

From the Division of Neurology,
Neurotec Department
Karolinska Institutet, Stockholm, Sweden

Genetic Mapping and Association Analysis in Multiple Sclerosis

Eva Åkesson



Stockholm 2005

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

Published and printed by Repro Print AB, Stockholm, Sweden

© Eva Åkesson, 2005

ISBN 91-7140-174-1

ABSTRACT

Multiple sclerosis (MS) is a chronic neurological disease. Epidemiological studies have shown evidence for both environmental and genetic involvement of the disease. The contribution of genes versus environment is complex, in that there is not one major gene causing the disease, but more likely several genes each contributing to a modest extent and most likely interacting in a complex manner.

Studies on genetic susceptibility in complex traits mainly rely on either of two principally different methods; Candidate gene analysis, where potential candidate genes are chosen based on the existing knowledge of pathogenic mechanisms; or genome-wide screens, where an unbiased selection of markers covering the genome with a density of about 10-15 cM between the markers (300-400 markers to cover the genome) is used for linkage analysis without previous assumption about the underlying pathology of the disease.

The high prevalence of multiple sclerosis (~130/100,000) in Scandinavia and the striking correlation between its global geographical distribution and the migration pattern of Northern Europeans suggest that important susceptibility alleles may initially have arisen in Scandinavia. In order to facilitate analysis of the disease in this founding population, a collaborative group was formed in 1994 (the Nordic MS Genetics Group) involving eight Nordic research centers with the aim to identify familial multiple sclerosis in Scandinavia in order to establish a resource for genetic analysis.

Study I was a genome-wide screen for linkage in 136 Nordic sib-pairs with multiple sclerosis that had been identified through the Nordic collaboration. Although the results showed no regions of genome-wide significance, 17 regions were identified where the lod score exceeded 0.7 (the 5% nominal significance level). These regions include 1q11-24, 2q24-32, 3p26.3, 3q21.1, 4q12, 6p25.3, 6p21-22, 6q21, 9q34.3, 10p15, 10p12-13, 11p15.5, 12q21.3, 16p13.3, 17q25.3, 22q12-13 and Xp22.3. In a genome wide screen of this density, only ten such regions would be expected by chance alone. The fact that our results exceed expectation suggests that some of these peaks are likely to be genuine.

One of the regions showing potential linkage in the Nordic screen was located on the short arm of chromosome 10. This region has also shown partially overlapping peaks in several of the genome-wide screens including the screens from Italy, Sardinia, United Kingdom, as well as the latest meta-analysis of all genome-wide screens performed in multiple sclerosis. In *study II* we explored this region by genotyping 13 microsatellites in 449 sib-pairs with MS from the Nordic populations as well as United Kingdom, Sardinia and Italy. This study showed increased support for suggestive linkage on 10p15 (MLS 2.5).

Study III is an association analysis of the gene coding for the alpha-receptor of IL-15 (IL-15RA). This gene is located in the area of the peak on 10p15 that was found to give suggestive linkage in Study II. In this study, six single-nucleotide polymorphisms (SNPs) were genotyped in 553 patients with multiple sclerosis and 530 healthy controls. The study did not show any association for IL-15RA to multiple sclerosis, neither in single-point analysis nor haplotype analysis, making it less likely that this is the gene responsible for the peak in this region in the linkage analysis.

In parallel with the published genome-wide screens for linkage in multiple sclerosis, a genome-wide linkage screen had been performed in a cohort of Icelandic families with multiple sclerosis. This screen showed a haplotype on chromosome 10p12, potentially associated with multiple sclerosis in these Icelandic families. As this region overlaps with our previous findings, we in *study IV* decided to investigate this chromosomal area in a cohort of Swedish patients with MS and controls. Although we could not show any definite significantly associated haplotypes, we detected a marginal significant core haplotype in the Swedish MS cohort overlapping a haplotype shared in four affected sibs in one Icelandic family, giving some support for continuing the genetic research in the region as one of the candidate regions on the short arm of chromosome 10 for susceptibility genes for MS.

LIST OF PUBLICATIONS

This thesis is based on the following articles and will be referred to by their Roman numerals:

- I. **Åkesson E**, Oturai A, Berg J, Fredrikson S, Andersen O, Harbo HF, Laaksonen M, Myhr KM, Nyland HI, Ryder LP, Sandberg-Wollheim M, Sorensen PS, Spurkland A, Svejgaard A, Holmans P, Compston A, Hillert J and Sawcer S (2002)
A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis
Genes and Immunity 3: 279-285

- II. **Åkesson E**, Corradu F, Marrosu MG, Massacesi L, Hensiek A, Harbo HF, Oturai A, Trojano M, Momigliano-Richiardi P, Cocco E, Murru R, Hillert J, Compston A and Sawcer S (2003)
Refining the linkage analysis on chromosome 10 in 449 sib-pairs with multiple sclerosis
Journal of Neuroimmunology 143: 31-38

- III. **Åkesson E**, Roos I, Lindgren CM, Kere J and Hillert J
An association study of IL-15 receptor alpha in multiple sclerosis
Manuscript

- IV. **Åkesson E**, Fossdal R, Peturson H, Thorlacius T, Guðnadóttir V, Hillert J, Stefansson K and Gulcher J
Fine-scale genotyping of a candidate-region on chromosome 10p12-11 in patients with multiple sclerosis
Manuscript

The papers are reprinted here with the kind permission of the publishers: Genes and Immunity, copyright 2002 Macmillan Publishers Ltd (Paper I) and Journal of Neuroimmunology, copyright 2003 Elsevier (Paper II).

CONTENTS

Introduction.....	1
Multiple Sclerosis.....	1
Immunology and Pathology.....	2
Epidemiology.....	5
Genetic Epidemiology.....	7
Complex Genetic Traits.....	9
Definition of a Complex Genetic Trait.....	9
Genetic Markers.....	9
Linkage Analysis.....	10
Association Analysis.....	11
Genome-wide Screens.....	11
Candidate Gene Approach.....	13
Linkage Disequilibrium (LD).....	13
Haplotypes.....	15
Statistical Significance in Association Analysis.....	16
Summary Genetic Analysis of Complex Traits.....	16
Multiple Sclerosis as a Complex Genetic Trait.....	17
Candidate Gene Studies in Multiple Sclerosis.....	17
Genome-wide Screens for Linkage in Multiple Sclerosis.....	18
Genome-wide Screens for LD in Multiple Sclerosis.....	23
Aims of Study I – IV.....	24
Materials and Methods.....	25
Patients and Controls.....	25
Sib-pair Families (Study I & II).....	25
Patients and Controls (Study III & IV).....	26
Genetic Analysis.....	26
DNA Extraction.....	26
Genetic Markers.....	26
Genotyping.....	27
HLA-typing.....	28
Genetic Maps.....	28
Statistical Analysis.....	28
Linkage Analysis (Study I & II).....	28
Association Analysis (Study III & IV).....	28
Estimation and Illustration of LD (Linkage Disequilibrium).....	29
Results and Interpretations.....	30
Study I.....	30
Study II.....	31
Study III.....	33
Study IV.....	34
General Discussion.....	37
Genome-wide Screens.....	37
Haplotype Analysis.....	39
Lack of Replication and Publication Bias.....	41
Concluding Remarks and Future Perspectives.....	43

Acknowledgements	44
References.....	47

LIST OF ABBREVIATIONS

APC	antigen presenting cell
APOE	apolipoprotein E
BBB	blood-brain barrier
bp	base pair
CD	cluster of differentiation molecule
CDCV	common disease common variant
CEPH	Centre d'Etude du Polymorphisme Humain
CNS	central nervous system
cM	centimorgan
CSF	cerebrospinal fluid
DNA	deoxyribonucleic acid
DZ	dizygotic (twin)
EAE	experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
EM	expectation maximization (algorithm)
HHV-6	human herpes virus 6
HLA	human leukocyte antigen
ht-SNP	haplotype-tagging SNP
IBD	identical by descent
IDDM	insulin dependent diabetes mellitus
IFN	interferon
Ig	immunoglobulin
IL	interleukin
kb	kilobase
LD	linkage disequilibrium
LOD	logarithm of odds
Mb	megabase
MHC	major histocompatibility complex
MLS	maximum lod score
MS	multiple sclerosis
MRI	magnetic resonance imaging
MZ	monozygotic (twin)
NCBI	National Center of Biotechnology Information
NPL	nonparametric linkage
PCR	polymerase chain reaction
PP-MS	primary progressive multiple sclerosis
RR-MS	relapsing-remitting multiple sclerosis
SNP	single nucleotide polymorphism
SP-MS	secondary progressive multiple sclerosis
TDT	transmission disequilibrium test
Th	T-helper cell
TNF α	tumour-necrosis factor alpha
UK	United Kingdom

INTRODUCTION

MULTIPLE SCLEROSIS

Multiple Sclerosis (MS) is a chronic neurological disease affecting the central nervous system (CNS). The disease mainly has its onset in young adults between the age of 20 and 40 years. There is a female predominance with about two thirds of the affected being women (reviewed in [1, 2]). The disease is heterogeneous. The disease course varies and is difficult to predict on the individual level. About 80% of the patients have a relapsing-remitting disease (RR-MS) at onset. Usually there is good recovery from the initial relapses/bouts, but over the time most patients will eventually move into a progressive phase with remaining neurological disability between the bouts: secondary progressive MS (SP-MS). About 15 to 20% of patients with MS have a progressive disease already from the disease onset without any relapses: primary progressive MS (PP-MS) [1, 2].

It has been debated whether primary progressive multiple sclerosis is a distinct different disease than the other forms of multiple sclerosis [3-5]. However, recent epidemiological data have indicated that the progressive phase of the disease is very similar regardless of the disease-course at onset as reviewed by Confavreux [6] and this view is probably the most accepted among neurologists today.

