.  
*Journal of Neuroimmunology* 2007, In Press

# CONTENTS

1	Introduction.....	1
1.1	Multiple sclerosis.....	1
1.1.1	Clinical characteristics .....	1
1.1.2	Diagnosis .....	1
1.1.3	Treatment of MS .....	2
1.1.4	Pathogenesis .....	2
1.1.5	Epidemiology .....	5
1.1.6	Genetic Epidemiology.....	6
1.2	Genetics of Mendelian diseases .....	8
1.3	Genetics of complex diseases.....	8
1.3.1	Strategies to identify genetic variations in complex diseases.....	8
1.4	Genetics in multiple sclerosis.....	14
1.4.1	Linkage screens in MS .....	15
1.4.2	Association studies in MS.....	16
2	Aims of present studies .....	18
3	Material and methods .....	19
3.1	Patients and controls.....	19
3.2	Genetic analysis.....	20
3.2.1	DNA extraction .....	20
3.2.2	Genetic markers.....	20
3.2.3	SNP discovery .....	20
3.2.4	Genotyping .....	21
3.2.5	Statistical analysis .....	22
3.3	Expression analysis.....	23
3.3.1	Preparation of PBMC and CSF-MC.....	23
3.3.2	mRNA preparation .....	23
3.3.3	Quantitative RT-PCR.....	23
3.3.4	Statistical analysis .....	24
4	Results and interpretations .....	25
4.1	Study I.....	25
4.2	Study II.....	26
4.3	Study III .....	28
4.4	Study IV .....	29
4.5	Study V .....	30
5	Discussion.....	32
5.1	<i>IL7R</i> in MS.....	32
5.2	How to succeed in MS genetics? .....	34
5.2.1	Selection of patients and controls .....	34
5.2.2	Design issues .....	35
5.3	Negative or positive ?.....	36
5.3.1	Multiple testing.....	37
5.3.2	Publication bias .....	37
5.4	Haplotypes – will they help us? .....	38
6	Concluding remarks and future perspectives.....	40
7	Acknowledgements .....	41

8 References ..... 44

## LIST OF ABBREVIATIONS

APOE	Apolipoprotein E
BBB	Blood brain barrier
bp	Base pair
CDCV	Common disease common variant
CDRV	Common disease rare variant
CNP	Copy number polymorphism
CNS	Central nervous system
CSF	Cerebrospinal fluid
DNA	Deoxyribonucleic acid
DZ	Dizygotic (twin)
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
EDSS	Expanded disability status scale
EM	Expectation maximization (algorithm)
GWA	Genome wide association
HHV-6	Human herpes virus 6
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
IFN	Interferon
Ig	Immunoglobulin
IL7	Interleukin 7
IL7R	Interleukin 7 receptor
kb	Kilobase
LAG3	Lymphocyte activation gene 3
LD	Linkage disequilibrium
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MHC	Major histocompatibility complex
MPO	Myeloperoxidase
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MSSS	Multiple sclerosis severity score
MZ	Monozygotic (twin)
NCBI	National centre of biotechnology information
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PPMS	Primary progressive multiple sclerosis
RNA	Ribonucleic acid
RRMS	Relapsing-remitting multiple sclerosis
SNP	Single nucleotide polymorphism
SPMS	Secondary progressive multiple sclerosis
tag-SNP	Tagging single nucleotide polymorphism
TDT	Transmission disequilibrium test
TNF	Tumor necrosis factor
VLA-4	Very late antigen 4



# 1 INTRODUCTION

## 1.1 MULTIPLE SCLEROSIS

The disease multiple sclerosis (MS) has been known and well documented since the 19<sup>th</sup> century. MS as a disease entity was described by J M Charcot in 1868 entitled “*la sclerose en plaques disseminées*” [1]. It is characterised as a chronic inflammatory disease affecting the central nervous system (CNS), with demyelination and axonal loss as a consequence. The disease is considered to be mediated through an autoimmune process, leading to breakdown of the myelin sheaths and to neuronal loss. The beneficial effect of immunomodulatory and immunosuppressive treatments further supports the autoimmune hypothesis (reviewed in [2]). MS is considered to be a complex disease, where both genetic and environmental factors contribute to the susceptibility and pathogenesis. In this thesis genetic factors influencing MS have been investigated in order to further dissect the genetic contribution to the disease.

### 1.1.1 Clinical characteristics

MS is most often diagnosed during young adulthood and is more common among women than men with a ratio of 2:1, an observation also made in other presumed autoimmune diseases. In a fraction of MS patients, up to 25%, the disease never affects daily living, whereas in 15% of cases, patients acquire severe disabilities within a short period of time after diagnosis (reviewed in [3]).

The clinical course of MS is divided into three main categories; relapsing-remitting MS (RRMS), secondary progressive MS (SPMS) and primary progressive MS (PPMS). The RRMS is the most common form at onset of the disease (80-90%), and a majority of these patients later develop SPMS involving a slow worsening of the symptoms. A small fraction of patients, (10-20%), present with a progressive form of the disease from onset, PPMS, without any relapses (reviewed in [4]). Today there is no consensus whether the progressive forms of MS, PPMS and SPMS, represent the same or different pathological processes, but recent studies suggest that these two entities could be considered the same, where progression is suggested to be an age-dependent process independent of previous relapse history [5].

MS patients present a broad range of symptoms which includes among others, motor disturbances, sensory disturbances, pain, coordination and balance disturbances, bladder dysfunction together with cognitive impairment and fatigue [3].

### 1.1.2 Diagnosis

Today, there are no laboratory tests or biomarkers available for a specific diagnosis of MS. Instead the MS diagnosis is based on a *combination* of the patient history, the clinical neurological examination and supporting laboratory tests. The present clinical guidelines with diagnostic criteria for MS, the McDonald criteria, were published 2001 [6], and revised in 2005 [7] and these criteria include the use of magnetic resonance imaging (MRI).

To fulfil the criteria of a MS diagnosis the history and neurological examination should reveal two or more episodes of neurological symptoms disseminated in time and space. Today, MRI is widely used to present evidence of dissemination of lesions in time and space. The MRI-image of an MS patient most often reveals multifocal white matter lesions of different age and size. Gadolinium enhancement is used to demonstrate active lesions with ongoing inflammation. Supporting laboratory tests include analysis of the cerebrospinal fluid (CSF) to detect a characteristic oligoclonal pattern of immunoglobulins (Ig) detectable in 95% of the patients [8].

To assess the clinical impact and disability of MS in patients, a number of different measures are available, all with somewhat different foci [9]. The most widely used instrument to assess disability and clinical outcome in MS patients is the Expanded Disability Status Scale (EDSS) [10]. The scale is ordinal, ranging from 0-10, where 0 represents “normal by neurological examination” and 10 represents “death due to MS”. Scoring of patients is based on the neurological examination.

### 1.1.3 Treatment of MS

The search for an effective treatment for MS has been a struggle for researchers and pharmaceutical companies. Even today there is no drug available to *cure* MS, the current treatments are mainly *disease modifying* agents. The Swedish Medical Products Agency has approved five pharmaceutical compounds for disease modifying treatment. The largest group is the three  $\beta$ -interferons; two IFN- $\beta$ 1a (Avonex® [11] and Rebif® [12]) and one IFN- $\beta$ 1b (Betaferon® [13]). The exact mechanisms by which these compounds exert their disease modifying actions are not identified. The effect could be achieved through different pathways; decreasing antigen-presentation, shifting from a Th1 to a Th2 response of the immune system, decreased production of tumor necrosis factor alpha (TNF- $\alpha$ ), and to some extent, limiting the trafficking of T-cells into the CNS [14].

Glatiramer acetate (Copaxone® [15, 16]), consists of synthetic polypeptides, which is suggested to alter the immune-regulatory balance in MS patients by shifting from a Th1- to a Th2-response and to down regulate inflammation [17].

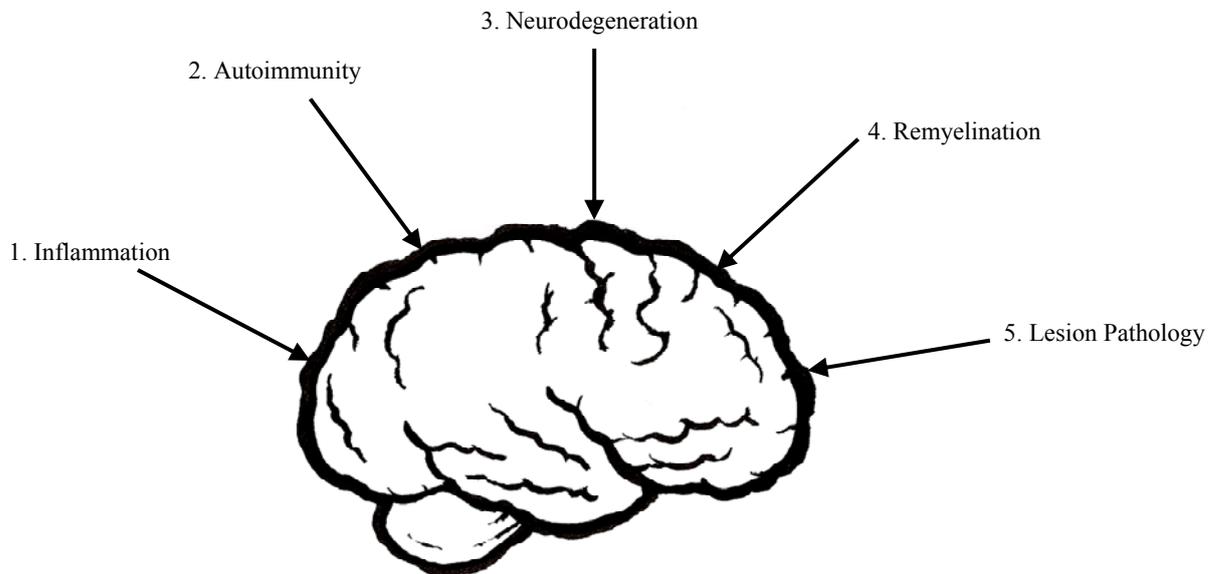
In 2004, a new compound was approved for treatment of MS, Natalizumab (Tysabri® [18]). Tysabri® is a monoclonal antibody, which exerts its action by blocking the trafficking of activated T cells from the periphery into the CNS, by VLA-4 molecules on lymphocytes. It has also shown to substantially reduce the relapse rate and progression of sustained disability. MRI findings also confirm that Tysabri® efficiently prevents the formation of new lesions [19].

### 1.1.4 Pathogenesis

MS is traditionally considered as a chronic inflammatory demyelinating disease of the CNS. The disease is associated with formation of focal lesions located in the white matter – the MS plaque. The plaques are characterized by a demyelinated area with glial scars, where axonal loss accompanies the demyelination process. During the demyelination process, the myelin sheath is stripped from the axon, leading to a

reduced conduction velocity of the axonal action potential. The recurrent bouts of inflammation lead to accumulation of CNS damage with neurological impairment as the result [20].

The biological processes that result in the development of MS are still poorly understood, but a number of different processes are considered to be of importance (Figure 1), these will be discussed in the following section.



*Figure 1. Summary of different mechanisms presumed to be of importance in the development and pathogenesis of MS. Each of the mechanisms will be presented in the following section.*

### ***Inflammation (1) and Autoimmunity (2)***

MS is presumed to be an autoimmune disease, where inflammation is the key player. By the 1960s adoptive transfer of experimental allergic encephalomyelitis (EAE), suggested the existence of autoimmune mechanisms in neuroinflammatory diseases [21]. Cellular infiltration into CNS and the reactivation of T- and B- cells by CNS antigens further support this hypothesis in addition to the genetic association of HLA class II (reviewed in [22]) and the effect of immunomodulatory agents in disease modifying treatment.

Initiation of CNS inflammation is assumed to start by activation of autoreactive myelin-specific T-cells in the periphery. The initial activation of the T-cells could be caused by either pathogens or self-proteins. Activation of lymphocytes by sequence or structural homology of pathogens to self-proteins of the CNS is referred to as “molecular mimicry” (reviewed in [23]). Upon activation, expression of integrins (VLA-4 and LFA-1) is up regulated. These integrins interact with endothelial adhesion molecules (ICAM-1 and VCAM-1) to facilitate cell recruitment into the CNS over the blood-brain-barrier (BBB) [24]. The BBB functions as a physical and metabolic barrier to protect the CNS by limiting immune cell trafficking into the brain. The properties of the BBB are changed during neuroinflammation, facilitating the recruitment of immune cells into the CNS [25].

The chronic inflammation starts when activated T-cells enter the CNS and are re-activated upon encountering their target antigens presented by microglial cells. This results in the release of pro-inflammatory molecules, including chemokines, cytokines and matrix-degrading enzymes, which will further facilitate the recruitment of immune cells into the CNS [24].

The recovery from relapses could to some extent be due to the existence of regulatory T cells (Tregs). These cells are specialised in counterbalancing the inflammatory responses and suppress the immune response. The role of Tregs in MS is further discussed in chapter 5.

MS has been considered a T-cell mediated disease; however, recent studies highlight the importance of B-cells. The presence of oligoclonal bands in the CSF of MS patients demonstrates an intrathecal (i.e. within the CNS) production of Igs and in addition, histopathological studies have identified B-cells in active demyelinating regions as well as in chronic lesions (reviewed in [26]).

### ***Neurodegeneration (3)***

In the past, demyelination was thought to be the main cause of neurological impairment in MS. In recent years, this view has been challenged, and the degree of neurodegeneration and axonal loss is proposed to correlate better with disability in MS [27]. Axonal loss has been recognized to occur early in disease, as well as secondarily to demyelination [28]. Primary neurodegeneration has been suggested as an initial event in lesion formation, where changes in the oligodendrocytes could be the cause [29], but ongoing axonal loss in older lesions has also been demonstrated.

The exact pathogenesis of axonal loss is not well understood, but two different pathways could be considered; (a) axonal loss as a consequence of inflammation and demyelination in plaques where the distribution of plaques and regions of axonal loss would then be correlated, or (b) axonal loss independent of inflammatory demyelination without any correlation with plaque load [30]. It is possible that axonal loss with time will be associated with irreversible neurological deficits.

MS has been considered as a white matter disease, but recent data have demonstrated lesions also in the grey matter, the most common being the cortical lesions. These type of lesions have less macrophage and T-cell infiltration compared to white matter lesions, thus suggesting that grey matter lesions are associated with less inflammation. Extensive cortical demyelination has also been associated with the progressive phase of MS as cortical demyelination are less abundant in RRMS [31].

### ***Remyelination and repair (4)***

The remyelination process is a spontaneous repair mechanism of the CNS, where new myelin sheaths are generated around demyelinated axons. The occurrence of remyelination in MS patients has been documented since the beginning of the 20<sup>th</sup> century. The remyelination process could in part restore the conduction properties of the axons and neurological function lost due to demyelination [32, 33]. The degree of

remyelination varies between lesions, and depends on the stage of progression of the lesion or pathological mechanisms underlying the lesion formation [34]. It has been observed that the extent of remyelination is directly correlated to the number of oligodendrocytes and macrophages in the lesion, where presence of oligodendrocytes was correlated with remyelination, and a negative correlation was observed for presence of macrophages in a lesion [35].

The remyelination process is most active during early disease phases, but decreases with time and progression. However, this view has been challenged, and extensive remyelination has also been observed in patients dying at an old age [36]. One study has also shown that as much as 40% of the MS lesions show signs of remyelination [37]. The loss of adequate remyelination could be due to a number of factors, and no specific cause has been identified, instead it is presumed that the environment within plaques does not favour the remyelination process [38].

The correlation of remyelination, disease severity and clinical outcome is not clear, but it has been suggested the remyelination is not responsible for the resolution of relapses early in disease. Instead it could be of importance in the later stages of functional recovery [39]. The resolution of relapses is instead facilitated by rearrangement of sodium channels (reviewed in [40]).

### ***Lesion pathology (5)***

Pathological studies of active MS lesions have revealed some heterogeneity in the immunopathological patterns between patients. The lesion patterns have been suggested to vary *between* patients and to be restricted to a single lesion type *within* patients.

Four different patterns (pattern I-IV) have been suggested to occur in MS lesions [34, 41]. Pattern I is described as macrophage associated demyelination, pattern II resembles pattern I, but with additional antibody and complement associated demyelination. Pattern III is defined by distal oligodendrocyte dystrophy and finally pattern IV, which resembles primary oligodendrocyte injury with secondary demyelination associated with macrophages. This view has been challenged by other scientists, stating that the lesion pattern may not be static, instead they suggest that patients may have lesions presenting features of more than one of the lesion pattern [29]. No association between the different lesion patterns and clinical features has been identified.

### **1.1.5 Epidemiology**

The fact that MS is a disease with an uneven global distribution over the world, has lead to extensive investigations of the epidemiology of MS. In 1993, Kurtzke [42] defined regions with varying prevalence of MS; high prevalence areas were considered more than 30 cases per 100,000 inhabitants, intermediate prevalence 5-30 cases per 100,000 and low prevalence less than 5 cases per 100,000. The high prevalence regions include Scandinavia, United Kingdom, northern Canada and USA, southern Australia and New Zealand. There is a correlation between the MS prevalence and the latitude, with the highest prevalence figures closer to the Arctic/Antarctic circles. Several

attempts have been made to elucidate the cause of this latitude gradient, but no specific cause has been identified. The view of MS as a complex disease is supported by these regional differences. The fact that within high prevalence areas, cluster of ethnic minorities report low prevalence of MS, such as Samis in Scandinavia and Hutterites in Canada further supports the interaction of genes and environment in MS [43].

