Immunogenetic studies in multiple sclerosis

Doctoral thesis by

Åslaug Rudjord Lorentzen

Department of Neurology, Faculty Division Ullevål University Hospital

University of Oslo

and

Institute of Immunology, Rikshospitalet, Oslo University Hospital

Oslo, June 2009

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 835

ISBN 978-82-8072-340-6

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2009.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

Table of Contents

1 Acknowledgement 5

2 Abbreviations 8

3 List of publications included 9

4 Introduction 11

4.1 Multiple sclerosis 11

4.1.1 Clinical aspects of MS 11

4.1.2 MS as an immune-mediated disease 14

4.1.3 Pathoimmunological aspects of MS 16

4.1.4 Genetic epidemiology and environmental factors in MS 17

4.2 Genetic concepts and strategies in studies of complex diseases 18

4.2.1 Basic genetic concepts 18

4.2.2 Strategies to identify disease associated or linked genes 19

4.2.3 Disease models for a common complex disease 21

4.2.4 GWAS - the recent approach in complex genetic studies 22

4.3 Genetic aspects of multiple sclerosis 24

4.3.1 The HLA complex - the early success 24

4.3.2 Search for other MS susceptibility loci - a challenge 25

4.3.3 New optimism in MS genetics - the GWAS success 28

5 Aim of the study 30

6 Summary of the results 30

Paper I 30

Paper II 31

Paper III 31

Paper IV 32

Paper V 33

7 Methodological considerations 34
7.1 The sample sets 34
7.2 Phenotype 35
7.3 Marker selection 38
7.4 Typing methods and accuracy 39
7.4.1 Microsatellite detection 40
7.4.2 SNP detection 40
7.4.3 KIR typing 41
7.4.4 HLA typing 41
7.5 Statistical considerations 42
7.5.1 Quality measurements 42
7.5.2 Effect size and statistical power 44
7.5.3 Increasing sample sets and meta-analyses 46
7.5.4 Significance level and multiple testing 47
7.5.5 The need for replication of true associations 48
8 General discussion 49
8.1 The difficulty of replicating claimed associations 49
8.2 Association to the HLA complex - not only the *HLA-DRB1* gene 54
8.3 The killer immunoglobulin-like receptor (KIR) genes in MS 58
8.4 GWASs and replicated findings 59
8.5 MS genes and the autoimmune model 64
8.6 Are there MS associated genes in non-immunological pathways? 67
9 The next generation of genetic studies in MS 70
10 References 73

1 Acknowledgement

This thesis is based on work performed at the Department of Neurology, Faculty division Ullevål University Hospital, University of Oslo (UiO), Norway and Institute of Immunology (IMMI), Rikshospitalet, Oslo University Hospital, Oslo, Norway during the period 2002-2009. I followed the Medical Student Research Program at UiO in the period 2002-2005, and from 2006 the Doctoral Degree Program at UiO. I was also given the opportunity for a research stay at the University of Cambridge, Department of Clinical Neurosciences, Addenbrooke's Hospital, Cambridge, UK, during winter and spring of 2007.

As a PhD student I received three years funding from the Research Council of Norway (166005/V50), as well as a scholarship for my research stay abroad, for which I am most grateful. I am also thankful for financial support from the Odd Fellow MS Society, Halvor Høies Fund and funds through the Multiple Sclerosis Society of Norway. Through the Oslo MS Genetics Group I have also received support from the South and Eastern Norway Regional Health Authority, the Scientific Advisory Council at Ullevål, Oslo University Hospital and the Norwegian Foundation for Health and Rehabilitation.

I want to thank the University of Oslo and the University of Cambridge, that have generously hosted me. During these years I have met many people who have inspired and educated me, and I thank them for having brought me to where I am today. Also, this research could not have been realized without MS patients and healthy volunteers who have so kindly contributed with blood samples and clinical data.

Hanne F. Harbo has done a fantastic job as my main supervisor and is a dear mentor. In a stimulating and enthusiastic way you have pushed me to do my very best. Thank you for all these years with continuous support and understanding. I am also very grateful to my other supervisor Benedicte A. Lie. As the leader for the Immunogenetics of autoimmune diseases group at IMMI she has introduced me to the genetics of autoimmune diseases in general and has been an inspiring guide in the massive jungle of the HLA complex. The two of you have been a brilliant team and I am indebted to you both professionally and personally.

I would like to thank the head of the Department of Neurology at Ullevål, Oslo University Hospital, Elisabeth G. Celius and all my other colleagues at the department. Cathrine Smestad and Elisabeth G. Celius have through the Oslo MS Genetics Group provided extensive clinical characterisation of the Oslo MS patients, fundamental for our MS research. Professor Chantal Tallaksen and Professor Helge J. Nordal are thanked for being my contact persons from UiO at the Department of Neurology, Ullevål. I also want to specifically mention the support and guidance I have had from Professor Anne Spurkland and Professor Frode Vartdal through these years. As co-authors they have provided constructive comments not only concerning MS genetics, but also scientific issues in general.

I have always felt included in the inspiring environment during my years at IMMI. I want to thank the head of the institute, Professor John Torgils Vaage and the previous head of the institute, Professor emeritus Erik Thorsby. Professor Thorsby has also contributed to my papers and I have appreciated his many comments. Thanks to all IMMI colleagues for all help I have received in my projects, in particular, the members and previous members of the Immunogenetics of autoimmune diseases group, alphabetically; Silja S. Amundsen, Morten C. Eike, Siri T. Flåm, Linda Haugse, Johannes E. R. Hov, Tom H. Karlsen, Angelina H. Maniaol, Espen Melum, Inger-Lise Mero, Gry N. Nordang, Hege D. Sollid, Marte K. Viken, Kristine Wiencke, Bente Woldseth. Especially, I need to thank Tom H Karlsen for the stimulating collaboration in my KIR-HLA study. I want to thank the excellent technicians who over the years have been somewhat involved in this work, listed chronologically; Ingebjørg Knutsen, Wenche Scheel Hamang, Monica Hals, Tone Aarskaug, Bente Woldseth, Maja Krezwinska and Hanne S Sæther. Ke-Zheng Dai, Trygve Holmøy, Anne-Lise Hestvik and Gjertrud Skorstad are thanked for challenging scientific discussions with focus on MS immunology. In 2008 Inger-Lise Mero started as a new PhD student in our MS genetic group. I have really appreciated your company both scientifically and personally. It is great to see that some of my projects now are continued and extended. Thanks also for helpful and enthusiastic comments when reading my thesis.

I feel honoured also to have had the opportunity to work with Professor Alastair Compston and Dr Stephen Sawcer as a guest researcher in their laboratory in Cambridge in 2007. I was warmly welcomed and received outstanding support, also from Amie Baker, An Goris and Maria Ban, as well as the other colleagues in the laboratory. I

enjoyed the Friday mornings with the Journal club and the Tuesday meetings at the department, and thank you all for making my stay so inspiring. I also thank my neighbour Mari Williams and her family. From chats over tea we became very good friends, and you and your family contributed to make our stay wonderful.

As a member of networks of excellent scientists like the Nordic MS Genetics Network and the International Multiple Sclerosis Genetics Consortium (IMSGC), I have experienced the great value of team work. Especially, I would like to thank the members of the Nordic MS genetics network who have been co-authors on the papers in this thesis (Jan Hillert, Ingrid Kockum, Jenny Link, Kjell-Morten Myhr, Tomas Olsson, Annette B. Oturai, Janna Saarela, Per S. Sørensen, Eva Åkesson (now Lindstrøm)). The list of co-authors from IMSGC is long, so none are named and nor forgotten. It has been a great privilege to be a part of this distinguished group of world leading experts in the field, for which I am very thankful.

In addition to local and international networks I will particularly like to thank my co-authors Marita Olsson at Karolinska Intitute, Stockholm, Sweden, for invaluable statistical expertise and Per O. Ekstrøm at Radiumhospitalet, Oslo University Hospital for support regarding the denaturant capillary electrophoresis method as well as other co-authors abroad; V. Ling, J. Y. Sun and D. Senitzer.

Finally, I would like to thank my family and friends. During my medical school and the PhD period you have always generously supported and encouraged me. You have reminded me that life is so much more than Eppendorf tubes and polymerase chain reactions. Thank you for your love and understanding. And last, I would say to my dearest; Øyvind, Eirik and Anne, as said by the Norwegian poet Sigmund Skard: “I want just one life – the life I live with you”.

Åslaug Rudjord Lorentzen
Oslo, 2009

2 Abbreviations

AH	ancestral haplotype
AITD	autoimmune thyroid disease
APC	antigen-presenting cell
AS	ankylosing spondylitis
BBB	blood brain barrier
bp	base pair
CD	cluster of differentiation molecule (e.g. CD4, CD8, CD28, CD226)
CD/CV	Common disease common variant
CD/RV	Common disease rare variant
CI	confidence interval
CIS	clinical isolated syndrome
CNS	central nervous system
CNV	copy number variation
CSF	cerebrospinal fluid
CTLA4	cytotoxic T-lymphocyte associated protein 4
DCE	denaturant capillary electrophoresis
DNA	deoxyribonucleic acid
EAE	experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
EDSS	expanded disability status scale
GSR	genotyping success rate
GWAS	genome wide association study
HLA	human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
ICOS	inducible T-cell co-stimulator
IFN-β	Interferon beta
Ig	Immunoglobulin
IL	Interleukin
IMSGC	International Multiple Sclerosis Genetics Consortium
kb	kilo base pairs
KIR	killer immunoglobulin-like receptor
LD	linkage disequilibrium
MAF	minor allele frequency
MHC	major histocompatibility complex
MRI	magnetic resonance imaging
MS	multiple sclerosis
MSSS	multiple sclerosis severity score
NK	natural killer
ns-SNP	non-synonymous SNP
OR	odds ratio
PCR	polymerase chain reaction
PP-MS	primary progressive multiple sclerosis
RFLP	restriction fragment length polymorphism
RR-MS	relapsing remitting multiple sclerosis
SNP	single nucleotide polymorphism
SP-MS	secondary progressive multiple sclerosis
SSO	sequence-specific oligonucleotide
SSP	sequence-specific primers
T1D	type 1 diabetes
TCR	T-cell receptor
TSAd	T-cell specific adaptor protein
UK	United Kingdom
US	United States of America
WTCCC	Wellcome Trust Case Control Consortium

3 List of publications included

Paper I: Lorentzen AR, Celius EG, Ekstrøm PO, Wiencke K, Lie BA, Myhr KM, Ling V, Thorsby E, Vartdal F, Spurkland A, Harbo HF: *Lack of association with the CD28/CTLA4/ICOS gene region among Norwegian multiple sclerosis patients.* J Neuroimmunol. 2005;166(1-2):197-201.

Paper II: Lorentzen AR, Smestad C, Lie BA, Oturai AB, Akesson E, Saarela J, Myhr KM, Vartdal F, Celius EG, Sørensen PS, Hillert J, Spurkland A, Harbo HF: *The SH2D2A gene and susceptibility to multiple sclerosis.* J Neuroimmunol. 2008;197(2):152-158

Paper III: Lorentzen AR, Karlsen TK, Olsson M, Smestad C, Mero IL, Woldseth B, Sun JY, Senitzer D, Celius EG, Thorsby E, Spurkland A, Lie BA, Harbo HF: *Killer immunoglobulin-like receptor ligand HLA-Bw4 protects against multiple sclerosis.* Ann Neurol 2009, in press

Paper IV: Ban M, Goris A, Lorentzen AR, Baker A, Mihalova T, Ingram G, Booth DR, Heard RN, Stewart GJ, Bogaert E, Dubois B, Harbo HF, Celius EG, Spurkland A, Strange R, Hawkins C, Robertson NP, Dudbridge F, Wason J, De Jager PL, Hafler D, Rioux JD, Ivinson AJ, McCauley JL, Pericak-Vance M, Oksenberg JR, Hauser S, Sexton D, Haines J, Sawcer S, the Wellcome Trust Case-Control Consortium (WTCCC) and Compston A: *Replication analysis identifies TYK2 as a multiple sclerosis susceptibility factor.* Eur J Hum Genet. 2009 in press

Paper V: International Multiple Sclerosis Genetics Consortium (IMSGC) (collaborators alphabetically sorted by country: Booth DR, Heard RN, Stewart GJ, Goris A, Dobosi R, Dubois B, Lorentzen AR, Celius EG, Harbo HF, Spurkland A, Olsson T, Kockum I, Link J, Hillert J, Ban M, Baker A, Sawcer S, Compston A, Mihalova T, Strange R, Hawkins C, Ingram G, Robertson NP, De Jager PL, Hafler DA, Barcellos LF, Ivinson AJ, Pericak-Vance M, Oksenberg JR, Hauser SL, McCauley JL, Sexton D, Haines J.) *The expanding genetic overlap between multiple sclerosis and type 1 diabetes.* Genes Immun. 2009;10(1):11-14.

4 Introduction

Multiple sclerosis (MS) has been studied for more than a hundred years, and Jean-Martin Charcot (1825-1893) was one of the first people who described the disease (Charcot JM 1868). MS affects the central nervous system (CNS) and is a common cause of neurological disability affecting young adults. In a newly published Norwegian study including patients from the Oslo area, the prevalence among individuals of native Norwegian ancestry was found to be 170/100 000, while the general prevalence was 148/100 000 (Smestad et al. 2008). This demonstrates that individuals of northern European ancestry have a high prevalence (Compston et al. 2005). The finding is in line with the known prevalence distribution worldwide, which shows an increased prevalence correlating with the distance from equator (Kurtzke et al. 1979). However, recent findings suggest that the latitude gradient is decreasing (reviewed in Alonso & Hernan 2008).

4.1 Multiple sclerosis

4.1.1 Clinical aspects of MS

The symptoms of MS depend on the location of the inflammatory and demyelinating lesions within the CNS, and are therefore very variable. In most of the patients the disease begins as episodic attacks of neurological deficits (MS attack, MS relapse or MS schub), which usually reach a plateau and resolve over days or weeks. This type of disease course is called relapsing remitting MS (RR-MS). Examples of symptoms are numbness and other sensory symptoms, spasticity, paresis, double vision, vision loss, ataxia or bladder control problems. Also more unspecific symptoms as depression, emotional lability, fatigue, cognitive impairment and pain are commonly seen. The RR-MS disease course often converts over time into a progressive disease, called secondary progressive MS (SP-MS). In some patients the relapsing symptoms are missing and the disease evolves as a primary progressive disease (PP-MS). The PP-MS frequency is suggested to be between 10-20% (Compston et al. 2005). MS is more commonly seen in females, with a female:male ratio of 2:1. Interestingly, a recent review by Alonso et al. reports that this ratio has increased in the last five decades (Alonso & Hernan 2008). The mean age at onset is in the early thirties, earliest in females with RR-MS and latest in PP-MS patients (Compston et al. 2005).

The MS diagnosis can often be made clinically, when episodes disseminating in time and neuroanatomical location are identified. However, both laboratory tests and magnetic resonance imaging (MRI) findings should support the diagnosis. In the cerebrospinal fluid (CSF) of MS patients one commonly finds a modest pleocytosis and raised intrathecal immunoglobulin G (IgG) synthesis, detected as oligoclonal bands by using isoelectric focusing or agarose gel electrophoresis. Approximately 95% of the MS patients have an increased CSF:serum ratio of IgG (Andersson et al. 1994). MRI is helpful both as a diagnostic tool and for following disease activity. On MRI scans, lesions are typically found in periventricular and subcortical regions as well as close to corpus callosum. MRI of both the brain and medulla may show multiple high signal lesions on a T2-weighted scan. However, such findings are not specific. Contrast enhancing lesions on a T1-weighted MRI scan (Gadolinium enhancement) is a sign of a permeable blood brain barrier (BBB), and thus a more specific sign of brain inflammation. Interestingly, clinical symptoms are not always correlated with the inflammation activity on MRI scans. RR-MS patients often have more activity on the MRI scan than clinically recognized (called silent lesions) (reviewed in Goodin 2006). A set of criteria (called Barkhof's criteria) are established for evaluation of MRI scans in MS (Barkhof et al. 1997). MRI data is often included in addition to other clinical information in different kinds of MS research. For instance, in a recent genome-wide association study (GWAS) of MS patients, MRI data were included in clinical subgroup analyses (Baranzini et al. 2009).

In 1983 Poser et al. published a set of MS criteria for research purposes (Poser et al. 1983). They defined the MS diagnosis in four categories; clinically definite MS, laboratory-supported definite MS, clinically probable MS and laboratory-supported probable MS. These criteria were later revised by "The International Panel on MS Diagnosis" (often called the McDonald criteria (McDonald et al. 2001)). MRI findings were then integrated into the criteria and it was recommended to use these categories; MS, possible MS and not MS. Later, the MS criteria have been evaluated (Miller et al. 2008; Palace 2009; Polman et al. 2005). By using the expanded disability status scale (EDSS), the disability of MS can be measured. The EDSS scoring system is widely used to follow the progression of disability in MS patients over time, both for evaluation of treatment results and for scientific purposes (Kurtzke 1983). The EDSS is based on the scores in the eight functional systems (FSs) (pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral and other), in addition to evaluation of walking

ability. An EDSS score of 0 means normal neurological examination, whereas a score of 10 is death due to MS. EDSS steps between 1.0 and 4.5 refer to an MS patient who is fully ambulant without walking aid. EDSS steps between 5.0 and 9.5 are defined by the impairment of ambulation. The EDSS score is commonly used; however, the inter-rater variability may bias a study (Compston et al. 2005). The EDSS score do not measure psychological signs or cognitive dysfunction, which may have a huge impact on the quality of life for MS patients. Recently, a global multiple sclerosis severity score (MSSS) was established, based on an algorithm which estimates the severity when only single assessment data are available (Roxburgh et al. 2005).

A MS schub or attack can be treated with corticosteroids in the acute phase. If a patient has more than two attacks during the last two-three years, immunomodulatory treatment has through the last decade been recommended. The recommendation guidelines are now changing towards an earlier start of immunomodulatory treatment. For patients with a clinical isolated syndrome (CIS) (one attack) with following silent MRI activity, treatment is now often recommended. Interferon beta (IFN-β) (Avonex®, Rebif®, Betaferon®) or glatiramer acetate (Copaxone®) are usually tried before monoclonal antibody therapies (reviewed in Feldmann & Steinman 2005; Linker et al. 2008). IFN-β and glatiramer acetate prevent approximately 30% of further attacks, however, it is still debated if the treatment delays the progression of the disease. In Norway, the monoclonal antibody natalizumab (Tysabri®) (Miller et al. 2003) has been registered for use in RR-MS since 2006. However, since severe side effects of natalizumab have been reported (as progressive multifocal leukoencephalopathy), this treatment is usually recommended to patients with insufficient effect of first-line medication or a severe disease onset (reviewed in Goodin et al. 2008). Another monoclonal antibody therapy, alemtuzumab (Mabcampath®, previously Campath-1H®), is now in its third phase of clinical trials and reports show very promising results (Coles et al. 2008). Mitoxantrone (Novantrone®), an immune suppressive medication widely used in cancer treatment, is recommended in MS patients with rapid secondary progression. In addition to the treatments mention above, new promising treatments are under development and studied in clinical trials (summarized in Linker et al. 2008). However, many MS patients do not fulfill the criteria for immunomodulatory treatments and mainly receive different kinds of symptomatic treatment.

4.1.2 MS as an immune-mediated disease

A review by Zhernakova and colleagues summarize that autoimmune diseases are believed to be caused by inappropriate destruction of normal tissue by the immune system, involving failure of the self-tolerance mechanism, whereas inflammatory disorders result from an excessive inflammatory response that is more harmful to the host than the exogenous antigens (Zhernakova et al. 2009). More concrete criteria have also been postulated to define an autoimmune disease (reviewed in Bar-Or 2008). First, presence of immune mediators (as auto-antigens) within the site of the pathologic lesion is required. Secondly, these immune mediators should not be present in people who do not have the illness. Thirdly, the presumed immune mediators should initiate the disease and fourth, when these are removed, therapeutic effects are seen. Also effects of anti-inflammatory and immunomodulatory treatments strongly support an immune-mediated condition.

It has been postulated that MS is an autoimmune disease affecting CNS (reviewed in McFarland & Martin 2007). Response of immunomodulatory drugs is often seen when treating MS patients, which supports the immune-mediated disease hypothesis. It has been difficult to identify auto-antigens in MS patients. Recently, Derfuss and others identified contactin-2 as an auto-antigen target by T cells, and they suggested that this could be important in development of grey matter pathology in MS (Derfuss et al. 2009). Quintana et al. found different auto-antibody patterns in serum of MS patients, suggesting that auto-antibody signatures are linked to different stages and pathologic processes in MS (Quintana et al. 2008). The findings in experimental autoimmune encephalomyelitis (EAE) (reviewed in Baxter 2007), an animal model for MS, strongly support an autoimmune or immune-mediated MS pathogenesis. In the EAE model, the animals are injected with myelin peptides (as myelin basic protein, proteolipid protein and myelin oligodendrocyte glycoprotein). An adjuvant is added to break the BBB. This induces an autoimmune response with activation of $CD4^+$ T cells specific for myelin peptides. The T cells are essential for development of the inflammatory, demyelinating lesions in CNS in the animal model. When these activated T cells are transferred between animals in the EAE model, the same CNS response can be induced (Paterson 1960).

The notion that one patient can have more than one immune-mediated disease more often than expected by chance, strengthens the hypothesis of common pathways for immune-

mediated diseases. Also MS patients may have co-existence of other immune-mediated diseases. Barcellos and colleagues showed that 26% of MS patients had at least one co-existing autoimmune disorder (Barcellos et al. 2006a). In that study, 64% of the families with a history of MS also reported other autoimmune disorders in one or more first-degree relatives. For both the MS patients and the family members the most common diseases were; Hashimoto thyroiditis, psoriasis and inflammatory bowel disease (including both Crohn's disease and ulcerative colitis). A study from Denmark supports these findings by showing an increased risk of ulcerative colitis in MS patients and first-degree relatives compared with the general population (relative risk 2.0 and 1.3, respectively) (Nielsen et al. 2008).