There are a variety of clinical symptoms and signs associated with multiple sclerosis including sensory disturbances, optic neuritis, limb weakness, unsteadiness due to ataxia, neurogenic bladder and bowel symptoms. Studies attempting to find factors predicting the disease course have shown indications that younger age at onset, female gender, optic neuritis or sensory disturbances as first bouts as well as complete recovery from individual episodes would predict a more benign course, while older age at onset, male gender, motor involvement together with only partial recovery from bouts would indicate a more severe course [7]. It has however also been shown that most important factor determining disability is onset of a progressive phase [8].

Diagnosing multiple sclerosis

The general basis in the diagnosis of multiple sclerosis is that the patient should have had two or more clinical bouts of neurological symptoms and signs affecting more than one anatomical site, i.e. the disease should show dissemination in time and space (anatomically). The definition of a relapse or “bout” is an episode of neurological symptoms with a duration of at least 24 hours. Clinical history and examination has been the basis in the diagnosis of multiple sclerosis for a long time. Other paraclinical investigations used for establishing the diagnosis include Magnetic Resonance Imaging (MRI), investigation of cerebrospinal fluid (CSF) and (to a lesser extent) evoked potentials.

MRI is nowadays regarded as “gold-standard” in the diagnosis of multiple sclerosis. A typical MRI-image of a patient with MS usually shows multifocal white matter lesions of different age and size. The lesions are often found in the periventricular white

matter, corpus callosum, cerebellum and brainstem as well as the spinal cord. Gadolinium-enhancement is regarded as an indication of active lesions (sites with inflammation). MRI-criteria to aid in the diagnosis have been established [9, 10]. Investigation of cerebrospinal fluid in patients with multiple sclerosis, shows abnormalities including presence of oligoclonal IgG bands on isoelectric focusing (without any concomitant bands in serum), and / or the presence of an elevated IgG index [11, 12]. Lymphocytic pleocytosis should generally not exceed 50/mm³. In some instances evoked potentials can be used to assist in diagnosis.

The heterogeneity of clinical symptoms and signs between the individual patients with multiple sclerosis together with the fact that there is not one single feature or test that is sufficient for the diagnosis has led to a need for clear diagnostic criteria. Stringent diagnostic criteria are also important in order to be able to design and evaluate clinical trials for new treatments. Through-out the years, three main criteria have been used. The first, the “Schumacher criteria” were published in 1965. They were later replaced by the “Poser-criteria” in 1983 [13] which are still widely used today. The latest revision of the diagnostic criteria of multiple sclerosis was published in 2001; the so called “McDonald-criteria” [14]. In these criteria, MRI has received a more prominent role, in establishing dissemination in time and space. The basis for diagnosis in all three criteria is the fact that, for the diagnosis of multiple sclerosis, the person should have neurological symptoms typical for multiple sclerosis with dissemination in time and space and other differential diagnoses should have been ruled out.

Current treatments

In Sweden there are currently four pharmaceutical products approved for disease modifying treatment in multiple sclerosis; three are β -interferons (IFN- β 1a: Avonex® [15] and Rebif® [16]; and IFN- β 1b: Betaferon® [17]) and the fourth is glatiramer acetate (Copaxone®) [18, 19]. (The modes of action of these substances are described later in this thesis.) The effect of the treatment is mainly a reduction in the number and severity of bouts. All substances registered as disease modifying treatments in MS today are approved only for disease with symptoms and signs indicating active inflammation for instance by frequent relapses. There is currently no substance registered for treatment of the progressive phase of the disease.

Very recently (Nov 24, 2004), a new drug, Natalizumab [20](Tysabri®), with a slightly different mode of action was approved in the United States by the FDA (U.S. Food and Drug Administration). <http://www.fda.gov/cder/drug/infopage/natalizumab/default.htm>

Immunology and Pathology

Multiple sclerosis is usually regarded as an autoimmune disease, where inflammatory factors play an important role in the disease pathogenesis. However, the exact events of these mechanisms are still the focus of intense research.

Studies in the animal model of multiple sclerosis; EAE (experimental autoimmune encephalomyelitis) have played an important role in deciphering the underlying immunological mechanisms behind MS. EAE is induced by immunization of the

animal (most often rodents; mice or rats) with myelin antigens and Freund's adjuvant. This initiates an inflammatory disease of the central nervous system leading to variable degrees of demyelination. The disease course and severity of EAE are strain dependent, which has been utilized for genetic linkage analysis.

T-cell mediated disease

Multiple sclerosis is considered as a predominantly T-cell mediated inflammatory disease. There are several lines of evidence supporting this; the probably most well-known is that EAE can be passively transferred by myelin-reactive T-cells [21], the so called "adoptive transfer" concept. Reviews by Noseworthy et al. [2] and O'Connor et al. [22] describe the chain of events involved in the disease process. The first step is an activation outside the CNS of autoreactive T-cells by for instance foreign microbes or self proteins. These autoreactive T-cells are specific for myelin antigen (e.g. myelin-associated protein (MAG), myelin basic protein (MBP), myelin oligodendrocyte glycoprotein (MOG) or proteolipid protein (PLP)). The expression of endothelial adhesion molecules (e.g. ICAM-1 and VCAM-1) is upregulated. Activated T-cells express integrins that bind to the adhesion molecules. This complex interaction between the T-cell and the adhesion molecules in the endothelial wall together with local chemokines and metalloproteinases enables the passage of T-cells through the otherwise tightly regulated blood-brain barrier (BBB) into the central nervous system (CNS) [23, 24]. Once inside the CNS, the T-cells encounter their target antigens again and are re-activated by local antigen presenting cells (APC). T-cells may be classified into Th1 or Th2 cells [25]. Th1 and Th2 cells each have a distinctive cytokine secretion pattern [26]. Th1 cells produce the pro-inflammatory cytokines: interferon-gamma (IFN γ), tumour necrosis factor alpha (TNF α) and interleukin-2 (IL-2), which are disease promoting. Th2 cells on the other hand produce anti-inflammatory cytokines such as IL-4, IL-5, IL-6, IL-10 and IL-13, which down-regulate local inflammation and regulate humoral immunity.

Antigen presentation

An important step in T cell mediated immunity is the antigen presentation. T cells are activated by an interaction between its T cell receptor, the MHC molecule on an antigen presenting cell (APC) and a peptide bound to the MHC (the "first signal"). However, the activation is dependent also on a second signal largely provided by an interaction between CD28 molecules on the T cell with the corresponding ligands on the antigen presenting cell: B7-1 (CD80) and B7-2 (CD86).

Major Histocompatibility Complex (MHC) class I molecules which are expressed on most nuclear cells in the body, present *endogenous* antigens to CD8⁺ T cells (cytotoxic T-cells). MHC class II molecules present *exogenous* antigens to CD4⁺ T cells (helper T-cells). MHC class II molecules are however only expressed by professional antigen presenting cells (APC), for instance dendritic cells, monocytes/macrophages and B cells. Microglia can also act as APCs [27].

The human MHC is also referred to as HLA (human leukocyte antigen). As discussed later in this thesis, genetic studies have shown association between multiple sclerosis and HLA class II (and class I) alleles.

Demyelination

Demyelination is the result of several mechanisms, including immune-mediated effects by inflammatory cytokines, macrophages or T cells, as well as an antibody-mediated damage to the myelin and complement-mediated injury. The destruction of the myelin sheath leads to reduced conduction velocity (loss of saltatory conduction) in the affected nerve, giving rise to clinical symptoms and signs typical of the disease. It is hypothesized that fast initial recovery from a relapse is due to an increase in the number of sodium-channels to restore the conduction, before remyelination can take place.

Th1-Th2

Multiple sclerosis has generally been regarded as a Th1 type mediated autoimmune disease [28], caused by an imbalance between Th1 and Th2 responses. It appears however that the underlying mechanism is not simply a shift from Th1 to Th2. Although a deviation from Th1 to Th2 has been shown to ameliorate EAE, studies also point towards an involvement of Th2 in the disease pathogenesis as well. One apparent evidence against a pure Th1 model is the fact, that when patients with multiple sclerosis were given anti-TNF α treatments, their disease was aggravated [29].

B-cells

Despite the strong evidence for involvement of T cells in the pathogenesis of multiple sclerosis, there is increasing support for the involvement of B cells as well (reviewed by Cross [30] and [31]). The presence of oligoclonal bands in the cerebrospinal fluid (and concomitant absence of oligoclonal bands in peripheral blood) has for a long time been part of the diagnostic procedure. However, the most persuasive evidence for a role of B cells and antibody mediated demyelination is based on histopathological studies showing that B cells and plasma cells are found in areas of active myelin destruction as well as in chronic lesions.

Mechanisms of disease modifying treatments

As described earlier, there are currently four different pharmaceutical products approved as disease modifying treatment in multiple sclerosis in Sweden; interferon- β (IFN- β 1a: Avonex $\text{\textcircled{R}}$ [15] and Rebif $\text{\textcircled{R}}$ [16]; and IFN- β 1b: Betaferon $\text{\textcircled{R}}$ [17]) and glatiramer acetate (Copaxone $\text{\textcircled{R}}$) [18, 19]. Interferon- β exerts its effect in several ways; decreased production of TNF α , decreased antigen presentation, a general shift towards a Th2 environment with increased secretion of IL-10. It also affects movement of T cells into the CNS by inhibiting the activity of T-cell matrix metalloproteinases. Glatiramer acetate is a synthetic polypeptide that promotes the proliferation of Th2 cytokines and inhibits antigen-specific T-cell activation.

Natalizumab, recently approved in USA (Tysabri $\text{\textcircled{R}}$), is a humanized monoclonal antibody against α 4 integrin [20]. Its mechanisms of action are by disturbing the

adhesion molecule interaction, thereby inhibiting the migration of leukocytes through the BBB, and by interfering with T-cell activation/re-activation.

Pathology / Atrophy

The typical histopathological features of MS lesions are multifocal, large, demyeliated plaques with reactive glial scar formation together with inflammatory infiltrates.