Over the years, reports describing epidemic MS have been presented, for example in Iceland [44, 45], the Shetland and Orkney Islands [46] and the Faroe Islands [47], but the findings have been questioned [48, 49]. Today, there is no consensus regarding the existence of epidemics in MS.

A number of migration studies have been performed in MS. These type of studies could be difficult to interpret and to draw general conclusions from, due to the fact that the migrating individuals rarely represent the general population, even so, the data is fairly consistent [50]. Individuals migrating from a high prevalence area to a low prevalence area, lower their risk of getting MS, whereas individuals migrating in the opposite direction, seem to retain their low risk of MS (reviewed in [51]). Also, these data suggest that environmental factors are of importance at the population level.

A correlation between the age of migration and the risk of MS has also been observed, where individuals migrating from a high risk area to a low risk area before the age of 15 acquire the risk of the new residence, whereas individual migrating after the age of 15 retain the risk of the area of origin (reviewed in [52]). These observations have lead to speculations about the importance of early life events in MS.

Several investigators have tried to pin-point potential environmental factors contributing to MS. The hygiene hypothesis, suggesting that the immature immune system needs to be challenged early in life in order to develop normally, has been proposed to be of importance in many autoimmune diseases including MS [53]. Viruses, including human herpes virus 6 (HHV-6) [54, 55] and Epstein-Bar virus (EBV) [56, 57], have been investigated and proposed to be involved in MS, but no consensus has been reached regarding the importance of these viruses. Sunlight exposure and the vitamin D status are other factors that have been suggested to influence the risk of MS. High levels of vitamin D in serum have been suggested to decrease the risk [58]. Further studies of the possible role of vitamin D in the causation of MS are needed to fully understand the mechanisms in action.

In a recent meta-analysis of environmental factors, smoking before onset of the disease, was identified as a risk factor for subsequent development of MS [59]. In relation to this, the authors speculate whether the increase in smoking among women may be an explanation for the increase in female predominance in MS observed in the recent years [60].

### **1.1.6 Genetic Epidemiology**

Epidemiological data for MS (see 1.1.5 Epidemiology) have suggested the disease to be caused by a combination of genetic factors and environmental agents. The genetic component in MS has been thoroughly investigated through studies of individuals

sharing genetic material to different degrees and their corresponding risk of developing MS. In general, a higher degree of genetic sharing corresponds to an increased risk of developing MS. Among patients, 15-20% have a close relative diagnosed with MS [61].

To establish the existence of a genetic component in MS, a number of twin studies have been performed, investigating the concordance rate among monozygotic (MZ) and dizygotic (DZ) twins [62-66]. The concordance rate among MZ twins varies between 24-30%, and among DZ twins from 2-5%, clearly demonstrating the effect of genetic sharing and the risk of MS. On the other hand, these data also point to the presence of risk factors other than genetic modulators, since a pure genetic cause of MS would present nearly 100% concordance in MZ twins (Figure 2).

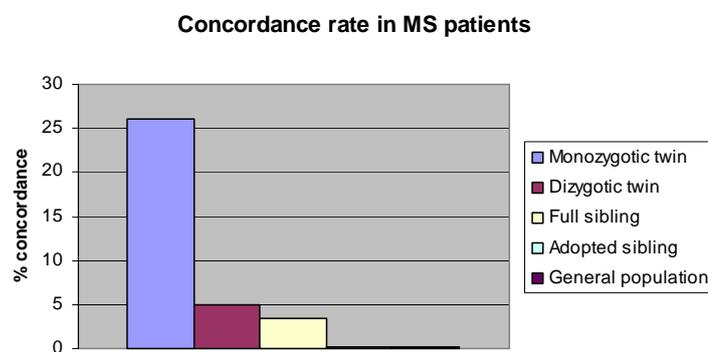


Figure 2. Concordance rates for individuals with different degree of genetic sharing, in relation to the MS patient [62-66].

The familial aggregation of MS has raised the question whether this is due to genetic sharing or the presence of environmental factors in these specific families. Studies of adopted individuals living in families with MS have not revealed any increased frequency of MS in these individuals as compared to the general population, thus suggesting that the familial aggregation is due to genetic factors rather than the shared environment [67].

On the other hand biological relatives to MS patients do have an increased risk of MS. The increase in risk is correlated to the degree of genetic sharing and children of parents, both suffering from MS, present an increased risk, as compared to children with only one affected parent [68, 69]. In one of these studies, the risk of MS was examined in 13,000 spouses to MS patients, without showing an increase in MS frequency, as compared to the general population [69].

In conclusion, present data clearly indicate the importance of genetic factors in the aetiology of MS, but environmental factors may also be of importance, presumably affecting the risk of MS at the population level rather than at the family level.

## 1.2 GENETICS OF MENDELIAN DISEASES

In 1865, Gregor Mendel, the “father of genetics”, published his work dealing with the inheritance of colour and form of garden peas. This was the starting point of modern genetics. Gregor Mendel also gave his name to a group of genetic diseases, “Mendelian” disorders, in which the inheritance pattern resembles the mode of inheritance he described in 1865. These diseases are most often monogenic – the cause of the disease is attributable to mutations in one single gene.

Linkage studies in families with several affected individuals have been successful in identifying the disease causing mutations in a number of Mendelian diseases. More than 1,822 genes for 1,200 Mendelian traits have been identified. Studies of monogenic disorders have greatly contributed to the understanding of pathogenic mutations and gene regulation and brought genetic research to its next challenge, to elucidate the genetics behind complex diseases [70].

## 1.3 GENETICS OF COMPLEX DISEASES

A disease is defined as a “complex disease” when combinations of genetic and environmental factors contribute to the aetiology of the disease. Individuals may have different genetic risk factors presented in different combinations, which all result in an increased risk of the disease. Each individual genetic component may not be sufficient to cause the disease and neither may the combination of variations sufficient to cause disease in *one* individual, be attributable to *all* cases. In addition, the genetic components most likely interact with environmental factors, making studies in complex diseases – complex.

In contrast to Mendelian disorders, the genetic modulators in complex diseases seldom *cause* the disease, instead they influence the *susceptibility*. Identifying genetic variations involved in a certain complex disease is a difficult and very expensive task, due to the need of extremely large sample sizes and top of the line laboratory equipment. Instead, other approaches have been developed to elucidate the genetic contribution, and recently, the genome wide association (GWA) approach has made the goal of capturing most disease associated variants come a bit closer.

### 1.3.1 Strategies to identify genetic variations in complex diseases

When studying complex genetics, there are a number of different approaches available, all with certain benefits. The different strategies will be further discussed below.

Most genetic studies are based on the hypothesis of “common disease common variant” (CDCV) stipulating that common genetic variants, present in a significant proportion of the population, both affected and unaffected individuals, are the cause of common diseases with a genetic component [71]. The validity of the CDCV hypothesis may be questioned, as rare variants may be as likely to confer risk to complex diseases. This issue will be further discussed in chapter 5.

### *Genetic markers*

Most genetic studies are performed using various genetic markers instead of direct sequencing. These markers may function as surrogate markers for linked and associated disease variants or be causative themselves. The markers most often utilized are microsatellites and single nucleotide polymorphisms (SNPs), but recently copy number polymorphisms (CNPs) have been introduced as a new group of markers.

#### *Microsatellites*

Microsatellites are tandemly repeated DNA sequences, generally di-, tri- or tetranucleotide repeats. The number of repetitions varies between markers, but up to 50 repeat units occurs. Microsatellites are more common in uncoding regions of the genome, making them less likely to be the direct cause of a disease; instead they are often used as indirect markers of a disease causing variation. However, there are exceptions to this, for example the “CGA” trinucleotide repeat in Huntington’s disease has been shown to be the direct cause of the disease [72].

Microsatellites are highly polymorphic, making them an excellent choice for linkage studies, which rely on recombination events. The mutation rate of microsatellites is high,  $10^{-3}$  to  $10^{-4}$  per locus per generation, more than 10,000 times higher than in non-repeated sequence [73, 74].

#### *Single nucleotide polymorphisms*

SNPs are defined as dinucleotide variations in the DNA sequence, where the rare allele is present in at least 1% of the population. Variations less frequent are instead regarded as mutations. Just like microsatellites, SNPs can be utilized as surrogate markers in genetic studies, but in contrast to microsatellites, SNPs are more likely to be directly causative, or in strong LD with a disease causative SNP. This is true, especially for SNPs located in coding regions of the genome.

SNPs are the most abundant DNA variations, around 1 SNP per 1200 base pairs on the average, and the distribution of SNPs in the genome is fairly even. This makes SNPs especially well suited for association studies. The biallelic nature of the SNP enables low-cost genotyping, another feature making SNPs an attractive choice in genetic studies.

#### *Copy number polymorphisms*

Recently, a new group of markers has been proposed to be important when trying to dissect complex diseases, copy number polymorphisms (CNPs). These markers are described as structural alterations in the genome ranging from kilobases (kb) to megabases (Mb) in size. CNPs include deletions, insertions, duplications and multi-site variations (reviewed in [75]). The distribution of CNPs in the human genome is still not fully understood, and the utilization of CNPs in genetic studies has just started. Further evaluation of the contribution of these markers to our understanding of complex diseases will be needed in order to understand the function and distribution of these polymorphisms.

### ***Linkage studies***

The linkage approach has its origin in genetic studies of rare Mendelian disorders and is based on *recombination events* in the genome. Linkage studies have also been extensively used to dissect complex diseases. The basic idea of linkage studies is to identify markers inherited through families in the same manner as the disease.

The studies have been most often performed as genomic screens, where about 250-700 microsatellites spread over the genome are genotyped. In this respect, linkage studies can be assumed to be unbiased. The studies are based on family materials, preferably large extended pedigrees with multiple affected individuals, although these are rare to find in complex diseases. Instead, affected sib-pairs are often the family material at hand in these type of studies [76]. An advantage of using affected sib-pairs is the statistical analysis; which is fairly simple and straightforward in comparison to analysis of extended pedigrees. Moreover, TDT analysis of trios within the cohort is possible.

The purpose of linkage studies is to follow the co-segregation of a marker and disease status. In the presence of linkage, the affected individuals are presumed to share a higher degree of DNA close to the disease gene than would be expected by chance [76].

Although extensive work has been put into linkage studies in complex diseases, the approach has not been as successful as hoped [77]. Very few disease causing genes have been identified in this way, and there are a number of reasons for this. First, the genetic contribution of each individual component may be too small in order to be captured in a linkage scan and secondly, in a 250 marker screen too few markers are genotyped and in addition, the power of the studies have not been sufficient (reviewed in [78]). Overall, linkage studies have usually less power to detect disease associated genes with modest effects than for example the association approach (described below) [79].

### ***Association studies***

In contrast to linkage studies, association studies are most often performed in groups of unrelated cases and controls; *case/control* studies. The basic idea of the association studies is to unravel differences in the genetic composition between patients and controls. Association studies using family-based materials, often *trios* consisting of two parents and one affected child, can also be performed. In these cases, the transmission disequilibrium test (TDT) is commonly used [80]. TDT compares the transmission versus non-transmission of an allele from heterozygous parents to affected offspring. In general, TDT has less power than case/control studies, but is also less sensitive to population stratification.

In this thesis, the studies have been focused on association analysis using the case/control strategy.

Until recently, association studies have been hypothesis driven, the so called *candidate gene approach*, where specific genes of interest are selected. This is still the most common way to perform an association study, but there is now also the possibility to do

a genome wide association study [78], which resembles linkage studies in the unbiased selection of markers genotyped (described further later in this chapter).

Practically all recent association studies have been performed using SNPs as the genetic markers, regardless of whether single genes or the whole genome are to be genotyped. The concept of association studies is to detect differences in allele frequencies of polymorphisms comparing groups of cases and controls. To test for an association, either direct or indirect association tests can be performed. The direct association test investigates SNPs with a functional consequence that are hypothesised to influence the risk of the disease, while indirect association analysis investigates markers in close genetic proximity to the actual causative variant. The possibility to perform indirect association studies relies on the presence of high linkage disequilibrium (LD) between the marker tested and the causative SNP (further discussed in the section “*Linkage Disequilibrium mapping*”).

In contrast to linkage studies, association studies have been somewhat more successful in identifying genes conferring risk to complex diseases [81-86]. A reason for the success can in part be ascribed to the international HapMap consortium. It was founded in 2002, and set out to “*determine the common pattern of DNA variations in the human genome*” in the “International HapMap Project” [87]. HapMap has aimed for a comprehensive map of SNPs, providing information of allele frequencies and LD structures in different human populations. In the project, a total of 270 individuals from three different ethnic groups (African, Asian and European ancestry) have been genotyped, thus capturing the most common haplotypes in these populations. All data are made public, accessible for all researchers, and thus provide valuable information for designing of genetic studies.

Recently, a new strategy to dissect complex diseases using association analysis has been developed, the genome wide association (GWA) approach. By taking advantage of microarray technologies, large numbers of SNPs can be simultaneously genotyped. At present, up to 500,000 SNPs (out of 5 millions validated SNPs) can be typed for each individual. In studies with adequate power, this will be helpful in search for genes with reasonable effect [88], but further fine mapping and confirmation will still be required.

To cover the genome, the selection of SNPs is an essential issue when designing the SNP arrays, and questions have been raised as to whether the selection and number of SNPs are sufficient in order to identify risk genes. Recently, the Wellcome Trust Case Control Consortium published a screen of 17,000 individuals for seven diseases (bipolar disease, coronary artery disease, Crohn’s disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes), presenting convincing evidence of association for a number of genes [89]. Two follow-up studies have already confirmed the initial findings for two of the diseases, type 1 diabetes [90] and Crohn’s disease [91].

In complex diseases these studies provide a careful validation of GWA as a method to identify genetic associations, and show that studies with sufficient power are successful in identifying genetic associations that survive the crucial confirmation step and represent true disease genes.

### ***Linkage disequilibrium mapping***

Linkage disequilibrium, LD, is defined as non-independent association of alleles at two or more loci on a chromosome within a given population. In the presence of high LD, combinations of alleles are inherited together more often than would be expected by chance. LD is often considered to be higher in isolated populations originating from a limited number of founders [92]. The extent of LD is largely influenced by population history; migration and population bottle-necks, but molecular events like recombination and mutation rate also contribute to the degree of LD. The term LD should not be confused with *linkage*, which describes the association of two loci on a chromosome with limited recombination between these loci.

There are a number of different measurements for LD [93], the two most important being  $D'$  and  $r^2$ .

$D$  is defined as;

$$D = f(AB) - f(A)f(B)$$

where  $f(AB)$  represents the observed haplotypes frequency of the haplotype AB, and  $f(A) f(B)$ , the random segregation of each allele.  $D$  represents the difference between the observed and expected frequencies of haplotypes.  $D$  is a function of the allele frequencies and normalization of the measure is often performed to facilitate interpretation. The most common normalization is the  $D'$  measure, defined as [94];

$$D' = D / D_{\max}$$

where  $D_{\max}$  is the maximum of  $D$  given the allele frequencies at the two loci.  $D'= 1$  is defined as complete LD, perfect disequilibrium and  $D'= 0$ , equilibrium. A disadvantage using  $D'$  as a measure of LD is the influence of small sample sizes and allele frequencies in SNPs with rare alleles. In these cases, the degree of LD is often overestimated and intermediate values of  $D' < 1$ , should be used with caution.

The measure  $r^2$  describes LD in terms of correlation between the alleles at two loci.

$r^2$  is defined by;

$$r^2 = D^2 / (f(A)f(a)f(B)f(b))$$

where A/a and B/b represents the alleles at two bi-allelic loci. In contrast to  $D'$ ,  $r^2$  is considering the difference in allele frequency of the two loci, avoiding an overestimated LD.

$r^2$  is also related to the power of association in genetic association studies. In order to achieve the same power at the marker locus as if the actual disease mutation was genotyped, the sample size should be increased by a factor of  $(1 / r^2)$ , where  $r^2$  represents the LD between the marker tested and the disease mutation.  $r^2 = 1$  represents

perfect LD, and intermediate measures  $r^2 < 1$ , are more easy to interpret than intermediate  $D'$  measures.

### *TagSNPs and Haplotypes*

The large number of SNPs identified within the human genome in combination with the knowledge of LD patterns, has made genetic association studies more appealing in the recent years. Tagging SNPs (tagSNPs) make it possible to study genomic regions by capturing genotype information for a large number of SNPs, but with a rather limited number of markers actually genotyped. The use of tagSNPs can also be beneficial in haplotype analysis, when comparing haplotype frequencies in cases and controls

The tagSNP approach utilizes LD in terms of  $r^2$  measures, where the genotyped tagSNP functions as a proxy for a number of other SNPs. TagSNPs are selected due to their LD ( $r^2$  measures) with other SNPs, inferring genotype information for these markers. In order to acquire high quality data, the cut-off for  $r^2$  between the tagSNP and captured SNPs should be conservative, at least  $r^2 > 0.8$ . Several different methods for selection of tagSNPs have been developed; even so, there is as yet no “gold-standard” for selection of tagSNPs. “Tagger” [95] is one method widely used and integrated in the Haploview software, but many additional options are available [96, 97].