A strong evidence for shared molecular mechanisms among immune-mediated diseases, is the genetic association with the human leukocyte antigen (HLA) complex found for most of the immune-mediated diseases, MS included (reviewed in Fernando et al. 2008; Thorsby & Lie 2005). The HLA molecules are surface molecules essential for a proper immune response (McCluskey & Peh 1999). HLA class I molecules are expressed on most nucleated cells, and about 10^4 to 10^6 class I molecules are expressed on the cell surface. The three classical class I loci (A, B and C) are all expressed at the same time. These molecules are specialized to present peptides synthesized within the host cells. The molecules can therefore present intracellular antigens, as viral and tumour antigens, for the T-cell receptor (TCR) on the $CD8^+$ T cells. The class I molecules act also as ligands for the killer immunoglobulin-like receptor (KIR) molecules, which are important receptors balancing the activation of natural killer (NK) cells (reviewed in Vivier et al. 2008) and some subset of T cells (van Bergen J. et al. 2004). The NK cells are important components of the innate immune system, and elimination of virus-infected and malignant cells are among their functions. Recently, there has been renewed interest in the NK cells as cytokine-producing cells and as potential regulators of adaptive immunity (reviewed in Orange & Ballas 2006). HLA class II molecules are found mainly on antigen-presenting cells (APCs). These molecules are specialized to present peptides derived from outside the host cells to the TCR on the $CD4^+$ T cells, such as bacterial fragments and other extracellular antigens. The HLA molecules are therefore key molecules both in innate and adaptive immunological pathways.

Among other key molecules in the immune response are the cluster of differentiation 28 (CD28) molecule, the cytotoxic T-lymphocyte-associated protein 4 (CTLA4) molecule and the inducible T-cell co-stimulator (ICOS) molecule (reviewed in Rudd & Schneider 2003). These molecules are all expressed on the T-cell surface. The CD28 molecule is expressed on naïve or immature T cells and act as a co-stimulatory molecule when the T cell is activated. Both the CD28 and CTLA4 molecules bind the same molecule on the APC, the B7 ligand (alias CD80 and CD86), but the CTLA4 molecule has 50-100-fold higher binding activity than the CD28 molecule. Whereas the CD28 molecule signal is essential for the initiation and progression of a T-cell response, interaction between the CTLA4 and the B7 molecule results in down-regulation of the ongoing immune response. The ICOS molecule is also important for the T-cell activation and proliferation, however, the exact function is more unclear.

The list of molecules involved in immune responses is extensive. Therefore, the list of candidate genes of interest in MS is long. However, the search for susceptibility loci in complex genetic disorders as MS and type 1 diabetes (T1D) has been more difficult than expected. With new techniques and collaborations worldwide, the genetics of immune-mediated diseases seems, however, to have taken a big step forward (reviewed in Gregersen & Olsson 2009; Lettre & Rioux 2008). Since at least 3-5% of the population worldwide suffers from immune-mediated diseases, a research breakthrough will have huge impact for many patients. Postulated similar pathways and shared molecular disease mechanisms create a combined research field for many diseases. Hopefully, within a few years, the impact of gene variants and the mechanisms in immune-mediated diseases will become clearer.

4.1.3 Pathoimmunological aspects of MS

Also pathologically, MS has been described as an inflammatory disease of the CNS. Demyelination and a variable extent of axonal injury are seen, as well as perivascular and parenchymal inflammatory infiltrates. The demyelination seen is thought to be driven by the inflammation (summarized in Compston et al. 2005). This is strongly supported by findings in the EAE model (reviewed in Wekerle 2008). The inflammation is not only seen in the white matter, but also in grey matter of CNS (Bo et al. 2003). In addition to signs of inflammation, signs of degeneration are seen (reviewed in Geurts & Barkhof 2008).

In active MS lesions, T cells and a smaller number of B cells, massive macrophage infiltration and microglia activation are seen. Based on such findings, the most established hypothesis suggests that the disease starts with activation of auto-reactive T cells in the peripheral immune system (reviewed in Holmoy & Hestvik 2008; Sospedra & Martin 2005). The auto-reactive T cells are usually guarded by regulatory T cells and NK cells. For reasons not known, the APCs probably present auto-antigens recognized by these T cells. The EAE model suggests that this antigen may be a myelin component, however, myelin auto-antigens are not yet identified in MS in humans.

After the T cells have been activated peripherally, the T cells penetrate the BBB. Activated T cells express adhesion molecules binding to cerebrovacular endothelial cells, which facilitates the penetration. Interestingly, natalizumab (Tysabri®, the humanized monoclonal antibody against the cellular adhesion molecule α4-integrin on the T cells), used in MS treatment, blockades the T cell's opportunity to adhere to the cerebrovascular endothelium (von Andrian & Engelhardt 2003). Also the humanized monoclonal antibody alemtuzumab (Mabcampath®, previously Campath-1H®), which binds CD52 expressed on lymphocytes and monocytes, leads to profound reduction in inflammatory activity (Coles et al. 2008). This suggests that these immune cells are strongly involved in CNS inflammation of MS patients.

After the T cells have passed the BBB a local reactivation of the T cells occurs in CNS, and the cells develop into T cell subtypes. Whereas $CD4^+$ T cells (the T-helper cell, subset Th1 and Th17) are probably the main inducers of the disease, the $CD8^+$ T cells may be more important for tissue damage. The pro-inflammatory Th1/Th17 T-cell subgroups may stimulate macrophages and microglia, which further attacks the brain tissue. These are possible mechanisms for the development of demyelination and axonal damage loss in MS (reviewed in McFarland & Martin 2007).

4.1.4 Genetic epidemiology and environmental factors in MS

Even though the cause of MS is still unknown, a substantial amount of data indicates that both environmental and genetic factors contribute to the disease. Studies of familial clustering in MS have shown that 15-20% of the patients also have relatives with the disease. When comparing the risk in a sibling of an affected individual with the risk in the general population, lamda (λ_s) can be calculated (Risch & Merikangas 1996). The λ_s

measurement includes not only the shared genetic factors but also the shared environmental factors as shown in the function λ_s=1+G+E+(GxE), where G=genetic factors and E=environmental factors (Guo 2000). In MS, λ_s is around 15 and at the same level as other immune-mediated diseases, such as T1D. The risk of getting MS is 0.1% in the general population, while 3-5% in first degrees and siblings of an affected individual and 25-30% in monozygotic twins (Ebers et al. 1995; Mumford et al. 1994). The fact that adopted siblings of a MS patient have the same risk as the general population proves that the familial clustering occurs on a genetic basis and is not caused by factors in the family environment (Dyment et al. 2006). However, the concordance rate in monozygotic twins of only 25-30% in MS, indicates that environmental factors are also important for development of the disease. Furthermore, genetics can not explain the decline in risk of developing MS for people immigrating from high risk to low risk areas, or vice versa (Gale & Martyn 1995). So, regional, environmental risk factors are probably of significance in the way that environmental factors may trigger an immune-mediated reaction in a genetically susceptible individual (Hafler 1999). A series of infectious agents, especially viruses, have been studied in MS. The most promising candidate is most likely the Epstein-Barr virus (EBV) (reviewed in Salvetti et al. 2009). Especially, individuals affected in late age by EBV-infection have increased risk of developing MS (Thacker et al. 2006). Interestingly, Serafini et al. identified EBV-infection in a substantial proportion of brain infiltrating B cells and plasma cells in almost all MS patients examined (Serafini et al. 2007). They suggested that EBV persistence and reactivation in the CNS play an important role in MS immunopathology, but these findings have not yet been replicated by other research groups. Also, low levels of Vitamin D, as well as smoking, have been suggested as risk factors for MS (reviewed in Ascherio & Munger 2007b; Ascherio & Munger 2007a). In conclusion, MS is a multi-factorial disease probably influenced by both multiple environmental and genetic risk factors.

4.2 Genetic concepts and strategies in studies of complex diseases

4.2.1 Basic genetic concepts

The genome variation makes every individual unique, but variations in the genome can also cause disease. The human genome consists of approximately three billion nucleotides or base pairs (bp). Roughly 25 000 genes are encoded. In the genome, billions of variants

are found, but it is estimated that only 10-14 millions are common variants, with a minor allele frequency (MAF) of more than 1-5% (Kruglyak & Nickerson 2001). *Polymorphisms or common sequence variants* are variations in the genome sequence. A locus is polymorphic in a population, if all alleles are seen in more than 1% of the total count (Wang et al. 2005). The most common variation is a *single nucleotide polymorphism* (SNP), which is a single bp variation between individuals at a specific locus. In the NCBI dbSNP Build 129 (http://www.ncbi.nlm.nih.gov) approximately 15 million SNPs are currently listed, however only 6 millions of these are validated. *Microsatellites* (also called variable number of tandem repeats) are short, repeated deoxyribonucleic acid (DNA) sequences usually made up of 2-5 nucleotides. In addition to these relatively common variations, larger repeats and structural variations are seen. Structural variations normally involve segments of DNA that are larger than one kilo base pair (kb). If the structural change between individuals differs in number, the structural variation is called *Copy Number Variation* (CNV) (reviewed in Beckmann et al. 2007). This could for instance be large deletions, insertions or duplications.

The consequence of a genetic variation depends on the location in the genome. The variation could for instance be in a protein-coding region and could therefore lead to different incorporated amino acids in the protein synthesis or a stop in the protein synthesis. If this leads to disease, this variation is a direct causal variant. Variation in a non-coding region could also influence the coding region, since non-coding regions may be regulatory sites for transcription of other genes and can also influence gene splicing.

An *allele* is the specific nucleotide or sequence variant for a polymorphic locus (e.g. A and a). The combination of the two alleles is termed the *genotype* for that specific locus. For a SNP marker, three genotypes are possible; homozygous for one allele, heterozygous for the two different alleles and homozygous for the other allele (e.g. AA, Aa, aa). A *haplotype* is a combination of alleles at different loci at the same chromosome. For a specific haplotype, the alleles from the different loci are inherited together on the same DNA strand.

4.2.2 Strategies to identify disease associated or linked genes

Broadly, there are two approaches for mapping and investigating genes causing complex (i.e. multifactorial) genetic diseases; one approach with a prior hypothesis (e.g. candidate

gene studies) and a hypothesis-free approach (e.g. genome-wide screen studies). These two strategies can again be addressed by two different genetic approaches; association or linkage. In candidate gene studies the investigated genes are selected based on their potential biological relevance to a disease. A huge number of candidate gene studies have been performed for most of the complex genetic diseases, however, until recently it has been difficult to point out susceptibility genes even with a prior hypothesis. The genome-wide screen studies which could be either linkage or genome-wide association studies (GWASs), are hypothesis-free. *Linkage studies* aim at establishing linkage between a genetic marker and a disease locus and use sample sets of families with several affected members (Lander & Kruglyak 1995). Linkage is a term used to describe two or more loci that are inherited together because of their usually close location on the same chromosome.

The goal of an *association study* is to find significant differences for a polymorphism between affected and unaffected individuals (risk-enhancing alleles or protective alleles). A c*ase-control design* is the most common for association studies. For a specific locus, the case-control study compares the allele frequencies of a set of unrelated affected individuals (the cases) to a set of unrelated unaffected individuals (the controls). Another study design is the *family-based association study*. In complex disease studies, trio families have often been used, which consist of the affected individuals and their parents. The advantage with trio-based association studies is that the "control group" is ethnically well-matched, since the parent's non-transmitted alleles are used as controls. However, the blood sampling is time consuming and more difficult to administrate than in a case-control study. The parents may not be available, since MS has a relatively late onset. Furthermore, for every affected case you have to genotype two individuals (the parents) as controls.

Alleles of neighbouring loci are often inherited together, a phenomenon called *linkage disequilibrium* (LD). Alleles at loci that are in LD will therefore be transmitted together to the offspring more often than expected by chance from their respective allele frequencies in the general population (Slatkin 2008). The recombination rates and the LD throughout the genome vary, but the LD is generally stronger when the loci are close to each other on the chromosome. Regions with high LD are often named conservative regions. Two different terms are commonly used to measure LD; the correlation coefficient r^2 and D'

(Devlin & Risch 1995; Pritchard & Przeworski 2001). When r^2=1, it is called perfect LD, and the alleles are completely correlated. For a two loci haplotype, both alleles will then always occur on the same haplotype (only two haplotypes are possible; e.g. the A-B haplotype and the a-b haplotype). In this case it is only necessary to genotype one of the markers to extract information for both markers in a bi-directional manner. When D'=1 it is called complete LD, compared to no LD when D'=0. D' measurement is uni-directional and only one of the two alleles is fully tagged by the other marker. When D'=1 one allele always occurs on the same haplotype as the other, but not visa versa.

The LD between nearby loci is used both in candidate gene and genome-wide approaches. Instead of typing all the variations in a DNA sequence, which would be exhaustive concerning cost and time, few variations can tag a region. If the genetic associated marker tested itself is not the causal variant, the marker is indirectly associated due to the LD with the causal variant. This implies that an associated polymorphism does not need to be the disease causing variant but rather a tag-maker (Johnson et al. 2001). In 2003 the Human Genome Project (HGP) (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) had sequenced the first human genome and in 2004 the first human reference sequence (NCBI Build 35) was produced by the International Human Genome Sequencing Consortium (http://genome.ucsc.edu/). During these years the information in the public available databases on common genetic variations has increased substantially. The HapMap project (http://www.hapmap.org/) was performed by genotyping individuals from four populations (of African, Asian and European ancestry). Based on these individual's genotypes, the LD pattern between alleles of different loci are calculated. By use of the HapMap database, release 2005 and 2007 (Frazer et al. 2007; The International HapMap Consortium 2005) and sophisticated software it is now easier to choose the most informative tagSNPs in a specific region.

4.2.3 Disease models for a common complex disease

Two models for the genetics in common complex diseases have been suggested. The first hypothesis is the "Common disease / common variants" (CD/CV) hypothesis (Pritchard & Cox 2002; Reich & Lander 2001), which suggests that 20-100 genes are involved in disease development and that each of these genes increase the risk only by a modest factor of 1.2-1.5 (Yang et al. 2005). This model suggests that the many genetic variants

which underlie complex diseases are common, and they may therefore be identified by a GWAS approach. The alternative model is the "Common disease / rare variants" (CD/RV) hypothesis or the heterogeneity hypothesis (Smith & Lusis 2002; Terwilliger & Weiss 1998). This hypothesis suggests the existence of hundreds possibly thousands of rare genetic variants. These rare and private variants are likely to have high penetrance, and each of the variants may increase the risk for disease by 10-20 times (Yang et al. 2005). Since these variants are rare, a SNP-tagging approach is difficult, and each of these rare variants have to be genotyped, for instance by sequencing. In favour of the CD/CV hypothesis it could be argued that the common alleles which now are disease predisposing, were in the past an advantage. An example is the body's ability to store fat, which in "older" times was an advantage, but which now is associated with lower survival. It could also be argued that the selection pressure for a common disease is moderate, since these diseases usually have a late at onset.

4.2.4 GWAS - the recent approach in complex genetic studies

Over the last couple of years, the research field of complex genetic diseases has taken a big step forward. With new high through-put techniques, decreased costs and increased size of case-control sample sets available, genome-wide association studies (GWASs) have been performed in many complex diseases. GWASs have mainly been performed in cases and controls, genotyping a dense set of SNPs genome-wide. The genotyped SNPs are supposed to tag a substantial proportion of the common variation in the genome. It is estimated that a few hundred thousand SNP markers could capture approximately 80% of the genome (Kingsmore et al. 2008; Wang et al. 2005). During the last few years, the numbers of GWASs have increased exponentially. The National Human Genome Research Institute provides an updated list of published GWASs (http://www.genome.gov/26525384). As shown in figure 1, the total numbers of GWASs in some selected immune-mediated diseases had reached 32 by the end of 2008.

An alternative approach to a GWAS, is a screening of only the SNPs leading to a protein-coding change in the genome. Non-synonymous SNP (nsSNP) screening might directly identify molecules of importance to disease development. In these kind of studies less numbers of SNPs need to be genotyped (approximately 15 000 compared to at least 500 000 in a GWAS) and the cost is therefore less.

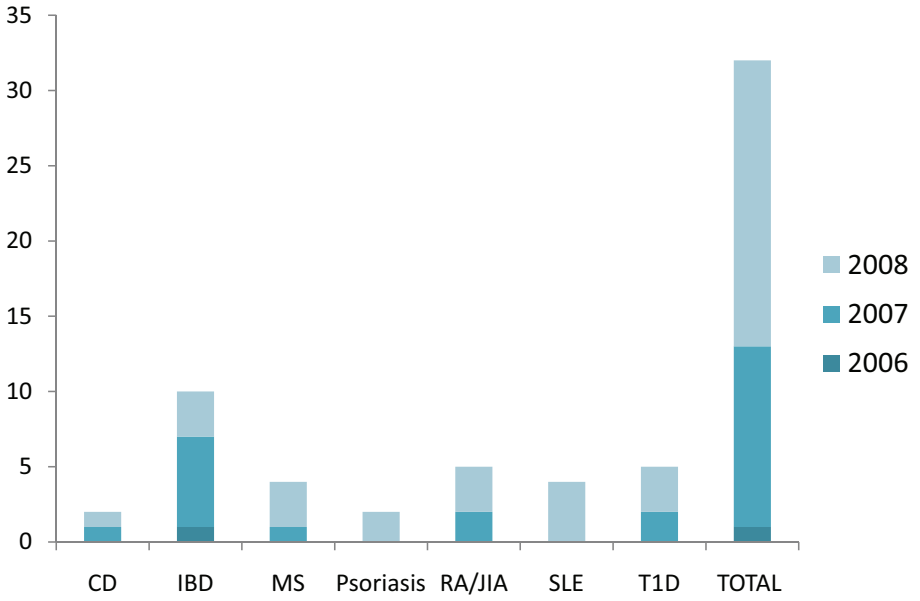

Figure: 1 Number of GWASs published in the years 2006, 2007 and 2008 for some selected immune-mediated diseases. The data in this figure is based on the updated list of GWASs (http://www.genome.gov/26525384)

Abbreviations: CD: celiac disease; IBD: inflammatory bowel disease including Crohn's disease and ulcerative colitis; MS: multiple sclerosis; RA: rheumatoid arthritis; JIA: juvenile idiopathic arthritis; SLE: systemic lupus erythematosus; T1D: type 1 diabetes

Both these approaches are hypothesis-free, in terms of not picking candidate genes or regions for a study *a priori*. However, since the nsSNP screen only focuses on the variations that give an amino acid change in the encoded proteins, this approach is based on the assumption that the studied nsSNPs are of functional importance. A lot of positive associations in GWASs and nsSNP screens are expected to be type I errors (false positive results). It has been widely discussed how to correct for multiple testing without loosing the true signals. It is often required that the associated SNPs have to be replicated in other sample sets as well as explored in more detail by fine mapping or resequencing to establish definite associations.

4.3 Genetic aspects of multiple sclerosis

4.3.1 The HLA complex - the early success

The major histocompatibility complex (MHC) is located on the short arm of chromosome 6 (6p21.3) and the region is divided into three; the class I region, which is the most telomeric region, followed by the class III and the class II gene region. The MHC spans 4 megabases and consists of approximately 300 loci and over 160 protein-coding genes. Many of these genes have immune-related functions (Horton et al. 2004). Extensive LD exists in the region, and large haplotype blocks can be defined (Alper et al. 2006; Blomhoff et al. 2006). Several of the genes encode classical immune response molecules, important in the antigen presentation. In humans, these genes are called the human leukocyte antigen (HLA) genes and most of them are highly polymorphic. These genes are located both in the class I (*HLA -A, -C, -B*) and class II (*HLA –DR, -DQ, -DP*) gene regions.

The first HLA-association with MS was reported for the HLA class I alleles; HLA-A3 (Naito et al. 1972) and HLA-B7 (Jersild et al. 1972). Later it was shown that the HLA class I associations were mainly secondary to an association in the HLA class II region (Compston et al. 1976; Olerup & Hillert 1991) due to strong LD in the region. The HLA-DRB1*1501-DQB1*0602 haplotype (often called the HLA-DR2 haplotype), is now well established as the strongest genetic association in MS. Since the two loci *HLA-DRB1* and *HLA-DQB1* are in strong LD in Europeans, it has been difficult to pin-point which of the two loci causes the primary association on the HLA-DR2 haplotype. However, by studying African American MS patients and controls who have less strong LD between these two loci, Oksenberg and colleagues found an independent effect of the *HLA-DRB1* locus and thus suggested this as causing the association (Oksenberg et al. 2004). This view seems now generally accepted, even though some studies have indicated that the *HLA-DQB1* locus could cause the primary association (Caballero et al. 1999; Spurkland et al. 1997; ves-Leon et al. 2007). Recently, it has been suggested that within the HLA class II region, epistatic interactions between loci may occur (*HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1*) and that both the HLA-DRB1*1501 and HLA-DQB1*0602 alleles each influence the MS risk in this way (Lincoln et al. 2009). Interestingly, in the study by Lincoln et al., incomplete haplotypes (carrying either DRB1*1501 or DQB1*0602) did not predispose to MS disease, thus the authors suggested that the HLA class II risk is

caused by the HLA-DR2 haplotype rather than one allele. The HLA-DR2 haplotype is especially frequent in Northern Europe. The allele frequency of HLA-DRB1*15 is found to be around 33-36% among Norwegian MS patients and 15-16% among Norwegian controls (odds ratio (OR) 2.9) (Harbo et al. 2004; Smestad et al. 2007). In almost all the studied MS populations in Europe there is an association to the HLA-DRB1*1501 allele (summarized in Ballerini et al. 2004). It has also been suggested that other *HLA-DRB1* alleles are associated with MS (Barcellos et al. 2006b; Dyment et al. 2005; Ramagopalan et al. 2007). The most established additional risk allele is HLA-DRB1*03 (Yeo et al. 2007). In Sardinia, where HLA-DRB1*1501 has a low frequency, the HLA-DRB1*03 is significantly associated with MS (Marrosu et al. 2001). It is estimated that the HLA class II association accounts for 20-60% of the genetic susceptibility in MS (Haines et al. 1998).

It has also been reported that genes located in the MHC outside the class II region are independently associated with MS. For instance a variation in the myelin oligodendrocyte glycoprotein (*MOG*) located telomeric to the *HLA-A* locus was found to be associated among Italian MS patients (D'Alfonso et al. 2008). Among Swedish patients, the HLA-A2 allele was found to decrease the risk of MS independently of the HLA-DR2 haplotype (Brynedal et al. 2007; Fogdell-Hahn et al. 2000). In a Norwegian study we found a higher genotypic relative risk in individuals carrying both HLA-DR2 and HLA-A3 compared to those who carry only HLA-DR2 or only HLA-A3 (Harbo et al. 2004). Recently, an independent effect of *HLA-C* was reported (Yeo et al. 2007). In this study the HLA-Cw*05 allele showed a protective effect after stratification for the *HLA-DRB1* associations (HLA-DRB1*15, *03 and *01 alleles). In addition to the classical HLA typing (*HLA-A*, *-B*, *-C*, *-DRB1* and *DQB1*), Yeo and colleagues genotyped 110 SNPs and 50 microsatellites to increase the information. When Lincoln et al. genotyped more than 1000 SNPs in the MHC complex in addition to *HLA-DRB1* and *DQB1* typing, they did not find any HLA class I effects (Lincoln et al. 2005). However, the study was not optimal to detect such associations since only 60% of the SNPs were successfully genotyped and the classical HLA class I loci (*HLA-A*, *-B* and *-C*) were not genotyped.