Multiple sclerosis is often defined as a demyelinating disease with relative sparing of axons. It was however shown already in the late 1800s [32] that there is axonal destruction as well. It has later been shown that axonal damage and transection occurs in both acute and chronic plaques as a consequence of demyelination. Axonal loss can indirectly be observed as atrophy (for instance on MRI). There is an increased rate of atrophy in the brain and spinal cord at all stages of the disease, evident already early in the disease course [33].

Lassmann and Lucchinetti et al. have performed detailed pathological studies [34, 35] on MS lesions and found heterogeneity in lesion “patterns”. They defined four patterns of lesions: Type I with macrophage mediated demyelination, Type II with antibody mediated demyelination, Type III with distal oligodendroglipathy and Type IV pattern involving primary oligodendrocyte damage with secondary demyelination. Pattern IV was only found in a few patients with a primary-progressive disease, whereas all patients with Devic’s neuromyelitis optica had pattern II. Otherwise there was no specific association any specific pattern and clinical disease course.

Epidemiology

The prevalence of multiple sclerosis varies considerably around the world. Kurtzke classified regions of the world according to prevalence: *low prevalence* was considered less than 5 cases per 100,000; an *intermediate prevalence* was 5 to 30 per 100,000, and *high prevalence* more than 30 per 100,000 inhabitants (reviewed in [36]). Generally, the prevalence is highest in northern Europe, southern Australia and the middle part of North America. The prevalence and incidence of multiple sclerosis varies with latitude with highest figures at the extremes of latitude in both the northern and southern hemispheres.

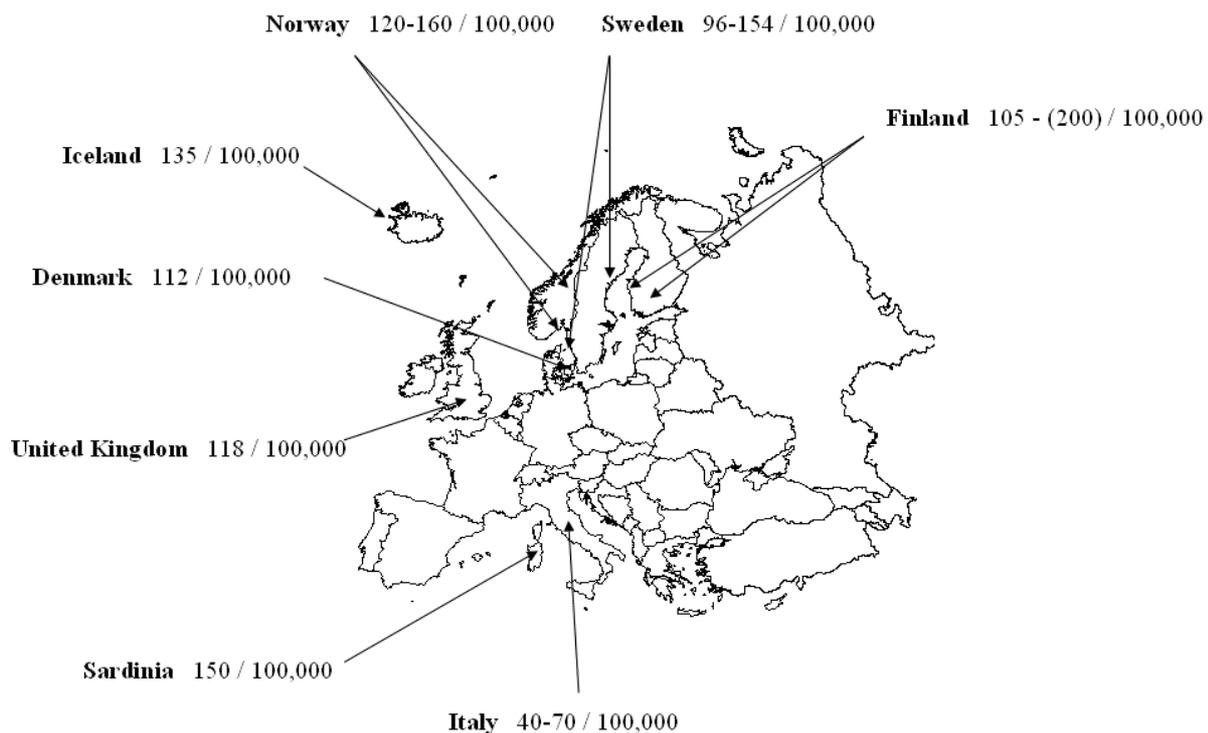


Figure 1 Prevalence of multiple sclerosis in the countries discussed in this thesis. The following publications are cited: Norway: Dahl et al. [37], Celius et al. [38], Sweden: Svenningsson et al. [39], Sundström et al. [40], Finland: Sarasoja et al. [41], Sumelahti et al. [42], Iceland: Benedikz et al. [43], Denmark: Koch-Henriksen et al. [44], United Kingdom: Fox et al. [45], Robertson et al. [46], Italy: Pugliatti et al. [47], Sardinia: Pugliatti et al. [47, 48].

The cause of the variation in the prevalence and incidence of multiple sclerosis worldwide is not completely understood. Environmental and genetic explanations have been put forward, and both factors probably have a role.

The high prevalence of multiple sclerosis in Scandinavia (Fig 1) and the striking correlation between its global geographical distribution and migration pattern of northern Europeans, could suggest that susceptibility genes may have arisen in Scandinavia and been disseminated by their descendants. This idea was suggested as the so-called “Viking hypothesis” by Poser [49, 50] favouring a genetic component of the disease. Kurtzke on the other hand [36] interpreted this as being caused by an “acquired, exogenous, environmental disease with a prolonged incubation period” originating from Scandinavia and spread overseas by migrations. Kurtzke partly bases his conclusion on studies of MS prevalence in the Faroe Islands. The studies of the Faroe Islands showed no cases of multiple sclerosis before 1943. The appearance of the first cases of MS occurred a few years after British troops had occupied the islands during World War II. Kurtzke hypothesize that these cases of multiple sclerosis were caused by a transmissible infection with a prolonged incubation period between acquisition and clinical expression with a thus-far unidentified agent.

Other “clusters” or “epidemics” have been claimed, for instance in Iceland [51, 52], Shetland and Orkney Islands [53] and Key West (Florida, USA) [54]. Many of these

reports including the “Faroe epidemic” have however been widely questioned and re-examination have often not supported the initial observation [55, 56]. It has also been put forward that “true” increase in prevalence rates are more likely caused by an improved ascertainment and increased awareness among physicians [56].

Interestingly, within areas with high prevalence of multiple sclerosis, there are some sub-populations in which multiple sclerosis is very rare, for instance the Sami population of northern Scandinavia [57] and the Maoris in New Zealand [58].

Studies of people who migrate from high- to low risk areas have shown that those migrating after the age of 15 keep the risk of their birthplace, while those migrating before the age of 15 acquire the lower risk of their new residence (reviewed in [59]). These migration studies have been regarded as support for an environmental factor responsible for the disease. The validity of the migration studies have however been criticized with the notion that migrants seldom are representative of the general population of the country which they have left [60].

Another factor that has been regarded as support of an environmental factor involved in disease pathogenesis is the fact that the prevalence of multiple sclerosis tends to follow latitudinal gradients with highest prevalence figures furthest away from the equator. This latitudinal gradient is supported by studies from both the northern hemisphere (USA) [61, 62] and the southern hemisphere (Australia) [63-65]. Critics however point to the importance of population ancestry rather than strict geographical factors [66].

Among the environmental agents studied in multiple sclerosis, most of the studies have investigated different viruses and the possible involvement in the disease. Some of the viruses most often discussed in this aspect are human herpes virus 6 (HHV-6) [67, 68], Epstein-Barr virus (EBV) [69-71] and Chlamydia pneumoniae [72, 73]. The results of the studies are somewhat conflicting and so far, no viruses have unequivocally been confirmed to be causative.

In a recently published article, Ebers [74] et al. found a seasonal birth effect, in that there was a higher risk of getting MS for people born in the month of May. They hypothesize that this could be explained by environmental factors, such as exposure to sunlight and its interaction with vitamin D.

Genetic Epidemiology

Different epidemiological methods have been applied in order to establish a genetic component in the aetiology of multiple sclerosis. About 15 to 20 % of patients with MS have one or more affected relative [75, 76]. Familial aggregation of a disease is often evaluated by studying the relatives of the affected person (proband) and establishing whether they are at higher risk of getting the disease than the normal population. This alone however, can not differentiate between shared environment and genetic aetiology.

Recurrence risk is one method often used to describe familial aggregation of a disease. Studies of relatedness in multiple sclerosis have shown that first, second and third degree relatives of people with multiple sclerosis are more likely to have the disease than the normal population. These recurrence risks vary with relatedness. One measurement often used to evaluate recurrence risk is λ_R (the ratio of the risk to relatives of type R (e.g. siblings, offspring) compared to the population risk. The most commonly used is λ_S (recurrence risk in siblings):

$$\lambda_S = \frac{\text{risk for siblings of an affected individual}}{\text{general population prevalence}}$$

In multiple sclerosis λ_S is estimated to be 20-40 (0.02-0.04/0.001) [77]. (Some examples of λ_S in other diseases are: $\lambda_S \sim 500$ for Cystic fibrosis; $\lambda_S 15$ for IDDM and $\lambda_S 4-5$ for Alzheimer's disease). In general, a $\lambda_S > 2.0$ is thought to indicate a genetic component. However, as λ_S is dependent on both the increased risk in related individuals and the general population risk, a strong effect in a very common disease will have quite a small λ_S .