Haplotypes can be described as combinations of alleles in high LD within a genomic region that are inherited together more often than would be predicted by chance. The existence of haplotypes in the human genome is a consequence of molecular mechanisms of sexual reproduction and the history of human evolution. The human population migrated out of Africa, and thus, haplotypes identified outside Africa represent a subset of all haplotypes, since not all genetic variations present in the ancestral population in Africa were brought outside. The frequency of haplotypes varies between populations due to the *migration pattern*, *random mating*, and *natural selection*, for example genetic bottlenecks, which reduce genetic variation and increase genetic drift (reviewed in [98]). The absolute definition of a haplotype is still not completely clear. In some cases haplotypes can be considered as a combination of alleles over long genomic distances, without presence of high LD. In this section, I will refer to haplotypes as combinations of alleles in regions with high LD.

Haplotype block structures have been identified throughout the genome, capturing most genetic variations within that particular region [99]. The length of haplotypes varies within the genome and each meiosis decreases the extent of the haplotype.

Haplotype association analysis may reveal presence of protective- or risk haplotypes and in addition, the distribution of haplotypes can be investigated, identifying differences in the distribution between cases and controls.

In order to perform haplotypes association analysis, the phase of the genotype data is essential, i.e. the parental origin of each allele should be known. In most association studies, based on unrelated case/control material, information regarding phase is not available. The haplotypes are instead inferred and haplotype frequencies are estimated from the genotype data. Estimation of haplotypes can be performed with different

methods, the most commonly used methods are the Expectation-Maximisation algorithm (EM) [100], Bayesian methods [101] and log-linear modelling [102]. The lack of phase information and the implications of using estimated haplotypes frequencies have been debated. The main issue is the robustness of the estimated haplotypes, and the reliability of the subsequent haplotype analysis.

Haplotype association tests include two main strategies; comparing the distribution of haplotypes between cases and controls identifying differences in distribution between the groups (global analysis), or comparing haplotype frequencies between cases and controls to identify predisposing or protective haplotypes (individual haplotype analysis). Often both analyses are performed in parallel.

### ***Linkage or association – benefits and weaknesses***

As previously mentioned, linkage screens have been rather unsuccessful in finding disease genes in complex disease for the reasons discussed above. Still, linkage screens can be beneficial in complex diseases. This is especially true for studies using extended families with many affected individuals. The linkage approach can take advantage of the increased genetic sharing in these extended families.

Although association studies have been successful in identifying a number of genes in complex diseases, there are issues to be aware of when performing these kind of studies.

The choice of statistical methods is critical for a correct interpretation of the data. Lately, large efforts have been put into the development of statistical methods capable of dealing with large quantities of genotype data although this issue is not yet resolved. In addition, the need for sufficient statistical power is essential. This depends largely on the allele frequencies, sample sizes, mode of inheritance and disease effect size, proving the need for large case/control materials [103].

Population heterogeneity is another important issue to address. It could potentially lead to population stratification with spurious association as a result [104]. Stringent inclusion criteria regarding ethnicity for patients and controls are needed in order to avoid this. As a complement *genomic control* can be performed in order to control for population stratifications within a case/control material (reviewed in [105]).

There are a very high number of genes *suggested* to be associated with a certain disease. The proportion of *true associated* gene variants does not correspond to this number though, proving the urgent need for independent replications to confirm true genetic associations [106].

## **1.4 GENETICS IN MULTIPLE SCLEROSIS**

Already in the 1890s, the existence of familial aggregation of MS was recognized, suggesting that genetic factors may play a role in MS. 80 years later in the 1970s, the first genetic association was reported, within the HLA complex [107, 108]. The HLA haplotype associated with MS was later specified to be the class II haplotype,

DRB1\*1501,DRB\*0101,DQA1\*0102,DQB1\*0602, often referred to as DRB1\*15 [109]. The importance of HLA class II in MS genetics has been demonstrated and replicated in numerous studies, and is now considered to be a genetic risk factor for MS. The class II molecules are involved in the presentation of externally derived antigens to CD4+ T-cells, thus, the class II association with MS strengthens the hypothesis that MS is an autoimmune, T-cell mediated disease.

In Northern European populations, 60% of the MS patients carry the risk haplotype, whereas only 30% of the controls do. In Sardinia however, the prevalence of DRB1\*15 is lower than in Northern Europe, 2.5% in MS patients and 1.5% in controls. The association with MS is still present but much weaker and it is mainly the DRB1\*03 and DRB1\*04 that are associated with MS in this population [110-112].

In a recent publication by Dymant et al. two class II alleles other than DRB1\*15, were reported to be associated with MS [113]. The allele DRB1\*17 was associated with an increased risk of MS, whereas the allele DRB1\*14 was identified as a protective allele. These results suggest that additional alleles may influence the risk of MS within the HLA class II locus.

A dose-effect for the risk of MS has been observed for carriers of DRB1\*1501, where individuals carrying one copy present a 2.7 fold increase in risk while two copies confer a 6.7 fold increase in risk of MS [114]. These observations have later been confirmed in a Swedish study, with slightly higher risk ratios observed for both carriers of one and two copies [115].

Several attempts have been made to associate DRB1\*15 with different clinical features. Association with female gender and early onset of the disease has been observed, but no association with clinical outcome or disease severity has been established [116-118].

HLA-A, located in the class I region, has been identified to harbour an independent association with MS [119, 120]. The allele HLA-A\*0201 has been suggested to confer a protective effect to MS, a finding recently replicated in a large case/control material [121].

#### **1.4.1 Linkage screens in MS**

A vast amount of research has been invested into linkage screens in MS. In total eleven microsatellite-based whole genome linkage screens have been performed in a number of populations [122-132]. None of the screens revealed any persuasive evidence for linkage. A meta-analysis, including all microsatellite screens, presented evidence for linkage in the MHC region, harbouring HLA-DRB1\*15, but no genome-wide linkage to other genomic regions was identified. Suggestive linkage was observed in two chromosomal regions, 17q21 and 22q13 [133].

Recently, the International Multiple Sclerosis Genetics Consortium (IMSGC) published a SNP based linkage screen [134]. DNA samples from 730 multiplex families in Australia, UK, USA and Scandinavia were collected and in total 2,692 individuals were genotyped for 4,506 SNPs. Apart from the MHC region, LOD 11.7, no other genomic

regions presented evidence of linkage with genome wide significance. Suggestive linkage was observed for three chromosomal regions, chromosome 19p13, 17q23 and 5q33.

Genome wide screens in multigenerational families with several affected individuals, have been carried out in MS [135-137]. Yet no genome wide significance has been identified for any region, but suggestive evidence for linkage has been shown on chromosome 9q and 12p12 [136, 137]. Further attempts to elucidate the genetic inheritance pattern of MS in multiplex families may be of importance in unravelling the genetic mechanisms.

The absence of additional identified linkage regions demonstrates the problems with low resolution and insufficient power in linkage screens, even when large sample sizes are genotyped in a significant number of markers. The genetic effect from each contributing factor is presumably too small to be captured in a linkage screen, suggesting that future genetic studies should be performed using association analysis.

#### **1.4.2 Association studies in MS**

Over the years, numerous studies focusing on non-HLA candidate genes for MS susceptibility, disease severity and progression have been performed, but with few exceptions, the results are inconclusive. Apart from the HLA associations no other genes have been confirmed to be associated with MS [2, 52, 138-141].

The selection of candidate genes has been focused on biological candidates involved mainly in inflammation and neurodegeneration, and positional candidates, located in chromosomal regions with presumed importance in autoimmune diseases. Many of the candidate studies have reported initial associations with MS, but attempts to confirm their genetic importance in MS have failed. The lack of success in this area can in part be attributed to the heterogeneity of disease. Different populations could display separate panels of susceptibility genes, which show a varying degree of overlap between populations. The uneven geographical distribution of MS further supports the notion that genetic heterogeneity is important to consider in genetic studies of MS. In addition, each single genetic risk factor confers only a modest risk effect, thus implying that a large number of different genetic factors are contributing to the development of MS. Due to the modest effect of each genetic modulator, the obtained odds ratios (ORs) will subsequently be modest. In the case of a complex disease like MS, the ORs that we can expect to find will in most cases be modest. Even a confirmed true association may not present  $OR < 1.4$ , even so it still represents a significant and important finding. This is important to keep in mind when assessing genetic studies in MS. Judging genetic findings solely on the basis of the magnitude of ORs and p-values may be misleading.

The number of candidate genes studies in MS is extremely high, and a full overview is impossible to present here. Instead some examples are presented representing interesting genes in MS, analysed in well-designed studies.

The gene *APOE* (apolipoprotein E), conferring risk to Alzheimer's disease [142], has received a great deal of attention in MS genetics. The involvement of this gene in

disease susceptibility, disease course and progression has been extensively studied (reviewed in [143]). It has been suggested that the *ApoEε4* allele is associated with a more severe disease progression, but the results are contradictory [143].

*PRKCA* (protein kinase C alpha), located on chromosome 17q22, is another gene suggested to predispose for MS [144]. The initial finding has been confirmed in a Finnish and Canadian population, but with different polymorphisms [145]. This discrepancy in associated polymorphisms needs further evaluation in order to fully understand the role of *PRKCA* in MS genetics. *PRKCA* has been shown to be involved in signal transduction in T-cell activation [146].

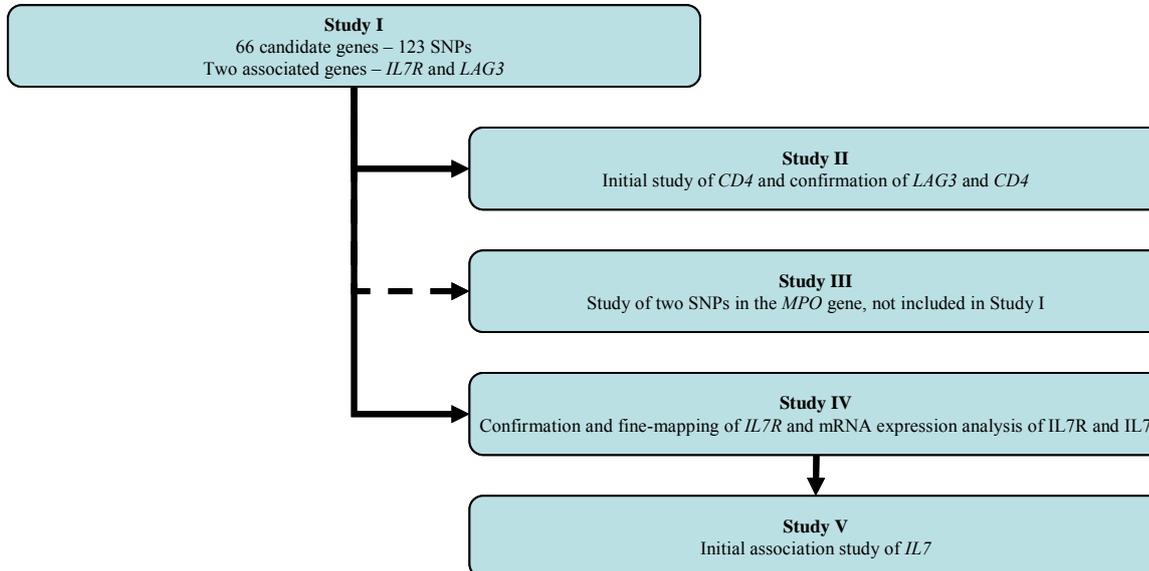
Recently, *MHC2TA* (MHC class II transactivator) was reported to confer risk to three autoimmune diseases, including MS [147]. Attempts to confirm the association with MS have been inconclusive, possibly due to insufficient power [148-150]. *MHC2TA* is involved in the assembly of several transcription factors at MHC promoters [151]. Further confirmation of this gene will be essential to elucidate the genetic involvement of *MHC2TA* in MS susceptibility.

In addition to the candidate gene approach represented by the above examples, the first GWA study for MS has recently been published [152]. In total 12,360 individuals were included in the analysis, identifying two genes apart from the HLA locus that presented significant association with MS, *IL2RA* and *IL7R*. The *IL7R* has been extensively studied and confirmed in MS, study IV and [153], whereas *IL2RA* still needs confirmation.

In conclusion, the search for MS genes has been a struggle throughout the years. The lack of positive findings has made researchers pessimistic about the ability to elucidate the genetic factors contributing to MS. But lately, the optimism in MS genetics has increased as a result of technical development and collaborations such as the HapMap project. These achievements will in the near future hopefully increase the number of successful attempts to identify genes of importance in MS susceptibility and clinical features.

## 2 AIMS OF PRESENT STUDIES

The overall aim of the five studies included in this thesis was to identify genetic variants contributing to the risk of MS in order to gain knowledge about the disease mechanisms.



### Study I

To investigate the genetic importance of 66 genes in the susceptibility to MS. Genes were selected based on chromosomal location and previous literature, as well as biological functions presumed to be of importance in autoimmune diseases.

### Study II

To confirm the genetic contribution of *LAG3*, identified to be associated with MS in study I. *CD4* was selected due to the close evolutionary relationship between the *LAG3* and *CD4*, as well as the chromosomal location, adjacent to the *LAG3* gene on chromosome 12.

### Study III

To evaluate the genetic influence of two promoter polymorphisms in the *MPO* gene in MS susceptibility and severity. One of the polymorphisms has previously been reported to be associated with MS,

### Study IV

To further investigate the role of *IL7R* located on chromosome 5 discovered to be associated with MS in study I. We wanted to confirm the initial findings, as well as refine the genetic data for *IL7R* in order to dissect the genetic association with MS in more detail.

### Study V

To investigate the genetic role of *IL7* on chromosome 8, the ligand to *IL7R*, in MS susceptibility, based on the results from study IV, implicating an important role of the *IL7R*.

### **3 MATERIAL AND METHODS**

A brief overview of the material and methods used in this thesis is presented here. For more detailed information, please refer to the individual papers.

#### **3.1 PATIENTS AND CONTROLS**

All patients with MS included in this thesis fulfilled the Poser criteria [154] for definite MS and/or MS according to the McDonald criteria [6] and informed consent, to participate in research, was obtained from all patients and controls and the work was approved by the local ethical committees.

The patients were recruited through the Neurology department at Karolinska University Hospital, Huddinge/Solna. An independent Nordic case/control material was used for confirmatory studies in study II and IV, fulfilling the criteria mentioned above, including informed consent and approval from the ethical committees .

In study I 672 MS patients and 672 controls were included. The control set consisted of 288 blood donors and 384 randomly selected non-related members from the Swedish twin registry, as well as 456 healthy controls in a second control set, used to evaluate the SNPs in the *HAVCR2* gene.

Study II included 920 MS patients and 778 controls, and 1,720 MS patients and 1,416 controls from the Nordic material.

In study III 871 MS patients and 532 controls were analysed.

In study IV, 1,210 MS patients and 1,234 controls consisting of blood donors from the Stockholm area were included, together with 1,820 patients and 2,634 controls from the Nordic material.

In study V, 1,210 MS patients and 1,234 controls were included, i.e. the same patients and controls as in study IV.

In the analysis of mRNA expression in study IV, peripheral blood mononuclear cells (PBMC) and cerebrospinal fluid (CSF) were collected from 75 MS patients, 65 patients with relapsing-remitting MS (53 patients in remission and 12 patients in relapse), and 10 patients with a progressive form of the disease, PPMS or SPMS. PBMC and CSF was collected from 48 patients with non-inflammatory other neurological disease (ONDs) (primarily patients diagnosed with headache) as a control group, along with PBMC from 20 healthy individuals.

## 3.2 GENETIC ANALYSIS

### 3.2.1 DNA extraction

Total genomic DNA was extracted from leukocytes using three different methods; salting out method [155], QiAMP DNA extraction kit (Qiagen GmbH, Germany) and PureGene (Gentra Systems, USA).

### 3.2.2 Genetic markers

In all studies (I-V), single nucleotide polymorphisms (SNPs) were selected as genetic markers.

In study I, we genotyped 123 SNPs, selected in 66 genes in a two-stage approach. A significance level of 8% was selected in order to be passed on to the second stage. Twenty-two genes survived to the second stage, and were genotyped in a larger group of patients and controls, in order to increase the statistical power to detect a true association. In addition, the number of SNPs in the genes and in flanking regions were increased. The SNPs were accessed via NCBI dbSNP ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), The SNP consortium, TSC, (<http://snp.cshl.org>) and proprietary databases of AstraZeneca.

In study II, three SNPs in the *LAG3* gene identified to be associated in study I were selected for confirmation analysis. In the *CD4* gene, nine SNPs were selected, evenly distributed over the gene for optimal coverage, from the NCBI dbSNP database. SNPs genotyped and validated by the HapMap consortium were prioritised.

In study III, two SNPs in the *MPO* gene (-463 and -129) were selected. One of the SNPs, (-463), has previously been associated with MS [156-159], and was therefore of special interest. Both SNPs have been suggested to influence the expression levels of *MPO* and they are located in the promoter region of the *MPO* gene.