4.3.2 Search for other MS susceptibility loci - a challenge

The HLA association found in MS can not explain all genetic susceptibility in MS (Haines et al. 1998). Therefore the search for other MS susceptibility loci has been

intense over the last few years. The early discovery of the HLA association was detected with just 36 cases (Jersild et al. 1973), and a highly significant association was shown with less than 200 cases (Olerup & Hillert 1991). This created a false belief that the relevant genes could be identified by testing only 100-200 cases and controls or 100 multiplex families (Oksenberg et al. 2008). Linkage screens were then performed, both with MS-affected sib pair families and trio families. The linkage studies were first performed by genotyping relatively few microsatellites (200-400 markers), since microsatellites are more polymorphic than SNPs and more informative in linkage studies (a low-density screen) (Ebers et al. 1996; Haines et al. 1996; Sawcer et al. 1996). The first linkage screens were small sized and proved to be underpowered, and some screens even had difficulties in finding the HLA association. Consortiums were established to perform larger studies, for instance, the Genetic Analysis of Multiple Sclerosis in EuropeanS (GAMES) (GAMES 2003). However, only the HLA association remained significant even in a larger high-density linkage screen performed by genotyping more than 5000 SNPs (Sawcer et al. 2005). A genome-wide LOD score for the HLA complex around 12 in this study indicated the limited power of the linkage studies to detect susceptibility loci outside the MHC in MS. Association studies have greater statistical power to detect common risks (the CD/CV hypothesis, see page 21-22) and an association approach was therefore chosen in the next generation of genome-wide studies in MS (reviewed in Wang et al. 2005).

In parallel to genome screens, a series of candidate gene studies have been performed in MS. A range of genes have been studied, from genes encoding myelin components and neuroprotective genes to genes involved in the immune response. Many of these studies have claimed associations with MS, but the findings have seldom convincingly been replicated in other sample sets. This was the case for the *CTLA4* gene, a gene located at chromosome 2q33 and expressed by T cells. The hypothesis was that polymorphisms in this gene, resulting in a dysfunction of the encoded molecule, could disrupt the immune response and contribute to immune-mediated disease. The *CTLA4* gene has been shown to be associated with several immune-mediated diseases (reviewed in Serrano et al. 2006). In a Norwegian MS case-control sample set association with the *CTLA4* +49 SNP was reported in 1999 (Harbo et al. 1999). A series of genetic analyses of the *CTLA4* gene and other closely located genes have thereafter been performed in MS, but diverse findings have been reported (for an overview see Teutsch et al. 2004). Interestingly, a new

polymorphism (CT60 SNP) in the *CTLA4* region was reported to influence gene splicing and thereby the relative abundance of soluble versus membrane-bound CTLA4 (Ueda et al. 2003). It was shown that a decrease in soluble CTLA4 protein resulted in an incomplete down-regulation of the immune response, a mechanism that could be involved in development of autoimmune disease. Association was shown for a haplotype including the CT60 SNP in autoimmune thyroid disease (AITD) and T1D (Ueda et al. 2003). Later, association with this gene region has been reported in celiac disease (Amundsen et al. 2004) and Addison's disease (Blomhoff et al. 2004). Regarding the *CD28* and *ICOS* genes, few studies have shown an association with immune-mediated diseases in spite of the important functions of these genes.

Another example of a candidate gene study in MS is the SH2 domain protein 2A gene (*SH2D2A)* gene, also investigated in this thesis. The *SH2D2A* gene, located at the chromosome 1q21 region was, identified by Spurkland et al. (Spurkland et al. 1998). It encodes the T-cell specific adaptor protein (TSAd) expressed in T cells (Dai et al. 2000; Sundvold et al. 2000) as well as NK cells (Nejad et al. 2004) and endothelial cells (Matsumoto et al. 2005; Wu et al. 2000). A microsatellite found in the gene's promoter region was studied in a Norwegian MS case-control sample set (Dai et al. 2001). Homozygosity for short alleles of this polymorphism (ie GA_{13} and GA_{16}) was associated with MS susceptibility in this study. A similar finding was found in another immune-mediated disease, juvenile rheumatoid arthritis (Smerdel et al. 2004), supporting that short alleles of the *SH2D2A* promoter polymorphism may contribute to the genetic susceptibility in immune-mediated diseases.

The list of other promising candidate genes in MS is long. Disappointingly, when genes found to be associated in other immune-mediated diseases have been investigated in MS, associations could seldom be replicated. For instance the *PTPN22* gene, a well-established susceptibility gene in numerous autoimmune diseases, has been investigated in MS without evidence of association (deJager 2006, Harbo 2006, Begovich 2005). Even though many studies have reported negative findings in MS, one can not conclude that all the genes examined in these studies are not involved in MS pathogenesis, since many of these studies have been underpowered.

4.3.3 New optimism in MS genetics - the GWAS success

The introduction of GWAS has been like a revolution for studies of complex diseases and led to a breakthrough in MS genetics. In July 2007 the first GWAS in MS was published (Hafler et al. 2007) by the International Multiple Sclerosis Genetics Consortium (IMSGC) (https://www.imsgc.org/). The original screen included 931 trio families from United Kingdom (UK) and United States of America (US), and the replication sample set was similarly sized. Around 400 000 SNP markers located genome-wide were investigated. This study was able to identify novel significantly associated susceptibility loci outside the HLA, which also in this GWAS was highly significantly associated with MS. The two main findings (outside the HLA) were the associations to the interleukin receptor genes (*IL2Rα* and *IL7R*). At the same time as this GWAS was published, two candidate gene studies of *IL7R* were published (Gregory et al. 2007; Lundmark et al. 2007). One of these reports included 600 cases and controls from the Oslo MS DNA-biobank (Lundmark et al. 2007). Interestingly, the *IL2Rα* gene association has also been found in other immune-mediated diseases like T1D (Lowe et al. 2007; Vella et al. 2005; WTCCC 2007), AITD (Brand et al. 2007) and celiac disease (Brand et al. 2007). It has therefore been postulated that this gene could mediate a shared susceptibility effect among immune-mediated diseases in general.

In 2007, also a nsSNP screen was published in MS by the Wellcome Trust Case Control Consortium (WTCCC) and the Australo-Anglo-American Spondylitis Consortium (Burton et al. 2007). Approximately 14 500 nsSNPs were genotyped in 1000 British cases from four different diseases (AITD, ankylosing spondylitis (AS), breast cancer and MS) and 1500 controls. Not surprisingly, the HLA gene region showed the strongest association. But also some other associations were reported, for instance the previously found association with the *IL7R* gene. The paper listed nine nsSNPs outside the HLA gene region as promising MS susceptibility SNPs (Burton et al. 2007). In an additional analysis the paper grouped the three immune-mediated diseases investigated (AITD, AS and MS) and reported possible combined associations for the *TYK2*, *C8B* and *IL17R* genes in these diseases.

In the first MS GWAS performed by IMSGC, a list of 14 loci were suggested to be associated to MS (Hafler et al. 2007). Several replication studies are reported recently and many of these gene regions are now firmly replicated and validated. Table 1 shows some

of the replicated MS-associated loci. Also other MS GWASs have been published recently and novel associations have been suggested (main findings listed in Table 1) (Aulchenko et al. 2008; Baranzini et al. 2009; Comabella et al. 2008). In conclusion, the GWAS revolution has led to the identification of a serious of genes outside the HLA complex in MS.

Table 1: Overview of recently reported non-HLA loci associated with MS

Gene	Chr	SNP	Located	RA	RAF	OR	References
IL2Rα (CD25)	10p15	rs2104286	Intronic	T	0.75	1.2	1, 2, 3, 4, 5, 6
IL7R (CD127)	5p13	rs6897932	Exon 6	C	0.75	1.2	1, 2, 3, 5, 6, 7, 8, 9
CD58 (LFA3)	1p13	> 1 SNP	Intronic			1.2	1, 4, 10
CLEC16A (KIAA0350)	16p13	> 1 SNP	Intronic			1.2	1, 4, 11,12, 13
RPL5#	1p22	rs6604026	Intronic	C	0.29	1.3	1, 4
FAM69A#	1p22	> 1 SNP	Intronic			1.1	1
EVI5#	1p22	> 1 SNP	Intronic			1.1	1, 14
CD226 (DNAM-1)	18q22	rs763361	Exon 7	T	0.47	1.1	11, 15
TYK2	19p13	rs34536443	Exon 21	C	0.95	1.3	16, 17
GPC5	13q32	rs9523762	Intronic	A	0.35	1.3	18
KIF1B	1p36	rs10492972	Intronic	C	0.27	1.3	19

Abbreviations: Chr: chromosome; SNP: single nucleotide polymorphism; RA: risk allele; RAF: risk allele frequency in unaffected controls; OR: odds ratio; IL2Rα: Interleukin 2 receptor alpha unit; IL7R: Interleukin 7 receptor; CD58: Clusters of differentiation molecule 58; CLEC16A: C-type lectin domain family 16, member A; RPL5: Ribosomal protein L5; FAM69A: Family with sequence similarity 69, member A; EVI5: Ecotropic viral integration site 5; CD226: Clusters of differentiation molecule 226; TYK2: Tyrosine kinase 2, GPC5: Glypican 5; KIF1B: Kinesin family member 1B. Comment: # *RPL5*, *FAM69A* and *EVI5* are located close to each other at chromosome 1p22, and are possibly the same signal

References: 1) (Hafler et al. 2007) 2) (Weber et al. 2008) 3) (IMSGC 2008) 4) (Rubio et al. 2008) 5) (Alcina et al. 2009) 6) (Akkad et al. 2009) 7) (Gregory et al. 2007) 8) (Lundmark et al. 2007) 9) (Alcina et al. 2008) 10) (De Jager et al. 2009) 11) (IMSGC 2009) 12) (Martinez et al. 2009) 13) (Zoledziewska et al. 2009) 14) (Hoppenbrouwers et al. 2008) 15) (Hafler et al. 2009) 16) (Burton et al. 2007) 17) (Ban et al. 2009) 18) (Baranzini et al. 2009) 19) (Aulchenko et al. 2008)

5 Aims of the study

The overall aim of this study has been to investigate genetic factors contributing to disease susceptibility as well as clinical outcome in MS. The study design has been based on two different approaches; the hypothesis-driven candidate gene approach and the hypothesis-free GWAS approach. Gene regions of importance to the immune system were regarded as of special interest for further studies.

The aim for **Paper I-III** and **Paper V** has been to investigate promising candidate genes in MS. Both the *CD28/CTLA4/ICOS* gene region (**Paper I**), the *SH2D2A* gene (**Paper II**) and the HLA class I genes and the KIR genes (**Paper III**), have previously been found of importance to MS or other autoimmune diseases, and we wanted to explore these regions more carefully. Furthermore, in a recent GWAS of T1D new susceptibility loci had been identified, and we aimed at studying these loci in MS in a large sample size (**Paper V**).

The aim for **Paper IV** was to replicate the most promising loci identified in the non-synonymous genome-wide SNP screen performed by WTCCC. Our Norwegian case-control sample set was included in a large combined analysis and thus a statistically well-powered sample set was achieved.

6 Summary of the results

Paper I

In the first study, the chromosome region 2q33, which encodes the CD28, CTLA4, and ICOS molecules, was investigated. These molecules are of regulatory importance in the immune system, and the corresponding genes are thus good candidate genes for immune-mediated diseases. An association to the *CTLA4* +49 SNP had been reported in Norwegian MS patients (Harbo et al. 1999), however, the involvement of this polymorphism in MS was not clear (summarized in Teutsch et al. 2004). In our study we aimed to extend our previous study by genotyping more markers in the gene region as well as increase the Norwegian sample size investigated. Six microsatellites (CD28-A,

CD28-B, SARA-43 (D2S307), SARA-1, SARA-31 and SARA-47) and three SNPs (*CTLA4* +49 (rs231775), CT60 (rs3087243) and CT61 (rs11571319)), were genotyped in 575 Norwegian MS patients and 551 controls. One of the SNPs (CT60) had been proposed to be the causal variant for the genetic effect in T1D and AITD (Ueda et al. 2003). No associations to MS susceptibility were observed for any of the markers analysed. When stratifying the sample set for HLA-DRB1*1501, gender, age at onset, disease course or familial aggregation association was neither found. Thus, this study could not confirm association with the *CD28/CTLA4/ICOS* gene region in MS. Our finding was in line with recent and more well-powered studies (Greve et al. 2008; Roxburgh et al. 2006).

Paper II

The background for study II was our previous report of an association to MS for the GA repeat sited in the *SH2D2A* gene promoter (Dai et al. 2001). The *SH2D2A* gene located at chromosome 1q23.1 encodes the TSAd molecule which is involved in regulation of T-cell activation. In this study we intended to replicate our previous finding in a large sample set as well as finemap the *SH2D2A* gene. In total 2128 Nordic (Danish, Finnish, Norwegian and Swedish) MS patients and 2004 controls were genotyped. By using a tagging strategy, the *SH2D2A* gene was finemapped in the Norwegian samples. Five tagSNPs (rs2768764, rs1800600, rs926103, rs909200, rs2768766) were chosen to tag the gene region. One of the tagSNPs was a non-synonymous SNP (rs926103) encoding a serine to asparagine substitution at amino acid position 52 in TSAd. In the Norwegians (624 MS patients and 562 controls) the haplotype, GA_{16} - rs926103*A was associated with MS (OR=1.4, P=0.004). The other Nordic sample sets were also genotyped for these two markers (the GA repeat polymorphism and the rs926103 SNP). A similar trend was observed in haplotype analysis among Danes (P=0.1). For single point analysis, when combining the OR between all the Nordic populations in a Cochran-Mantel-Haenszel analysis, association was found for the GA_{16} allele (OR=1.15 (1.04-1.28) and P=0.007). We conclude that the *SH2D2A* gene may contribute to susceptibility to MS.

Paper III

Traditionally, the HLA class I genes have been investigated in light of their ability to present antigen to the TCR on the T cells. But the HLA class I (HLA –A, -B, -C) variants

also serve as ligands for the KIRs expressed on NK cells and some subsets of T cells. KIR genes are members of the immunoglobulin (Ig) superfamily located in the leukocyte receptor complex (LRC) at chromosome 19q13 (Barrow & Trowsdale 2008). We investigated HLA class I alleles defined by their KIR binding motifs and evaluated whether these influenced MS susceptibility or severity, alone or in combination with their corresponding KIR genes. A sample set of 631 Norwegian MS patients and 555 Norwegian controls were typed for *HLA-A, -B, -C* and *–DRB1* alleles as well as the presence or absence of genes encoding inhibitory (*KIR2DL1, KIR2DL2, KIR2DL3, KIR2DL5, KIR3DL1, KIR3DL2, KIR3DL3*) and activating (*KIR2DS1, KIR2DS2, KIR2DS3, KIR2DS4, KIR2DL4, KIR2DS5, KIR3DS1*) KIRs. The frequency of the HLA-Bw4 specificity, which binds the KIR3DL1, was significantly reduced in MS patients as compared with controls (41.4% versus 55.1%, $P=4.6x10^{-6}$), even in individuals who did not carry any of the known HLA class II susceptibility alleles DRB1*1501 and DRB1*03, or the protective DRB1*01 allele (P=0.002). No significant differences in gene carrier frequencies of inhibitory and activating KIRs in MS patients as compared with controls were seen. However, our data indicate that MS patients who carry the inhibitory *KIR2DL2* and the activating *KIR2DS2* genes have a more severe disease than patients not carrying these genes. Also a trend of association was seen in the statistical interaction analysis between the KIR-HLA receptor-ligand pair KIR2DL1-HLA-C2. In conclusion, carriage of the ligand of the inhibitory KIR3DL1 receptor, HLA-Bw4, was found to protect against MS in an *HLA-DRB1* independent manner.

Paper IV

This paper aimed to replicate the findings identified in the recent genome-wide non-synonymous SNP screen (nsSNP screen) performed by WTCCC (Burton et al. 2007) including genotyping of 12 374 nsSNP. This is a consortium paper, including samples collected by research groups in five different countries (Australia, Belgium, Norway, UK and the US), thereby increasing the power of this study. A number of possible MS candidate genes were found in the nsSNP screen. Our study describes the extended analysis of 17 of these loci performed in samples from 4234 MS patients, 2983 controls and 2053 trio families. Among these, 638 Norwegian MS cases and 1026 Norwegian controls were genotyped. The SNPs were chosen among the 1% SNPs with lowest P-value in the nsSNP screen. Further interesting loci based on possible immunological

function were favoured even though their P-values were not amongst the lowest (but still in the 1% group). The previously reported MS-associated nsSNP in the *IL7R* gene was also genotyped and showed association also in our study. When combining all available data, we found strong association for one of the 17 loci investigated. This MS-associated nsSNP (rs34536443) is situated in exon 21 of the *TYK2* gene located at 19p13 ($P=2.7x10^{-6}$, OR=1.32 (1.17–1.47)) and cause an amino acid change from proline to alanine (P1104A). This SNP could have a functional role in MS, for which further studies are needed.

Paper V

In recent years it has been shown that some susceptibility genes predispose to more than one immune-mediated disease (for instance the *CTLA4* and *PTPN22* loci), and a common pathway for autoimmune susceptibility genes has been suggested. This is supported by familial clustering of immune-mediated diseases. In this paper, we tested seven SNPs, that were shown to be associated with T1D, in a large MS cohort consisting of 2369 trio families, 5737 cases and 10 296 unrelated controls collected in six different countries (Australia, Belgium, Norway, Sweden, UK, US). These seven SNPs were located in genes encoding molecules with a possible immunological function. The Norwegian sample set consisted of 644 MS cases and 1023 controls. Two of the seven genotyped SNPs showed association with MS in our study; rs12708716, an intronic SNP in the C-type lectin domain family 16, member A (*CLEC16A*) gene ($P=1.6x10^{-15}$) (located at chromosome 16p13) and rs763361, an intronic SNP in the *CD226* gene ($P=5.4x10^{-8}$) (located at chromosome 18q22). In addition the SNP rs3184504 in exon 3 in the SH2B adaptor protein 3 (*SH2B3*) gene (located at chromosome 12q24) showed a borderline genome-wide association in the total sample set ($P=4.4x10^{-6}$). Thus, in this study we confirm association of two novel MS susceptibility loci and support that some loci are shared amongst immune-mediated disease.

7 Methodological considerations

7.1 The sample sets

The Norwegian MS cases and their parents as well as the healthy controls used in this thesis are volunteers who have given informed written consent according to approved protocols by the Regional Medical Research Ethics Committee and the Data Inspectorate. Similar approvals have been given for the samples from other countries. The different case-control sample sets as well as trio families that have been include in our studies are listed in table 2 and 3 below.

Table 2: Overview of the case-control sample sets studied in this thesis

Case-control sample set	Study I		Study II		Study III		Study IV		Study V	
	case	cont	case	cont	case	cont	case	cont	case	cont
Norway	575	551	624	562	631	555	638	1026	644	1023
Belgium							368	374	398	461
Denmark			620	553						
Finland			240	240						
Sweden			644	649					1014	1173
UK							1871	1583	2370	7078
US							1357	543	1311	561
WTCCC							979	1470		
TOTAL	**575**	**551**	**2128**	**2004**	**631**	**555**	**5213**	**4453**	**5737**	**10296**

Table 3: Overview of the trio sample sets studied in this thesis

Trio sample set	Study IV	Study V
	Trio	Trio
Australian	372	356
Belgium	113	113
UK	693	1143
US	875	757
TOTAL	**2053**	**2369**

Abbreviations for Table 2 and 3: case: number of cases; cont: number of controls; trio: number of trio samples; **Study I:** *CD28/CTLA4/ICOS* gene region; **Study II:** *SH2D2A* gene; **Study III:** *KIR-HLA*; **Study IV:** *TYK2* gene (nsSNP screen); **Study V:** *CLEC16A* and *CD226* gene (T1D overlap); UK: United Kingdom; US: United States of America

In all studies included in this thesis (**Paper I-V**), samples from the Oslo MS DNA-biobank have been used. This biobank contains blood samples and extracted DNA from MS patients living in Oslo or its suburban areas. The collection started around 1992 and continues to this date. The majority of the patients have been recruited through the Department of Neurology, Ullevål, Oslo University Hospital, by neurologists working at

the department (organized by Dr Elisabeth G. Celius and Dr Hanne F. Harbo). In addition to the DNA samples from the Oslo MS DNA-biobank, patients from the Nordic MS Genetics Network (Denmark, Finland, Sweden and Norway (additional samples from Bergen before the Norwegian MS registry and Biobank was establish) (**Paper I**, **II** and **V**) and the International MS Genetics Consortium (IMSGC, https://www.imsgc.org/) (**Paper IV and V**) have been included.

To minimize possible population stratification, it is important that the controls used in genetic studies are well-matched with the cases, particularly concerning ethnicity. The Norwegian controls were all randomly collected among healthy blood donors recruited through the Norwegian Bone Marrow Donor Registry (NBMDR; http://www.nordonor.org/), placed at the Institute of Immunology, Rikshospitalet, Oslo University Hospital. The controls were healthy at the time of collection and are of Norwegian ancestry. NBMDR includes approximately 28 000 individuals, of which 58% are female (a ratio of 1.4 in 2008). Most of the control samples used in our studies were collected during the years 1992-1994. The distribution of the *HLA-A*, *-B* and *-DRB1* alleles in these controls has previously been compared with the allele distribution in the total NBMDR. No significant differences for allele frequencies above 0.05 were found between our controls and the total NBMDR, suggesting that the HLA distribution among the Norwegian controls used in these studies reflect the general Norwegian population. On a group level, the controls have been sex-matched to have more females than males due to the expected excess of females in the MS group.

To avoid running out of Norwegian DNA samples, whole-genome amplification (WGA) was performed using the GenomiPhi® kit (GE Healthcare, Global Headquarters, Chalfont St Giles, UK) prior to genotyping. The genotyping using amplified DNA has been validated with very good consistency between genotyping with genomic and amplified DNA (please see the sections “KIR-typing“ and “HLA-typing” page 41-42). Amplified DNA was used for genotyping in **Paper II-V**.