Identical (monozygotic – MZ) twins are genetically identical, whereas fraternal (dizygotic – DZ) twins share on average half of their genes, just like siblings. Comparing concordance rates between MZ and DZ twins gives an estimation of the role of genetic factors involved in the disease. If the concordance rate of a disease is 100% in MZ and 25-50% in DZ twins, the disease would be regarded as strictly genetic and probably due to a single gene. In diseases where genetic factors are important, although not the only underlying factor, the concordance rate for MZ twins is greater than for DZ twins. In twin studies, ascertainment is an important issue, and the studies should preferably be population-based. Several twin studies have been performed in multiple sclerosis, most of which show a higher concordance rate in MZ twins (25-30%) compared to DZ twins (2-5%) [74, 78, 79] indicating the involvement of genetic factors.

Adoption-studies in multiple sclerosis [80] have shown that an adoptive relative raised together with a person affected by multiple sclerosis is no more likely to develop multiple sclerosis than the general population. This indicates that the familial aggregation of multiple sclerosis is related to genetic factors rather than to shared familial environment.

Half-siblings share on average 25% of their genetic background compared to 50% in full siblings. This difference enables a test of the effect of genetic sharing on recurrence risk. One can also compare half-siblings reared together with or apart from the index patient. In multiple sclerosis, the half-sibling recurrence risk (1.32%) is significantly less than that for full-siblings (3.46%) in the same family [81] giving support to genetic sharing instead of environmental factors.

Two studies [82, 83] have shown a significantly higher risk for off-springs of two persons with multiple sclerosis compared to that for offspring with only one affected parent (30.5% compared with 2.49% respectively in the Canadian study [83]). In a

study of 13,000 spouses of multiple sclerosis patients, Ebers et al. [83] showed that the rate of multiple sclerosis among them was no different from that expected, based on the general population risk.

COMPLEX GENETIC TRAITS

Definition of a Complex Genetic Trait

The term “complex genetic disease” refers to any phenotype that does not exhibit classic Mendelian inheritance attributable to a single locus. Most of our common diseases such as diabetes mellitus, hypertension etc. are regarded as complex genetic diseases. The underlying background to complex diseases is most probably an interplay between several susceptibility genes interacting with other factors for instance environment and probably stochastic factors as well. The genes involved in complex genetic traits are often referred to as susceptibility genes to distinguish them from the causative genes in Mendelian monogenic diseases.

Today over 1,200 gene variations responsible for human diseases have been identified, almost all of which are Mendelian diseases. For complex genetic diseases or traits however the work has been less rewarding. In Mendelian diseases where a disease “runs in a family” it is not uncommon to find large pedigrees with many affected family members. In complex diseases, large pedigrees with many affected family-members are only rarely encountered. If a large pedigree with an assumed complex disease is found, the first question to ask is whether the members of this pedigree actually have the complex disease in question or whether they are affected by another less common Mendelian trait. In these situations it is particularly important to consider other differential diagnoses.

Genetic Markers

Genetic markers reflect natural sequence variations in the genome. The two main markers used for genetic analysis today are microsatellites and SNPs (single nucleotide polymorphisms).

Microsatellites

Microsatellites are short (2-6 bp), tandemly repeated DNA sequences that are widely spread in the genome. Based on the size of the tandem repeat unit, the microsatellites are divided into di-, tri- and tetranucleotide repeats. Dinucleotides are most frequent, followed by tetranucleotides while trinucleotides are the least common form. Microsatellites are highly polymorphic and have been widely used for genetic analysis, especially in genome wide screens. The predominant mutational mechanism underlying the development of microsatellites is thought to be DNA polymerase slippage during replication. The mutation rate is much higher than for SNPs: up to 10^{-3} per site per generation [84] compared to an average of 10^{-8} for SNPs [85]. The rate of mutation however differs between genetic locations.

Single-nucleotide polymorphisms (SNPs)

Polymorphisms are variations in the genome sequence. The most abundant and commonly used polymorphisms are the SNPs (single nucleotide polymorphism). SNPs are single base pair positions in the genomic DNA at which different alleles exist in normal individuals. The effect of a SNP is dependent on its location in the genome; e.g. whether located in coding sequences, non-synonymous/synonymous etc. SNPs can be disease-causing, but are most commonly a part of the human variation. They are more common than any other polymorphisms and occur at a frequency of approximately 1 in 1,250 base pairs throughout the genome. The bi-allelic nature of SNPs has made them amenable to high-throughput genotyping. There are currently (Dec 4th, 2004) 10 million (5 million validated) SNPs identified and publicly available at NCBI – dbSNP build 123 (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi).

Genetic analysis is commonly divided into “Linkage analysis” and “Association analysis”.

Linkage Analysis

Classical linkage analysis



In multigenerational pedigrees with many affected family-members, *classical linkage analysis* (model-dependent) is commonly used to find the gene responsible for the disease in this specific family. Linkage is the tendency for two loci on the same chromosome to be inherited together more often

than by chance alone. Linkage analysis can be very powerful, provided that one can specify the correct model (mode of inheritance) in terms of parameters such as penetrance and disease allele frequency. As mentioned above, large multigenerational families with a complex trait are rare, therefore the focus has instead turned towards smaller nuclear families, especially affected sib-pairs. In this case, no mode of inheritance can be determined and classical linkage analysis is not fully adequate. Instead, non-parametric (or model-free) methods have been developed.

Non-parametric sib-pair methods were first developed by Penrose already in 1935 (reviewed by Holmans [86]). Over the years, Penrose’s method has been further developed but the basis is still the same: if there is linkage, affected individuals will be more similar in those parts of the genome close to a disease susceptibility gene than would be expected by chance [86].

Allele sharing methods



The sib-pair methods are often referred to as *allele-sharing methods*. The most common allele-sharing method is the “affected sib-pair test”, which compares the observed number of affected sib-pairs sharing zero, one or two alleles identical by descent (IBD) with that expected under no linkage: $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$

(for 0, 1 or 2 alleles IBD respectively). IBD-status can however often not be determined unequivocally, for instance when genotyping data from parents and grandparents is not available. Therefore, the most commonly used programs today are based on a likelihood-ratio method developed by Risch [87, 88], which maximizes the likelihood of the data with respect to the probabilities of pairs sharing 0, 1 or 2 alleles IBD.

Risch's method was further developed by Holmans [89] who restricted the maximization to the set of IBD probabilities consistent with possible genetic models (the "possible triangle" method). This method is implemented in SPLINK [90] and MAPMAKER/SIBS [91].

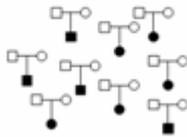
Association Analysis

Case-control study



Association studies test whether a particular allele occurs at higher frequency among affected than unaffected individuals. Basically, there are two main types of association studies. The first is *case-control studies*, in which a comparison is performed between the allele frequencies in a set of unrelated affected individuals to that in a set of unrelated controls.

Family-based (TDT)



The other type of association study is the *family-based approach*; most commonly the TDT-test (transmission disequilibrium testing) in which trios (an affected individual + both parents) are studied. TDT investigates whether the frequency of alleles transmitted from heterozygous parents to affected offspring is significantly different to the frequency of the non-transmitted alleles [92]. In general, family-based methods are less powerful than case-control studies, but because of the intra-familial comparisons, they are less susceptible to population stratification, which potentially can be a problem in case-control studies.

From a methodological point of view genetic studies can be categorized into *genome-wide screens* and *candidate gene approaches*.

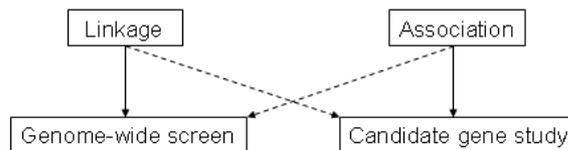


Figure 2 Different applications of linkage and association analysis. Hitherto, genome-wide screens have mainly been based on linkage. In the near future it is likely that more studies (especially for *complex* genetic traits) will be based on association. Linkage has been used for candidate gene studies, however today most candidate gene studies are based on association analysis.

Genome-wide Screens

Genome-wide screens for linkage have been successfully used to locate disease-causing genes for monogenic/Mendelian diseases. Although the genetics is more complicated in complex genetic diseases, performing a genome-wide screen is usually the first step towards gene identification here as well. Genome-wide screens or scans are performed by typing a set of microsatellite markers evenly spread along the genome. The markers should preferably be highly polymorphic to increase the likelihood of informative meioses. An average marker-density of one microsatellite every 10cM (a total of 300-

400 markers) has been accepted as a “gold standard” for genome-wide screens. Today there are commercially available pre-designed “Genome-mappings sets”. Wherever a region of interest is found the scientist usually saturates the region with a denser set of markers in an extended number of families.

Genome-wide screening has the advantage that it is performed in a hypothesis independent manner and therefore does not require any prior knowledge of the underlying disease mechanisms.

There are different ways of performing genome-wide screens in order to use the data as efficiently as possible in a cost-effective way. Hauser [93] suggested dividing the work into two stages. In the first stage an unbiased genome-wide screen is performed. Regions of interest from the first stage is then followed-up by either typing additional families or family members in these regions, or by adding additional markers in the region of interest (or both). One can choose whether or not to include other family members than the affected sib-pairs in the screening phase [89, 93]. Parents or unaffected siblings contribute to the information content by adding information about phase. However, this gain in information also has a cost as there is an increase in genotyping effort. Therefore one approach is to genotype these additional family members (where available) in a second stage. Apart from contributing to information about the phase, genotyping unaffected relatives also helps in detecting genotyping errors.

In the case of pedigrees with a seemingly Mendelian mode of inheritance, parametric methods are more powerful and should be used for analysis. In other cases, such as studies with affected sib-pairs, parametric methods can still be used, however, if there is no clear mode of inheritance, non-parametric methods are more commonly used. To further increase the information from the data and to facilitate the localization of the peaks, multi-point analysis, can be performed.

Statistical significance levels in genome-wide screens

One important issue often discussed regarding genome-wide screens is how to define the level of statistical significance when so many statistical tests are performed at the same time. Lander and Kruglyak stated in an often cited article from 1995 [94] the difference between point-wise significance levels and genome-wide significance levels. The point-wise or nominal significance level is the probability that one would encounter a peak with such LOD-score *at a specific locus* just by chance. The genome-wide significance is the probability that one would encounter such a LOD score *somewhere in the whole genome scan*.