In study IV, three SNPs in the *IL7R* gene, shown to be associated with MS in study I, were genotyped to confirm the initial associations. In addition twelve other SNPs were selected in order to fine-map the LD block harbouring the *IL7R* gene. Ten of these SNPs were selected due to their tagging properties, allowing inferring genotypes from 69 SNPs in total. To further pinpoint any functional variations in the gene, two non-synonymous SNPs were added for genotyping. Due to the dense SNP map of the *IL7R* gene in HapMap, the selection of SNPs were primarily based on the HapMap consortium genotype information in the CEU population (Utah residents with north- and western European ancestry).

In study V, nine SNPs were selected based on their ability to tag for a total of 23 SNPs in the *IL7* gene. The selection was based on genotype information from the HapMap consortium and NCBI dbSNP.

### 3.2.3 SNP discovery

SNP discovery in the *LAG3* gene was performed to follow up an initial finding in this gene, due to the lack of additional markers at the time of the study. All coding

sequences of the gene, as well as its promoter and 5'- and 3'-untranslated regions, were amplified in 96 subjects, by 16 separate polymerase chain reactions (PCRs). Denaturing high-performance liquid chromatography (DHPLC) was then performed using the Transgenomic WAVE System (Transgenomic, Omaha, Neb, USA). PCR products were then separated on a preheated reverse-phase column (DNASep; Transgenomic). Individuals detected by DHPLC as being heterozygous were then sequenced using ABI PRISM Big Dye Terminator (Applied Biosystems, Foster City, CA, USA), and the sequencing products were analyzed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems), in order to detect the nature and position of the polymorphism in the amplified fragment.

Two novel SNPs, located in noncoding sequences of *LAG3*, were discovered in this manner, these SNPs were later registered by others in dbSNP under the identification numbers rs2365095 and rs7488113.

### **3.2.4 Genotyping**

In this thesis, three different SNP genotyping methods were applied; Pyrosequencing – short sequencing via primer extension, restriction-enzyme genotyping and matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) - mass spectrometry of allele-specific primer extension products (Sequenom Inc., San Diego, USA). Each of the methods will be described below. Appropriate controls were included in all genotyping experiments.

#### ***Pyrosequencing***

All genotyping in study I and the SNP -129, in study III, were performed using the Pyrosequencing method [160] according to the protocol provided by the manufacturer (Biotage, Uppsala, Sweden). Primers, both PCR primers and sequencing primers were designed using the Oligo 5.0 software.

#### ***Restriction-enzyme genotyping***

In study III, the -463 polymorphisms, was genotyped using a restriction-enzyme assay. A 350-bp DNA fragment was amplified and the following PCR product was digested using 5 U of the restriction enzyme *AciI* (New England Biolabs, England). The reaction was incubated in 37° for 5 hours, before separation on a 2.5% agarose gel with ethidium bromide, identifying the genotypes.

#### ***MALDI-TOF***

The genotyping in study II, IV and V were performed by the Mutation Analysis Facility (MAF) at Karolinska Institutet using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom Inc., San Diego, USA), of allele-specific primer extension products. Two different protocols were used, hME, in study II and the confirmatory part of study IV, and iPLEX for the fine-mapping in study IV and for all SNPs genotyped in study V.

All amplification-reactions within respective protocol were run under the same conditions. The PCR and allele-specific primer extension were performed using the Mass EXTEND reagents kit and primers were designed using the SpectroDESIGNER software (Sequenom Inc., San Diego, USA). The primer extension products were analysed using a MassARRAY mass spectrometer and the resulting mass spectra was analysed with SpectroTYPER software (Sequenom).

### **3.2.5 Statistical analysis**

#### ***Single point analysis***

In all studies (I-V), a test of Hardy-Weinberg equilibrium (HWE) was performed in order to assure that allele frequencies in the control group conformed to HWE. In study I, two-sided p-values were calculated comparing carriage counts in patients and controls using a  $\chi^2$ -test, without correction for multiple testing (GraphPad, Instat). In study II and III, two-sided p-values were calculated using Fischer's exact test, with no correction for multiple testing (GraphPad, Instat). In study IV, the single-point calculations for two-sided p-values were based on a  $\chi^2$ -test, comparing carriage counts in patients and controls (GraphPad, Instat). Bonferroni correction was used to correct for multiple testing. In study V, two-sided p-values were calculated using Fischer's exact test (GraphPad, Instat) and Bonferroni correction was used to adjust for multiple comparisons. Power analysis for detecting association was performed for all studies, assuming a two-sided significance level of 0.05 and one-sided power for detecting an OR at least 1.5 for study I, and 1.3 in study II-V, without adjustment for multiple comparisons.

#### ***Odds ratio calculation***

For study I-V, odds ratios for single SNPs were calculated using Instat (Graphpad, Instat), with the approximation of Woolf for the 95% confidence interval (CI). In study IV, Mantel-Haenzsel combined ORs were calculated using Instat (Graphpad, Instat).

#### ***Meta-analysis***

The meta-analysis in study III, was performed using the method of Woolf [161].

#### ***Logistic regression***

Logistic regression was performed for single SNPs in study IV, to test for independent associations of SNPs with MS (SAS Institute Inc). In study IV and V, test for interaction between HLA-DRB1\*1501 and SNPs in *IL7R* and *IL7* were performed using logistic regression routines available in the R-package.

#### ***Haplotype analysis***

In study I, the pair-wise measures of LD was obtained using EH [162]. Based on the LD measures the haplotype structure was defined, a cut-off of  $|D'|$  greater than 0.85 was required. Estimation of haplotype frequencies based on the identified haplotype blocks was performed using the SNPHAP software (<http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt>). Test for significance of the

haplotypes was performed using the T1 statistic of CLUMP with 10,000 simulations [163]. The distribution of estimated haplotype frequencies was tested for significance.

In study II and V  $|D'|$  measures of LD between pairs of SNPs was calculated using the Haploview software, based on the EM algorithm [164]. The algorithm by Gabriel et al. implemented in Haploview was used for the generation of haplotype blocks. Test of haplotype association was performed using the same software, with 10,000 permutations [99].

Construction and estimation of haplotypes in study IV, was performed using the Haploview software [99, 164]. Due to the limitations of the EM algorithm, the robustness of the same analysis was performed with the PHASE software [165] (Bayesian based), with virtually identical results as compared to Haploview. Sliding window analysis of the haplotype block was performed using the Haploview software. To test for haplotype association, the same software was used, with 10,000 permutations.

### **3.3 EXPRESSION ANALYSIS**

This section relates to the methods used in study IV.

#### **3.3.1 Preparation of PBMC and CSF-MC**

Peripheral blood was collected using sodium citrate-containing cell preparation tubes (Vacutainer CPT, Becton Dickinson and Company, USA), and CSF was sampled in siliconized glass tubes, immediately centrifuged and the pellet stored in  $-70^{\circ}\text{C}$  until use. PBMC were separated by density gradient centrifugation, and washed twice with Dulbecco's PBS. Ninety-five percent of the cells were viable, assessed using trypan blue exclusion. Cell pellets from PBMC were stored in  $-70^{\circ}\text{C}$  until use.

#### **3.3.2 mRNA preparation**

Total RNA was extracted from lysed cell pellets (PicoPure RNA isolation kit, Arcturus Bioscience, USA). The quality of the RNA was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and only samples with completely pure RNA were used in the following RT-PCR analysis.

#### **3.3.3 Quantitative RT-PCR**

Reverse transcription was performed using 1-5 ng (10  $\mu\text{l}$ ) of the total RNA, 0.1  $\mu\text{g}$  random hexamers (Gibco BRL) and 200 U of Superscript Reverse Transcriptase (Gibco BRL). Real-time PCR was performed using a Byroad iQ5 iCycler Detection System (BioRad Laboratories, Ltd). All primers were designed with the Primer Express Software (Perkin Elmer). GAPDH was used as endogenous control, and all samples were run in duplicate. The PCR efficiency was above 90% in all PCR runs. A comparative threshold cycle (CT) was used to determine gene expression relative to the control (calibrator). Hence, steady-state mRNA levels were expressed as  $n$ -fold difference relative to the calibrator.

### **3.3.4 Statistical analysis**

Differences in relative mRNA levels in CSF and PBMCs were tested for significance with the non-parametric Wilcoxon's signed-ranks test. Analysis of correlation between mRNA and genotype was performed using the Kruskal Wallis rank sum test.

## 4 RESULTS AND INTERPRETATIONS

### 4.1 STUDY I

In the first study, performed in collaboration with AstraZeneca R&D 66 genes were investigated for association with MS susceptibility. In total 123 SNPs were genotyped in up to 672 MS patients and 672 controls, in a two-stage study-design. Thirty-four of the genes were genotyped for a single SNP and the remaining 32 genes for two or more SNPs. The selection of genes was based on chromosomal location suggested to influence susceptibility for MS and other autoimmune diseases [133, 166, 167] as well as biological and functional properties considered to be of importance in MS. Out of the 66 genes, 21 were selected only on the basis of the biological function or previous genetic associations reported for MS [168-172]. As well as being aimed at understanding the disease processes *per se*, the study was an attempt to validate potential novel drug targets.

After completion of the first stage (149-288 MS patients and 84-288 controls), 22 genes showed association with susceptibility to MS at the 8% significance level. All of the 22 genes were transferred to the second stage of the study, to confirm the preliminary associations. To increase the power of the analysis, the number of cases and controls were increased at this stage. In addition, more markers were added within genes and in flanking regions of associated genes.

After finishing the second stage of the study, two genes showed a significant association with MS, the *LAG3* gene on chromosome 12p13 and the *IL7R* gene located on chromosome 5p13. Both genes presented three or more significantly associated SNPs ( $p < 0.05$ ). Linkage disequilibrium (LD) analysis of the *LAG3* gene did not identify any LD block within the gene, and no haplotype analysis was performed. The *IL7R* gene on the other hand presented an altered distribution of the haplotypes between MS patients and controls. In the *IL7R* gene we also identified two associated haplotypes, one haplotype conferring risk of MS and one protective haplotype.

Both genes are important players of the immune-regulatory mechanisms in the immune system. LAG3 (CD223) is an MHC class II ligand that is evolutionary related to CD4, and downregulates the activated T cells by blocking the binding of CD4 [173]. The interaction of the MHC class II and LAG3 during inflammation may also be of importance in the activation of antigen-presenting cells [174].

IL7R is a member of the haematopoietic receptor family and is a type 1 membrane glycoprotein capable of binding alpha-helical cytokines [175]. The IL7R complex consists of the IL7R alpha chain (IL7R- $\alpha$ ) and the common cytokine-receptor gamma chain (CD132), where IL7R- $\alpha$  transduces transmembrane signals through the recruitment of intracellular messengers to its cytoplasmic tail, while CD132 activates this transduction [175, 176] (Figure 3). It is expressed by cells of the lymphoid lineage, in which it has essential non-redundant functions. IL7R is required for the development of  $\gamma\delta$  T- cells but not natural killer (NK) cells, in addition, NK T-cells do require signals from IL7R for proper development [177].

IL7R- $\alpha$  also binds to the thymic stromal lymphopoietin receptor (TSLPR) to form another receptor complex, where TSLP is the ligand [178, 179] (Figure 3). This complex has been shown to be of importance in the maturation and activation of dendritic cells (DCs) (reviewed in [180]), and decreased levels of TSLP have been shown to be associated with Crohn's disease, suggesting TSLP to be of importance in autoimmune diseases [181].

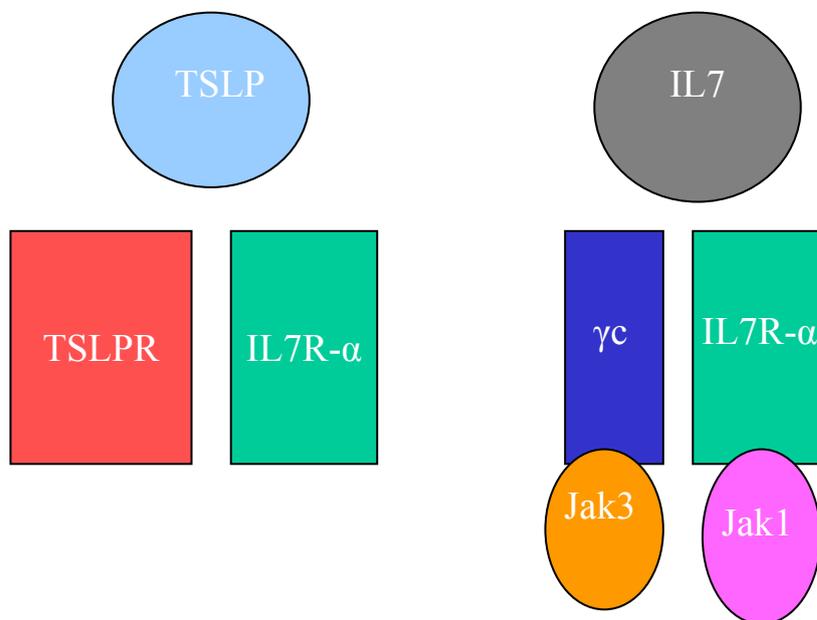


Figure 3. The IL7R- $\alpha$  is part of two different receptor complex. The IL7R- $\alpha$  binds to  $\gamma$ c to form the IL7R, where IL7 is the ligand. IL7R- $\alpha$  also binds to TSLPR, where TSLP is the ligand.

Signalling via IL7R induces somatic recombination of T-cell receptor and immunoglobulin genes, and IL7R is expressed on both mature T and B cells. The receptor is essential for the human T-cell proliferation and survival [182]. IL7R has also recently been suggested to function as a marker for regulatory T-cells (Tregs), where absence of IL7R on CD4<sup>+</sup>CD25<sup>+</sup> T-cells could function as a Treg marker [183].

The need for large case/control materials when performing association studies in complex diseases has been illustrated in this study. Of the 22 genes that showed an initial association in the first stage, only two genes, *IL7R* and *LAG3*, survived the second stage, and were considered to be associated with MS in this population. The lack of sufficient power in many genetic studies has resulted in an increasing number of suggested MS genes, where very few have been confirmed in independent studies. The limited number of markers studied in each gene in the present study, the possibility of false negative findings, and genetic variants of the non-associated genes in the present study may still be of importance in MS susceptibility [184].

## 4.2 STUDY II

In order to validate and confirm the association of *LAG3* identified in the previous study, a follow-up study was performed. The *CD4* gene was included in the study as we

were unable to exclude that the association with *LAG3* was in fact attributable to a signal arising from the neighbouring *CD4* gene due to the LD between the two genes. The evolutionary relationship between them and a previously observed association between the *CD4* gene and MS also encouraged us to study both genes [185].

*LAG3* and *CD4* were analysed in two independent case/control materials, a Swedish and a Nordic (Norway, Denmark and Finland) material. The Swedish material consisted of 920 MS patients and 778 controls, and the independent Nordic material was comprised of 1,720 MS patients and 1,416 controls.

In *CD4*, nine SNPs were initially genotyped in the Swedish case/control material to test for association with MS. The markers were selected to be evenly distributed over the gene region. Three of these markers showed a significant association ( $p < 0.05$ ) with MS, encouraging us to test these markers in the independent Nordic case/control material. None of the initial associations were replicated in the larger Nordic material. The ethnic similarity between the populations as well as the high power, due to the large sample size indicates that the initial association detected was likely due to a type I statistical error. Pair-wise LD was calculated using the  $|D'|$ , and two LD blocks were identified within *CD4*. The haplotype analysis did not reveal any difference in distribution of the haplotypes between cases and controls, and subsequently no risk or protective haplotypes were identified.

In order to seek confirmation for the preliminary associations in the *LAG3* gene, three SNPs were selected for genotyping based on previous results from study I. None of the markers genotyped were confirmed when analysed in the independent and larger Nordic material. One of the SNPs, rs879849, did present a significant p-value ( $p = 0.01$ ), but a deviation from HWE was observed in the control group, indicating a false positive result, and therefore it was considered as a non-significant finding.

The genes studied, *CD4* and *LAG3*, had increased prior probability in a genetic context to be involved in MS and they are both involved in the regulation of immune responses, acting on the regulation of activated T cells [173].

The *CD4* gene is located on chromosome 12pter-p12, and the protein it encodes is expressed on mature T-helper cells together with the T-cell receptor (TCR) [186, 187]. During antigen recognition *CD4* interacts with non-antigen-binding regions of the MHC class II molecules [188]. Further, a low expression of *CD4* on the cell surface results in impaired TCR response to antigenic stimulation, and reduced efficiency on immune responses [189].

The *LAG3* gene is located on chromosome 12p13.32 and is evolutionarily related to *CD4*. *LAG3* is an MHC class II ligand and downregulates activated T-cells by blocking the binding of *CD4* [173]. The binding of *LAG3* to MHC class II interferes with the binding of *CD4* to MHC class II, thereby blocking the stimulatory signals mediated by *CD4* [190-192]. The interaction of the MHC class II and *LAG3* during inflammation may also be of importance in the activation of antigen-presenting cells [174]. Furthermore, during inflammation both *LAG3* and MHC class II are up-regulated, indicating the important role of *LAG3* in inflammatory processes [193].