7.2 Phenotype

In genetic studies an overall goal is to identify genes that influence disease phenotypes and use of well-defined phenotypes are essential. To restrict misclassification, all MS patients included in the studies presented here, have been diagnosed in neurological

departments by neurologists. The MS criteria according to Poser (Poser et al. 1983) or McDonalds (McDonald et al. 2001) have been used in the clinical assessment. In a review by Sawcer, he points out that the total amount of misclassification is probably low in MS studies (1-2%) (Sawcer 2008). On the other hand, for large sample sets (like 10 000 cases and controls), the misclassification could be as much as 25% without dramatically reducing the power, based on the common variants hypothesis (OR=1.3 and MAF>0.1) (Sawcer 2008). In the Norwegian collection, MS patients with a Norwegian ancestry were preferred, to avoid possible ethnical genetic heterogeneity.

It could, however, be argued that MS patients are a heterogeneous group. The first symptoms of the disease, the localisation in CNS, the age at onset and age at diagnosis may be very variable between patients. Also the disease course and progression are variable in MS. A heterogeneous phenotype may contribute to more noise than a homogenous phenotype. Characterisation of the patients according to clinical parameters is therefore essential. However, a sub-grouping strategy in the statistical analysis may also run the risk of losing power, since each test-group will include less number of cases.

In this thesis, the following parameters have been used for sub-phenotype analysis:

1) *Gender* (**Paper I and II**): Since the disease affects females more often than males (ratio 2-3:1), sub-grouping by gender is often suggested. Many immune-mediated diseases have an overrepresentation of affected females (for instance AITD), and sub-grouping by gender is frequently used.

2) *Age at onset* (**Paper I-III**): This parameter is often used, since the genetic burden is suggested to be heavier among patients with early age at onset than for patients with late age at onset. Confavreux et al. showed that age at onset correlated with disability in the way that patients with younger age at onset also had earlier “milestone” disability (Confavreux & Vukusic 2006). However, it could be argued that the age at onset parameter is unreliable. The first symptom of the MS disease is very variable. Some symptoms are for instance more recognisable both for the patients and the doctors and hence the age at onset may vary accordingly. Therefore, some registries rather record age at diagnosis. An argument against use of age at diagnosis is the variable delay before patients consult a doctor. It has been observed recently that the time gap between age at

onset and age at diagnosis has decreased, suggesting that which of these parameters to use is less important.

3) *Initial disease course* (**Paper I-III**): Initial disease course is usually defined as RR-MS or PP-MS. However, the disease course may change over time and especially for RR-MS patients the relapsing phase often converts to a progressive phase, SP-MS. It may be difficult to exactly pin-point when the patient converts to the progressive phase. We have therefore only included the initial disease course in our studies. It should also be mentioned that the recruitment of MS patients to research projects could favour one clinical subgroup, for instance the RR-MS patients. This is the case in Denmark, where the recruitment of MS patients to research projects is done in treatment clinics. Therefore, most of the Danish patients recruited to our studies, are RR-MS patients. This is different from our recruitment in Oslo, where all patients in the Department of Neurology, Ullevål, Oslo University Hospital, will be asked to join research projects. However, this will not affect the results as long as the sub-group analysis evaluates these differences.

4) *Presence or absence of familial cases of MS* (**Paper I**): In study I, 485 MS patients were available for clinical evaluation. Among these 17% had familial cases of MS (data not shown in the study). This is in line with the report form the Oslo MS registry. Among 782 patients, 134 patients (21%) knew other MS cases in the family (C Smestad, personal communication).

5) *MSSS* (**Paper III**): In study III, the parameter MSSS was included. The MSSS was described in 2005 by Roxburgh and colleagues (Roxburgh et al. 2005). By using clinical data from 9 892 MS patients from eleven countries, the Oslo MS registry included, Roxburgh et al. established an MSSS algorithm, which is based on the EDSS score from each patient as well as the disease duration. Against this background data, a global MSSS reference table was established. The global MSSS table is a helpful tool for analyses of disease progression when only single assessment data is available. In our Oslo MS registry many patients were given an EDSS score, and time from age at onset to the measured EDSS score had been noted. Thus, we could on this background estimate as a measure of disease progression (MSSS) in our patients.

7.3 Marker selection

We have used a candidate gene approach when selecting the markers investigated in both **Paper I-III** and partly in **Paper V**. However, the selecting strategy has slightly changed during the progress of the studies. The candidate genes investigated in **Paper I** (*CD28/CTLA4/ICOS*) were, in addition to the functional aspect, good candidates due to previous association found to the gene region both in MS and other immune-mediated diseases. The markers were selected based on previous or ongoing studies (Ling et al. 2001; Ueda et al. 2003; Wiencke et al. 2006). Before the publicly available databases were launched, as the UCSC genome browser (http://genome.ucsc.edu/) and the HapMap (http://www.hapmap.org/), few tools were available to choose which markers to investigate.

In the study behind **Paper II,** we used publicly available databases to find the most informative SNP markers to finemap the *SH2D2A* gene. However, due to financial restrictions, we genotyped only a small fraction of the known variation in this candidate gene region in Norwegian samples. We used the tagging approach to select a minimal number of SNPs that cover as much as possible of the genetic variation in a gene region (reviewed by Balding 2006; Carlson et al. 2004). The tagging approach is based on the fact that most neighbouring SNPs are highly correlated. TagSNPs allow us therefore to indirectly test the genetic variations across the gene region of interest or the entire genome (as in GWASs). The tagSNPs are selected based on the correlation coefficient (r^2) values between alleles at variable sites. By using the tagging tool program Haploview (http://www.broad.mit.edu/haploview/) (de Bakker et al. 2005), five tagSNPs in the *SH2D2A* gene region were chosen. The r^2 between the tagSNPs and not genotyped variants were >0.8 and the MAF of the tagSNPs were estimated to be >0.1. Lately, new HapMap versions have been released where more SNPs are reported in the database, improving the possibilities for better genetic coverage (Frazer et al. 2007).

Over recent years large-scale genotyping technologies have improved (http://www.illumina.com/ and http://www.affymetrix.com/), and also statistical software has become more advanced. However, the tagging approach will still be essential in association studies until whole-genome sequencing becomes routinely available at a low cost. Both study IV and V are based on results from large-scaled studies, respectively a genome-wide nsSNPs screen (**Paper IV**) and an ordinary GWAS (**Paper V**). The SNPs

tested in our studies are tagSNPs from the original studies chosen both because of their significant association to disease, as well as possible impacts in immune pathways. Finemapping or resequencing is needed in future studies to decide whether the associated SNPs are causal variants or SNPs indirectly correlated to the causal variants.

7.4 Typing methods and accuracy

Genotyping errors due to poor DNA sample quality, bad primer/probe design or unfavourable PCR conditions are important issues for all genetic laboratory studies. As shown in the time line for this PhD project, different methods have been used in this thesis (Figure 2B).

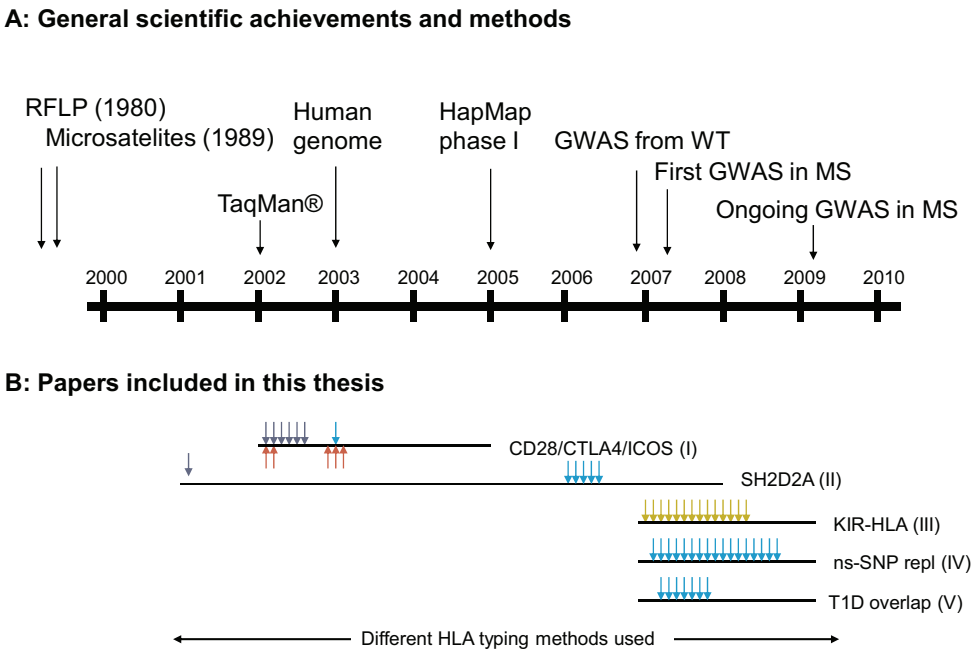

Figure 2 Time line for scientific achievements and methods as well as my PhD project

Abbreviations and information: Each arrow corresponds to each marker investigated, while the colours correspond to typing method; microsatellites (violet), RFLP and DCE (pink), TaqMan® (blue), SSP-PCR (yellow). RFLP: restriction fragment length polymorphism; DCE: denaturant capillary electrophoresis; SSP-PCR: sequence-specific primer - polymerase chain reaction

During the project period there has been a considerable development in the field, both concerning knowledge about polymorphisms, laboratory technologies, publicly available genetic databases and statistical software (Figure 2A). My projects have followed this development, and therefore different methods and approaches have been used in this thesis (Figure 2B).

7.4.1 Microsatellite detection

At the start of the project period (2001-2002), the markers commonly used in genetic studies were microsatellites, both as a tool for linkage and mapping of a candidate gene regions (Nakamura et al. 1987; Weber & May 1989). This was thought to be the optimal approach, since microsatellites are often highly polymorphic. A general worry regarding the genotyping success rate (GSR) as well as the genotyping accuracy, appeared to be an issue for these types of studies (Broquet & Petit 2004). A GSR of 90-95% has previously been classified as a good rate, but it is now recommended that the rate should be a least 98-99%. It was also observed that analyses of "difficult" markers with a low success rate, could be skewed because one or some of the alleles could be more difficult to amplify than others (allele dropouts or "false" alleles). Reference samples and duplicated samples as well as the GSR, give important information regarding genotyping quality. In our studies, the GSR for the microsatellites in **Paper I** (CD28-A, CD28-B, SARA-43, SARA-1, SARA-31 and SARA-47) were >95%. The GA repeat in the *SH2D2A* promoter region (**Paper II**) had a genotype success rate above 94% in all sample sets used. In family data a good quality check is to look for Mendelian errors. This option is not possible in case-control sample sets and makes it even more important to have a high genotyping quality standard. As mentioned under "Quality measurements", page 42-43), both testing for Hardy Weinberg Equilibrium (HWE) and differences in GSR between the cases and controls, are good measurements for quality, which we have used in our studies.

7.4.2 SNP detection

Regarding the SNP detection, different methods have been used in the studies included in this thesis. As discussed in **Paper I**, method accuracy could differ between methods. For instance, the genotyping results from the PCR-restriction fragment length polymorphism (PCR-RFLP) method (Botstein et al. 1980; Wyman & White 1980) deviated much compared with the results performed by denaturant capillary electrophoresis (DCE)

(Fischer & Lerman 1983) or the TaqMan® technology (De Jong et al. 2003; de Kok et al. 2002). The difference in accuracy between TaqMan® and DCE was only 1% compared with 13% between PCR-RFLP and the two other methods. The PCR-RFLP method has more steps and is relying on manual genotype calling, so the risk for errors is much higher than for more automatic techniques. TaqMan® is now a method frequently used, especially if few SNPs are to be genotyped in a moderate sized cohort. TaqMan® technology has become more robust during the past few years and more assays are also pre-tested before sending it to the customers. Most of the TaqMan® assays used in this thesis are pre-tested. Obviously, the accuracy is even more important in small sample sets. If the MAF is low, an error could change the results dramatically. In addition to more precise techniques, the techniques are now more efficient and with a high-throughput potential. For instance, the genotyping format in TaqMan® has been changed from genotyping 96 wells (96-format) in one run, to 384 wells (384-format) in one run. When analysing 384 instead of 96 wells, outliers are easier to detect and the chance for misclassification is therefore lower. For large-scale genotyping, both in terms of more SNPs as well as more samples to be genotyped, SNPlex and GWAS-related technology could be more beneficial, both concerning time and cost.

7.4.3 KIR typing

By using the multiplex PCR sequence-specific primer (SSP) approach the presence or absence of 14 KIR genes encoding inhibitory and activating KIRs was determined in **Paper III** (Sun et al. 2004). This technique was time consuming and based on manually calling of the gels, hence it was important that two independent observers interpreted the gels. Inconsistencies were resolved by discussions or retyping of the samples in question. The KIR typing protocol in our laboratory was also validated by 100% consistency obtained when genotyping International Histocompatibility Workshop cell-line DNAs with known outcome. These cell-lines were also subjected to WGA before KIR genotyping since we used amplified DNA for our cases and controls.

7.4.4 HLA typing

Concerning HLA typing there has also been a change of preferred methods during the last years. We have used both serological HLA typing (by using lymphotype HLA-ABC 120 from Biotest AG (Dreieich, Germany) (Vartdal et al. 1986), the sequence-specific

oligonucleotide (SSO) method (Dynal RELI™ SSO from Invitrogen, Paisley, UK), kits with sequence-specific primers (OLERUP SSP™ from Qiagen, Hilden, Germany) and a sequencing-based approach (Sayer et al. 2004). The most recently developed sequencing-based method is more automatic as well as more specific regarding allele detections and is now the method of first-choice in our laboratory. However, many of our samples have been typed using the other methods prior to this PhD project period. As addressed in **Paper III,** the resolution level has been downscaled and alleles assigned a two-digit level to be consistent between all MS patients and controls, since four-digit resolution (sequenced based method) is only available for a proportion of our sample set. HLA typing is expensive and laborious, and we did not find it necessary to spend restricted resources to retype all our samples using more than one method. Still, many samples have been typed by more than one method, which has served as an excellent quality control of our typing.

The possibility of allele dropouts is a concern when using WGA-DNA. We have been aware of this possible pitfall and have regenotyped 37 of the WGA samples which were genotyped as homozygote using sequencing based HLA typing. These 37 samples showed exactly the same results as with the SSO method. Furthermore, several samples have been genotyped with sequencing-based HLA typing both in WGA and genomic DNA. For the *HLA DRB1* locus, 192 samples were genotyped both using genomic and WGA-DNA and only two samples showed inconsistencies. Also in the TaqMan® assays performed on WGA-DNA samples, we have genotyped some genomic samples as well, to test for allele dropouts regarding the use of WGA-DNA. This test showed 100% consistency between the genotyping of WGA- and genomic DNA.

7.5 Statistical considerations

7.5.1 Quality measurements

In addition to the GSR and the genotyping accuracy, the test for deviations from Hardy Weinberg Equilibrium (HWE) is used as a data quality check. Deviations from HWE among controls at the significance level $\alpha<10^{-3}$, would normally discard the marker from the data set investigated. Usually deviations from HWE are seen when genotyping errors have skewed the genotype distribution, but also extensive inbreeding, population stratification and strong selection (migration or hidden subdivision) could provide deviations. Skewed HWE in the cases could also be caused by a true association to a

disease (Nielsen et al. 1998). It should also be noted that HWE could be due to common deletions, which could influence the PCR amplification resulting in a misclassification of a heterozygote as a homozygote, or segmental duplications (Conrad et al. 2006), which could be of importance in disease (Bailey & Eichler 2006).

In **Paper I** and **II**, all markers were in HWE in the MS cases and controls in all populations tested. In **Paper III**, HWE for the KIR genes could not be calculated, since the KIR genes investigated were only typed for presence or absence of the genes. The HLA loci often deviate from HWE in healthy controls, and strong HLA disease associations cause deviations from HWE among cases. The HWE calculation was therefore of little practical use for this paper. For **Paper IV** and **V**, all included markers were in HWE. One marker was excluded for further genotyping and analysis in study IV (rs11080149), since the SNP showed deviations from HWE.

Another quality test is to test whether the GSR is significantly different between the cases and the controls in a study. By coding all the observed genotypes as one and the missing genotypes as zero, a simple test for association of this variable will show association if the cases compared to the controls have a different number of missing genotypes (Balding 2006). In **Paper IV** and **V** this test was performed. While the top five SNP markers in study IV did not show any significant difference between cases and controls regarding the genotyping failure rate, the associated SNP (rs763361) in study V showed significant difference of missing genotypes between cases and controls. However, the associated SNP marker differed with only 1.5%. The difference was too small to account for the observed case-control difference in the allele frequency.

For large-scale studies, the quality measurements are of great importance. A set of different quality steps are defined. The genotyping call rate for each SNP should be high; >0.95 for markers with MAF >0.1 and even higher if MAF is decreased (>0.99). The SNP markers should also be in HWE. For each individual the total genotyping call for all markers tested should be high (>0.95). DNA quality check, as well as a test for gender is included in the pre-testing of the samples before genotyping. Tests for individual relatedness are performed and if relatedness between samples is identified, only one of the samples is included. Also tests for ethnic ancestry are performed, and individuals with other ethnic ancestry than the rest of the case-control samples are excluded. Finally, the

use of data imputation can replace the missing genotypes with predicted genotypes based on observed genotypes on neighbouring loci (SNPs). These quality control methods are widely used in large-scale studies.

7.5.2 Effect size and statistical power

Over the past few years, it has become clear that most of the MS susceptibility loci outside the HLA complex are of moderate effect size. The search for novel susceptibility loci has to include as many as a hundred common variants (reviewed in Oksenberg et al. 2008; Sawcer 2008). The difficulties to find and even replicate associated loci can be explained by low power of previous studies due to small sample sets. As shown in figure 3 below, for susceptibility loci with a MAF=0.1 and a modest effect size range (OR range 1.1-1.2) a large sample set is essential to achieve necessary statistical power. The power refers to the probability that a study will detect a true association.

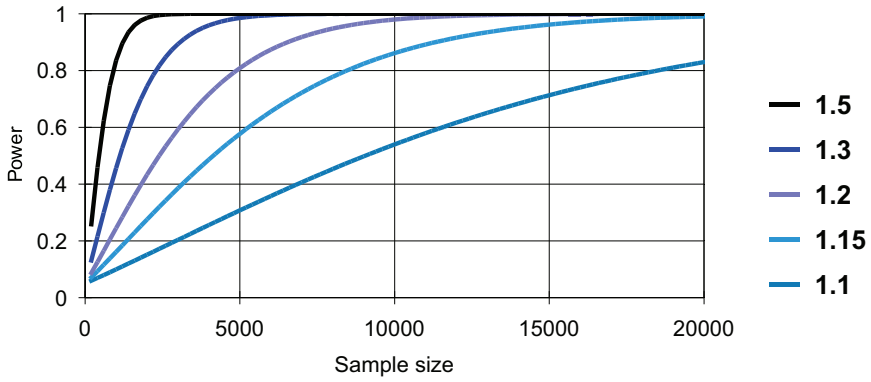

Figure 3: The correlation between power, sample size (2n) and odds ratio (OR) for a marker with MAF=0.1

The coloured lines show the different OR range 1.1-1.5. The significance level α=0.05. The graph is performed in the PS power calculator version 2.1.31 (Dupont & Plummer, Jr. 1990)

Prior to genetic studies, a power calculation should be performed. Many genetic power calculators are freely available, as the PS power calculator by Dupont et al. (Dupont & Plummer, Jr. 1990). Table 4 in the general discussion (page 50), gives an overview of the power in our studies for markers previously investigated. As described in the review by Sawcer based on the formula employed by Risch & Merikangas, the change in power is small when the risk allele frequency (RAF) are above 0.1 if the other conditions are the

same (sample size and effect size) (Risch & Merikangas 1996; Sawcer 2008). As seen in figure 4, for a sample size of 2000 cases and controls, the power decreases when the MAFs (or RAFs) are low. Thus, the sample size has to increase to still have the same power.

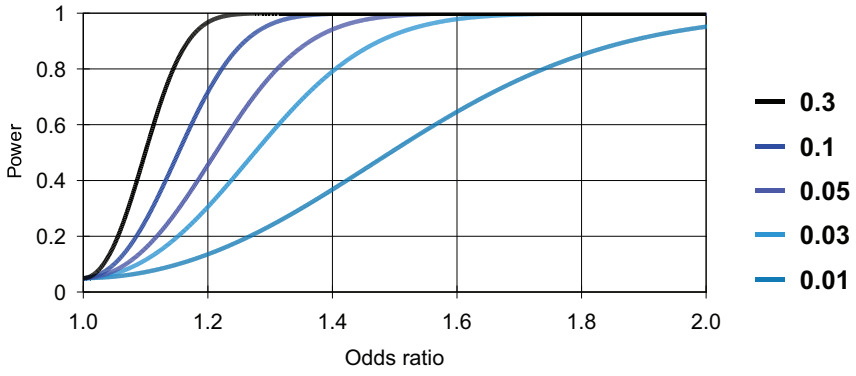

Figure 4: The correlation between power and odds ratio (OR) for a sample set of 2000 cases and controls, when the minor allele frequency (MAF) vary

The coloured lines show the different MAF range 0.01-0.3. The significance level α=0.05. The graph is performed in the PS power calculator version 2.1.31 (Dupont & Plummer, Jr. 1990)

Even when the effect size is higher, as postulated for the CD/RV hypothesis (page 21-22) it may be difficult to detect these susceptibility variations, since the amount of samples needed is high. In **Paper IV**, the MS-associated SNP in the *TYK2* gene had a MAF below 0.05 in the general population. To replicate this finding is a challenge since a very high number of samples are needed.

Note also that for the tagging SNP approach, in both candidate studies as well as GWASs, the chance to detect a disease-causing variant is dependent on the correlation of the tagSNP to the disease allele. Lower correlation between a tagSNP and a causal variant results in less power if all the other conditions are fixed. For poorly tagged causal variants with low correlation to a tagSNP, the sample size has to increase to still reach the same power.

The fact that initial findings tend to overinflate the effect size, could indicate that the effect size found in replication studies may be closer to the "true" effect size, a phenomenon called the "Winner's Curse' phenomena" (Lohmueller et al. 2003). If a sufficiently large replication sample set can not replicate the finding, it may be that the true effect is more moderate than first expected.