Lander and Kruglyak [94] stated the following guidelines for assigning genome-wide significance:

- *Suggestive linkage* – statistical evidence that would be expected to occur one time at random in a genome scan.
- *Significant linkage* - statistical evidence expected to occur 0.05 times in a genome scan (that is with probability 5%)

- *Highly significant linkage* - statistical evidence expected to occur 0.001 times in a genome scan.
- *Confirmed linkage* – significant linkage from one or a combination of initial studies that has subsequently been confirmed in a further sample, preferably by an independent group of investigators. For confirmation, a nominal p-value of 0.01 should be required.

In the case of sib pair studies the first three categories would correspond to point-wise lod scores of 2.2; 3.6 and 5.4. These levels have later been regarded as slightly too conservative. Sawcer et al. [95] performed simulations based on the first UK genome-wide screen in MS. These simulations were based on real data and were slightly less conservative. They also reported a maximum lod score level (MLS) for nominal significance of $MLS > 0.7$, which is consistent with the level predicted theoretically [89].

Candidate Gene Approach

Instead of performing an unbiased non-hypothesis driven screen of the whole genome in search for susceptibility genes, the candidate gene approach focuses on genes chosen on the basis of existing knowledge of the pathogenic mechanisms. Association studies can be based on either positional or functional candidate genes (or both). Positional means that the gene of interest is located in a region of a peak from a genome-wide screen. Functional candidates are chosen based on their expected involvement in the disease process or pathways (based on results from other biological studies). Candidate gene studies are today almost exclusively based on association analysis.

Association analysis can be performed as either a direct test of association, i.e. it is the polymorphisms in question that may have the functional consequences that gives rise to the phenotype (or disease) in question. The alternative is to perform an indirect test, which means that the marker tested is in very close genetic proximity to the variant or mutation that is responsible to the functional outcome. Indirect testing relies on so called linkage disequilibrium (LD).

Linkage Disequilibrium (LD)

Linkage Disequilibrium (LD) is the non-random association of alleles at adjacent loci, i.e. two specific alleles at two closely located loci are found together on the same chromosome more often than would be expected by chance. The extent of LD is dependent on several factors both on the molecular level such as recombination and mutation rate as well as demographic and evolutionary factors such as migration, population growth and admixture between populations. LD is expected to be higher in populations derived from relatively few founders, so called “population isolates”. Example of populations that in this sense have been regarded as population isolates are Sardinians [96], French Canadians [97] and some parts of Finland [98, 99].

There are two main ways of measuring the level of linkage disequilibrium, both of which are based on Lewontin’s D [100]:

- $|D'|$ (the absolute value of D')
- r^2

The first frequently used measure of linkage disequilibrium is $|D'|$ as illustrated in Fig. 3. D is the difference between the *observed* haplotype frequency “A-B”, and the frequency it would be expected to have if the alleles were randomly segregating [$f(A1)f(B1)$]. D is obtained by dividing D with D_{\max} (the maximum D possible for a given set of allelic frequencies at the two loci).

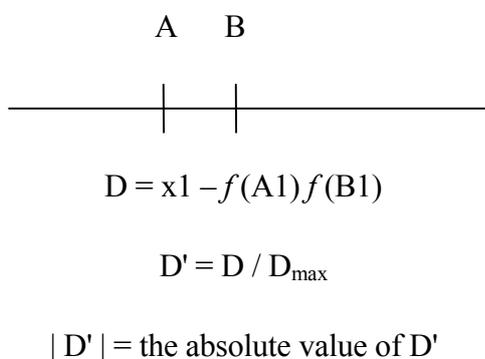


Figure 3a Two loci named “A” and “B” are located close to each other on the same chromosome. The formula above describes the calculation of $|D'|$. The term “ $x1$ ” is the observed haplotype frequency of the haplotype “A-B”. $f(A1)$ is the frequency of allele A1 and $f(B1)$ the frequency of allele B1.

$|D'| = 1$ is called complete LD. One main drawback is that $|D'|$ values can be highly inflated, in the case of small sample sizes. It is also sensitive to allele frequencies, and can be inflated for SNPs with rare alleles. Intermediate values of $|D'| < 1$ can be rather difficult to interpret.

The other commonly used measure of linkage disequilibrium is r^2 (or Δ^2).

$$r^2 = \frac{D^2}{f(A1)f(A2)f(B1)f(B2)}$$

Figure 3b Formula for calculating r^2 . Based on the same loci “A” and “B” as above, $f(A1)$ and $f(A2)$ are the alleles at locus “A” while $f(B1)$ and $f(B2)$ are the alleles at locus “B”. (“Genetics of Populations” 2nd edition Philip W. Hedrick. Jones and Bartlett Publishers 2000, p398-402)

r^2 is the correlation of alleles at the two loci as described above. $r^2 = 1$ is called perfect LD. The term r^2 has the advantage that it does not show the same inflation for rare alleles or small sample sizes. Intermediate values of r^2 are more easy to interpret than for $|D'|$.

The degree of LD between two alleles is dependent on how old the two polymorphisms are i.e. when they appeared in the population and the degree of recombination between them. Reich et al. [101] estimated the mean $|D'|$ in the genome to be 60kb by studying unrelated individuals from Utah, of north European descent. A study by Gabriel et al. [102] indicated that in European or Asian population half of the human genome exists in blocks of 44 kb or larger, while in African or African-American populations the figure was 22 kb.

Markers that are located close to each other, generally have higher LD than those located far apart. It has however recently been shown, that there is a high degree of variation in the extent of LD, so that “the variation overwhelms the mean”. There are regions with blocks of markers with high LD, so called “Haplotype blocks”. These block are broken up by areas with high recombination rates, “recombination hot-spots”.

Haplotypes

As a result of linkage disequilibrium, alleles located close to each other tend to be inherited together. A *haplotype* consists of a number of alleles that are inherited from a single parent. The length and number of involved alleles of the haplotypes varies between regions of the genome and is decreased by each generation (meiosis).

During the last few years there has been an increased interest in the use of haplotypes in the analysis of complex genetic traits. Empirical studies show that common haplotypes can capture most of the genetic variation in the region. If we for example have four bi-allelic markers (SNPs): A, B, C and D, they could theoretically form $2^4 = 16$ possible haplotypes. In a haplotype block however, about three to five haplotypes account for more than 90% of the observed chromosomes in a population.

Haplotypes can be deduced in family-based analysis by studying the inheritance patterns from parents to offspring. In case-control studies however, where the “phase” is unknown, the haplotype distribution has to be inferred by statistical methods. One of the most commonly used statistical algorithm for inferring haplotypes in phase-unknown data is the EM-algorithm (Expectation-Maximization), which is implemented in several haplotype-analysis programs.

The genome-wide extent of LD and the haplotype distribution in the genome has become the focus of intense studies in the hope that it might facilitate whole-genome association studies in complex human diseases. The success of this idea is highly dependent on a comprehensive knowledge of the patterns of LD throughout the genome. To facilitate this, “The International HapMap Project” [85] (<http://www.hapmap.org/>) was founded in 2002. This is an international consortium with the goal: “to determine the common patterns of DNA sequence variation in the

human genome and to make this information freely available in the public domain". Through the HapMap project 600,000 "common" SNPs will be genotyped in 270 individuals from four different populations (European ancestry, Yoruban (Nigeria), Japanese and Han Chinese), followed by additional genotyping in selected regions.

Statistical Significance in Association Analysis

The appropriate level of p-value in order to declare statistical significance is widely debated. Traditionally, when performing a test, a significance level of $\alpha < 0.05$ has been accepted as a level for declaring significance. This however, assumes that there is only one test performed which is rarely the case in genetic analysis, where many markers are tested, often under different models and with different sub-group analyses. In a study where k independent analyses are performed, the type I error rate is $1-(1-\alpha)^k$. In order to correct for the increase in risk of type I errors when multiple tests are performed, the "Bonferroni correction" is commonly used. In Bonferroni correction, the significance level is divided by the number of independent tests performed (α / k). This correction, however assumes that the tests performed are independent of each other, which is not the case when haplotypes or markers in LD are analysed. The level of significance is discussed further in "General Discussion".

Summary Genetic Analysis of Complex Traits

Many factors contribute to the complexity of complex genetic traits. There is not a straight forward correlation between genotype and phenotype; i.e. carrying a susceptibility allele does not fully predict disease status. There are individuals who carry the susceptibility allele, but who do not get the disease (*reduced penetrance*). There are also individuals who get the disease without having the susceptibility allele (*phenocopies*).

Two other important factors are locus- and allelic heterogeneity. In *locus heterogeneity*, a disease is caused by mutations or polymorphisms at more than one locus. *Allelic heterogeneity* on the other hand means that there are several alleles at the same locus that are causing the disease. Association analysis will be hampered by both of these forms of heterogeneity, while linkage analysis only will suffer if there is locus heterogeneity.

However, in some complex diseases, for instance breast cancer, families with many affected members have been found. In these specific families, the disease onset is usually very early in life. By focusing on families with early onset, scientists managed to find genes (BRCA1 and BRCA2) associated with hereditary susceptibility to breast cancer (reviewed in [103]). These genes are responsible for the disease in some of the families with early onset, but is not responsible for the sporadic cases of breast cancer. This illustrates one difficulty in what conclusion that can be drawn from pedigrees with many affected individuals (i.e. how representative a single-gene from a Mendelian pedigree is for the general "sporadic" population of patients). It is however important to note that even if the identified gene is not causing the disease in the majority of the patients; it can still give important information about the underlying disease mechanisms.

In conclusion the background of complex genetic traits is intricate. Apart from the already mentioned factors such as phenocopies, reduced penetrance, locus- and allelic heterogeneity, there are other factors influencing the phenotype of a complex genetic trait, such as clinical heterogeneity and complex interactions, both between genes and with environmental factors, and stochastic events.