In this study we present data that, once more, illustrates the importance of access to large case/control material and replication of initial genetic associations. Even though the initial genotypings of both *CD4* and *LAG3* were well-powered, we failed to replicate the preliminary findings in a second independent material, pin-pointing the difficulties in adequate and correct evaluations of initial genetic association findings [184]. The major conclusion of this study would be that even though both *CD4* and *LAG3* proteins may still be of importance in MS pathogenesis, we failed to find evidence for a genetic contribution of these genes to MS susceptibility.

### 4.3 STUDY III

Several genes have been investigated for their genetic involvement in MS, and even though sometimes, a number of investigators have focused on a particular gene, no consensus regarding the genetic association has been made. One such example is the *MPO* gene, located on chromosome 17q23.1, a region suggested to be linked with MS [125, 156]. A number of studies have investigated the genetic influence of a specific marker (-463) in relation to MS, without reaching a conclusive result. This could be mainly due to the lack of sufficient power in the individual studies [156-159, 194]. In study III, we therefore wanted to explore the genetic association with MS, analysing two SNPs (-463 and -129) located in the promotor region of the *MPO* gene in a larger Swedish case/control material consisting of 871 MS patients and 532 controls, aiming to reach sufficient power to provide a more definitive answer. Both SNPs have been suggested to influence the expression and protein levels of MPO, and were thus appealing candidates [195].

Neither of the SNPs showed any significant association with MS with regard to susceptibility, and we conclude that these SNPs are not influencing the susceptibility to MS. In an attempt to summarize the efforts put into this gene, a meta-analysis was performed for the -463 polymorphism. This analysis included four studies, including the present one. Adding the data from the four different studies together, the negative result was further strengthened, since no significant association could be detected in the meta-analysis.

MPO is an enzyme that is expressed by myeloid cells, catalysing the production of hypochlorous acid. The enzyme is also involved in inflammatory processes, where activated neutrophils can release active MPO, causing tissue damage through strong oxidative activity [196]. Due to the tissue damaging effects of MPO, we wanted to investigate associations with the genotyped SNPs and disease severity. To test for associations with disease severity, the Multiple Sclerosis Severity Scale (MSSS), was used [197]. Data regarding EDSS, disease duration along with genotype information for the two SNPs was collected for each patient, in total 508 MS patients. We could not detect any association between *MPO* genotypes and disease severity, suggesting that these markers do not influence the disease severity.

Due to the small number of SNPs analysed in this study, the conclusions drawn only represent a fraction of the *MPO* gene. The importance of other genetic variations in the gene can therefore not be ruled out, and further functional consequences may be of importance both in the susceptibility of MS, but possibly more likely, due to the functional properties of the molecule, in disease progression and severity,

#### 4.4 STUDY IV

In the light of the present status in the genetic field of MS, genes identified with a primary association with MS need confirmation in independent materials in order to acquire the status of a true MS gene. Regarding *IL7R*, two Australian reports have raised the attention to this gene in MS susceptibility [198, 199] along with the results from study I.

To further investigate the genetic role of the *IL7R* gene in MS, an expanded confirmatory study was performed. To seek confirmation for the three associated SNPs from study I, a Nordic case/control material consisting of 1,820 MS patients and 2,634 controls was obtained. This material was completely independent from the material used in study I. All three markers were confirmed to be associated with MS, and the haplotype analysis was consistent with the previous haplotype findings, strengthening the observation of a true association.

Due to the limitation in the number of identified SNPs at the time of performing study I, the SNP coverage of the *IL7R* gene was not complete. In order to acquire genotype data from the whole gene and the surrounding LD block harbouring the gene, an expanded fine mapping tagSNP study was performed. According to data from the HapMap CEU population, *IL7R* is located within a tight LD block based on  $r^2$ -measures, not harbouring any additional genes. To map the complete gene, and more precisely map the risk-modifying variants, twelve additional SNPs, ten tagSNPs and two non-synonymous SNPs were genotyped in a Swedish case/control material (1,210 MS patients and 1,234 controls). The three previously confirmed SNPs were included in order to increase the comparability of haplotypes between studies IV and I. The fine mapping effort revealed three associated SNPs, surviving correction for multiple testing. A SNP in exon 6 (rs6897932), shown to be important for the splicing of *IL7R* in a paper published back-to-back with study IV [153], showed the most significant association ( $p=0.001$ ).

Trying to further pin-point the association effect and to test whether any SNPs were independently associated with MS, logistic regression analysis was performed which showed that the co-dominant model of inheritance showed the best model fit. The SNP in exon 6 was identified to best explain the genetic association for the *IL7R* gene and MS, consistent with the previous single-point data.

Along with the single-point analysis, haplotype analysis was performed. One associated haplotype was identified; the most common haplotype conferred a protective effect. The positively associated 3-marker haplotype confirmed in the Nordic material was

split into four different haplotypes, each insignificantly more common among cases than controls. In addition, a sliding-window analysis, including four markers in each window of the haplotype block revealed a slightly stronger protective haplotype effect, including the previous 3-marker haplotype together with the most associated SNP from this stage.

Functional analysis including mRNA expression of the IL7R and the ligand IL7 was performed in CSF and peripheral blood mononuclear cells (PBMCs) from 75 MS patients and 48 non-inflammatory controls (OND) together with 20 healthy controls contributing with PBMCs only. IL7R was in general more highly expressed in CSF than PBMC regardless of the status of the individual. Further, a significant increase in expression of both IL7R and IL7 were detected when comparing CSF of MS and OND patients. The existence of a functional difference between MS patients and OND patients increases validity of claiming *IL7R* as true MS gene. No correlation could be detected between mRNA expression and genotype.

The identification of non-HLA genes with an implication for the risk of MS has been difficult. Until recently, the only confirmed genetic risk factor for MS has been the HLA-DRB1\*1501, illustrating the difficulties in identifying genes modifying the risk for complex diseases. The reason for this is probably weak influences of each genetic factor relative to the limitation in access to large case/control groups, increasing the risk of type I and type II errors. With the *IL7R* we present an exception to this dilemma, with a gene that has been analysed and confirmed in three independent populations (Australian, Nordic and American) including study IV [153, 199], showing significant differences between MS patients and controls. Evidence for differences in expression levels between MS patients and controls and the altered distribution of soluble and membrane bound form of IL7R, further confirms the importance of this molecule in MS [153, 198].

#### 4.5 STUDY V

The importance of the *IL7R* in MS was demonstrated in study I and IV and in order to further understand the role of the IL7R complex, we wanted to investigate the potential genetic contribution of the ligand, IL7, to the risk of MS. *IL7* is located in chromosome 8q12-q13 and harbours six exons. IL7 binds to the alpha chain of the IL7R, which consists of the IL7R alpha chain and the common cytokine receptor gamma chain  $\gamma_c$  (Figure 3) [200]. It is a non-redundant cytokine, essential for the human T-cell development and survival, but also important in the proliferation and survival of mature T-cells in the periphery [201]. The mechanism behind IL7 production is not clear, but IL7 has been reported to be expressed in stromal tissue, epithelial cells, in the thymus and in the bone marrow [202].

In total nine SNPs were selected based on genotype information from the HapMap consortium. The markers were selected to acquire genotype information for a larger number of SNPs, representing all parts of the *IL7* gene, using the tagSNP approach.

All SNPs were genotyped in a large Swedish case/control material consisting of 1,210 MS patients and 1,234 controls. None of the nine SNPs showed any significant association with MS after correction for multiple testing. Frequencies of haplotypes were estimated and haplotype association tests were performed, without revealing any significant difference between cases and controls.

From this data we conclude that any genetic contribution of *IL7* to MS susceptibility is not detectable in our material. Although negative, this finding may in fact be a supportive result in the story of the *IL7/IL7R* complex in MS. The lack of association for the *IL7* gene is in accordance with the functional studies of the *IL7/IL7R* complex suggesting that signalling is regulated by a differential expression of the receptor and not the ligand (reviewed in [203]). The hypothesis of the importance of the *IL7R* pathway in MS has been strengthened by these results, and thus supporting our conclusion that the *IL7R* pathway is of importance in MS.

## 5 DISCUSSION

This thesis is based on five papers, all with a focus on genetic association studies in MS. The evolution in this field over the time of my postgraduate education has been remarkable. The release of the sequence of the human genome and as a consequence the establishment of the HapMap consortium has taken geneticists to a whole new world of possibilities.

As a reflection of the rapid development in the field, my time as a PhD-student mirrors this process. In study I, at that time the biggest candidate screen in MS, we assessed 66 genes, genotyping in total 123 SNPs in a total of 672 MS patients and as many controls. With our current knowledge of the nature of the human genome, it was a rather modest attempt to elucidate the genetics of MS. Four years later, in study IV and V, the number of SNPs genotyped have increased dramatically and the size of the case/control materials used has increased with almost 100%. These studies are in many aspects still small genetic studies, but the evolution from study I to V in this thesis reflects the trends and development in the field to some extent. But one should not neglect early studies as a rule, small studies may still be of importance. The fact that the importance of *IL7R* in MS, was first described in a small study which was later replicated, shows that even the small studies can be of great importance [198, 199].

### 5.1 *IL7R* IN MS

The journey of *IL7R* in the MS genetics world began in 2003 [199] and is starting to reach its final conclusion in 2007. This five-year journey of *IL7R* does say something about the difficulties of identifying and confirming genes beyond doubt in a complex disease. In the case of *IL7R*, the genetic evidence for an important role in MS susceptibility is mounting. Recently, three studies, including study IV, were published presenting data that *IL7R* is indeed the first non-HLA gene identified in MS [152, 153]. The identification of this gene represents quite a breakthrough in MS genetics, since no unequivocally confirmed genes have been identified since the HLA associations reported in early 1970's.

A number of SNPs have been associated with MS in *IL7R* although fine-mapping efforts in study IV as well as in the accompanying paper by Gregory et al [153], points out one SNP in exon 6, rs6897932, to be the most important. A joint analysis of the three recent studies of *IL7R* in MS [152, 153, 204] presents an impressive p-value for this specific SNP,  $p=5.84 \times 10^{-12}$ , providing an association with genome wide significance [205]. The location of the SNP within a coding region, and with a non-synonymous change causing an amino acid substitution, makes this SNP the most appealing candidate for the possible disease modifying effects of *IL7R*. Haplotype analysis has also pointed out the importance of this SNP, since haplotypes including this SNP present the strongest evidence for association.

An increasing amount of data suggest functional alterations of *IL7R* in MS patients, thus supporting an important biological role for the molecule in the disease aetiology. In addition to the genetic association, Gregory et al. [153] present data concerning a

splice variant of *IL7R*, where exon 6 is skipped. The risk allele of the exon 6 SNP shows an increase of the soluble form (spliced variant) of the receptor in relation to the membrane bound form. The soluble form of IL7R lacks the transmembrane domain, but maintains the ability to bind IL7, although the biological relevance of this complex remains to be elucidated (reviewed in [206]). The increase of soluble IL7R could result in an impaired signalling function, whereas an increase of the soluble receptor could reduce important survival signals for proliferation and maintenance of the T-cell population. Cell survival and proliferation mediated by IL7R is maintained by increasing expression of survival proteins like B cell lymphoma 2 (BCL-2) and myeloid-cell leukaemia sequence 1 (MCL1) and altering the distribution of death proteins like BCL-2-associated X protein (BAX) and BCL-2-antagonist of cell death (BAD) (reviewed in [207]). A shift in this balance could result in disturbed cell survival, and in a possible accumulation of auto reactive cells. The importance of IL7R in immunological homeostasis was recognized early, due to the severe effect of specific mutations in *IL7R*, causing a particular form of severe combined immunodeficiency (SCID) [208, 209].

In addition, the IL7R has lately received attention as a potential marker for regulatory T cells, where a low expression of IL7R by CD4+CD25+ T-cells (Treg) correlates with high expression of FoxP3 [183]. Tregs are important players in the immune system, acting as a balance between pro- and anti-inflammatory mechanisms protecting from autoimmunity (reviewed in [2]). The observation that MS patients have an altered function of their Tregs in the periphery could be an important key to the understanding of the autoimmune mechanisms involved in MS. The suppressive function of the Tregs has been shown to be altered in MS patients compared to controls in PBMC, although the frequency of the cells did not differ between the groups, suggesting an impaired function of the Tregs in MS [210, 211], but other studies could not detect any alterations of the Treg functions [212, 213]. In contrast to PBMC, in CSF the frequency of Tregs has been shown to be increased in MS patients compared to ONDs. The increase of Tregs in the CSF, is suggested to be due to recruitment of Tregs from the periphery into the CSF, rather than expanding of Tregs once they reach the CSF [214]. These observations indicate the importance of Tregs in MS pathogenesis and the relationship between Tregs and IL7R present an appealing hypothesis of the function of IL7R in MS.

An altered expression of IL7R on Tregs could be one explanation for the decreased suppression properties of Tregs derived from MS patients. The increase of IL7R expression observed in the CSF but not PBMC, in study IV should be further investigated to understand the specific role of an altered expression of IL7R in different compartments. The data presented so far related to IL7R expression in MS, are not consistent. Gregory et al. [153] do present an altered expression of IL7R in the PBMCs, and Booth et al. present differential expression of IL7R in PPMS compared to RRMS in the periphery [198], whereas we could not detect any difference in expression level of IL7R in PBMCs when comparing MS patients, OND patients and healthy individuals.

In summary, these findings suggest an important role of the Tregs in the pathogenesis of MS, where the *IL7R* association may play an important role in modulating the function of the Tregs.

The utilization of IL7R as a possible drug target for MS treatment is a difficult issue. The unique properties of the receptor, being a non-redundant cytokine receptor make modulation of its biological function challenging. In human immunodeficiency virus (HIV) infection the possibility of targeting the IL7/IL7R complex for therapy has been discussed, but the situation in HIV is somewhat different from MS. Their main concern regarding using IL7 therapeutics for HIV is the risk of increased viral infectivity and the risk of autoimmunity (reviewed in [206]). Both these concerns would be even more problematic in MS, due to the autoimmune nature of the disease and the proposed influence of virus infections in MS pathogenesis. Using IL7/IL7R as a drug target in MS would need further biological understanding of this complex and the signalling pathways in order to modulate this system in a beneficial manner for MS patients.

Even though the genetic data support an important role of *IL7R* in MS, the exact biological and functional consequences of the genetic findings are still to be elucidated. A genetic association could be regarded as a statistical finding until the biological effect of the genetic variant has been demonstrated and understood.

In general, the association of *IL7R* strengthens the hypothesis of MS as an autoimmune and T-cell mediated disease. It also gives credit to the disease modifying treatments used as of today in MS care, since the present treatments all have mechanisms of action developed on the basis of MS as an autoimmune and T-cell mediated disease.

## **5.2 HOW TO SUCCEED IN MS GENETICS?**

The development of MS genetics has moved from almost exclusively focusing on linkage studies to the present situation, with total devotion to the association approach. One could always argue that this is just a trend and that linkage will regain its glory, but linkage will most likely not dominate the MS genetics again. The shift towards association studies can be attributed to a number of reasons, where lack of power and low resolution in linkage screens are the main reasons. An exception to this situation is extended pedigrees with multiple affected individuals, where a specific inheritance patterns can be assumed. In these cases, the linkage approach may still be considered useful, and provide important information, which hopefully can be applied to the general MS population.

The overall trend in genetics of complex diseases goes towards larger case/control data sets, and extensive genotyping. As previously mentioned, the evolution has been rapid, and the requirements of the genetic studies today are challenging.

### **5.2.1 Selection of patients and controls**

The outmost important component in the study design is the selection of cases and controls. Inclusion criteria for affected individuals need to be stringent, and should be in accordance with the current diagnostic criteria, to avoid dilution of the patient group by including non-MS patients.

Selection of the control group may be even more difficult, but just as important. Due to the geographical and ethnic differences in the distribution of MS, the selection of controls may be a critical issue in many studies. In addition, allele frequencies in general differ between populations, *population heterogeneity*, and therefore matching for ethnicity is crucial to obtain reproducible results. A debate as to whether the control group should be matched to the patients with regard to age and gender is ongoing. One could argue that a perfect matched control group will give the most reliable results since most confounding factors have been accounted for. On the other hand, a perfectly matched control set will not reflect the general population, in which you want to estimate the effect of an association. The matched control set will also limit the possibility to identify interactions between gender and genotype for example. The answer to this dilemma is not clear, and the debate regarding selection of control groups will continue.

Using genomic control is another strategy to avoid spurious associations due to population heterogeneity. This approach includes additional genotyping and difficulties in the selection of unbiased markers could be an issue. The approach to control for population heterogeneity was recently utilized in the large GWA study by the Wellcome Trust Case Control Consortium proving this strategy to be valid [89].

The choice of material to study is another important issue to consider; case/control materials and family materials like trios and sib-pairs both have their benefits but also disadvantages. The use of family materials provide important information regarding phase, valuable in haplotype analysis and are less sensitive to population stratifications but on the other hand, have less power and require more genotyping, in contrast to the case/control materials. The ultimate choice would be a combination of the two. This would allow for correct haplotype analysis, and still provide sufficient power to detect associations.