7.5.3 Increasing sample sets and meta-analyses

There has been increasing knowledge about the importance of large sample sizes in studies of complex diseases over the last few years, leading to a demand for larger sample collections to be included in genetic studies. Also a more extensive use of meta-analyses in collaborations between research groups has increased the power in many studies. In the papers included in this thesis, we have used both strategies. Even though a high power is a great advantage in a meta-analysis, population stratifications could bias the results as well as dilute associations that may be found in subgroups of patients.

A test for heterogeneity, like a Breslow Day test, is often recommended before combining genetic data from case-control sample sets collected in different populations. The Breslow-Day statistic tests whether different clusters have different ORs and thus is a test of homogeneity. The Cochran-Mantel-Haenszel test combines the ORs between the different sample sets, adjusting for control variables (K) in 2x2xK stratified tables. One advantage with this statistical approach, is that each population-specific set of samples is treated as a separate stratum. This allows the RAF to vary without influencing the total effect size, as long as the trend for associations in all populations goes in the same direction.

If heterogeneity is seen between the populations investigated, this can often be explained by true differences in population structures. The patient collection could also be biased, in terms of being different for variables like disease course, sex or other clinical parameters. The diagnostic criteria should be generally agreed on and be easy to follow, to restrict the misclassification to a minimum. As pointed out by Seldin & Amos, case-control samples may be overlapping between study reports and thus the studies may not be independent replications of each other (Seldin & Amos 2009). Such studies may also be difficult to include in meta-analysis. Furthermore, the authors discuss the use of the same "general controls" sample sets available from studies of different diseases. They say, that in large-

size control sample sets the controls probably reflect the general population; however, cases and controls should be scored according to genetic background.

7.5.4 Significance level and multiple testing

When testing the significance of a genetic association, hypothesis testing is performed. Two hypotheses are suggested; the null hypothesis (H_0) states that there are no differences between the cases and controls, while the alternative hypothesis (H_1) states that there are differences between the two groups. If a significant difference is found between the groups, the H_0 hypothesis is rejected. Traditionally, with a P- value below 0.05 it has been acceptable to reject H_0. However, 5% of the tests will by chance have a P-value under this limit, resulting in a type I error (false positive finding). In this case, the H_0 hypothesis is incorrectly rejected (Plenge & Rioux 2006). Therefore, when testing many genetic markers in a study, a "multiple testing problem" occurs. Fore instance, when testing 20 markers for association, one marker will just by chance be significantly associated, but is a false positive finding (Balding 2006).

Correction for multiple testing is therefore required, and different approaches could be used. The most used method is the Bonferroni correction, which corrects for multiple testing by multiplying the nominal P-values with the number of testes performed in the study. This correcting method has been used in **Paper I-III**. However, it can be argued that this correction strategy is too conservative and overestimates the number of independent hypotheses. For example, both in **Paper I** and **II**, small gene regions were investigated (262kb and 18kb, respectively) and since many of the markers within a gene region are in LD, the statistical tests performed are not fully independent. The weight-free Bonferroni correction may therefore be too strict in such studies, and one may in this way miss true positive results. In study II we chose to replicate only one of the tagSNPs in the Nordic MS samples since, most of the tagSNPs tested in the Norwegians were in strong LD with each other (see Figure 2 in **Paper II**). Thereby we restricted the number of statistical tests performed. Also by grouping multi-allelic markers the numbers of comparisons can be minimized. For instance, the *SH2D2A* GA repeat marker was grouped in to GA_{16} and GA_{other} in study II.

An alternative to the weight-free Bonferroni correction approach is the Bayesian correction approach, which incorporates the prior probability of a hypothesis. This

approach will favour a true association, by taking into account prior knowledge regarding the loci. By using this approach less stringent P-values could be considered as significant (Wacholder et al. 2004).

There has been much debate concerning the significance level for GWASs. The large GWAS performed by WTCCC recommended a Bayesian approach and suggested that a P-value at 10^{-7} should be regarded as significant in GWAS (WTCCC 2007). This significance level accounts for the very small prior probability that any given locus or region is truly associated with disease (Wang et al. 2005). But it is also quite clear that associations with a lower degree of significant levels could represent real effects, and that this strict threshold will cause false negative findings.

7.5.5 The need for replication of true associations

Even though a study has sufficient statistical power, independent replication is often regarded as the final step to confirm or refuse the claimed association. Most large studies (as GWASs) therefore apply a two stage-approach; the original screening and an independent replication. If an association is true, it should be replicable in a large, independent sample set.

Until recently, most candidate studies have not been independently replicated. This may be due to lack of power in the replication study or a false positive finding in the original study. However, also technical artefacts leading to genotyping errors, inappropriate matching of cases and controls and clinical heterogeneity across study populations could dilute a true association. Also failure to correct appropriately for multiple testing in the initial study may explain failures in replication.

Conversely, under-powered studies with negative results do not rule out that there are genetic variants which influence disease in the studied gene region. As pointed out by Hafler et al. in a commentary to the first published MS GWAS, under-powered, promising candidate loci may have been wrongly rejected and may ultimately prove to be genuine (Hafler et al. 2007). In conclusion, well-powered studies are needed to exclude or confirm true susceptibility genes.

8 General discussion

In this thesis, the studies have been based on two different approaches. The three first studies (**Paper I-III**) are based on a candidate gene approach, while **Papers IV** and **V** are based on previously published genome-wide screens, a non-synonymous SNP screen (Burton et al. 2007) and a GWAS (Todd et al. 2007), respectively. MS researchers were enthusiastic in the decades after the early identification of the HLA association (Jersild et al. 1972; Naito et al. 1972), but experienced series of disappointments when attempting to explore true MS susceptibility loci outside the HLA complex. Our understanding regarding the effect size of MS-associated loci has changed dramatically during this PhD period, which is reflected in a recent trend of large collaborative studies in order to increase the power to detect associations. After the "GWAS revolution" the genetics of MS is proven to involve more than the HLA complex. The novel susceptibility genes found to be associated with MS so far are involved in immunological pathways. This supports the notion that MS as an immune-mediated disease, share immunological pathways with other immune-mediated diseases.

8.1 The difficulty of replicating claimed associations

In both **Paper I** and **II** the aims of our studies were to investigate in more dept candidate gene regions, which we had previously shown to be associated with MS (Dai et al. 2001; Harbo et al. 1999). Both the *CD28/CTLA4/ICOS* gene region (**Paper I**) as well as the *SH2D2A* gene (**Paper II**) encode proteins of importance for T-cell regulation (Granum et al. 2006; Rudd & Schneider 2003). In mice models, both mice lacking the CTLA4 or TSAd molecule, show spontaneous development of autoimmunity in the form of lethal autoimmune multiorgan destruction (Tivol et al. 1995; Waterhouse et al. 1995) or lupus-like autoimmune disease (Drappa et al. 2003). These genes were therefore found to be good candidates for further investigation.

However, the loci aimed to study appeared to have small or moderate risk effects. The knowledge about the modest effect size of MS-associated loci and thus the need for large sample sizes to detect these effects, has come gradually. In **Paper I** and **II** the Norwegian case-control sample set was of similar size as our original studies, which had shown association to MS in approximately 300 cases and 300 controls. Looking back, this

replication size is now considered to be relatively small. As shown in table 4, the power of the Norwegian replication set in study I can now be calculated to be only 52% for the *CTLA4* +49 marker. The table gives an overview of the power in all the studies in this thesis for markers previously investigated by others. The power calculations are based on effect sizes (ORs) and the RAFs in the controls reported by others (see references in the table) prior to our studies as well as the sample sizes of our studies.

Table 4: Calculated power for our sample sets for the markers investigated based on odds ratios (ORs) and risk-allele frequencies (RAFs) reported in previous studies as listed in the table

Study	Previously investigated markers	OR	Ref	RAF	Cases / controls in our studies	Power in our studies
I	*CTLA4* +49 (rs231775)	1.3	1	0.65	273 / 269	**52%**
	CT60 (rs3087243)	1.15 / 1.5	2	0.53	575 / 551	**38 / 100%**
II	$GA_{16/other}$ repeat	1.3	3	0.19	1815 / 1727	**99%**
	rs926103	1.2	4	0.33	1504 / 1442	**92%**
III	HLA-Bw4 $^{-/-}$	2.0	5	0.45	631 / 555	**100%**
	HLA-Bw4 $^{-/}$ stratified				233 / 389	**78%**
IV	*TYK2* (rs34536443)	1.3 / 1.7	6	0.96	5213 / 4453	**92 / 100%**
V	*CLEC16A* (rs12708716)	1.2	7	0.78	5737 / 10296	**100%**
	CD226 (rs763361)	1.2	7	0.47	5737 / 10296	**100%**

Abbreviations and information: For study I (*CTLA4* +49) and II (*SH2D2A* promoter $GA_{16/other}$ repeat) only half of the Norwegian sample set was a replication sample set, not genotyped previously. In study II, the rs926103 data in the Norwegian sample set was used as reference for the power calculation for the other Nordic samples.

RAF: risk allele frequency. The HLA-Bw4 $^{-/-}$ was stratified for HLA-DRB1*1501. The power was calculated using the PS power calculator version 3.0.4 in allelic or phenotypic models (Dupont & Plummer, Jr. 1990). Note that the power can be slightly different if calculated using other genetic calculators. Significance level was set at α=0.05

References: 1) = (Harbo et al. 1999), 2) = (Ueda et al. 2003) investigated AITD OR=1.5, and T1D OR=1.15, 3) = (Dai et al. 2001), 4) = (Lorentzen et al. 2008) (in this thesis), 5) = (Karlsen et al. 2007) investigated primary sclerosing cholangitis, 6) = nsSNP screen from WTCCC (Burton et al. 2007), 7) = (Todd et al. 2007) investigated T1D

The first reported associations tend to overinflate the effect size, a phenomenon called the "Winner's Curse' phenomena" (Lohmueller et al. 2003). This makes it difficult to replicate a finding in a sample set of similar size to the original sample set, as in our case.

If a sufficiently large replication sample set can not replicate a finding, the true effect might be more moderate than first expected. Then the replication study will show a false negative finding. On the other hand, the first study could report a false positive finding, while the replication shows the true negative result.

In **Paper I** and **II**, in addition to trying to replicate or refute markers associated in previous studies, more markers in the specific gene regions were genotyped to try to pin-point the associations more carefully. For the *CTLA4* gene region (**Paper I**), no associations were found neither to our previously associated marker *CTLA4* +49 (rs231775), nor to new markers genotyped in the region. Especially the SNP CT60 (rs3087243) was of interest, since this marker had shown association with both T1D and AITD, and was believed to be associated with differences in the levels of soluble versus membrane-bound CTLA4 molecules (Ueda et al. 2003). It was suggested that the *CTLA4* gene increased the susceptibility to immune-mediated diseases in general. It was also hypothesized that the CTLA4 pathway could be of general interest in autoimmunity. However, our study of Norwegian MS cases and controls did not find any significant association in the *CTLA4* gene region including the CT60 SNP. This is in line with recent publications (Bagos et al. 2007; Greve et al. 2008; Roxburgh et al. 2006). Interestingly, Barcellos and colleagues, who studied MS patients and their families with the focus on immune-mediated diseases in general, found an association to the CT60 SNP only in MS patients from families with increased prevalence of immune-mediated diseases (Barcellos et al. 2006a). This supports the idea of general susceptibility loci in autoimmune diseases.

Similarly, for **Paper II**, no convincing association was shown for the previously claimed associated marker (the *SH2D2A* promoter GA repeat polymorphism) in a new sample set of Norwegian MS cases and controls. We then genotyped more SNPs mapping the gene region (five tagSNPs covering approximately 17.5kb gene region; rs2768764, rs1800600, rs926103, rs909200, rs2768766). The tagSNP closest to the GA repeat polymorphism in the promoter region showed the strongest association in the total Norwegian sample set (rs1800600; OR=1.3 (1.1-1.5) and P=0.01), when including both new and previously typed samples. The *SH2D2A* study was then extended to include Nordic sample sets (Danes, Swedes and Finns) to further increase the power of the study (see Table 4, page 50). These samples were genotyped for the GA repeat polymorphism as well as one of the five tagSNPs, the nsSNP rs926103 in exon 3 of the *SH2D2A* gene. Only one tagSNP was

chosen to restrict the analysis, since the LD was strong between the SNPs, and an nsSNP may directly explain a biological effect. The LD between the two tagSNPs rs926103 and rs1800600 was strong (D'=0.99 and r^2=0.82). A Cochran-Mantel-Haenszel test combining the OR between the populations (Norwegians, Danes, Swedes and Finns) showed OR=1.15 (1.04-1.28) and P=0.007 for the GA_{16} allele in all sample sets combined, and OR=1.1 (0.99-1.24) and P=0.09 in the independent replication set alone. This shows that the GA_{16} allele can not be excluded to be of importance for MS susceptibility, albeit with a weak effect.

Hundreds of SNPs are located in the *CD28/CTLA4/ICOS* gene region (262kb) and above fifty SNPs are located in the *SH2D2A* gene region (18kb). Thus, in **Paper I** and **II** we were unable to genotype all variations in the two gene regions due to high financial costs. Therefore we chose some markers that were thought to capture some of the variations in the gene regions through LD. In the *CD28/CTLA4/ICOS* study we used six microsatellites which are in general often more informative than SNPs due to a high variation in addition to three SNPs. These two projects were initiated when the databases for choosing genetic markers were more limited than today. In addition, the technology has later progressed markedly. Thus, if the study had been initiated today, with current knowledge of effect size and power, and having the advantage of improved tagging tools and reduced genotyping costs, our strategy would probably have been slightly different. However, the most likely causal variant in T1D and AITD in the *CTLA4* gene region, the CT60 SNP (Ueda et al. 2003), was genotyped in our study and showed no association in MS, in line with other publications (Greve et al. 2008). In the *SH2D2A* study (**Paper II**) a tagging approach was performed. An advantage with a tagging strategy is that the cost as well as the laboratory work could be decreased. When defining the range of the *SH2D2A* gene region wider than in our study (25kb), three to five novel tagSNPs are now suggested by the tagging tool program Haploview version 3.32 (http://www.broad.mit.edu/mpg/haploview/). This exemplifies that we can not rule out that other variants not properly tagged in study I and II are associated with MS, an that our studies may have lost these signals.

In **Paper II**, a prominent finding was the differences in the allele frequencies between the Nordic controls (Norwegian, Danish, Swedish and Finnish) used in the study. Among the Nordic MS cases the allele frequencies were relatively similar (for both the GA repeat polymorphism and the SNP rs926103), while frequencies among the Nordic controls

differed more. However, the Breslow-Day test for homogeneity of the OR was not significant (P=0.09 and P=0.4 respectively for the two markers), meaning that a combined analysis still could be performed. The Danish controls carried the GA_{16} allele (0.21 allele frequency) and rs926103*A allele (0.35 allele frequency) with similar frequencies as the Norwegians (0.20 and 0.33, respectively), while the Swedes and Finns differed more (0.25 and 0.40 in Swedish controls, while 0.24 and 0.40 in Finnish controls, respectively). This allele frequency difference between the Norwegian and Swedish controls was surprisingly large. The Finnish population on the other hand, could be expected to have a somewhat different genetic structure due to population bottlenecks (Service et al. 2006). Also the fact that the Finnish sample set was half the size of the Swedish set makes the frequency estimates among the Finns more uncertain. For comparison, the frequency of the rs926103*A allele was 0.36 among UK MS patients and 0.35 among UK controls (OR=1.07 (0.95-1.21) and P=0.2, calculated in EpiInfo version 5) in the nsSNP screen performed by the WTCCC (Burton et al. 2007). This is in line with our Norwegian and Danish data, supporting our conclusion, that we can not exclude a weak effect of this locus in MS.

The problem with control frequencies that deviate and drive an association in a significant direction has also been observed in other studies. For instance, the *KIF1B* association in MS was first reported by Aulchenko and colleagues (Aulchenko et al. 2008). In this paper, the *KIF1B* gene was studied in a small isolated cohort (45 MS samples and 195 controls), as well as replicated in three independent case-control sample sets. Interestingly, when other groups, ourselves included, genotyped the claimed associated SNP (rs10492972), no association was found. A total of approximately 5000 cases and 5000 controls have so far been genotyped for this SNP within the IMSGC without replicating this association (S Sawcer, personal communication). The most prominent difference between the sample sets in the paper by Aulchenko et al. compared to the IMSGC replication is that the allele frequencies were different among the controls in the two studies. In the paper by Aulchenko et al. the allele frequency in the combined controls was 0.27, while it was 0.32 in the IMSGC controls. Also in these two studies the Norwegian and Swedish controls differed in allele frequency (0.32 among Norwegians in the IMSGC replication versus 0.29 among Swedes in the Aulchenko paper). Many of the Swedish controls used in the *SH2D2A* study were also included in the *KIF1B* study. Therefore, we cannot rule out that differences in the control sets from Sweden and

Norway could bias also our *SH2D2A* study and drive our combined analysis in a false negative direction.

The WTCCC GWAS, looked into geographic differences for the genetic markers investigated (WTCCC 2007). The study was performed by genotyping 2000 individuals from seven different diseases (in total 14 000 individuals) and a shared control sample set of 3000 healthy individuals. All individuals were from the UK. The main finding regarding the population structure among the UK samples, was that the differences were small and that none of the disease-associated SNPs were located in regions with large population variations. Thirteen regions showed strong geographical variations, and the predominant pattern was a north-south gradient in the SNPs allele frequencies. However, none of the genomic regions identified with geographic variations in this GWAS where located close to the *SH2D2A* gene region studied in **Paper II**.

8.2 Association to the HLA complex - not only the *HLA-DRB1* gene

The discovery of the HLA molecule's function has been of great importance. In the 1960's and 1970's it was observed that differences in the HLA molecules between a donor and a recipient could cause rejection of organ transplants. At the same time as the transplant immunology developed, genetic associations to specific HLA loci were found for different immune-mediated diseases (reviewed in Klein & Sato 2000). Later, it was shown that especially HLA class II genes carry a strong risk for immune-mediated diseases. Different immune-mediated diseases show different primary HLA loci associations and also different allele associations (reviewed in Fernando et al. 2008; Thorsby & Lie 2005). The HLA allele most frequently shared in autoimmunity is HLA-DRB1*03. It has been suggested that both common and disease-specific pathological mechanisms are involved in the HLA associations. The strong HLA associations for most immune-mediated diseases have convincingly been replicated when applying the GWASs approach. In the GWAS performed by the WTCCC in 2007, the most strongly associated SNPs in T1D and RA were located in the HLA complex ($P<10^{-115}$ and $<10^{-26}$, respectively) (WTCCC 2007).

The fact that the strongest genetic associations are located in the HLA complex for most immune-mediated diseases, indicates that HLA gives the largest genetic contribution in these diseases. In MS it has been proposed that 20-60% of the genetic variation could be

explained by the HLA complex (Haines et al. 1998). The tight LD between the MS-associated HLA-DRB1*1501 and HLA-DQB1*0602 alleles has been well-known for more than a decade. In addition to effects of other HLA-DRB1 alleles (Barcellos et al. 2006b; Dyment et al. 2005), there are also effects from genes outside the HLA class II region. The HLA class I genes are good candidates for an additional HLA effect in MS. Traditionally, the HLA class I genes have been investigated in light of their ability to present antigen to TCRs on T cells. But the HLA class I (HLA –A, -B,-C) variants have also been shown to serve as ligands for inhibitory and activating KIRs expressed on NK cells (Vivier et al. 2008) and some subsets of T cells (van Bergen J. et al. 2004). In **Paper III**, we studied whether HLA class II independent associations could be detected for HLA class I variants relevant to their KIR binding capacity. Our study (Paper III) is the first study specifically addressing this in MS. Our main finding was a significantly reduced frequency of the HLA-Bw4 specificity in MS patients compared with controls. All *HLA-B* alleles can be grouped into the two groups; HLA-Bw4 and HLA-Bw6. This is described in detail in Paper III.

The gene encoding HLA-Bw4 is associated with protection from MS in our study. Figure 2, panel A shows the HLA-Bw4 molecule as as a ligand for KIR3DL1. The interaction of HLA-Bw4 with KIR3DL1 will lead to an inhibitory signal limiting the activation of the NK cell. Panel B shows the HLA-Bw6 molecule, which is not known to be a ligand for any inhibitory KIR receptors. Hence the HLA-Bw6 does not bind to KIR3DL1 and the balance is altered in favour of activation of the NK cell. The frequency of patients homozygous for the HLA-Bw6 group of alleles (i.e. lacking the HLA-Bw4 group) was significantly higher in MS patients compared to the healthy controls in our study (OR=1.74 (1.37-2.20) and $P=4.6x10^{-6}$). However, since some of the alleles in the HLA-Bw6 group are situated on the HLA-DRB1*1501 and HLA-DRB1*0301 haplotypes, it could be argued that the *HLA-B* effect seen in our study, simply mirrored the *HLA-DRB1* effect. After removing the susceptibility alleles DRB1*15 and DRB1*03 and the possibly protective DRB1*01 allele from the statistical analysis, the observed HLA-Bw6 association was still significant (OR=2.22 (1.35-3.64) and P=0.002). Thus, the *HLA-B* observed was shown to be independent of known HLA class II associations in MS.

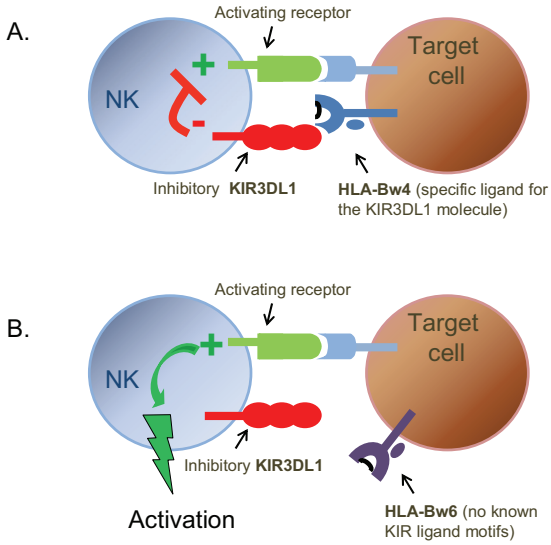

Figure 2: Schematic figure of how the MS associated HLA-B KIR ligand may influence natural killer cell activation. The figure is from Paper III

When grouping the *HLA-A* and *HLA-C* loci according to the KIR ligands properties, association to neither *HLA-A* nor *HLA-C* was found. The two loci have both been suggested to add to the genetic risk in MS in addition to *HLA-DRB1* (Fogdell-Hahn et al. 2000; Harbo et al. 2004; Yeo et al. 2007). Recently, both Brynedal et al. and Yeo et al. demonstrated that after controlling for HLA DRB1*1501, significant associations in MS were found to HLA-A*02 (Brynedal et al. 2007) and HLA-Cw*05 (Yeo et al. 2007).