MULTIPLE SCLEROSIS AS A COMPLEX GENETIC TRAIT

As described in the paragraph about genetic epidemiology, multiple sclerosis fulfills the "criteria" of being a complex genetic disease. A large number of candidate genes have been studied. Apart from studies in the HLA-complex, most candidate genes that have been claimed to be associated with susceptibility to multiple sclerosis have not been subsequently confirmed. Unfortunately, many studies have used small sample sizes and therefore suffer from lack of power to detect the effect we could expect to find for a gene associated with multiple sclerosis. Several recent reviews [104-107] have described and listed candidate gene studies in multiple sclerosis.

Candidate Gene Studies in Multiple Sclerosis

Already in the early 1970's, an association between multiple sclerosis and HLA complex was shown [108-110]. This association has been replicated and confirmed several times (reviewed in [110]). About 60% of people with multiple sclerosis have DRB1*1501,DQA1*0102,DQB1*0602, commonly called DR15, compared to 25-30% of the general population. These figures are based on people of Northern European descent. In Sardinia however, the main HLA-class II association is with DR3 (DRB1*0301,DQA1*0501,DQB1*0201), with 32% in patients and 22% in controls, followed by DR4 (DRB1*0405,DQA1*0501,DQB1*0301), with 6.6% in multiple sclerosis and 3.2% in controls. Interestingly, DR15 is also associated in the Sardinian multiple sclerosis-population, but to a much lesser extent and the prevalence is much lower (2.5% in multiple sclerosis and 1.5% in controls) [111-113].

In relation to clinical characteristics, DR15 has been shown to be associated with female gender and younger age at onset, but not to disease course [114-116]. Barcellos et al. [117] has demonstrated a dose effect of the HLA-class II haplotype DR15 (DRB1*1501,DQB1*0602) on the risk of multiple sclerosis. Carriage of one or two copies of the DR15-haplotype conferred a 2.7-fold and 6.7 fold risk respectively, compared to carrying no DR15-haplotype. These results have been confirmed in a Swedish population where a 3.0-fold and 8.3-fold risk respectively was shown [118].

A number of groups have also found indication of a second susceptibility locus including both alleles conferring an increased risk as well as protective alleles in the HLA-region class I region [119-121].

Another candidate gene that has been extensively studied in several neurological diseases, especially in relation to neurodegeneration is apolipoprotein E (APOE). It is well established that homozygous carriage of allele $\epsilon 4$ is associated with an increased risk of developing Alzheimer's disease [122, 123]. Studies in multiple sclerosis have shown conflicting results regarding disease susceptibility, however in studies looking at

disease-severity or progression instead of susceptibility; some studies show that the $\epsilon 4$ allele confers an increased risk of severe disease [124, 125]. (Both of these studies used MRI parameters indicating tissue destruction as an indicator of disease severity.) Another study, looking at clinical parameters, did not find any association with disease severity [126].

Many other candidate gene studies have been performed in multiple sclerosis, but the results have been either negative, conflicting or not confirmed in other populations.

Genome-wide Screens for Linkage in Multiple Sclerosis

The first three genome-wide screens for linkage performed in multiple sclerosis were published “back-to-back” in *Nature Genetics* 1996 [127-129]. Three independent research groups had studied American, British and Canadian families with multiple sclerosis. The first reaction on the results of these screens was somewhat negative as no regions with linkage of genome-wide significance were seen in any of the three screens. There were however, more regions with suggestive and potential linkage than expected by chance.

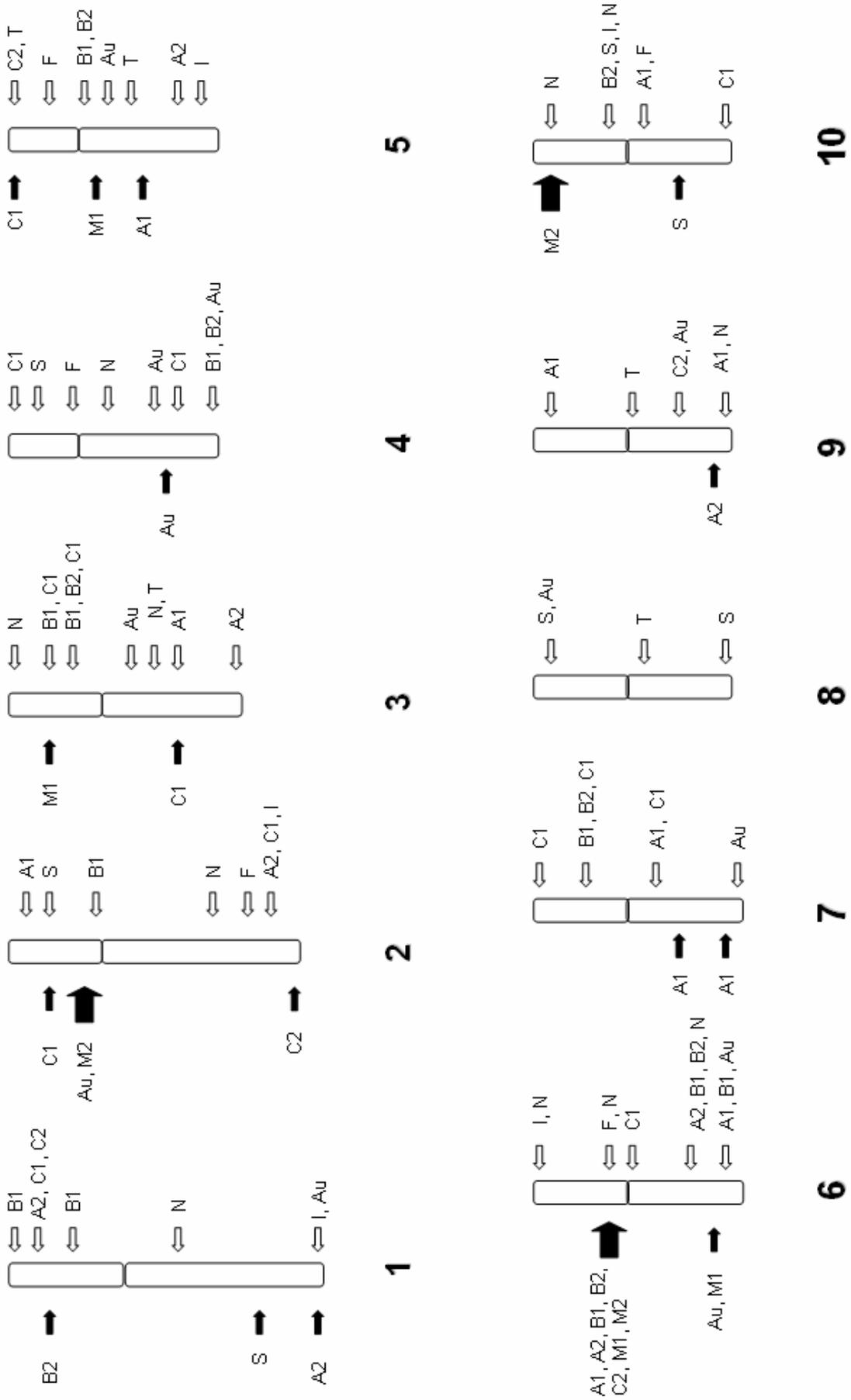
To date (Dec 2004), 12 full genome-wide screens for linkage from different populations have been published [127-138]. Two meta-analyses have been performed; the first one was based on the raw-data from the three original screens of the American, British and Canadian families [139], a total of 241 families (381 pairs) with multiple sclerosis. The following meta-analysis [140] is the largest to date and is based on the raw-data from 719 families (944 pairs) from the initial three screens together with the screens from Finland, Sardinia, Italy, the Nordic countries, Turkey and Australia. In contrast to the previous screens and the first meta-analysis, the second meta-analysis could show genome-wide significant linkage for the HLA-region on chromosome 6p21.