### **5.2.2 Design issues**

We have the knowledge to design high quality studies, in theory, but access to sufficient data sets, good quality phenotypic data and funding are still limiting factors. One key component in designing genetic studies are the issues of sample size and sufficient power. Lack of power hampers many studies, with subsequent inconclusive results. Aiming for accurate sample sizes will make large case/control sets crucial to being successful in finding genetic modulators in MS. The recent GWAs reported for a number of diseases including MS [89, 152], present data that indicate that case/control materials of 1,000/1,000 may be a “threshold” for a minimum sample size in order to achieve reasonable power to detect associations. Access to even larger materials (2,000 cases/2,000 controls) substantially increase the power, even so, in order to actually claim full power to reach p-values with genome wide significance, even larger materials at least 10,000 cases /10,000 controls may be required [215].

In the light of our current knowledge, the majority of genetic risk factors in MS will contribute with a small, but important piece of the puzzle. The odds ratios observed for genetic modulators in MS will subsequently be modest, in many cases not exceeding  $OR < 1.4$ , even though the importance of the identified genetic factor may be central.

This fact illustrates clearly the need for properly powered studies in order to capture these fairly small effects and the importance of acknowledge confirmed, true associations even when the OR's are not extreme.

Other important aspects to consider are the allele frequencies of the genotyped markers, LD structure in the area of interest as well as strict inclusion criteria for patients and controls. In addition, seeking confirmation for initial findings is necessary to rule out or confirm candidate genes at an early stage. In the perfect world, all studies should be subjected to confirmation effort regardless of a positive or negative initial finding, in order to assure that accurate and reliable genetic data are reported.

The studies performed today are most often designed based on the hypothesis of “*common disease- common variant*” (CDCV). The CDCV hypothesis stipulates that common genetic variants, present in a significant proportion of the population, both affected and unaffected individuals, are the cause of common diseases with a genetic component [71]. The extensive support for CDCV may in part be due to the fact that “this is what is possible to study”, i.e. it is in a way convenient to apply this strategy, but this view may be challenged. Arguing the validity of CDCV points out an important issue in genetic research, the importance and contribution of rare alleles in complex disease “*common disease- rare variant*” (CDRV). To evaluate rare variants, substantial data sets are needed, up to hundred of thousands of cases and controls. As of today, many geneticists consider it nearly impossible to perform studies focusing on rare variants due to these conditions. Collecting large, well-characterized case/control materials in collaboration could be beneficial in attempts to understand the role of rare variants in MS. These studies will have to be carefully controlled for possible population stratification in order to succeed, but in the complex world of complex genetics, it would be arrogant to ignore the importance of either common or rare variants. Most likely a combination of the two will contribute to the susceptibility to MS and our mission is to understand this interplay, even though it may be an almost impossible task, we cannot ignore it.

### **5.3 NEGATIVE OR POSITIVE ?**

A critical point in MS genetics is *when* a gene should be regarded as a disease gene or be excluded to be of importance. As discussed in the previous section, differences in allele frequencies among populations suggest that population specific associations could contribute to the genetic susceptibility. Another issue is the level of significance selected. Do we as scientists accurately interpret the results as significant or non-significant in a biological context, or is the statistical significance a poor predictor of biological validity and instead leads us in the wrong direction?

In this thesis, we present data for a number of candidate genes. Due to the lack of statistical power and limited number of SNPs genotyped in study I, the “negative” genes may still be of genetic importance, yet undiscovered, and one should not regard these genes as excluded and as such, not as true negative findings. In contrast, the second study could be regarded more conclusive in this respect. The initial associations in *LAG3* and *CD4* were subjected to genotyping in a second independent population to

seek confirmation. The lack of confirmation in this larger, well-powered and independent data set could justify the conclusion that these genes are not of significant importance in MS susceptibility in these populations.

Selection of a more conservative significance level in the initial stage of this study could have saved us from expensive genotyping in the confirmation step, but one should be cautious to rely exclusively on the magnitude of p-values. A more strict selection would, as just stated, have saved us from genotyping *CD4* and *LAG3* and still reach the same conclusion. On the other hand, in the case of *IL7R*, a more conservative cut-off level would have made us regard the results as non-significant, and an important true association would have been missed. How to assess significance remains difficult, due to the modest effect of each risk factor and subsequently even fairly modest p-values may account for true associated variants.

### 5.3.1 Multiple testing

The large number of markers analysed in a study, will occasionally produce significant results merely due to the large number of comparisons made. To overcome this problem, different strategies to correct for multiple testing have been developed. Adjusting multiple comparisons using the Bonferroni correction is regarded by many geneticists to be too conservative, causing true associations to be regarded as non-significant after correction. The Bonferroni correction assumes the tests to be independent of each other, which could be difficult to determine due to LD, making Bonferroni correction best suited for correction of single-point data.

Less conservative methods suitable for single-point and haplotype analysis include permutation tests and simulation of the data set. Using independent case/control materials to confirm preliminary findings is maybe the most accurate method, but also the most expensive. The lack of consensus in this matter has led us to report uncorrected p-values along with corrected p-values using different correction methods in all studies included in this thesis. Overall I find correction by replication to be the best method since it reflects the actual biological situation without taking the magnitude of the p-values into account. With regard to the modest p-values achieved for *IL7R* in study I, the other methods of correction would have made us drop the initial finding as negative after correction for multiple testing.

It is not just in individual studies that multiple testing is an issue. The fact that initial findings with modest effect are replicated in number of studies, at best providing modest p-values, could generate false-positives simply due to the large number of tests performed globally. This is more difficult to correct for since the actual number of analysis performed for a specific marker is not known.

### 5.3.2 Publication bias

Numerous genetic studies with negative results have been published during the years, but an even larger number of negative studies may never be published due to the lack of interest in negative findings from journals and the researchers themselves, *publication bias* [216]. This represents a major issue since publication is the major source for researchers to gain access to research data from the scientific community. Reports of

negative findings are essential to scientists in order to validate and make decisions about up-coming studies, since negative findings represent important knowledge and contribute to the future directions in research. The journals do have an obligation to publish well designed and conclusive reports regardless if the findings are negative or positive, and researchers should be encouraged to publish these findings as well.

#### **5.4 HAPLOTYPES – WILL THEY HELP US?**

The observation that the human genome is organized in blocks, where certain blocks are inherited together more often than would be predicted by chance due to LD, have influenced the strategies of analysis. The focus of genetic analysis has turned from analysis of single markers to haplotypes. A number of reasons have been presented in favour of using haplotypes to improve genetic analysis as compared to single markers only; the haplotypes can be argued to define functional units of a gene by protein sequence corresponding to haplotypes, population genetic variations are structured into haplotypes and using haplotypes increase the statistical power by reducing the number of tests (reviewed in [217]). Utilizing haplotypes minimizes the genotyping effort without losing too much power.

The utilization of haplotype analysis also has issues that need to be addressed. First, it assumes that the CDCV hypothesis is valid since the haplotype analysis will not cover rare variants, secondly, analysing case/control data rarely includes information regarding phase of the genotypes which instead must be inferred, introducing possible errors and thirdly, haplotype analysis relies on high LD.

Even though almost all scientists in complex genetics have performed haplotype analysis, the aims are not clear in all cases. In the situation of an identified haplotype association, the geneticists rarely settle with that. Instead, we want to further dissect the haplotype and ultimately find the disease-causing variant within that haplotype rather than accept the haplotype as the significant finding. At present, the single-point analysis goes hand in hand with haplotype-based analysis. Both are important contributors to the progress of MS genetics. The haplotype analysis could be used in different ways; it could serve as a strategy to scan genomic regions to detect associations that could be further corroborated, or as a way to identify genetic elements influencing biological mechanisms.

In the studies included in this thesis, haplotype analyses have been performed for all studies but study III. In the case of *IL7R*, the haplotype analysis further strengthened the conclusion that this gene was associated with MS. Still the actual cause of the risk modifying properties of the association is not clear. The T-allele of the most associated SNP was also unique to the protective haplotype identified, but the relation between the single SNP and the haplotype needed further evaluation in a functional setting in order for us to draw firm conclusions about the origin and cause of the association effect.

In study II, the haplotype analysis of the *CD4* gene, did not reveal any association in contrast to the single point data in the Swedish material. This observation was not replicated and the gene was considered not associated with MS in accordance with the

haplotype data. But one could think of a situation where single markers show significant association in absence of a haplotype effect. This could be due to mainly two reasons. First, the actual, true disease modifying variant has been captured, acting on its “own”, explaining the lack of haplotype effect and secondly, which may be a more common situation, the single-point data are false positives and thus do not present a haplotype association. The analysis and interpretation of haplotype data will most certainly be further developed and improved, but at present, a combination of single-point- and haplotype analysis is the best approach to dissect genetic data.

## 6 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

In this thesis we have made an attempt to increase the knowledge concerning the genetic component in MS by investigating candidate genes using the association approach.

In the initial stage, three genes showed evidence for association with MS, *LAG3*, *IL7R* and *CD4*. Upon seeking confirmation of these results in an independent case/control material, only the *IL7R* gene remained associated with MS. The solid confirmation of *IL7R* as a genetic risk factor for MS susceptibility represents a breakthrough in the search for “MS genes”.

With the rapid development and achievements in the field of complex genetics during the last few years, the future looks bright. The challenge of genetic studies has moved from genotyping in the laboratory to development of clever analysis. It is almost impossible to predict the future, but one can have dreams and hopes of what the future in genetic research will hold. So far, genetic research has been focused on identifying single genes contributing to an altered susceptibility to MS. The interactions of genes, *gene-gene interactions*, and genes and environment, *gene-environmental interactions*, have not yet been very well investigated. In the future, these interactions will hopefully receive the attention they deserve. In addition, the genotyping will probably be replaced by rapid whole genome or regional sequencing, where the full individual variation will be identified. Further, international collaborations will be essential in establishing joint bio-banks to enable access to large sample sets with high quality phenotypic data. The past, represented by small studies performed by single research groups will be replaced by the signature of the future, large collaborative studies, set out to solve the mystery of MS.

The near future will be an interesting time in MS genetics, and especially for researchers devoted to *IL7R*. The answer to the question of what the function of *IL7R* is in MS may be just around the corner, opening up a whole new set of opportunities to get closer to the truth about MS. Step by step we are going to find the truth!

*My own personal dream is that we in the near future have identified the majority of genetic risk factors for MS and understood their biological function, transferred this knowledge into drug development and MS patients will no longer be “patients”.*

## 7 ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor, **Professor Jan Hillert**, for giving me the opportunity to do a PhD in your research group. It has been a privilege to be part of your group, and you have taught me so many important things, not only about MS but also how to perform excellent research. Your enthusiasm for MS genetics is inspiring and you always know how to encourage your students to do that little extra.

I would also thank my co-supervisor, **Doctor Hugh Salter**, my boss at AstraZeneca. You made me understand that nothing is impossible, and made me realize that doing a PhD is great! Thank you for “letting me out” of AstraZeneca to do my thing over these years.

Many thanks to **Professor Maria Anvret**, who gave me the opportunity to get involved in MS genetics in the first place.

I would like to express my deepest gratitude to all patients and healthy volunteers who have participated in the studies included in my thesis; without your contribution this research would not have been possible.

I want to thank all members at the Division of Neurology:

**Kerstin Imrell** for your enthusiasm in research, and for travel company to conferences all over the world. **Boel Brynedal** for amusing discussions regarding ancient haplotypes and excellent ski company in the Rocky Mountains. **Iza Roos** for truly inspiring discussions, not only about science, but everything that matters in life. **Kristina Duvefelt** for all collaborations and fun moments during the last years, and for all genotypes...! **Eva Åkesson** for your patience with my and my questions during the thesis writing. **Thomas Masterman** for all help during what ever writing I have tried to do, and your true commitment to research. **Jenny Link** for bringing fresh ideas into our genetics group. **Ingrid Kockum**, you always have an extra minute to discuss “urgent” issues. **Viriginja Karrenbauer** for your never ending energy. **Kristina Gottberg** for introducing new perspectives of MS research. **Cecilia Svarén-Quiding**, for sharing the love for Västergötland, the best part of Sweden! **Anna Fogdell-Hahn** for your commitment to science and for fruitful discussions. **Annelie Mattisson** for all the early breakfasts and nice chats!

Also many thanks to **Jenny Ahlqvist**, **Anna Cunningham**, **Christina Sjöstrand**, **Eva Johansson**, **Marjan Jahanpanah**, **Kosta Kostulas**, **Sverker Johansson**, **Helena Modin**, **Susanna Mjörnheim**, **Raymond Press**, **Malin Lundqvist**, **Ajith Sominanda**, **Sanjaya Adikari**, **Leszek Stawiarz**, **Anna Steinberg**, **Mathula Thangarajh**, **Erik Nelson**, **Pernilla Andersson**, **Lisa Grönlund**, **Anna Hillert**, **Merja Kanerva**, **Faezeh Vejdani** and **Anna Mattsson** and all nurses at the out-patient clinic for help with sampling from patients and controls.

**Gunnel Larsson** and **Yvonne Sjöлинд** for your excellent help with practical matters.

The MS group in Solna for fruitful collaborations:

**Maja Jagodic and Johan Öckinger** for interesting discussions regarding design issues and genotyping and for the chemokine collaboration. **Mohsen Khademi, Ellen Iacobaeus, Erik Wallström** and **Professor Tomas Olsson** for the collaboration that took us all the way to Nature Genetics!

All members of the Nordic MS genetics group for excellent collaboration: **Hanne F Harbo, Anne Spurkland, Frode Vartdal, Elisabeth G Celius, Kjell-Morten Myhr, Oluf Andersen, Magnhild Sandberg-Wollheim, Jorma Ilonen, Leena Peltonen, Janna Saarela, Annette Oturai, Per Solberg-Sorensen, Arne Svejgaard** and **Lars P Ryder**.

My colleagues at AstraZeneca R&D, Discovery, where it all began: **Zhiping Zhang** for introducing me to genetics and teaching me all about genotyping! **Dennis Hellgren**, you always take time to explain the mystery of bioinformatics. **Tesfai Emahazion, Gunnar Falk** and **Anna Wallin** for all amusing and frustrating moments in the genetics lab. **Henric Zazzi** for your patience with me during MAGDA sessions. Also many thanks to **Maria Nilsson** for making me feel so welcome when I first came to AstraZeneca, and for being a friend ever since.

Also many thanks to ALL of my colleagues at the former department of “Molecular Sciences” at AstraZeneca R&D.

**Gudrun Jonasdottir** – you make statistics understandable!

**Cecilia Lindgren, Astrid Lindstedt, Linda Berglind** and **Ville Mäkelä** at the Mutation Analysis Facility for all the genotypes!

Many warm thanks to:

My singing friends in “Ray of Light”, **Nina Fröidh, Åsa Fröidh, Karin Nyström** and **Gunilla Thorsén**, you are all fantastic, singing with you gives me energy!

**Shideh O’Brien** and **Carina Raynoscheck**, for being my friends and for always believing in me!

**Helena Almqvist** and **Malin Lindborg**, “Stad i ljus”, say no more, thank you for all dinners, parties and fun!

My Skara-girls, **Kristina von Oelreich** and **Jenny Hassel** for being true friends no matter what!

My Linköping friends **Marie Waltersson-Ahnström, Ninna Karlsson** and **Christer Larsson**. It’s twelve years since we all started out at the Linköping university, you all have a special place in my heart!

My best friends, **Ida** and **Per Thunberg**, having you in my life is a privilege, I love you!

My godson **Elias Thunberg**, you give my life new perspectives, and you are so precious to me!

My dear parents-in-law, **Catarina** and **Börje Persson**, for always being encouraging and supporting in every way possible!

My sister-in-law, **Annika Ekdahl** and her husband **Lars Ekdahl** for all fun moments, especially the frustrating hunts for the easter eggs...

My grand-parents, **Allan** and **Greta Torstensson**. *Mormor och Morfar, tack för all uppmuntran och alla värmande ord, ni är fantastiska!*

My dear brother **Mattias Svensson**, you are the best little brother one can imagine, and **Kristina Nyström** for all the fun weekends!

My wonderful parents, **Lena** and **Bo Svensson**. *Kära Mamma och Pappa, jag har er båda att tacka för allt, utan ert engagemang och uppmuntran hade den här avhandlingen aldrig skrivits, ni har alltid trott på min förmåga, även när jag tivit som mest, jag älskar er!*

**Mattias**, my dear husband, you are my everything! Thank you for putting up with me during the last months, without you and your endless love I would not have made it all the way, I love you so much!