The HLA class I and II regions are very polymorphic, and the *HLA-B* locus is the most polymorphic seen in humans. Extensive research has investigated the HLA complex carefully. There is an extensive LD in the region, and haplotype blocks have been defined. The blocks are very variable in length, depending on which alleles at each HLA locus are included (Blomhoff et al. 2006). The HLA-DRB1*1501 and *0301 alleles are both on conserved ancestral haplotypes (AH) (AH 7.1 and AH 8.1, respectively) (Alper et al. 2006; Thorsby & Lie 2005). Both AH 7.1 and AH 8.1 are risk haplotypes for MS (AH 7.1; HLA A*03-Cw*07-B*07-DRB1*1501-DQB1*0602 and AH 8.1; HLA A*01-

Cw*07-B*08-DRB1*0301-DQB1*0201). These haplotypes are common in Europeans, and especially AH 8.1 is also found to be associated in many immune-mediated diseases, for instance T1D (Thorsby & Lie 2005).

The extensive LD in the HLA complex has also made it difficult to conclude which variants are causal in MS. In a SNP screen of the HLA complex in MS no other risk loci than *HLA-DRB1* were identified (Lincoln et al. 2005). However, the gene region was not explored comprehensively in this study. The study had a high genotyping error rate (40% drop out of SNPs attended to be genotyped) and some regions were poorly tagged (for instance few tagSNPs in the HLA class I region). This combined with the fact that HLA loci are multi-allelic, in contrast to SNPs which are di-allelic, may explain why this SNP screen could not detect the HLA class I associations (Lincoln et al. 2005). Thus, there are still unanswered questions regarding associations in the HLA complex.

The 1000 Genomes Project, a deep catalogue of human genetic variation, was recently initiated (http://www.1000genomes.org/). The intension is to create the most detailed and useful picture of human genetic variation to date. The project involves sequencing of the genomes of approximately 1200 people from around the world, HapMap samples included. The project also includes classical HLA class I and II typing. The data are made available through freely accessible public databases. With this new tool, combining sequencing and classical HLA typing data from many populations, the HLA gene region will hopefully be easier to explore.

Recently, an editorial by Trachtenberg was published in Annals of Neurology (http://www3.interscience.wiley.com/journal/76507645/home) with the title "Understanding the Role of Natural Killer Cell Receptors and their HLA Ligands in Multiple Sclerosis" as a comment to our KIR-HLA paper (**Paper III**). Based on our findings Trachtenberg suggests, that a deficit in HLA-Bw4 could decrease the function of NK cells, resulting in a diminished response to infectious diseases and increased susceptibility to MS. Furthermore, the author points out that HLA-Bw4 homozygous individuals have a strong protection against HIV infection compared to individuals that are HLA-Bw4/Bw6 heterozygous or HLA-Bw6 homozygous (Flores-Villanueva et al. 2001). A recent paper studying cancer development in MS patients showed an overall reduction in cancer risk for MS patients (Bahmanyar et al. 2009). This finding strengthens

our hypothesis that the KIR-HLA interaction is involved in MS pathogenesis. Interestingly, it has been shown that the ligands for inhibitory KIRs are important to "educate" the NK cells to act more rapidly (also called licensing). In a study by Kim et al. they showed that the responsiveness to tumor stimulation in KIR3DL1^{+} individuals was increased in individuals homozygote for HLA-Bw4 (Kim et al. 2008). In contrast, individuals lacking KIR3DL1, showed no differences to tumor stimulation. From our point of view, it is interesting that this association is found on the HLA-Bw4 and thus the ligand side of the interaction. However, both the HLA and the KIR genes are highly polymorphic and the mechanisms for possible involvement in MS need to be further investigated.

8.3 The killer immunoglobulin-like receptor (KIR) genes in MS

The KIR genes in combination with their corresponding ligands have been shown to be of importance in several diseases, for example cancer, infectious diseases, autoimmunity and pregnancy complications (reviewed in Kulkarni et al. 2008; Parham 2005). We were the first to investigate possible involvement of the KIR genes in MS (**Paper III**). A lower gene carrier frequency was seen for the inhibitory *KIR2DL1* gene among MS patients compared with the healthy controls (95.7% versus 98.4%, P=0.01), but after Bonferroni correction for multiple testing, no significant difference was seen ($P_{corrected}$=0.1). More interestingly, our data indicate a possible association between severity of MS disease, measured by MSSS, and the presence of *KIR2DL2*. MS patients carrying the inhibitory *KIR2DL2* gene had a higher median MSSS (5.61; range 0.13-9.99, n=204) compared with MS patients not carrying *KIR2DL2* (MSSS=4.55; range 0.13-9.98, n=254), indicating a more severe disease (P=0.009). These findings did neither remain statistically significant after correction for multiple testing ($P_{corrected}$=0.3), but it can be argued that applying Bonferroni correction is too conservative since the KIRs genes are in strong LD. Nevertheless, the observation merits further study and replication.

The KIR molecules are thought to be important in balancing the activity of the NK cells. Schematically, two main subsets of NK cells exist. The cytotoxic (CD56^{dim}) NK cells are important components of the innate immune system and involvement in elimination of virus-infected and malignant cells. The subset producing cytokines and chemokines (CD56^{bright}) may act as a bridge towards the adaptive immune system (Vivier et al. 2008). Recently, there has been a renewed interest in NK cells related to MS development

(Lunemann & Munz 2008). A recent study by De Jager et al. investigated the cytometric profiles of circulating blood cells in RR-MS, CIS and healthy controls (De Jager et al. 2008). A reduced frequency of the NK cell profile corresponding to the cytokine producing NK cells ($CD56^{bright}$ cells) was found among RR-MS and CIS patients compared with the healthy controls. The finding is supported by others who found that reduced MS disease activity during the last trimester in pregnancy, was associated with a significant increase of circulating $CD56^{bright}$ NK cells in blood (Airas et al. 2008). Simultaneously, the proportion of circulating $CD56^{dim}$ NK cells was clearly reduced. Also studies of MS treatments (IFN-β and daclizumab, a IL2Rα target treatment (Zenapax®)) have examined the $CD56^{bright}$ NK cell proportion in peripheral blood and found an expansion of these cells in the blood of RR-MS patients following these treatments (Bielekova et al. 2006; Saraste et al. 2007). Bielekova et al. could also show a relationship between the daclizumab-induced expansion of $CD56^{bright}$ NK cells and its therapeutic effect. These studies give important evidence for a key role for NK cells in immunoregulatory pathways in MS pathogenesis and progression.

8.4 GWASs and replicated findings

The year 2007 has already been mentioned as a milestone in MS genetics as well as other complex diseases (WTCCC 2007) due to the introduction of GWASs. In July 2007 the first GWAS was published in MS (Hafler et al. 2007). When it was realized that the linkage approach did not have the power to detect MS-susceptibility loci with modest effect sizes, the association approach was chosen. Association studies have more power to detect modest risks, and the case-controls sample sets are also easier to obtain (Risch 2000). Finally, novel associations were found with high degree of significance, however, the effect size was as expected of modest for each locus identified outside the HLA complex. In the following years many of the MS-associated SNPs identified in the first MS GWAS have been replicated and validated. The list of replicated and validated gene regions associated with MS is now constantly growing (see Table 1, page 29 for the list of references to MS GWASs as well as replication studies). Out of the 13 suggested loci (outside the HLA complex) in the 2007 MS GWAS, seven of them have now been replicated (*IL2Rα, IL7R, CD58, CLEC16A, RPL5, FAM69A, EVI5).* Three of these genes; *RPL5*, *FAM69A* and *EVI5* are located close to each other at chromosome 1p22 and could cause the same signal. Both **Paper IV** and **V** have contributed to this list of validated

SNPs. Without doubt, the large sample sizes in study IV and V have been of great importance for the ability to replicate associations. Even though many research groups have increased their sample sizes through new collections (the Oslo MS Genetics Group as well), collaborations are needed to achieve the number of patients and controls included in study IV and V (see Table 2 and 3 page 34). Advantages of collaborative projects are not only increased sample sizes, but also interplay between researchers with different expertise.

The nsSNP screen by WTCCC showed evidence of association for in total 21 nsSNPs in the MS dataset, but the screen did not include a MS replication sample set (Burton et al. 2007). Thus, independent replications of the findings were needed. Since our collaborators in UK had been involved in the original nsSNP screen, we had an advantage of having early access to the results. This formed the basis of **Paper IV**, where a two-stage approach was applied. In the first stage, samples from Belgium, Norway and UK were genotyped for 17 nsSNPs picked among the most strongly associated SNPs in the original screen (both based on MS results alone as well as combined autoimmune disease results). NsSNPs in interesting pathways were favoured even though their P-values were not among the lowest in the data sets. In the second stage, the five most strongly associated SNPs from stage one were genotyped in samples from Australia, Sweden, UK and US. Only a few SNPs were chosen in this stage. Furthermore, a combined analysis including the two replication sample sets as well as data from the original WTCCC screen was performed.

The most strongly associated nsSNP in our study (**Paper IV**) was the SNP (rs6897932) in the *IL7R* gene at chromosome 5q13. This SNP was also recently identified by using a candidate gene approach (Gregory et al. 2007; Lundmark et al. 2007) as well as a genome-wide approach (Burton et al. 2007; Hafler et al. 2007), so this SNP proved to be a positive control for our replication study. This illustrates that screens only based on nsSNPs are capable of identifying genuine associations (Burton et al. 2007). Interestingly, the associated *IL7R* SNP (rs6897932, C/T) is located in exon 6 and codes an amino acid change from tyrosine to isoleucine (T244I). This amino acid change has shown functional effect on gene expression (Gregory et al. 2007). Two splicing variants exist, either a soluble form of the protein (skipped exon 6) or a membrane-bound IL7R (exon 6 included). The protein is a trans-membrane protein, expressed in T cells. The MS-

associated C allele results in increased soluble form of the protein, which may reduce the function of the protein. In MS patients elevated levels of IL7R and IL7 mRNA in CSF have been found (Lundmark et al. 2007). This supports that the IL7R molecule is involved in MS pathogenesis.

The second most strongly associated nsSNP in the combined analysis in **Paper IV**, was the SNP in the tyrosine kinase 2 gene (*TYK2)*. The nsSNP (rs34536443, C/G) codes an amino acid change from proline to alanine (P1104A) in exon 21. This SNP was found associated in the combined analysis between the three immune-mediated diseases in the nsSNP screen performed by WTCCC (AITD, AS and MS, P-value at 10^{-4}) and was therefore selected in our study. In the original screen the P-value for the comparison between MS patients and controls was 0.004. Also in systemic lupus erythematosus (SLE), associations to the *TYK2* gene have been shown (Cunninghame Graham et al. 2007; Sigurdsson et al. 2005; Suarez-Gestal et al. 2009). However, the SLE studies have not genotyped the MS-associated SNP, but other SNPs close by. The strongest SNP signal in SLE was found to the nsSNP rs2304256 (V362F) (MAF=0.27-0.32 in unaffected controls) (Sigurdsson et al. 2005; Suarez-Gestal et al. 2009). When looking at the genotypes for our MS-associated SNP (rs34536443) in the CEU HapMap samples (http://www.hapmap.org), only two individuals among 90 CEU samples has the C/G genotype, the rest the G/G genotype. This illustrates the low MAF for the MS-associated SNP. In the CEU samples the calculated LD between the MS-associated SNP (rs34536443) and the SLE-associated SNP (rs2304256) is r^2=0.044 (Haploview version 4.1, http://www.broad.mit.edu/mpg/haploview/). Thus, there is no evidence for LD between the MS SNP and the SLE SNP. However, the SNPs in the *TYK2* gene reported in the HapMap database (HapMap Release 27, Phase II + III, February 09, on NCBI B36 assembly, dbSNP b 126), seem so far to be in relatively strong LD (measured by D'). The SNP database at NBCI (http://www.ncbi.nlm.nih.gov/) shows that 329 SNPs are located in the *TYK2* gene and 23 of them are nsSNPs. It is still unclear whether the MS-associated *TYK2* SNP is the causal variant or indirectly associated. It should also be mentioned that a second nsSNP (rs35018800, A928V) in the *TYK2* gene was associated in the combined analysis between the immune-mediated diseases in the nsSNP screen (AITD, AS and MS, P-value at 10^{-6}). However, the MAF for this ns-SNP was only 0.01 and we did not genotype this SNP in our replication study.

The *TYK2* gene is a promising candidate gene for immune-mediated disease, and it encodes the TYK2 protein which is part of the Janus kinase family of non-receptor protein tyrosine kinases (Ghoreschi et al. 2009). The receptor is involved in the pro-inflammatory immune responses and has a crucial role in signal transduction for a wide range of cytokines (Shaw et al. 2006). Mutations leading to loss of the TYK2 function give severe combined immunodeficiency, with reduced Th1 and increased production of Th2 cytokines (Minegishi et al. 2006; Watford & O'Shea 2006). The MS-associated SNP in the *TYK2* gene has a major allele frequency of 95.3% in our combined control population and increases the MS risk with an OR of 1.3. To detect associations with a low frequent minor allele (MAF<0.05) the sample size needs to be very large, as discussed in the methodological considerations page 44-46. From a functional point of view, Nejentsev and colleagues found by resequencing the interferon induced with helicase C domain (*IFIH1*) gene, that associations to rare variants consistently protected from T1D (Nejentsev et al. 2009). Associations to *IFIH1* alleles carried by the majority of the population (common variants) predisposed to disease. They further suggested that these rare risk variants were selected negatively due to disrupted antiviral function, rather than selected by their protection against T1D. This could also be the case for the MS-associated SNP in the *TYK2* gene, as the gene has role in the immune function.

In **Paper V**, two MS susceptibility loci were confirmed. The strongest signal was the association to a SNP in the C-type lectin domain family 16, member A (*CLEC16A)* gene (alias *KIAA035*0) located at chromosome 16p13. All the seven SNPs investigated in **Paper V** were selected based on previous reported associations to T1D (Todd et al. 2007). Our hypothesis was that the two diseases had some degree of overlap regarding the genetic predisposition to immune-mediated diseases (Zhernakova et al. 2009). The *CLEC16A* SNP rs12708716 (intron 18) was previously found associated to T1D (Hakonarson et al. 2007; Todd et al. 2007). In the paper by Todd et al. the exons in the *CLEC16A* gene were sequenced without finding support for novel SNPs which could explain the association to T1D better than the rs12708716 SNP (Todd et al. 2007). In comparison, in the first MS GWAS, a SNP (rs6498169) in intron 21, of the *CLEC16A* gene was associated with MS (Hafler et al. 2007) and was later replicated in an Australian case-control sample set (Rubio et al. 2008). A study by Zoledziewska and colleagues found association to another SNP (rs725613) in the *CLEC16A* gene both among T1D and MS patients (Zoledziewska et al. 2009). The latter study was performed in Sardinia,

where the co-occurrence of T1D and MS is common. The associated SNP (rs725613) in the paper by Zoledziewska et al. was, in perfect LD (r^2=1) with the "T1D-SNP" (rs12708716) identified by Todd and colleagues (Todd et al. 2007). Between the "MS-SNP" (rs6498169) (first identified in the MS GWAS (Hafler et al. 2007) and the "T1D-SNP" (rs12708716) there is some degree of LD (D'=1, r^2=0.3 in the CEU HapMap samples (http://www.hapmap.org/). Thus, all evidence point at CLEC16A as a susceptibility genes for several immune-mediated diseases.

It is important to underline that GWASs are picking up a signal in a region tagged by an associated SNP. Consequently, the associated signal could either be the causing variant, or more likely, an indirect signal of association in the region. Thus, after a locus is identified in GWASs, much work remain to establish which variant is the causative variant (Plenge & Rioux 2006). Hence, it is difficult to conclude whether the causal variant in the *CLEC16A* gene region in MS and T1D will be the same. Also another autoimmune disease, Addison disease, has shown association to the *CLEC16A* gene region (Skinningsrud et al. 2008). However, both Skinningsrud et al. and Todd et al. point out that the *CLEC16A* gene is located close to other genes (*SOCS1* and *CIITA*), and the primary signal could come from these genes (Skinningsrud et al. 2008; Todd et al. 2007). For the *CLEC16A* gene, the function is yet not known. Other C-lectin molecules are important in recognition of carbohydrates (Robinson et al. 2006). Since the CLEC16A protein has a small C-lectin domain (consisting of less than 20 amino acids), it is more likely that the effect of this molecule is related to other domains rather than the C-lectin domain. Todd and colleagues have suggested that a tyrosine-based activation motif (ITAM) in exon 12 could be of importance for the function of the gene. However, more functional studies are required to establish the function of CLEC16A and how this molecule is implicated in immune-mediated diseases.

Our study (**Paper V**) also found association to the SNP rs763361 located in the *CD226* gene, encoding the CD226 molecule also known as DNAX accessory molecule 1 (DNAM-1). This membrane-bound surface molecule is expressed on haematopoietic cells such as NK and T cells and has a role in NK cell cytotoxicity as well as Th1 mediated immune responses (Shibuya et al. 1996). Studies in mice have shown that anti-CD226 monoclonal antibodies delay onset and severity of EAE (Dardalhon et al. 2005). The T1D-associated nsSNP in exon 7 in the *CD226* gene changes the amino acid at position

207 from glycine to serine (G207S). In the study by Todd et al. they suggested that this nsSNP disrupt a splice site enhancer or a silencer, and may alter the RNA splicing or alter the signalling cascade by affecting the known phosphorylation site on the molecule (Todd et al. 2007). Further finemapping studies by Hafler et al. also suggested that this nsSNP is the causal variant in the *CD226* gene (Hafler et al. 2009). However, due to the tagging approach it can not be excluded that another SNP in strong LD could be the causal variant or that SNPs not properly tagged could cause independent effects in the gene.

The GWASs have until now not been able to capture the entire genome completely, so some gene regions may have been poorly mapped. However, the growing list of novel MS-associated SNPs identified in GWASs, demonstrates a successful approach (Table 1, page 29). As mentioned by Hafler et al. the associated SNPs identified in the first MS GWAS, could just explain a small proportion (probably less than 0.2%) of the risk variants involved in the development of MS (Hafler et al. 2007). The fact that there is a huge discrepancy between the number of validated MS susceptibility loci found in the GWASs and the estimated number from the CD/CV model, has changed our focus to also consider the alternative hypothesis (CD/RV) (Wang et al. 2005). The CD/RV model suggests that genetic heterogeneity consists of many rare variants, which each confer a high risk. Recently, it has been shown in T1D that in addition to associations to common genetic variants discovered in GWASs, rare variants could also occur in these associated gene regions (Nejentsev et al. 2009). Nejentsev et al. resequenced the *IFIH1* gene, associated in T1D with a common SNPs, and found independent associations to rare variants. Interestingly, Nejentsev and colleagues showed for the *IFIH1* gene that these rare variants predicted biological effects as truncation of the protein or affection of essential splicing positions or highly conserved amino acids. In conclusion, both the CD/CV and CD/RV hypotheses are supported by the present GWAS data.

8.5 MS genes and the autoimmune model

Since the recently discovered non-HLA genes have low effect sizes in MS, they are unsuitable as clinical markers. However, identification if these genes is important since this may generate new hypotheses about disease mechanisms (reviewed in Gregersen & Olsson 2009). Not surprisingly, the MS susceptibility genes identified so far are genes related to regulation of the immune response (the *HLA complex*, *IL2Rα*, *IL7R, CLEC16A* and *TYK2* genes, among others).

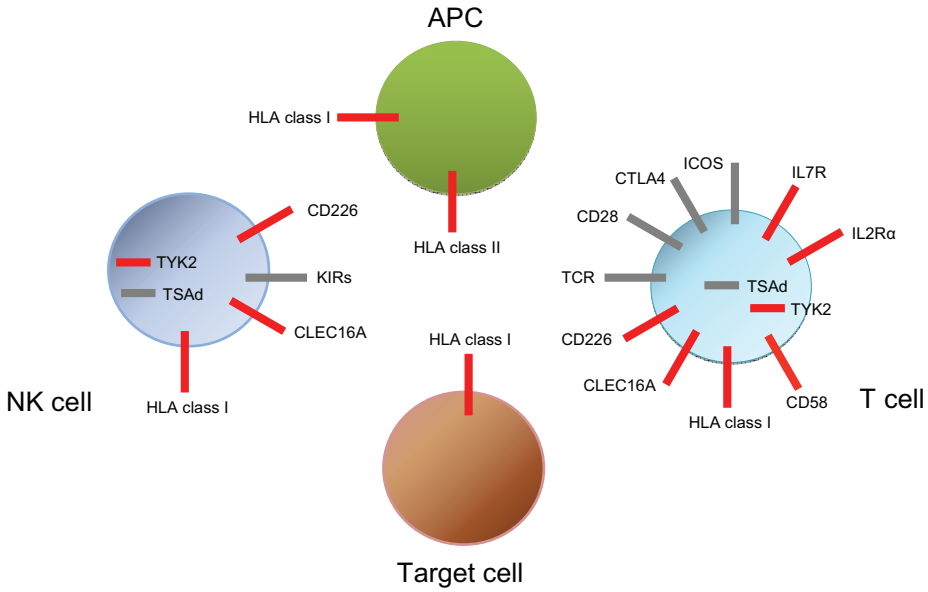

Figure 5: Schematic drawing showing how some MS-associated genes are expressed (marked red) in immune cells. Molecules that are encoded by genes not substantially proven to be implicated in MS susceptibility are marked grey.