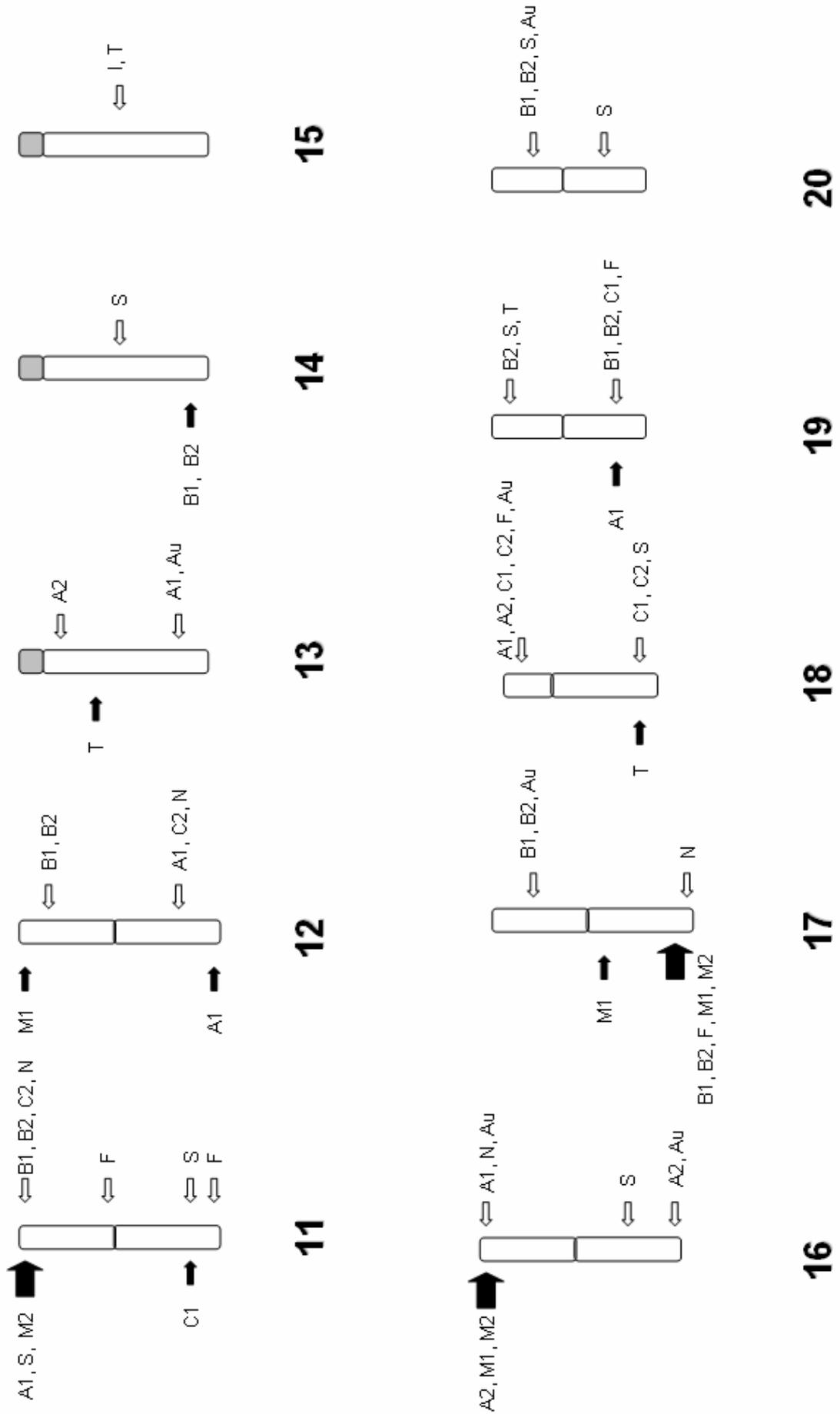
Table 1 (p19) The table shows all published genome-wide screens for linkage in out-bred populations in multiple sclerosis (Dec, 2004). The columns show the population studied, year of publication, number of families/sib-pairs genotyped in the first stage of respective screen. Some of the screens were performed in two stages (with different study-design between the studies). The number of families/sib-pairs included in the second stage (where applicable) is noted in the right-most column. Further, the number of markers typed in the screening phase is given, together with the analysis method used in each study. Depending on study-design and choice of analysis program, each study used different criteria for calling a peak “suggestive” (Level A) or “potential” (Level B) linkage. Note that as these criteria vary, it is not possible to directly compare them

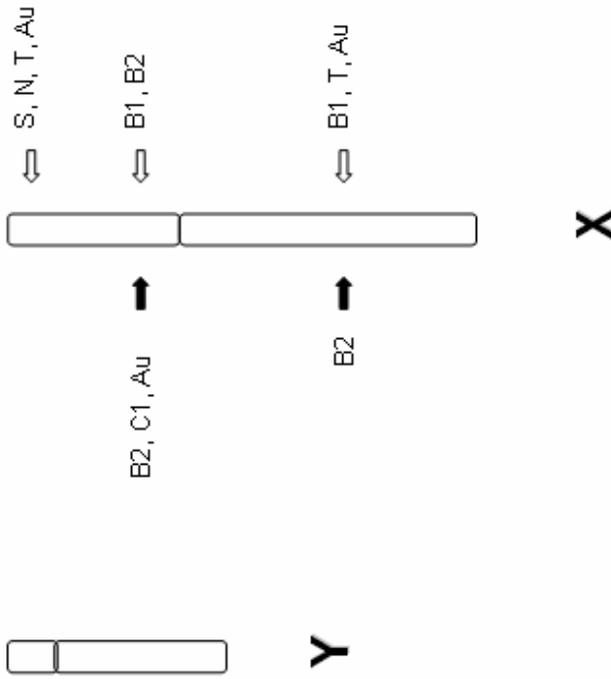
Figure 4 (p20-22) An illustration the location of peaks from the genome-wide screens for linkage listed in table 1. Each study is depicted with a letter referring to the first column in table 1 (the references are also listed at the end of this figure). The open arrows on the right side of each respective chromosome refer to peaks of “potential linkage”, while the filled arrows on the left side represent peaks of “suggestive” linkage. The larger filled arrows show the peaks from the latest meta-analysis [140].

	Population	Year	Families/sib-pairs	Markers	Analysis method	Level A	A	Level B	B	Comments
A1	US 1	1996	52 families (81 sib-pairs)	443	FASTLINK, SIBPAL, SimIBD	*1 (see below)	6	*1 (see below)	14	Stage 2:23 fam (45 sib-pairs)
A2	US/France 2	2004	246 families (466 affected relative pairs) US 1 & 2 are partially overlapping.	390	FASTLINK, HOMOG, ALLEGRO	*2 (see below)	4	*2 (see below)	9	
B1	UK 1	1996	129 families	311	MAPMAKER/SIBS	MLS>=1.8	6	MLS>=0.7	19	Stage 2:98 families
B2	UK 2	2003	226 families UK 1 & 2 are partially overlapping	353	MAPMAKER/SIBS	MLS>=1.8	5	MLS>=0.7	20	
C1	Canada 1	1996	100 sib-pairs from 61 families	257	ASPEX	MLS>1.0	5	MLS>0.5	20	Stage 2:114 fam (122 sib-pairs)
C2	Canada 2	2004	219 sib-pairs Canada 1 & 2 are partially overlapping	498	ASPEX	MLOD>2.27	1	MLOD>1.0	9	Stage 2:333 sib-pairs
F	Finland	1997	16 pedigrees from isolated population	328	GENEHUNTER (<i>no point for stage 2</i>)			NPLall>1.0	10	Stage 2:5 families
S	Sardinia	2001	49 multiplex families	327	MAPMAKER/SIBS	MLS>=1.8	3			
I	Italy	2001	40 multiplex families	322	GENEHUNTER-PLUS			MLS>=0.7	8	
N	Nordic	2002	136 sib-pairs	399	MAPMAKER/SIBS			MLS>=0.7	17	Stage 2:unaffected relatives
T	Turkey included 43 families (30%) with consanguinity	2003	43 multiplex families	392	GENEHUNTER-PLUS	MLS>=1.8	2	MLS>=0.7	14	Stage 2:unaffected relatives
Au	Australia	2002	54 sib-pairs	397	MAPMAKER/SIBS	MLS>=1.8	4	MLS>=0.7	18	
M1	Meta MS I	2001	241 families (381 sib-pairs)	257-442	GENEHUNTER	NPL>2.0	8			
M2	Meta MS II Meta MS I & II are partially overlapping	2003	719 families (944 sib-pairs)	257-463	GENEHUNTER	NPL>=2.1	7			

*1 Criteria for US 1: any marker must reach the nominal criteria (a lod score of 1 or better, and P values of 0.05 or less) on 3 of 3 tests for level A and 2 of 3 tests for level B
 *2 Criteria for US/France 2: Primary criteria: heterogeneity LOD (HLOD) >2.0 and multipoint Z scores >2.0. Secondary criteria: HLOD or Z score >2.0







21

22

⇐ Region reaching "Level B"

➔ Region reaching "Level A"

➔ Region from Meta MS II

A1	US 1	Haines et al 1996
A2	US/France 2	Kenealy et al 2004
B1	UK 1	Sawcer et al 1996
B2	UK 2	Hensiek et al 2003
C1	Canada 1	Ebers et al 1996
C2	Canada 2	Dyment et al 2004
F	Finland	Kuokkanen et al 1997
S	Sardinia	Coraddu et al 2001
I	Italy	Broadley et al 2001
N	Nordic	Åkesson et al 2002
T	Turkey	Eraksoy et al 2003
Au	Australia	Ban et al 2002
M1	Meta I	The Transatlantic Multiple Sclerosis Cooperative 2001
M2	Meta II	GAMES & The Transatlantic Multiple Sclerosis Cooperative 2003

Apart from the genome-wide screens in out-bred populations, two genome-wide screens in multigenerational pedigrees with multiple sclerosis have been published. Vitale et al. genotyped 613 markers in 18 members (7 affected and 11 unaffected) of a multigenerational family of Pennsylvania Dutch background [141]. In this family, the disease seemed to segregate with an autosomal dominant inheritance pattern. A peak of suggestive linkage (maximum multipoint LOD score 2.71) was found on 12p12. Modin et al. studied a family of Middle Eastern origin, with a possible autosomal recessive inheritance and found a LOD score of 2.29 on 9q22 [142].

Genome-wide Screens for LD in Multiple Sclerosis

Alongside with the disappointments of the initial genome-wide screens, the thoughts of novel ways of performing unbiased screens emerged. In Cambridge, UK, Stephen Sawcer and Alastair Compston founded a collaborative project called “Genetic Analysis in Multiple sclerosis in Europeans” or “GAMES”. The aim of this project was to take advantage of the linkage disequilibrium in the genome. The GAMES-project was based on the idea of genotyping 6000 microsatellite markers evenly spread across the genome in order to perform an indirect genome-wide screen [143, 144].

To make the genotyping possible, the study was based on typing pools of DNA [145]. The first study was performed in the British population [146] by typing four pools of DNA from cases (n=216), controls (n=219) and trio families index patients (n=745 affected) and their parents (n=1490). This experiment was thereafter followed by similar studies in pools from the participating countries (18 groups from 17 countries) [147-163]. Each study contributed one pool of DNA from patients with MS and one pool of DNA from control individuals (some groups also used trio-families). The pools contained DNA from around 200 individuals, with concentrations carefully measured to make sure that each individual gave equal contribution to the pool.

Over 80% of the screens found significant association with at least one marker located in the HLA-region, which can serve as a positive validation or “proof of principle” of the method. The analysis in the individual groups also provided ranking of markers. The main problem encountered in the GAMES study was that the density of 6000 microsatellites (on average 1 / 500 kb), which at the time of the design of the study was regarded as an appropriate density, turned out not to be sufficient [144]. Recent studies have shown that the extent of LD is shorter than expected and that there is a huge and unpredictable variation in LD across the genome. Another problem was the artifacts inherent in microsatellites. These artifacts turned out to be technically more challenging than initially anticipated. The GAMES-project has however provided the basis for many future collaborative studies.