## 8 REFERENCES

1. Charcot, J., *Histologic de la sclerose en plaque*. Gaz Hôp, 1868. **41**: p. 554-566.
2. Hafler, D.A., et al., *Multiple sclerosis*. Immunol Rev, 2005. **204**: p. 208-31.
3. Compston, A. and A. Coles, *Multiple sclerosis*. Lancet, 2002. **359**(9313): p. 1221-31.
4. Noseworthy, J.H., et al., *Multiple sclerosis*. N Engl J Med, 2000. **343**(13): p. 938-52.
5. Koch, M., et al., *Progression in multiple sclerosis: further evidence of an age dependent process*. J Neurol Sci, 2007. **255**(1-2): p. 35-41.
6. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis*. Ann Neurol, 2001. **50**(1): p. 121-7.
7. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria"*. Ann Neurol, 2005. **58**(6): p. 840-6.
8. Freedman, M.S., et al., *Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement*. Arch Neurol, 2005. **62**(6): p. 865-70.
9. Amato, M., Portaccio Emilio, *Clinical outcome measures in multiple sclerosis*. J Neurol Sci, 2007(In press).
10. Kurtzke, J.F., *Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)*. Neurology, 1983. **33**(11): p. 1444-52.
11. Jacobs, L.D., et al., *Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG)*. Ann Neurol, 1996. **39**(3): p. 285-94.
12. *Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group*. Lancet, 1998. **352**(9139): p. 1498-504.
13. *Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group*. Neurology, 1995. **45**(7): p. 1277-85.
14. Billiau, A., B.C. Kieseier, and H.P. Hartung, *Biologic role of interferon beta in multiple sclerosis*. J Neurol, 2004. **251 Suppl 2**: p. II10-4.
15. Johnson, K.P., et al., *Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group*. Neurology, 1995. **45**(7): p. 1268-76.
16. Comi, G., M. Filippi, and J.S. Wolinsky, *European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer acetate on magnetic resonance imaging--measured disease activity and burden in patients with relapsing multiple sclerosis. European/Canadian Glatiramer Acetate Study Group*. Ann Neurol, 2001. **49**(3): p. 290-7.
17. Ziemssen, T. and W. Schrempf, *Glatiramer acetate: mechanisms of action in multiple sclerosis*. Int Rev Neurobiol, 2007. **79**: p. 537-70.
18. Miller, D.H., et al., *A controlled trial of natalizumab for relapsing multiple sclerosis*. N Engl J Med, 2003. **348**(1): p. 15-23.
19. Miller, D.H., et al., *MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS*. Neurology, 2007. **68**(17): p. 1390-401.
20. Keegan, B.M. and J.H. Noseworthy, *Multiple sclerosis*. Annu Rev Med, 2002. **53**: p. 285-302.
21. Paterson, P.Y., *Transfer of allergic encephalomyelitis in rats by means of lymph node cells*. J Exp Med, 1960. **111**: p. 119-36.
22. Hillert, J., *Human leukocyte antigen studies in multiple sclerosis*. Ann Neurol, 1994. **36 Suppl**: p. S15-7.
23. Libbey, J.E., L.L. McCoy, and R.S. Fujinami, *Molecular mimicry in multiple sclerosis*. Int Rev Neurobiol, 2007. **79**: p. 127-47.

24. McQualter, J.L. and C.C. Bernard, *Multiple sclerosis: a battle between destruction and repair*. J Neurochem, 2007. **100**(2): p. 295-306.
25. Engelhardt, B., *Regulation of immune cell entry into the central nervous system*. Results Probl Cell Differ, 2006. **43**: p. 259-80.
26. Klawiter, E.C. and A.H. Cross, *B cells: no longer the nondominant arm of multiple sclerosis*. Curr Neurol Neurosci Rep, 2007. **7**(3): p. 231-8.
27. Neumann, H., *Molecular mechanisms of axonal damage in inflammatory central nervous system diseases*. Curr Opin Neurol, 2003. **16**(3): p. 267-73.
28. Bitsch, A., et al., *Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation*. Brain, 2000. **123** ( Pt 6): p. 1174-83.
29. Barnett, M.H. and J.W. Prineas, *Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion*. Ann Neurol, 2004. **55**(4): p. 458-68.
30. DeLuca, G.C., et al., *The contribution of demyelination to axonal loss in multiple sclerosis*. Brain, 2006. **129**(Pt 6): p. 1507-16.
31. Bo, L., et al., *Grey matter pathology in multiple sclerosis*. Acta Neurol Scand Suppl, 2006. **183**: p. 48-50.
32. Smith, E.J., W.F. Blakemore, and W.I. McDonald, *Central remyelination restores secure conduction*. Nature, 1979. **280**(5721): p. 395-6.
33. Jeffery, N.D. and W.F. Blakemore, *Locomotor deficits induced by experimental spinal cord demyelination are abolished by spontaneous remyelination*. Brain, 1997. **120** ( Pt 1): p. 27-37.
34. Lucchinetti, C., et al., *Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination*. Ann Neurol, 2000. **47**(6): p. 707-17.
35. Lucchinetti, C., et al., *A quantitative analysis of oligodendrocytes in multiple sclerosis lesions. A study of 113 cases*. Brain, 1999. **122** ( Pt 12): p. 2279-95.
36. Patrikios, P., et al., *Remyelination is extensive in a subset of multiple sclerosis patients*. Brain, 2006. **129**(Pt 12): p. 3165-72.
37. Barkhof, F., et al., *Remyelinated lesions in multiple sclerosis: magnetic resonance image appearance*. Arch Neurol, 2003. **60**(8): p. 1073-81.
38. Franklin, R.J., *Why does remyelination fail in multiple sclerosis?* Nat Rev Neurosci, 2002. **3**(9): p. 705-14.
39. Chari, D.M., *Remyelination in multiple sclerosis*. Int Rev Neurobiol, 2007. **79**: p. 589-620.
40. Waxman, S.G., *Axonal conduction and injury in multiple sclerosis: the role of sodium channels*. Nat Rev Neurosci, 2006. **7**(12): p. 932-41.
41. Lassmann, H., W. Bruck, and C. Lucchinetti, *Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy*. Trends Mol Med, 2001. **7**(3): p. 115-21.
42. Kurtzke, J.F., *Epidemiologic evidence for multiple sclerosis as an infection*. Clin Microbiol Rev, 1993. **6**(4): p. 382-427.
43. Pugliatti, M., et al., *The epidemiology of multiple sclerosis in Europe*. Eur J Neurol, 2006. **13**(7): p. 700-22.
44. Kurtzke, J.F., K.R. Gudmundsson, and S. Bergmann, *Multiple sclerosis in Iceland: I. Evidence of a postwar epidemic*. Neurology, 1982. **32**(2): p. 143-50.
45. Cook, S.D., et al., *Multiple sclerosis and distemper in Iceland 1966-1978*. Acta Neurol Scand, 1980. **61**(4): p. 244-51.
46. Cook, S.D., et al., *Declining incidence of multiple sclerosis in the Orkney Islands*. Neurology, 1985. **35**(4): p. 545-51.
47. Kurtzke, J.F. and K. Hyllested, *Multiple sclerosis: an epidemic disease in the Faeroes*. Trans Am Neurol Assoc, 1975. **100**: p. 213-5.
48. Poser, C.M., et al., *Analysis of the 'epidemic' of multiple sclerosis in the Faroe Islands. I. Clinical and epidemiological aspects*. Neuroepidemiology, 1988. **7**(4): p. 168-80.
49. Poser, C.M. and P.L. Hibberd, *Analysis of the 'epidemic' of multiple sclerosis in the Faroe Islands. II. Biostatistical aspects*. Neuroepidemiology, 1988. **7**(4): p. 181-9.
50. Gale, C.R. and C.N. Martyn, *Migrant studies in multiple sclerosis*. Prog Neurobiol, 1995. **47**(4-5): p. 425-48.
51. Marrie, R.A., *Environmental risk factors in multiple sclerosis aetiology*. Lancet Neurol, 2004. **3**(12): p. 709-18.

52. Compston, A., *Genetic epidemiology of multiple sclerosis*. J Neurol Neurosurg Psychiatry, 1997. **62**(6): p. 553-61.
53. Fleming, J.O. and T.D. Cook, *Multiple sclerosis and the hygiene hypothesis*. Neurology, 2006. **67**(11): p. 2085-6.
54. Challoner, P.B., et al., *Plaque-associated expression of human herpesvirus 6 in multiple sclerosis*. Proc Natl Acad Sci U S A, 1995. **92**(16): p. 7440-4.
55. Sola, P., et al., *Human herpesvirus 6 and multiple sclerosis: survey of anti-HHV-6 antibodies by immunofluorescence analysis and of viral sequences by polymerase chain reaction*. J Neurol Neurosurg Psychiatry, 1993. **56**(8): p. 917-9.
56. Sundstrom, P., et al., *An altered immune response to Epstein-Barr virus in multiple sclerosis: a prospective study*. Neurology, 2004. **62**(12): p. 2277-82.
57. Ascherio, A. and M. Munch, *Epstein-Barr virus and multiple sclerosis*. Epidemiology, 2000. **11**(2): p. 220-4.
58. Munger, K.L., et al., *Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis*. Jama, 2006. **296**(23): p. 2832-8.
59. Giovannoni, G. and G. Ebers, *Multiple sclerosis: the environment and causation*. Curr Opin Neurol, 2007. **20**(3): p. 261-8.
60. Orton, S.M., et al., *Sex ratio of multiple sclerosis in Canada: a longitudinal study*. Lancet Neurol, 2006. **5**(11): p. 932-6.
61. Sadovnick, A.D., P.A. Baird, and R.H. Ward, *Multiple sclerosis: updated risks for relatives*. Am J Med Genet, 1988. **29**(3): p. 533-41.
62. Sadovnick, A.D., et al., *A population-based study of multiple sclerosis in twins: update*. Ann Neurol, 1993. **33**(3): p. 281-5.
63. Mumford, C.J., et al., *The British Isles survey of multiple sclerosis in twins*. Neurology, 1994. **44**(1): p. 11-5.
64. Hansen, T., et al., *Concordance for multiple sclerosis in Danish twins: an update of a nationwide study*. Mult Scler, 2005. **11**(5): p. 504-10.
65. Ebers, G.C., et al., *A population-based study of multiple sclerosis in twins*. N Engl J Med, 1986. **315**(26): p. 1638-42.
66. Willer, C.J., et al., *Twin concordance and sibling recurrence rates in multiple sclerosis*. Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12877-82.
67. Ebers, G.C., A.D. Sadovnick, and N.J. Risch, *A genetic basis for familial aggregation in multiple sclerosis*. Canadian Collaborative Study Group. Nature, 1995. **377**(6545): p. 150-1.
68. Robertson, N.P., et al., *Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis*. Lancet, 1997. **349**(9065): p. 1587-90.
69. Ebers, G.C., et al., *Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring*. Canadian Collaborative Study Group. Ann Neurol, 2000. **48**(6): p. 927-31.
70. Antonarakis, S.E. and J.S. Beckmann, *Mendelian disorders deserve more attention*. Nat Rev Genet, 2006. **7**(4): p. 277-82.
71. Lander, E.S., *The new genomics: global views of biology*. Science, 1996. **274**(5287): p. 536-9.
72. *A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes*. The Huntington's Disease Collaborative Research Group. Cell, 1993. **72**(6): p. 971-83.
73. Ellegren, H., *Microsatellite mutations in the germline: implications for evolutionary inference*. Trends Genet, 2000. **16**(12): p. 551-8.
74. Ellegren, H., *Microsatellites: simple sequences with complex evolution*. Nat Rev Genet, 2004. **5**(6): p. 435-45.
75. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome*. Nat Rev Genet, 2006. **7**(2): p. 85-97.
76. Holmans, P., *Affected sib-pair methods for detecting linkage to dichotomous traits: review of the methodology*. Hum Biol, 1998. **70**(6): p. 1025-40.
77. Altmuller, J., et al., *Genomewide scans of complex human diseases: true linkage is hard to find*. Am J Hum Genet, 2001. **69**(5): p. 936-50.
78. Hirschhorn, J.N. and M.J. Daly, *Genome-wide association studies for common diseases and complex traits*. Nat Rev Genet, 2005. **6**(2): p. 95-108.

79. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases*. Science, 1996. **273**(5281): p. 1516-7.
80. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)*. Am J Hum Genet, 1993. **52**(3): p. 506-16.
81. Hugot, J.P., et al., *Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 599-603.
82. Ogura, Y., et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*. Nature, 2001. **411**(6837): p. 603-6.
83. Hampe, J., et al., *Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations*. Lancet, 2001. **357**(9272): p. 1925-8.
84. Sigurdsson, S., et al., *Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus*. Am J Hum Genet, 2005. **76**(3): p. 528-37.
85. Bottini, N., et al., *A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes*. Nat Genet, 2004. **36**(4): p. 337-8.
86. Begovich, A.B., et al., *A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis*. Am J Hum Genet, 2004. **75**(2): p. 330-7.
87. *The International HapMap Project*. Nature, 2003. **426**(6968): p. 789-96.
88. Barrett, J.C. and L.R. Cardon, *Evaluating coverage of genome-wide association studies*. Nat Genet, 2006. **38**(6): p. 659-62.
89. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls*. Nature, 2007. **447**(7145): p. 661-78.
90. Todd, J.A., et al., *Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes*. Nat Genet, 2007. **39**(7): p. 857-64.
91. Parkes, M., et al., *Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility*. Nat Genet, 2007. **39**(7): p. 830-2.
92. Varilo, T. and L. Peltonen, *Isolates and their potential use in complex gene mapping efforts*. Curr Opin Genet Dev, 2004. **14**(3): p. 316-23.
93. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping*. Genomics, 1995. **29**(2): p. 311-22.
94. Lewontin, R.C., *The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models*. Genetics, 1964. **49**(1): p. 49-67.
95. de Bakker, P.I., et al., *Efficiency and power in genetic association studies*. Nat Genet, 2005. **37**(11): p. 1217-23.
96. He, J. and A. Zelikovsky, *MLR-tagging: informative SNP selection for unphased genotypes based on multiple linear regression*. Bioinformatics, 2006. **22**(20): p. 2558-61.
97. Stram, D.O., *Software for tag single nucleotide polymorphism selection*. Hum Genomics, 2005. **2**(2): p. 144-51.
98. Conrad, D.F., et al., *A worldwide survey of haplotype variation and linkage disequilibrium in the human genome*. Nat Genet, 2006. **38**(11): p. 1251-60.
99. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome*. Science, 2002. **296**(5576): p. 2225-9.
100. Excoffier, L. and M. Slatkin, *Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population*. Mol Biol Evol, 1995. **12**(5): p. 921-7.
101. Stephens, M. and P. Donnelly, *A comparison of bayesian methods for haplotype reconstruction from population genotype data*. Am J Hum Genet, 2003. **73**(5): p. 1162-9.
102. Morris, A., A. Pedder, and K. Ayres, *Linkage disequilibrium assessment via log-linear modeling of SNP haplotype frequencies*. Genet Epidemiol, 2003. **25**(2): p. 106-14.
103. Zondervan, K.T. and L.R. Cardon, *The complex interplay among factors that influence allelic association*. Nat Rev Genet, 2004. **5**(2): p. 89-100.
104. Freedman, M.L., et al., *Assessing the impact of population stratification on genetic association studies*. Nat Genet, 2004. **36**(4): p. 388-93.

105. Devlin, B., K. Roeder, and L. Wasserman, *Genomic control, a new approach to genetic-based association studies*. *Theor Popul Biol*, 2001. **60**(3): p. 155-66.
106. Dahlman, I., et al., *Parameters for reliable results in genetic association studies in common disease*. *Nat Genet*, 2002. **30**(2): p. 149-50.
107. Jersild, C., A. Svejgaard, and T. Fog, *HLA-A antigens and multiple sclerosis*. *Lancet*, 1972. **1**(7762): p. 1240-1.
108. Jersild, C., et al., *Histocompatibility-linked immune-response determinants in multiple sclerosis*. *Transplant Proc*, 1973. **5**(4): p. 1791-6.
109. Hillert, J. and O. Olerup, *Multiple sclerosis is associated with genes within or close to the HLA-DR-DQ subregion on a normal DR15,DQ6,Dw2 haplotype*. *Neurology*, 1993. **43**(1): p. 163-8.
110. Marrosu, M.G., et al., *Multiple sclerosis in Sardinia is associated and in linkage disequilibrium with HLA-DR3 and -DR4 alleles*. *Am J Hum Genet*, 1997. **61**(2): p. 454-7.
111. Marrosu, M.G., et al., *DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population*. *Hum Mol Genet*, 1998. **7**(8): p. 1235-7.
112. Marrosu, M.G., et al., *Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia*. *Hum Mol Genet*, 2001. **10**(25): p. 2907-16.
113. Dymont, D.A., et al., *Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance*. *Hum Mol Genet*, 2005. **14**(14): p. 2019-26.
114. Barcellos, L.F., et al., *HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course*. *Am J Hum Genet*, 2003. **72**(3): p. 710-6.
115. Modin, H., et al., *Modes of action of HLA-DR susceptibility specificities in multiple sclerosis*. *Am J Hum Genet*, 2004. **74**(6): p. 1321-2.
116. Celius, E.G., et al., *Sex and age at diagnosis are correlated with the HLA-DR2, DQ6 haplotype in multiple sclerosis*. *J Neurol Sci*, 2000. **178**(2): p. 132-5.
117. Hensiek, A.E., et al., *HLA-DR 15 is associated with female sex and younger age at diagnosis in multiple sclerosis*. *J Neurol Neurosurg Psychiatry*, 2002. **72**(2): p. 184-7.
118. Masterman, T., et al., *HLA-DR15 is associated with lower age at onset in multiple sclerosis*. *Ann Neurol*, 2000. **48**(2): p. 211-9.
119. Fogdell-Hahn, A., et al., *Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease*. *Tissue Antigens*, 2000. **55**(2): p. 140-8.
120. Harbo, H.F., et al., *Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis*. *Tissue Antigens*, 2004. **63**(3): p. 237-47.
121. Brynedal, B., et al., *HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis*. *PLoS ONE*, 2007. **2**: p. e664.
122. Ebers, G.C., et al., *A full genome search in multiple sclerosis*. *Nat Genet*, 1996. **13**(4): p. 472-6.
123. Haines, J.L., et al., *A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex*. *The Multiple Sclerosis Genetics Group*. *Nat Genet*, 1996. **13**(4): p. 469-71.
124. Sawcer, S., et al., *A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22*. *Nat Genet*, 1996. **13**(4): p. 464-8.
125. Kuokkanen, S., et al., *Genomewide scan of multiple sclerosis in Finnish multiplex families*. *Am J Hum Genet*, 1997. **61**(6): p. 1379-87.
126. Broadley, S., et al., *A genome screen for multiple sclerosis in Italian families*. *Genes Immun*, 2001. **2**(4): p. 205-10.
127. Coraddu, F., et al., *A genome screen for multiple sclerosis in Sardinian multiplex families*. *Eur J Hum Genet*, 2001. **9**(8): p. 621-6.
128. Akesson, E., et al., *A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis*. *Genes Immun*, 2002. **3**(5): p. 279-85.
129. Ban, M., et al., *A genome screen for linkage in Australian sibling-pairs with multiple sclerosis*. *Genes Immun*, 2002. **3**(8): p. 464-9.