Abbreviation**s:** APC: antigen presenting cell; CD28: clusters of differentiation molecule 28; CD58: clusters of differentiation molecule 58; CD226: clusters of differentiation molecule 226; CLEC16A: C-type lectin domain family 16, member A; CTLA4: cytotoxic T-lymphocyte associated protein 4; HLA: human leuckocyte antigen; ICOS: inducible T-cell co-stimulator; IL2Rα: interleukin 2 receptor alpha unit; IL7R: interleukin 7 receptor; KIR: killer immunoglobulin-like receptor; NK cell: natural killer cell; TCR: T-cell receptor; TSAd: T-cell specific adaptor protein; TYK2: Tyrosine kinase 2

These findings support the notion that several immune-mediated diseases may share the same genetic associations. It has been hypothesized that some common genetic variants may contribute to a general immune-mediated phenotype (Hafler et al. 2007). A recent review by Zhernekova et al. suggested that shared genetic associations among immune-mediated diseases could be grouped in four main immune pathways (Zhernakova et al. 2009). These shared pathways are; 1) T-cell differentiation, 2) Immune cell activation and signalling, 3) Innate immunity and TNF signalling, 4) Other categories. Also Gregersen & Olsson have categorized the genetic findings in immune-meditated diseases and suggest

similar categories (Gregersen & Olsson 2009). Figure 5 shows schematically how some of the MS-associated genes are expressed in immune cells (marked red). Also some molecules that are encoded by genes not substantially proven to be implicated in MS susceptibility, are marked in the figure for comparison (marked grey).

Table 5: Overview of recent reportedly MS-associated loci, with references to publications reporting the same gene region as associatied in other immune-mediated diseases.

Gene	Chr	CD	IBD	Pso	RA	SLE	T1D	References
HLA class II	6p21	X	X		X	X	X	1,2
HLA class I	6p21			X			X	2
IL2Rα	10p15	X					X	3, 4, 5, 6
IL7R	5p13						X	7
CD58	1p13							
CLEC16A	16p13		X				X	7, 8, 9, 10, 11
RPL5#	1p22							
FAM69A#	1p22							
EVI5#	1p22							
CD226	18q22						X	7,
TYK2	19p13					X		12, 13, 14
GPC5	13q32							
KIF1B	1p36							

Abbreviations and information: CD: celiac disease; IBD: inflammatory bowel disease including Crohn's disease and ulcerative colitis; Pso: psoriasis; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; T1D: type 1 diabetes; IL2Rα: Interleukin 2 receptor alpha unit; IL7R: Interleukin 7 receptor; CD58: Clusters of differentiation molecule 58; CLEC16A: C-type lectin domain family 16, member A; RPL5: Ribosomal protein L5; FAM69A: Family with sequence similarity 69, member A; EVI5: Ecotropic viral integration site 5; CD226: Clusters of differentiation molecule 226; TYK2: Tyrosine kinase 2, GPC5: Glypican 5; KIF1B: Kinesin family member 1B;
Comment: # RPL5, FAM69A and EVI5 are located close to each other at chromosome 1p22, and are possibly the same signal.
References: 1) (Fernando et al. 2008) 2) (Thorsby & Lie 2005) 3) (Brand et al. 2007) 4) (WTCCC 2007) 5) (Lowe et al. 2007) 6) (Vella et al. 2005) 7) (Todd et al. 2007) 8) (Hakonarson et al. 2007) 9) (Martinez et al. 2009) 10) (Zoledziewska et al. 2009) 11) (Marquez et al. 2009) 12) (Sigurdsson et al. 2005) 13) (Cunninghame Graham et al. 2007) 14) (Suarez-Gestal et al. 2009)

Interestingly, many of the gene regions now confirmed to be associated with MS have also been identified in other immune-mediated diseases. Table 5 above lists the MS-associated genes, and refers to publications which show association to the same gene

regions in other immune-mediated diseases. If different immune-mediated diseases share common pathways, collaborations between different kinds of research-groups are even more important. For instance, since many of these genes have modest effects, cases from different diseases can be grouped together as a wider phenotype to increase the power to detect shared susceptibility genes. This strategy was successfully used for the identification of the *TYK2* gene in the nsSNP screen by WTCCC (Burton et al. 2007).

The National Human Genome Research Institute provides an updated list of published GWASs (http://www.genome.gov/26525384). As mention (see Figure 1, page 23), during the years 2006-2008, 32 GWASs were reported for some selected immune-mediated diseases. Roughly, 120 gene regions were suggested among these diseases, and eleven of the regions were shared with another autoimmune disease. Even though some of the suggested associations may have been false positive findings, 10% of the gene regions were shared among some of the immune-mediated diseases. Overall, there are accumulation evidence for shared susceptibility genes in immune-mediated diseases.

Grouping the genetic associations found in MS could give clues to the underlying mechanisms (Zhernakova et al. 2009). In this thesis, we identified a protective effect in MS for the HLA-Bw4 group (the ligand for the KIR3DL1 (**Paper III**), association to the *CLEC16A* gene and the *CD226* gene (**Paper V**) and association to the *TYK2* gene (**Paper IV**). The three first genes mentioned here, are all important for NK cell functions. Based on these findings, studies of innate immunity may be considered as of particular interest. In MS, EBV is a promising candidate for an environmental factor that could trigger underlying genetic predisposition. The involvement of innate immunity provides links to environmental triggers. Improved knowledge about these mechanisms might provide new tools for disease prevention (Zhernakova et al. 2009).

8.6 Are there MS associated genes in non-immunological pathways?

Although until now only genes in immunological pathways are validated and replicated as MS-associated loci, there are possibly also other mechanisms for MS-disease development. As earlier mentioned, the *KIF1B* gene is suggested to be associated with MS (Aulchenko et al. 2008), a gene that is neuronally expressed. The encoded protein is hypothesized to be implicated in the axonal transport of mitochondria and synaptic vesicle precursors. However, the MS-associated SNP (rs10492972) is so far not replicated

in other MS cohorts. Since MS is a disease of the CNS, mechanisms shared with other CNS diseases may also be expected. Pathological findings point towards both an immunological and a degenerative component of MS disease, suggesting to that shared pathways with degenerative diseases are likely. Such data are however warranted.

Identified MS-associated loci outside the HLA region are mainly found by a hypothesis-free approach using GWASs and ns-SNP screens. As discussed in the methodological section (see page 47-48), an expected amount of false positive findings may make it difficult to identify the true associations among the highest-ranked associations. To restrict the amount of false positive findings, a P-value below 10^{-7} is usually applied in GWASs. Consequently, some true findings will have a P-value over this limit and fall into the false negative group. On this background, it is often challenging to select which results should be selected for follow-up analysis. Due to cost restrictions, not all initial associations can be selected for replication studies, and the selection of SNPs for the replication phase could be biased. Many will favour SNPs in loci of immunological importance. This was the case in the MS GWAS. For instance, the *IL7R* and *IL2Rα* had not the lowest P-values in the study, but were selected for replication partly because of the known importance of the encoded molecules (Hafler et al. 2007). Thus, our current list of MS-associated loci is probably biased towards genes of importance in immunological pathways.

Compared to diseases like T1D and Crohn's disease, MS has not yet been studied in many GWASs (Figure 1, see also updated list at http://www.genome.gov/26525384). It has been seen for both T1D and Crohn's disease that different GWASs may identify some different susceptibility genes, even though many of the same associations are replicated. This may be due to slightly different study designs, for instance genotyping of different tagSNPs. Recently also meta-analyses of published GWASs have been available. Barrett et al. found 40 novel susceptibility loci in T1D when combining the data from two previously published GWASs (Barrett et al. 2009). The meta-analysis increased the sample size (up to 7514 T1D cases and 9045 controls), and thereby increased the power to detect novel susceptibility genes. In the near future meta-analysis of available GWAS data will also be performed in MS, and the numbers of new MS-associated loci are thereby expected to increase.

Another problem in the interpretation of GWASs is the bias towards identifying common alleles. The power to detect common alleles is larger than to detect rare alleles, as discussed in the methodological considerations (page 44-46). Missense SNPs are more often chosen for replication, which could bias a study. SNPs in protein-coding regions may have relevance for the function of proteins, but we also know that non-coding genetic variation also could influence for instance regulatory sites for transcription. Other genetic variations are not properly detected by most GWAS. For instance, until recently most GWASs did not include screening of CNVs. CNVs and somatic genetic changes are for instance suggested to explain the relatively discordance between twins in autoimmune diseases (Bruder et al. 2008), and studies need to explore this suggestion more carefully.

9 The next generation of genetic studies in MS

Approximately one year ago, Oksenberg et al. wrote in their review that "the tools finally exist to identify the full set of genes influencing the pathogenesis to MS" (Oksenberg et al. 2008). As pointed out in the reviews by Wang et al. and Hirschhorn & Daly, at least 10 000 cases and controls will be needed to finally conclude which are the common genetic risk factors for a complex genetic disease (Hirschhorn & Daly 2005; Wang et al. 2005). With this sample size problems, related to genotyping and phenotype errors, misclassifications and confounding factors are less.

Based on today's knowledge, the "ultimate" GWAS project in MS has been planned for some years and is now ongoing. This has been possible by international collaborative efforts through the IMSGC, which includes groups from Australia, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Norway, Poland, Spain, Sweden, the UK and the US. The study, which is a part of the second phase of the Wellcome Trust Case Control Consortium study (WTCCC2), will genotype as much as 11 000 cases using the Illumina 610 Quad chip, performed at the Sanger Institute, UK (http://www.sanger.ac.uk/). The chip includes assays for nearly 700 000 genetic variants, most of them SNPs, but also a set of CNVs. In addition, a replication sample set of similar size is collected. This next generation MS GWAS is therefore very well-powered and should be able to identify common risk alleles with an OR above 1.2. The results of this project will hopefully substantially influence our understanding of this enigmatic disease. There is optimism that this study will identify a series of MS candidate genes. But as in other GWASs, the novel SNPs and suggested gene regions need to be validated and further studied (for instance by resequencing or finemapping) to finally conclude which are the causal variants. Thereafter, the mechanisms of how the associated variants are involved in MS pathogenesis need to be analysed in extensive functional studies. So even though results from the "ultimate" MS GWAS is soon available, there is still more to do for MS researchers for years to come.

Last but not least, we need to consider how this new genetic information may have an impact on the quality of life of patients suffering from MS. Will this work help the patients? In large collaborative projects it is possible to study specific subgroups of

patients without losing too much power. We need to define how the genotypes associated relate to the different phenotypes of the disease, and how clinical characteristics of MS relate to molecular mechanisms involved in MS pathogenesis. This will be an important step towards development of better therapies. For instance, patients may be grouped into treatment responders or non-responders, and genetic markers might indicate which treatment should be initiated. Hopefully, in the near future newly acquired knowledge will have a positive impact for MS patients and their quality of life (Lettre & Rioux 2008).

10 References

Airas L, Saraste M, Rinta S, Elovaara I, Huang YH & Wiendl H. Immunoregulatory factors in multiple sclerosis patients during and after pregnancy: relevance of natural killer cells. *Clinical and Experimental Immunology* 2008; 151(2): 235-243.

Akkad DA, Hoffjan S, Petrasch-Parwez E, Beygo J, Gold R & Epplen JT. Variation in the IL7RA and IL2RA genes in German multiple sclerosis patients. *Journal of Autoimmunity* 2009; 32(2): 110-115.

Alcina A, Fedetz M, Ndagire D, Fernandez O, Leyva L, Guerrero M et al. The T244I variant of the interleukin-7 receptor-alpha gene and multiple sclerosis. *Tissue Antigens* 2008; 72(2): 158-161.

Alcina A, Fedetz M, Ndagire D, Fernandez O, Leyva L, Guerrero M et al. IL2RA/CD25 gene polymorphisms: uneven association with multiple sclerosis (MS) and type 1 diabetes (T1D). *PLoS.ONE.* 2009; 4(1): e4137

Alonso A & Hernan MA. Temporal trends in the incidence of multiple sclerosis: a systematic review. *Neurology* 2008; 71(2): 129-135.

Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA & Yunis EJ. The haplotype structure of the human major histocompatibility complex. *Human Immunology* 2006; 67(1-2): 73-84.

Amundsen SS, Naluai AT, Ascher H, Ek J, Gudjonsdottir AH, Wahlstrom J et al. Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease. *Tissue Antigens* 2004; 64(5): 593-599.

Andersson M, varez-Cermeno J, Bernardi G, Cogato I, Fredman P, Frederiksen J et al. Cerebrospinal fluid in the diagnosis of multiple sclerosis: a consensus report. *Journal of Neurology, Neurosurgery and Psychiatry* 1994; 57(8): 897-902.

Ascherio A & Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Annals of Neurology* 2007a; 61(4): 288-299.

Ascherio A & Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Annals of Neurology* 2007b; 61(6): 504-513.

Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, Broer L, Jafari N, Hillert J et al. Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. *Nature Genetics* 2008; 40(12): 1402-1403.

Bagos PG, Karnaouri AC, Nikolopoulos GK & Hamodrakas SJ. No evidence for association of CTLA-4 gene polymorphisms with the risk of developing multiple sclerosis: a meta-analysis. *Multiple Sclerosis* 2007; 13(2): 156-168.

Bahmanyar S, Montgomery SM, Hillert J, Ekbom A & Olsson T. Cancer risk among patients with multiple sclerosis and their parents. *Neurology* 2009; 72(13): 1170-1177.

Bailey JA & Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat.Rev.Genet.* 2006; 7(7): 552-564.

Balding DJ. A tutorial on statistical methods for population association studies. *Nat.Rev.Genet.* 2006; 7(10): 781-791.

Ballerini C, Guerini FR, Rombola G, Rosati E, Massacesi L, Ferrante P et al. HLA-multiple sclerosis association in continental Italy and correlation with disease prevalence in Europe. *Journal of Neuroimmunology* 2004; 150(1-2): 178-185.

Ban M, Goris A, Lorentzen AR, Baker A, Mihalova T, Ingram G et al. Replication analysis identifies TYK2 as a multiple sclerosis susceptibility factor. *European Journal of Human Genetics* 2009; In press

Bar-Or A. The immunology of multiple sclerosis. *Seminars in Neurology* 2008; 28(1): 29-45.

Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Human Molecular Genetics* 2009; 18(4): 767-778.

Barcellos LF, Kamdar BB, Ramsay PP, DeLoa C, Lincoln RR, Caillier S et al. Clustering of autoimmune diseases in families with a high-risk for multiple sclerosis: a descriptive study. *Lancet Neurol.* 2006a; 5(11): 924-931.

Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Human Molecular Genetics* 2006b; 15(18): 2813-2824.

Barkhof F, Filippi M, Miller DH, Scheltens P, Campi A, Polman CH et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain* 1997; 120 (11): 2059-2069.

Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature Genetics* 2009; In press

Barrow AD & Trowsdale J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunological Reviews* 2008; 224: 98-123.

Baxter AG. The origin and application of experimental autoimmune encephalomyelitis. *Nat.Rev.Immunol.* 2007; 7(11): 904-912.

Beckmann JS, Estivill X & Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat.Rev.Genet.* 2007; 8(8): 639-646.

Bielekova B, Catalfamo M, Reichert-Scrivner S, Packer A, Cerna M, Waldmann TA et al. Regulatory CD56(bright) natural killer cells mediate immunomodulatory effects of IL-

2Ralpha-targeted therapy (daclizumab) in multiple sclerosis. *Proc.Natl.Acad.Sci.U.S.A* 2006; 103(15): 5941-5946.

Blomhoff A, Lie BA, Myhre AG, Kemp EH, Weetman AP, Akselsen HE et al. Polymorphisms in the cytotoxic T lymphocyte antigen-4 gene region confer susceptibility to Addison's disease. *J Clin.Endocrinol.Metab* 2004; 89(7): 3474-3476.

Blomhoff A, Olsson M, Johansson S, Akselsen HE, Pociot F, Nerup J et al. Linkage disequilibrium and haplotype blocks in the MHC vary in an HLA haplotype specific manner assessed mainly by DRB1*03 and DRB1*04 haplotypes. *Genes Immun.* 2006; 7(2): 130-140.

Bo L, Vedeler CA, Nyland HI, Trapp BD & Mork SJ. Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *Journal of Neuropathology and Experimental Neurology* 2003; 62(7): 723-732.

Botstein D, White RL, Skolnick M & Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 1980; 32(3): 314-331.

Brand OJ, Lowe CE, Heward JM, Franklyn JA, Cooper JD, Todd JA et al. Association of the interleukin-2 receptor alpha (IL-2Ralpha)/CD25 gene region with Graves' disease using a multilocus test and tag SNPs. *Clin.Endocrinol.(Oxf)* 2007; 66(4): 508-512.

Broquet T & Petit E. Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology* 2004; 13(11): 3601-3608.

Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, de Stahl TD et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics* 2008; 82(3): 763-771.

Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson E, Palmgren J et al. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS.ONE.* 2007; 2(7): e664.

Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A et al. Association scan of 14, 500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nature Genetics* 2007; 39(11): 1329-1337

Caballero A, ves-Leon S, Papais-Alvarenga R, Fernandez O, Navarro G & Alonso A. DQB1*0602 confers genetic susceptibility to multiple sclerosis in Afro-Brazilians. *Tissue Antigens* 1999; 54(5): 524-526.

Carlson CS, Eberle MA, Kruglyak L & Nickerson DA. Mapping complex disease loci in whole-genome association studies. *Nature* 2004; 429(6990): 446-452.

Charcot JM. Histologie de la sclérose en plaques. *Gaz.Hop.civils et militaires* 1868.

Coles AJ, Compston DA, Selmaj KW, Lake SL, Moran S, Margolin DH et al. Alemtuzumab vs. interferon beta-1a in early multiple sclerosis. *New England Journal of Medicine* 2008; 359(17): 1786-1801.

Comabella M, Craig DW, Camina-Tato M, Morcillo C, Lopez C, Navarro A et al. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500, 000 single nucleotide polymorphisms. *PLoS.ONE.* 2008; 3(10): e3490.

Compston A, Lassmann H, McDonald I, Miller D, Noseworthy J, Smith K & Wekerle H. McAlpine's multiple sclerosis. 2005 Fourth edition, Churchill Livingstone Elsevier, Philadelphia.

Compston DA, Batchelor JR & McDonald WI. B-lymphocyte alloantigens associated with multiple sclerosis. *Lancet* 1976; 2(7998): 1261-1265.

Confavreux C & Vukusic S. Age at disability milestones in multiple sclerosis. *Brain* 2006; 129(Pt 3): 595-605.

Conrad DF, Andrews TD, Carter NP, Hurles ME & Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genetics* 2006; 38(1): 75-81.

Cunninghame Graham DS, Akil M & Vyse TJ. Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families. *Rheumatology.(Oxford)* 2007; 46(6): 927-930.

D'Alfonso S, Bolognesi E, Guerini FR, Barizzone N, Bocca S, Ferrante D et al. A sequence variation in the MOG gene is involved in multiple sclerosis susceptibility in Italy. *Genes Immun.* 2008; 9(1): 7-15.

Dai KZ, Harbo HF, Celius EG, Oturai A, Sorensen PS, Ryder LP et al. The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. *Genes Immun.* 2001; 2(5): 263-268.

Dai KZ, Vergnaud G, Ando A, Inoko H & Spurkland A. The SH2D2A gene encoding the T-cell-specific adapter protein (TSAd) is localized centromeric to the CD1 gene cluster on human Chromosome 1. *Immunogenetics* 2000; 51(3): 179-185.

Dardalhon V, Schubart AS, Reddy J, Meyers JH, Monney L, Sabatos CA et al. CD226 is specifically expressed on the surface of Th1 cells and regulates their expansion and effector functions. *Journal of Immunology* 2005; 175(3): 1558-1565.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ & Altshuler D. Efficiency and power in genetic association studies. *Nature Genetics* 2005; 37(11): 1217-1223.

De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L et al. The role of the CD58 locus in multiple sclerosis. *Proc.Natl.Acad.Sci.U.S.A* 2009; 106(13): 5264-5269.

De Jager PL, Rossin E, Pyne S, Tamayo P, Ottoboni L, Viglietta V et al. Cytometric profiling in multiple sclerosis uncovers patient population structure and a reduction of CD8low cells. *Brain* 2008; 131(7): 1701-1711.

De Jong WS, De Jong DM & Bodis M. A fluorogenic 5' nuclease (TaqMan) assay to assess dosage of a marker tightly linked to red skin color in autotetraploid potato. *Theor.Appl.Genet.* 2003; 107(8): 1384-1390.

de Kok JB, Wiegerinck ET, Giesendorf BA & Swinkels DW. Rapid genotyping of single nucleotide polymorphisms using novel minor groove binding DNA oligonucleotides (MGB probes). *Human Mutation* 2002; 19(5): 554-559.

Derfuss T, Parikh K, Velhin S, Braun M, Mathey E, Krumbholz M et al. Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals. *Proc.Natl.Acad.Sci.U.S.A* 2009; 106(20): 8302-8307.

Devlin B & Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995; 29(2): 311-322.

Drappa J, Kamen LA, Chan E, Georgiev M, Ashany D, Marti F et al. Impaired T cell death and lupus-like autoimmunity in T cell-specific adapter protein-deficient mice. *Journal of Experimental Medicine* 2003; 198(5): 809-821.

Dupont WD & Plummer WD, Jr. Power and sample size calculations. A review and computer program. *Control Clin.Trials* 1990; 11(2): 116-128.

Dyment DA, Herrera BM, Cader MZ, Willer CJ, Lincoln MR, Sadovnick AD et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Human Molecular Genetics* 2005; 14(14): 2019-2026.

Dyment DA, Yee IM, Ebers GC & Sadovnick AD. Multiple sclerosis in stepsiblings: recurrence risk and ascertainment. *Journal of Neurology, Neurosurgery and Psychiatry* 2006; 77(2): 258-259.

Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C et al. A full genome search in multiple sclerosis. *Nature Genetics* 1996; 13(4): 472-476.

Ebers GC, Sadovnick AD & Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* 1995; 377(6545): 150-151.

Feldmann M & Steinman L. Design of effective immunotherapy for human autoimmunity. *Nature* 2005; 435(7042): 612-619.

Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS.Genet.* 2008; 4(4): e1000024.

Fischer SG & Lerman LS. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc.Natl.Acad.Sci.U.S.A* 1983; 80(6): 1579-1583.

Flores-Villanueva PO, Yunis EJ, Delgado JC, Vittinghoff E, Buchbinder S, Leung JY et al. Control of HIV-1 viremia and protection from AIDS are associated with HLA-Bw4 homozygosity. *Proc.Natl.Acad.Sci.U.S.A* 2001; 98(9): 5140-5145.

Fogdell-Hahn A, Ligers A, Gronning M, Hillert J & Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* 2000; 55(2): 140-148.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449(7164): 851-861.

Gale CR & Martyn CN. Migrant studies in multiple sclerosis. *Progress in Neurobiology* 1995; 47(4-5): 425-448.

GAMES. A meta-analysis of whole genome linkage screens in multiple sclerosis. *Journal of Neuroimmunology* 2003; 143(1-2): 39-46.