130. Eraksoy, M., et al., *A whole genome screen for linkage in Turkish multiple sclerosis*. J Neuroimmunol, 2003. **143**(1-2): p. 17-24.
131. Hensiek, A.E., et al., *Updated results of the United Kingdom linkage-based genome screen in multiple sclerosis*. J Neuroimmunol, 2003. **143**(1-2): p. 25-30.
132. Kenealy, S.J., et al., *A second-generation genomic screen for multiple sclerosis*. Am J Hum Genet, 2004. **75**(6): p. 1070-8.
133. *A meta-analysis of whole genome linkage screens in multiple sclerosis*. J Neuroimmunol, 2003. **143**(1-2): p. 39-46.
134. Sawcer, S., et al., *A high-density screen for linkage in multiple sclerosis*. Am J Hum Genet, 2005. **77**(3): p. 454-67.
135. A, D.D., et al., *A genome-scan in a single pedigree with a high prevalence of multiple sclerosis*. J Neurol Neurosurg Psychiatry, 2007.
136. Modin, H., et al., *Genome-wide linkage screen of a consanguineous multiple sclerosis kinship*. Mult Scler, 2003. **9**(2): p. 128-34.
137. Vitale, E., et al., *Linkage analysis conditional on HLA status in a large North American pedigree supports the presence of a multiple sclerosis susceptibility locus on chromosome 12p12*. Hum Mol Genet, 2002. **11**(3): p. 295-300.
138. Hauser, S.L. and J.R. Oksenberg, *The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration*. Neuron, 2006. **52**(1): p. 61-76.
139. Dyment, D.A., G.C. Ebers, and A.D. Sadovnick, *Genetics of multiple sclerosis*. Lancet Neurol, 2004. **3**(2): p. 104-10.
140. Sawcer, S., *A new era in the genetic analysis of multiple sclerosis*. Curr Opin Neurol, 2006. **19**(3): p. 237-41.
141. Oksenberg, J.R. and L.F. Barcellos, *Multiple sclerosis genetics: leaving no stone unturned*. Genes Immun, 2005. **6**(5): p. 375-87.
142. Siest, G., et al., *Apolipoprotein E: an important gene and protein to follow in laboratory medicine*. Clin Chem, 1995. **41**(8 Pt 1): p. 1068-86.
143. Pinholt, M., J.L. Frederiksen, and M. Christiansen, *The association between apolipoprotein E and multiple sclerosis*. Eur J Neurol, 2006. **13**(6): p. 573-80.
144. Barton, A., et al., *Association of protein kinase C alpha (PRKCA) gene with multiple sclerosis in a UK population*. Brain, 2004. **127**(Pt 8): p. 1717-22.
145. Saarela, J., et al., *PRKCA and multiple sclerosis: association in two independent populations*. PLoS Genet, 2006. **2**(3): p. e42.
146. Baier, G., *The PKC gene module: molecular biosystematics to resolve its T cell functions*. Immunol Rev, 2003. **192**: p. 64-79.
147. Swanberg, M., et al., *MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction*. Nat Genet, 2005. **37**(5): p. 486-94.
148. Akkad, D.A., et al., *Promoter polymorphism rs3087456 in the MHC class II transactivator gene is not associated with susceptibility for selected autoimmune diseases in German patient groups*. Int J Immunogenet, 2006. **33**(1): p. 59-61.
149. O'Doherty, C., et al., *The MHC2TA -168A/G and +1614G/C polymorphisms and risk for multiple sclerosis or chronic inflammatory arthropathies* Tissue Antigens, 2007. doi:10.1111/j.1399-0039.2007.00876.x.
150. Martinez, A., et al., *Role of the MHC2TA gene in autoimmune diseases*. Ann Rheum Dis, 2007. **66**(3): p. 325-9.
151. Nagarajan, U.M., A. Bushey, and J.M. Boss, *Modulation of gene expression by the MHC class II transactivator*. J Immunol, 2002. **169**(9): p. 5078-88.
152. *Risk Alleles for Multiple Sclerosis Identified by a Genomewide Study*. N Engl J Med, 2007.
153. Gregory, S.G., et al., *Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis*. Nat Genet, 2007.
154. Poser, C.M., et al., *New diagnostic criteria for multiple sclerosis: guidelines for research protocols*. Ann Neurol, 1983. **13**(3): p. 227-31.
155. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells*. Nucleic Acids Res, 1988. **16**(3): p. 1215.

156. Chataway, J., et al., *A screen of candidates from peaks of linkage: evidence for the involvement of myeloperoxidase in multiple sclerosis*. J Neuroimmunol, 1999. **98**(2): p. 208-13.
157. Nagra, R.M., et al., *Immunohistochemical and genetic evidence of myeloperoxidase involvement in multiple sclerosis*. J Neuroimmunol, 1997. **78**(1-2): p. 97-107.
158. Nelissen, I., et al., *PECAMI, MPO and PRKARIA at chromosome 17q21-q24 and susceptibility for multiple sclerosis in Sweden and Sardinia*. J Neuroimmunol, 2000. **108**(1-2): p. 153-9.
159. Zakrzewska-Pniewska, B., et al., *Association of apolipoprotein E and myeloperoxidase genotypes to clinical course of familial and sporadic multiple sclerosis*. Mult Scler, 2004. **10**(3): p. 266-71.
160. Nordfors, L., et al., *Large-scale genotyping of single nucleotide polymorphisms by Pyrosequencingtrade mark and validation against the 5'nuclease (Taqman((R))) assay*. Hum Mutat, 2002. **19**(4): p. 395-401.
161. Woolf, B., *On estimating the relation between blood group and disease*. Ann Hum Genet, 1955. **19**(4): p. 251-3.
162. Zhao, J.H., D. Curtis, and P.C. Sham, *Model-free analysis and permutation tests for allelic associations*. Hum Hered, 2000. **50**(2): p. 133-9.
163. Sham, P.C. and D. Curtis, *Monte Carlo tests for associations between disease and alleles at highly polymorphic loci*. Ann Hum Genet, 1995. **59**(Pt 1): p. 97-105.
164. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps*. Bioinformatics, 2005. **21**(2): p. 263-5.
165. Stephens, M., N.J. Smith, and P. Donnelly, *A new statistical method for haplotype reconstruction from population data*. Am J Hum Genet, 2001. **68**(4): p. 978-89.
166. Becker, K.G., et al., *Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases*. Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9979-84.
167. Oksenberg, J.R., et al., *Multiple sclerosis: genomic rewards*. J Neuroimmunol, 2001. **113**(2): p. 171-84.
168. Feakes, R., et al., *Interleukin 1 receptor antagonist (IL-1ra) in multiple sclerosis*. J Neuroimmunol, 2000. **105**(1): p. 96-101.
169. Giedraitis, V., et al., *Cloning and mutation analysis of the human IL-18 promoter: a possible role of polymorphisms in expression regulation*. J Neuroimmunol, 2001. **112**(1-2): p. 146-52.
170. Schrijver, H.M., et al., *Association of interleukin-1beta and interleukin-1 receptor antagonist genes with disease severity in MS*. Neurology, 1999. **52**(3): p. 595-9.
171. Kotze, M.J., et al., *Analysis of the NRAMP1 gene implicated in iron transport: association with multiple sclerosis and age effects*. Blood Cells Mol Dis, 2001. **27**(1): p. 44-53.
172. Hensiek, A.E., et al., *Osteopontin gene and clinical severity of multiple sclerosis*. J Neurol, 2003. **250**(8): p. 943-7.
173. Bruniquel, D., et al., *Regulation of expression of the human lymphocyte activation gene-3 (LAG-3) molecule, a ligand for MHC class II*. Immunogenetics, 1998. **48**(2): p. 116-24.
174. Andrae, S., et al., *Maturation and activation of dendritic cells induced by lymphocyte activation gene-3 (CD223)*. J Immunol, 2002. **168**(8): p. 3874-80.
175. Fry, T.J. and C.L. Mackall, *Interleukin-7: from bench to clinic*. Blood, 2002. **99**(11): p. 3892-904.
176. Lai, S.Y., J. Molden, and M.A. Goldsmith, *Shared gamma(c) subunit within the human interleukin-7 receptor complex. A molecular basis for the pathogenesis of X-linked severe combined immunodeficiency*. J Clin Invest, 1997. **99**(2): p. 169-77.
177. He, Y.W. and T.R. Malek, *Interleukin-7 receptor alpha is essential for the development of gamma delta + T cells, but not natural killer cells*. J Exp Med, 1996. **184**(1): p. 289-93.

178. Pandey, A., et al., *Cloning of a receptor subunit required for signaling by thymic stromal lymphopoietin*. *Nat Immunol*, 2000. **1**(1): p. 59-64.
179. Park, L.S., et al., *Cloning of the murine thymic stromal lymphopoietin (TSLP) receptor: Formation of a functional heteromeric complex requires interleukin 7 receptor*. *J Exp Med*, 2000. **192**(5): p. 659-70.
180. Ziegler, S.F. and Y.J. Liu, *Thymic stromal lymphopoietin in normal and pathogenic T cell development and function*. *Nat Immunol*, 2006. **7**(7): p. 709-14.
181. Rimoldi, M., et al., *Intestinal immune homeostasis is regulated by the crosstalk between epithelial cells and dendritic cells*. *Nat Immunol*, 2005. **6**(5): p. 507-14.
182. Ye, S.K., et al., *Induction of germline transcription in the TCRgamma locus by Stat5: implications for accessibility control by the IL-7 receptor*. *Immunity*, 1999. **11**(2): p. 213-23.
183. Liu, W., et al., *CD127 expression inversely correlates with FoxP3 and suppressive function of human CD4+ T reg cells*. *J Exp Med*, 2006. **203**(7): p. 1701-11.
184. Trikalinos, T.A., et al., *Establishment of genetic associations for complex diseases is independent of early study findings*. *Eur J Hum Genet*, 2004. **12**(9): p. 762-9.
185. Xu, C., et al., *Linkage analysis in multiple sclerosis of chromosomal regions syntenic to experimental autoimmune disease loci*. *Eur J Hum Genet*, 2001. **9**(6): p. 458-63.
186. Krummel, M.F. and M.M. Davis, *Dynamics of the immunological synapse: finding, establishing and solidifying a connection*. *Curr Opin Immunol*, 2002. **14**(1): p. 66-74.
187. Doyle, C. and J.L. Strominger, *Interaction between CD4 and class II MHC molecules mediates cell adhesion*. *Nature*, 1987. **330**(6145): p. 256-9.
188. Wang, J.H., et al., *Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule*. *Proc Natl Acad Sci U S A*, 2001. **98**(19): p. 10799-804.
189. Feito, M.J., et al., *CD4 dependence of activation threshold and TCR signalling in mouse T lymphocytes*. *Scand J Immunol*, 1997. **45**(2): p. 166-74.
190. Triebel, F., et al., *LAG-3, a novel lymphocyte activation gene closely related to CD4*. *J Exp Med*, 1990. **171**(5): p. 1393-405.
191. Baixeras, E., et al., *Characterization of the lymphocyte activation gene 3-encoded protein. A new ligand for human leukocyte antigen class II antigens*. *J Exp Med*, 1992. **176**(2): p. 327-37.
192. Huard, B., et al., *T cell major histocompatibility complex class II molecules down-regulate CD4+ T cell clone responses following LAG-3 binding*. *Eur J Immunol*, 1996. **26**(5): p. 1180-6.
193. Avicé, M.N., et al., *Lymphocyte activation gene-3, a MHC class II ligand expressed on activated T cells, stimulates TNF-alpha and IL-12 production by monocytes and dendritic cells*. *J Immunol*, 1999. **162**(5): p. 2748-53.
194. Kantarci, O.H., et al., *Association of a myeloperoxidase promoter polymorphism with multiple sclerosis*. *J Neuroimmunol*, 2000. **105**(2): p. 189-94.
195. Piedrafita, F.J., et al., *An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element*. *J Biol Chem*, 1996. **271**(24): p. 14412-20.
196. Hoy, A., et al., *Growing significance of myeloperoxidase in non-infectious diseases*. *Clin Chem Lab Med*, 2002. **40**(1): p. 2-8.
197. Roxburgh, R.H., et al., *Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity*. *Neurology*, 2005. **64**(7): p. 1144-51.
198. Booth, D.R., et al., *Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis*. *J Mol Med*, 2005. **83**(10): p. 822-30.
199. Teutsch, S.M., et al., *Identification of 11 novel and common single nucleotide polymorphisms in the interleukin-7 receptor-alpha gene and their associations with multiple sclerosis*. *Eur J Hum Genet*, 2003. **11**(7): p. 509-15.

200. Ziegler, S.E., et al., *Reconstitution of a functional interleukin (IL)-7 receptor demonstrates that the IL-2 receptor gamma chain is required for IL-7 signal transduction.* Eur J Immunol, 1995. **25**(2): p. 399-404.
201. Jiang, Q., et al., *Cell biology of IL-7, a key lymphotrophin.* Cytokine Growth Factor Rev, 2005. **16**(4-5): p. 513-33.
202. Wiles, M.V., P. Ruiz, and B.A. Imhof, *Interleukin-7 expression during mouse thymus development.* Eur J Immunol, 1992. **22**(4): p. 1037-42.
203. Mazzucchelli, R. and S.K. Durum, *Interleukin-7 receptor expression: intelligent design.* Nat Rev Immunol, 2007. **7**(2): p. 144-54.
204. Lundmark, F., et al., *Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis.* Nat Genet, 2007.
205. Harley, J.B., *IL-7Ralpha and multiple sclerosis risk.* Nat Genet, 2007. **39**(9): p. 1053-1054.
206. Sasson, S.C., J.J. Zaunders, and A.D. Kelleher, *The IL-7/IL-7 receptor axis: understanding its central role in T-cell homeostasis and the challenges facing its utilization as a novel therapy.* Curr Drug Targets, 2006. **7**(12): p. 1571-82.
207. Khaled, A.R. and S.K. Durum, *Lymphocyte: cytokines and the control of lymphoid homeostasis.* Nat Rev Immunol, 2002. **2**(11): p. 817-30.
208. Puel, A., et al., *Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency.* Nat Genet, 1998. **20**(4): p. 394-7.
209. Roifman, C.M., et al., *A partial deficiency of interleukin-7R alpha is sufficient to abrogate T-cell development and cause severe combined immunodeficiency.* Blood, 2000. **96**(8): p. 2803-7.
210. Viglietta, V., et al., *Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis.* J Exp Med, 2004. **199**(7): p. 971-9.
211. Haas, J., et al., *Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis.* Eur J Immunol, 2005. **35**(11): p. 3343-52.
212. Putheti, P., et al., *Circulating CD4+CD25+ T regulatory cells are not altered in multiple sclerosis and unaffected by disease-modulating drugs.* J Clin Immunol, 2004. **24**(2): p. 155-61.
213. Venken, K., et al., *Secondary progressive in contrast to relapsing-remitting multiple sclerosis patients show a normal CD4+CD25+ regulatory T-cell function and FOXP3 expression.* J Neurosci Res, 2006. **83**(8): p. 1432-46.
214. Feger, U., et al., *Increased frequency of CD4+ CD25+ regulatory T cells in the cerebrospinal fluid but not in the blood of multiple sclerosis patients.* Clin Exp Immunol, 2007. **147**(3): p. 412-8.
215. Altshuler, D. and M. Daly, *Guilt beyond a reasonable doubt.* Nat Genet, 2007. **39**(7): p. 813-5.
216. Olson, C.M., et al., *Publication bias in editorial decision making.* Jama, 2002. **287**(21): p. 2825-8.
217. Clark, A.G., *The role of haplotypes in candidate gene studies.* Genet Epidemiol, 2004. **27**(4): p. 321-33.