Geurts JJ & Barkhof F. Grey matter pathology in multiple sclerosis. *Lancet Neurol.* 2008; 7(9): 841-851.

Ghoreschi K, Laurence A & O'Shea JJ. Janus kinases in immune cell signaling. *Immunological Reviews* 2009; 228(1): 273-287.

Goodin DS. Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: have we been overly harsh in our assessment? *Annals of Neurology* 2006; 59(4): 597-605.

Goodin DS, Cohen BA, O'Connor P, Kappos L & Stevens JC. Assessment: the use of natalizumab (Tysabri) for the treatment of multiple sclerosis (an evidence-based review): report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2008; 71(10): 766-773.

Granum S, Sundvold-Gjerstad V, Dai KZ, Kolltveit KM, Hildebrand K, Huitfeldt HS et al. Structure function analysis of SH2D2A isoforms expressed in T cells reveals a crucial role for the proline rich region encoded by SH2D2A exon 7. *BMC.Immunol.* 2006; Jul 13: 7-15.

Gregersen PK & Olsson LM. Recent advances in the genetics of autoimmune disease. *Annual Review of Immunology* 2009; 27: 363-391.

Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A et al. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nature Genetics* 2007; 39(9): 1083-1091.

Greve B, Simonenko R, Illes Z, Peterfalvi A, Hamdi N, Mycko MP et al. Multiple sclerosis and the CTLA4 autoimmunity polymorphism CT60: no association in patients from Germany, Hungary and Poland. *Multiple Sclerosis* 2008; 14(2): 153-158.

Guo SW. Familial aggregation of environmental risk factors and familial aggregation of disease. *American Journal of Epidemiology* 2000; 151(11): 1121-1131.

Hafler DA. The distinction blurs between an autoimmune versus microbial hypothesis in multiple sclerosis. *J.Clin.Invest* 1999; 104(5): 527-529.

Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL et al. Risk alleles for multiple sclerosis identified by a genomewide study. *New England Journal of Medicine* 2007; 357(9): 851-862.

Hafler JP, Maier LM, Cooper JD, Plagnol V, Hinks A, Simmonds MJ et al. CD226 Gly307Ser association with multiple autoimmune diseases. *Genes Immun.* 2009; 10(1): 5-10.

Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. *Nature Genetics* 1996; 13(4): 469-471.

Haines JL, Terwedow HA, Burgess K, Pericak-Vance MA, Rimmler JB, Martin ER et al. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group. *Human Molecular Genetics* 1998; 7(8): 1229-1234.

Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007; 448(7153): 591-594.

Harbo HF, Celius EG, Vartdal F & Spurkland A. CTLA4 promoter and exon 1 dimorphisms in multiple sclerosis. *Tissue Antigens* 1999; 53(1): 106-110.

Harbo HF, Lie BA, Sawcer S, Celius EG, Dai KZ, Oturai A 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): 237-247.

Hirschhorn JN & Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat.Rev.Genet.* 2005; 6(2): 95-108.

Holmoy T & Hestvik AL. Multiple sclerosis: immunopathogenesis and controversies in defining the cause. *Curr.Opin.Infect.Dis.* 2008; 21(3): 271-278.

Hoppenbrouwers IA, Aulchenko YS, Ebers GC, Ramagopalan SV, Oostra BA, van Duijn CM et al. EVI5 is a risk gene for multiple sclerosis. *Genes Immun.* 2008; 9(4): 334-337.

Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK et al. Gene map of the extended human MHC. *Nat.Rev.Genet.* 2004; 5(12): 889-899.

IMSGC. Refining genetic associations in multiple sclerosis. *Lancet Neurol.* 2008; 7(7): 567-569.

IMSGC. The expanding genetic overlap between multiple sclerosis and type I diabetes. *Genes Immun.* 2009; 10(1): 11-14.

Jersild C, Svejgaard A & Fog T. HL-A antigens and multiple sclerosis. *Lancet* 1972; 1(7762): 1240-1241.

Jersild C, Svejgaard A, Fog T & Ammitzboll T. HL-A antigens and diseases. I. Multiple sclerosis. *Tissue Antigens* 1973; 3(4): 243-250.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di GG et al. Haplotype tagging for the identification of common disease genes. *Nature Genetics* 2001; 29(2): 233-237.

Karlsen TH, Boberg KM, Olsson M, Sun JY, Senitzer D, Bergquist A et al. Particular genetic variants of ligands for natural killer cell receptors may contribute to the HLA associated risk of primary sclerosing cholangitis. *Journal of Hepatology* 2007; 46(5): 899-906.

Kim S, Sunwoo JB, Yang L, Choi T, Song YJ, French AR et al. HLA alleles determine differences in human natural killer cell responsiveness and potency. *Proc.Natl.Acad.Sci.U.S.A* 2008; 105(8): 3053-3058.

Kingsmore SF, Lindquist IE, Mudge J, Gessler DD & Beavis WD. Genome-wide association studies: progress and potential for drug discovery and development. *Nat.Rev.Drug Discov.* 2008; 7(3): 221-230.

Klein J & Sato A. The HLA system. Second of two parts. *New England Journal of Medicine* 2000; 343(11): 782-786.

Kruglyak L & Nickerson DA. Variation is the spice of life. *Nature Genetics* 2001; 27(3): 234-236.

Kulkarni S, Martin MP & Carrington M. The Yin and Yang of HLA and KIR in human disease. *Seminars in Immunology* 2008; 20(6): 343-352

Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33(11): 1444-1452.

Kurtzke JF, Beebe GW & Norman JE, Jr. Epidemiology of multiple sclerosis in U.S. veterans: 1. Race, sex, and geographic distribution. *Neurology* 1979; 29(9 Pt 1): 1228-1235.

Lander E & Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 1995; 11(3): 241-247.

Lettre G & Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Human Molecular Genetics* 2008; 17(R2): 116-121.

Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M et al. A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nature Genetics* 2005; 37(10): 1108-1112.

Lincoln MR, Ramagopalan SV, Chao MJ, Herrera BM, Deluca GC, Orton SM et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc.Natl.Acad.Sci.U.S.A* 2009; 106(18): 7542-7547.

Ling V, Wu PW, Finnerty HF, Agostino MJ, Graham JR, Chen S et al. Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches. *Genomics* 2001; 78(3): 155-168.

Linker RA, Kieseier BC & Gold R. Identification and development of new therapeutics for multiple sclerosis. *Trends in Pharmacological Sciences* 2008; 29(11): 558-565.

Lohmueller KE, Pearce CL, Pike M, Lander ES & Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genetics* 2003; 33(2): 177-182.

Lorentzen AR, Smestad C, Lie BA, Oturai AB, Akesson E, Saarela J et al. The SH2D2A gene and susceptibility to multiple sclerosis. *Journal of Neuroimmunology* 2008; 197(2): 152-158.

Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nature Genetics* 2007; 39(9): 1074-1082.

Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallstrom E, Khademi M et al. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nature Genetics* 2007; 39(9): 1108-1113.

Lunemann JD & Munz C. Do natural killer cells accelerate or prevent autoimmunity in multiple sclerosis? *Brain* 2008; 131(Pt 7): 1681-1683.

Marquez A, Varade J, Robledo G, Martinez A, Mendoza JL, Taxonera C et al. Specific association of a CLEC16A/KIAA0350 polymorphism with NOD2/CARD15(-) Crohn's disease patients. *European Journal of Human Genetics* 2009; In press.

Marrosu MG, Murru R, Murru MR, Costa G, Zavattari P, Whalen M et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Human Molecular Genetics* 2001; 10(25): 2907-2916.

Martinez A, Perdigones N, Cenit M, Espino L, Varade J, Lamas JR et al. Chromosomal region 16p13: further evidence of increased predisposition to immune diseases. *Annals of the Rheumatic Diseases* 2009; In press.

Matsumoto T, Bohman S, Dixelius J, Berge T, Dimberg A, Magnusson P et al. VEGF receptor-2 Y951 signaling and a role for the adapter molecule TSAd in tumor angiogenesis. *EMBO Journal* 2005; 24(13): 2342-2353.

McCluskey J & Peh CA. The human leucocyte antigens and clinical medicine: an overview. *Rev.Immunogenet.* 1999; 1(1): 3-20.

McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology* 2001; 50(1): 121-127.

McFarland HF & Martin R. Multiple sclerosis: a complicated picture of autoimmunity. *Nat.Immunol.* 2007; 8(9): 913-919.

Miller DH, Khan OA, Sheremata WA, Blumhardt LD, Rice GP, Libonati MA et al. A controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine* 2003; 348(1): 15-23.

Miller DH, Weinshenker BG, Filippi M, Banwell BL, Cohen JA, Freedman MS et al. Differential diagnosis of suspected multiple sclerosis: a consensus approach. *Multiple Sclerosis* 2008; 14(9): 1157-1174.

Minegishi Y, Saito M, Morio T, Watanabe K, Agematsu K, Tsuchiya S et al. Human tyrosine kinase 2 deficiency reveals its requisite roles in multiple cytokine signals involved in innate and acquired immunity. *Immunity.* 2006; 25(5): 745-755.

Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH & Compston DA. The British Isles survey of multiple sclerosis in twins. *Neurology* 1994; 44(1): 11-15.

Naito S, Namerow N, Mickey MR & Terasaki PI. Multiple sclerosis: association with HL-A3. *Tissue Antigens* 1972; 2(1): 1-4.

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987; 235(4796): 1616-1622.

Nejad S, Bryceson Y, Dissen E, Sundvold-Gjerstad V, Naper C, Rolstad B et al. cDNA cloning of a rat orthologue of SH2D2A encoding T-cell-specific adaptor protein (TSAd): expression in T and NK cells. *Immunogenetics* 2004; 56(5): 338-342.

Nejentsev S, Walker N, Riches D, Egholm M & Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009; 324(5925): 387-389.

Nielsen DM, Ehm MG & Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics* 1998; 63(5): 1531-1540.

Nielsen NM, Frisch M, Rostgaard K, Wohlfahrt J, Hjalgrim H, Koch-Henriksen N et al. Autoimmune diseases in patients with multiple sclerosis and their first-degree relatives: a nationwide cohort study in Denmark. *Multiple Sclerosis* 2008; 14(6): 823-829.

Oksenberg JR, Baranzini SE, Sawcer S & Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat.Rev.Genet.* 2008; 9(7): 516-526.

Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *American Journal of Human Genetics* 2004; 74(1): 160-167.

Olerup O & Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* 1991; 38(1): 1-15.

Orange JS & Ballas ZK. Natural killer cells in human health and disease. *Clinical Immunology* 2006; 118(1): 1-10.

Palace J. Guidelines for differential diagnosis of suspected multiple sclerosis. *Nat.Clin.Pract.Neurol.* 2009; 5(3): 134-135.

Parham P. MHC class I molecules and KIRs in human history, health and survival. *Nat.Rev.Immunol.* 2005; 5(3): 201-214.

Paterson PY. Transfer of allergic encephalomyelitis in rats by means of lymph node cells. *Journal of Experimental Medicine* 1960; 111(119-136.

Plenge R & Rioux JD. Identifying susceptibility genes for immunological disorders: patterns, power, and proof. *Immunological Reviews* 2006; 210: 40-51.

Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of Neurology* 2005; 58(6): 840-846.

Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Annals of Neurology* 1983; 13(3): 227-231.

Pritchard JK & Cox NJ. The allelic architecture of human disease genes: common disease-common variant...or not? *Human Molecular Genetics* 2002; 11(20): 2417-2423.

Pritchard JK & Przeworski M. Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* 2001; 69(1): 1-14.

Quintana FJ, Farez MF, Viglietta V, Iglesias AH, Merbl Y, Izquierdo G et al. Antigen microarrays identify unique serum autoantibody signatures in clinical and pathologic subtypes of multiple sclerosis. *Proc.Natl.Acad.Sci.U.S.A* 2008; 105(48): 18889-18894.

Ramagopalan SV, Morris AP, Dyment DA, Herrera BM, Deluca GC, Lincoln MR et al. The inheritance of resistance alleles in multiple sclerosis. *PLoS.Genet.* 2007; 3(9): 1607-1613.

Reich DE & Lander ES. On the allelic spectrum of human disease. *Trends in Genetics* 2001; 17(9): 502-510.

Risch N & Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273(5281): 1516-1517.

Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; 405(6788): 847-856.

Robinson MJ, Sancho D, Slack EC, LeibundGut-Landmann S & Reis e Sousa. Myeloid C-type lectins in innate immunity. *Nat.Immunol.* 2006; 7(12): 1258-1265.

Roxburgh RH, Sawcer S, Maranian M, Seaman S, Hensiek A, Yeo T et al. No evidence of a significant role for CTLA-4 in multiple sclerosis. *Journal of Neuroimmunology* 2006; 171(1-2): 193-197.

Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology* 2005; 64(7): 1144-1151.

Rubio JP, Stankovich J, Field J, Tubridy N, Marriott M, Chapman C et al. Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes Immun.* 2008; 9(7): 624-630.

Rudd CE & Schneider H. Unifying concepts in CD28, ICOS and CTLA4 co-receptor signalling. *Nat.Rev.Immunol.* 2003; 3(7): 544-556.

Salvetti M, Giovannoni G & Aloisi F. Epstein-Barr virus and multiple sclerosis. *Current Opinion in Neurology* 2009; 22(3): 201-206

Saraste M, Irjala H & Airas L. Expansion of CD56Bright natural killer cells in the peripheral blood of multiple sclerosis patients treated with interferon-beta. *Neurol.Sci.* 2007; 28(3): 121-126.

Sawcer S. The complex genetics of multiple sclerosis: pitfalls and prospects. *Brain* 2008; 131(Pt12): 3118-3131

Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A et al. A high-density screen for linkage in multiple sclerosis. *American Journal of Human Genetics* 2005; 77(3): 454-467.

Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J et al. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nature Genetics* 1996; 13(4): 464-468.

Sayer DC, Whidborne R, De SD, Rozemuller EH, Christiansen FT & Tilanus MG. A multicenter international evaluation of single-tube amplification protocols for sequencing-based typing of HLA-DRB1 and HLA-DRB3, 4, 5. *Tissue Antigens* 2004; 63(5): 412-423.

Seldin MF & Amos CI. Shared susceptibility variations in autoimmune diseases: a brief perspective on common issues. *Genes Immun.* 2009; 10(1): 1-4.

Serafini B, Rosicarelli B, Franciotta D, Magliozzi R, Reynolds R, Cinque P et al. Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *Journal of Experimental Medicine* 2007; 204(12): 2899-2912.

Serrano NC, Millan P & Paez MC. Non-HLA associations with autoimmune diseases. *Autoimmun.Rev.* 2006; 5(3): 209-214.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nature Genetics* 2006; 38(5): 556-560.

Shaw MH, Freeman GJ, Scott MF, Fox BA, Bzik DJ, Belkaid Y et al. Tyk2 negatively regulates adaptive Th1 immunity by mediating IL-10 signaling and promoting IFN-gamma-dependent IL-10 reactivation. *Journal of Immunology* 2006; 176(12): 7263-7271.

Shibuya A, Campbell D, Hannum C, Yssel H, Franz-Bacon K, McClanahan T et al. DNAM-1, a novel adhesion molecule involved in the cytolytic function of T lymphocytes. *Immunity.* 1996; 4(6): 573-581.

Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *American Journal of Human Genetics* 2005; 76(3): 528-537.

Skinningsrud B, Husebye ES, Pearce SH, McDonald DO, Brandal K, Wolff AB et al. Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. *J.Clin.Endocrinol.Metab* 2008; 93(9): 3310-3317.

Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat.Rev.Genet.* 2008; 9(6): 477-485.

Smerdel A, Dai KZ, Lorentzen AR, Flato B, Maslinski S, Thorsby E et al. Genetic association between juvenile rheumatoid arthritis and polymorphism in the SH2D2A gene. *Genes Immun.* 2004; 5(4): 310-312.

Smestad C, Brynedal B, Jonasdottir G, Lorentzen AR, Masterman T, Akesson E et al. The impact of HLA-A and -DRB1 on age at onset, disease course and severity in Scandinavian multiple sclerosis patients. *European Journal of Neurology* 2007; 14(8): 835-840.

Smestad C, Sandvik L, Holmoy T, Harbo HF & Celius EG. Marked differences in prevalence of multiple sclerosis between ethnic groups in Oslo, Norway. *Journal of Neurology* 2008; 255(1): 49-55.

Smith DJ & Lusis AJ. The allelic structure of common disease. *Human Molecular Genetics* 2002; 11(20): 2455-2461.

Sospedra M & Martin R. Immunology of multiple sclerosis. *Annual Review of Immunology* 2005; 23: 683-747.

Spurkland A, Brinchmann JE, Markussen G, Pedeutour F, Munthe E, Lea T et al. Molecular cloning of a T cell-specific adapter protein (TSAd) containing an Src homology (SH) 2 domain and putative SH3 and phosphotyrosine binding sites. *Journal of Biological Chemistry* 1998; 273(8): 4539-4546.

Spurkland A, Celius EG, Knutsen I, Beiske A, Thorsby E & Vartdal F. The HLA-DQ(alpha 1*0102, beta 1*0602) heterodimer may confer susceptibility to multiple sclerosis in the absence of the HLA-DR(alpha 1*01, beta 1*1501) heterodimer. *Tissue Antigens* 1997; 50(1): 15-22.

Suarez-Gestal M, Calaza M, Endreffy E, Pullmann R, Ordi-Ros J, Domenico SG et al. Replication of recently identified systemic lupus erythematosus genetic associations: a case control study. *Arthritis Res.Ther.* 2009; 11(3): R69.

Sun JY, Gaidulis L, Miller MM, Goto RM, Rodriguez R, Forman SJ et al. Development of a multiplex PCR-SSP method for Killer-cell immunoglobulin-like receptor genotyping. *Tissue Antigens* 2004; 64(4): 462-468.

Sundvold V, Torgersen KM, Post NH, Marti F, King PD, Rottingen JA et al. T cell-specific adapter protein inhibits T cell activation by modulating Lck activity. *Journal of Immunology* 2000; 165(6): 2927-2931.

Terwilliger JD & Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology* 1998; 9(6): 578-594.

Teutsch SM, Booth DR, Bennetts BH, Heard RN & Stewart GJ. Association of common T cell activation gene polymorphisms with multiple sclerosis in Australian patients. *J Neuroimmunol.* 2004; 148(1-2): 218-230.

Thacker EL, Mirzaei F & Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. *Annals of Neurology* 2006; 59(3): 499-503.

The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; 437(7063): 1299-1320.

Thorsby E & Lie BA. HLA associated genetic predisposition to autoimmune diseases: Genes involved and possible mechanisms. *Transplant Immunology* 2005; 14(3-4): 175-182.

Tivol EA, Borriello F, Schweitzer AN, Lynch WP, Bluestone JA & Sharpe AH. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity.* 1995; 3(5): 541-547.

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature Genetics* 2007; 39(7): 857-864.

Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G et al. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; 423(6939): 506-511.

van Bergen J., Thompson A, van der SA, Ottenhoff TH, Gussekloo J & Koning F. Phenotypic and functional characterization of CD4 T cells expressing killer Ig-like receptors. *Journal of Immunology* 2004; 173(11): 6719-6726.

Vartdal F, Gaudernack G, Funderud S, Bratlie A, Lea T, Ugelstad J et al. HLA class I and II typing using cells positively selected from blood by immunomagnetic isolation--a fast and reliable technique. *Tissue Antigens* 1986; 28(5): 301-312.

Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, Widmer B et al. Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. *American Journal of Human Genetics* 2005; 76(5): 773-779.

ves-Leon SV, Papais-Alvarenga R, Magalhaes M, Alvarenga M, Thuler LC & Fernandez O. Ethnicity-dependent association of HLA DRB1-DQA1-DQB1 alleles in Brazilian multiple sclerosis patients. *Acta Neurologica Scandinavica* 2007; 115(5): 306-311.

Vivier E, Tomasello E, Baratin M, Walzer T & Ugolini S. Functions of natural killer cells. *Nat.Immunol.* 2008; 9(5): 503-510.

von Andrian UH & Engelhardt B. Alpha4 integrins as therapeutic targets in autoimmune disease. *New England Journal of Medicine* 2003; 348(1): 68-72.

Wacholder S, Chanock S, Garcia-Closas M, El GL & Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute* 2004; 96(6): 434-442.

Wang WY, Barratt BJ, Clayton DG & Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat.Rev.Genet.* 2005; 6(2): 109-118.

Waterhouse P, Penninger JM, Timms E, Wakeham A, Shahinian A, Lee KP et al. Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* 1995; 270(5238): 985-988.

Watford WT & O'Shea JJ. Human tyk2 kinase deficiency: another primary immunodeficiency syndrome. *Immunity.* 2006; 25(5): 695-697.

Weber F, Fontaine B, Cournu-Rebeix I, Kroner A, Knop M, Lutz S et al. IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. *Genes Immun.* 2008; 9(3): 259-263.

Weber JL & May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 1989; 44(3): 388-396.

Wekerle H. Lessons from multiple sclerosis: models, concepts, observations. *Annals of the Rheumatic Diseases* 2008; 67 (Suppl 3) 56-60.

Wiencke K, Boberg KM, Donaldson P, Harbo H, Ling V, Schrumpf E et al. No major effect of the CD28/CTLA4/ICOS gene region on susceptibility to primary sclerosing cholangitis. *Scandinavian Journal of Gastroenterology* 2006; 41(5): 586-591.

WTCCC. Genome-wide association study of 14, 000 cases of seven common diseases and 3, 000 shared controls. *Nature* 2007; 447(7145): 661-678.

Wu LW, Mayo LD, Dunbar JD, Kessler KM, Ozes ON, Warren RS et al. VRAP is an adaptor protein that binds KDR, a receptor for vascular endothelial cell growth factor. *Journal of Biological Chemistry* 2000; 275(9): 6059-6062.

Wyman AR & White R. A highly polymorphic locus in human DNA. *Proc.Natl.Acad.Sci.U.S.A* 1980; 77(11): 6754-6758.

Yang Q, Khoury MJ, Friedman J, Little J & Flanders WD. How many genes underlie the occurrence of common complex diseases in the population? *International Journal of Epidemiology* 2005; 34(5): 1129-1137.

Yeo TW, De Jager PL, Gregory SG, Barcellos LF, Walton A, Goris A et al. A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Annals of Neurology* 2007; 61(3): 228-236.

Zhernakova A, van Diemen CC & Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat.Rev.Genet.* 2009; 10(1): 43-55.

Zoledziewska M, Costa G, Pitzalis M, Cocco E, Melis C, Moi L et al. Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia. *Genes Immun.* 2009; 10(1): 15-17.