AIMS OF STUDY I – IV

Study I

To perform a genome-wide screen for linkage in 136 sib-pairs with multiple sclerosis from four Nordic countries (Denmark, Finland, Norway and Sweden).

Study II

To refine the linkage in an area on chromosome 10, found in study I, by linkage analysis in four European populations (Italy, Sardinia, United Kingdom and the Nordic countries).

Study III

To study a potential candidate gene, interleukin 15 receptor alpha (IL-15RA) located in the refined region from study II.

Study IV

Fine-mapping of a region on chromosome 10 in a Swedish population based on results from a genome-wide screen for linkage in Icelandic families with multiple sclerosis.

MATERIALS AND METHODS

PATIENTS AND CONTROLS

All patients with multiple sclerosis in study I-IV fulfilled the criteria for definite multiple sclerosis according to the Poser-criteria [13] and/or McDonald-criteria [14]. All patients (study I-IV), their family members (in study I & II) and controls (study III & IV) gave informed consent to participate in research. Ethical permission was obtained from the local ethical committee.

Sib-pair Families (Study I & II)

For study I, a total of 136 families with affected sibling pairs were included in the statistical analysis (Table 2). From the beginning 138 sib-pairs were genotyped, but after performing SIBERROR-test (as described under “Statistical analysis”), two pairs had to be excluded. All families had been identified by the members of “The Nordic MS Genetics Group” (see Acknowledgements for participating groups). Where available, DNA was also obtained from parents or unaffected siblings.

Nordic	Danish	Swedish	Norwegian	Finnish	TOTAL
Families					
Affected sibling pairs	50	41	33	12	136
- <i>n</i> with both parents	8	7	2	0	17
- <i>n</i> with one parent	13	7	4	0	24
- <i>n</i> with no parents	29	27	27	12	95
- <i>n</i> of unaffected siblings	0	4	0	20	24

Table 2 Family characteristics of sibling pairs in the Nordic linkage screen (study I).

Study II contained the families from study I plus sib-pair families from United Kingdom, Italy and Sardinia (Table 3).

	UK (UK1/UK2)	Sardinian	Italian	Nordic	TOTAL
Families					
- total	226 (129/97)	49	38	136	449
- number with pairs	215 (122/93)	46	37	136	434
- number with three affected sibs	11 (7/4)	3	1	0	15
- families with both parents	59 (40/19)	22	21	17	119
- families with one parent	72 (34/38)	21	13	24	130
- families with no parents	95 (55/40)	6	4	95	200
Individuals					
- total	765 (447/318)	220	163	354	1502
- affected siblings	463 (265/198)	101	77	272	913
- unaffected siblings	112 (68/44)	54	31	24	221
- parents	190 (114/76)	65	55	58	368

Table 3 Family characteristics of the individual populations and total number of individuals genotyped in study II.

Patients and Controls (Study III & IV)

The patients with multiple sclerosis were diagnosed by neurologists at the Karolinska University Hospital – Huddinge and Solna sites (tertiary referral centres in Stockholm). The corresponding control group consisted of ethnically matched blood donors residing in the Stockholm area. All patients and controls were Caucasians residing in the Stockholm area (Sweden), who were born and raised in either Sweden or one of the neighboring Nordic countries (Denmark, Finland or Norway).

GENETIC ANALYSIS

DNA Extraction

Total genomic DNA was extracted from leukocytes using either the salting-out method described by Miller [164], QIAmp DNA extraction kit (QIAGEN; Hilden, Germany) or Puregene (Gentra systems).

Genetic Markers

Microsatellites

In study I, a total of 399 microsatellites from the Applied Biosystems Medium Density Linkage Mapping Set (LMS-MD10) were used for the genome-wide screen. All microsatellites were dinucleotide repeats and the mean heterozygosity was 78%, mean PIC 76%. For study II, 13 microsatellites (dinucleotides) from Applied Biosystems High Density Linkage Mapping Set (LMS-HD5) were genotyped (mean heterozygosity 78%). An additional 5 microsatellites (dinucleotides) located on the long arm of chromosome 10 were genotyped only in the Sardinian families.

In study IV, 42 di- or tetranucleotide microsatellite were genotyped (Table 1 in Paper IV).

Single-nucleotide polymorphisms (SNPs)

In study III, eleven SNPs in the gene coding for IL-15RA were initially selected from dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) and run in a test-sample. Priority was given for markers that had been annotated more than once and showed a high heterozygosity. Six of these SNPs were polymorphic and had high success rates (>85%) in the test-run and were therefore selected for further analysis.

In study IV, six SNP located in the “core”-region of the haplotype-analysis were selected (Table 1 in Paper IV).

Genotyping

In study I & II, polymerase chain reaction (PCR) was performed by either of three methods: (1) “Touch Down” PCR [165, 166] on either GRI Tetrad or a Hybaid Omnigene thermal-cycler, (2) an 877 Integrated Thermal Cycler (Applied Biosystems) with up to two markers in each reaction or (3) “True Allele” (Applied Biosystems) PCR on a GRI Tetrad machine using ABI recommended conditions.

The PCR products were electrophoresed on either a 373A Sequencing machine (Applied Biosystems) using 6% acrylamide slab gels, or a 3700 Genetic Analyser (Applied Biosystems) by capillary electrophoresis. Internal size standard was used for fragment sizing. In order to ensure accurate allele calling across the various gels, alleles were normalized against a CEPH-individual (1347-02). Allele calling was performed using the GENESCAN and GENOTYPER software.

In study IV, microsatellite genotyping at decode Genetics was performed in the following way: The PCR reactions were set-up on Cyberlab C-400 (Cyberlab, Inc; Brookefield, Conn USA), run on MJR PCR machines (MJ Research Inc., Boston, Mass, USA) and pooled on Cyberlab C-400, with a similar protocol for each marker. The pooled DNA products were separated and detected on an Applied Biosystems model 3700 or 3730 Sequencers. Alleles were called automatically with the DAC (deCODE Allele Caller) program, and the program deCODE GT was used to fractionate the readings according to quality and to edit the called genotypes [167].

The SNP-genotyping in study III was performed by the “Mutation Analysis Facility” (MAF) core facility at Karolinska Institutet (C. Lindgren, J Kere) using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom Inc., San Diego, California) [168, 169]. Assays for the PCR and associated extension reaction were designed using the SpectroDESIGNER software (Sequenom Inc., San Diego, California). All amplification reactions were run under the same conditions. Allele-specific primer extensions were conducted using 5 pmol of a extension primer and the Mass EXTEND Reagents Kit and cleaned using SpectroCLEANER (Sequenom Inc., San Diego, California) on a MULTIMEK 96 automated 96-channel robot (Beckman Coulter, Fullerton, California). Primer extension

products were loaded onto positions of a 384-element chip with a nanoliter pipetting system (SpectroCHIP, SpectroJet; Sequenom) and analysed using a MassARRAY mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). The resulting mass spectra were processed and analysed for peak identification using the SpectroTYPER RT 2.0 software (Sequenom). For each SNP, two independent scorers confirmed all genotypes. In study IV the SNP-genotyping at decode Genetics was performed using an in-house developed fluorescence based method (not published).

HLA-typing

In study I, each of the affected individuals was tested for the presence or absence of the DR15 alleles as described in Paper I.

Genetic Maps

In the Nordic screen (study I), the Genethon map [170] was used. The other populations included in study II had in their respective original screens used the following genetic maps; the Italian and Sardinian screens had used marker positions from the location database (LDB) [171] together with the map used by Davies et al [172], while the UK screen used the Genethon-map [170] and the map by Davies et al [172]. In study II, the Marshfield map [173] was used for all population groups. In Study IV, marker locations were defined according to the deCODE map [174] and the NCBI build 34 physical location.

STATISTICAL ANALYSIS

Linkage Analysis (Study I & II)

Non-parametric (model-independent) linkage analysis was performed using the MAPMAKER/SIBS program [91] (version 2.0). This program was also used to calculate the information extraction along each chromosome. The SPLINK program [90] (version 1.07) was used to estimate the allele frequencies from the data. As no parents were genotyped in the first stage of the genome-wide screen in study I, the SIBERROR program [175] was used to test for the presence of pedigree errors. In the SIBERROR analysis, one sib-pair was identified as identical twins and one other pair as unrelated. These two pairs were therefore excluded from the analysis, leaving 136 affected sib-pairs used in the analysis. The analysis of possible interaction with HLA in study I was performed by Peter Holmans [176]. The TRANSMIT program [177] (version 2.5.2) was used for transmission disequilibrium testing in study II.

Association Analysis (Study III & IV)

In study III, genotype frequencies in cases and controls were compared using a χ^2 -analysis. There were no deviations from Hardy-Weinberg Equilibrium for any SNP in the control group. Haplotype frequencies were estimated by the Expectation Maximization (EM)-algorithm [178] and haplotype p-values were obtained by Monte-Carlo simulations (performed by G. Jonasdottir).

In study IV, both single-point and haplotype analysis were performed by the NEMO-program as described in [179]. Haplotype-analysis in NEMO is based on the EM-

algorithm. To assess the significance of the haplotype data after correcting for multiple testing, a randomization test was carried out using the same genotype data.

Estimation and Illustration of LD (Linkage Disequilibrium)

In study III, linkage disequilibrium (LD) between the markers were quantified by the measure D' [100], obtained by Haploview [180] version 2.05: (<http://www.broad.mit.edu/personal/jcbarret/haploview>)

In study IV, D' and corresponding p-values were obtained by NEMO [179]. For the microsatellites, we applied an extension to the definitions of D' and its statistical significance for biallelic markers (described in Paper IV). The data for 252 SNPs genotyped in 90 CEPH individuals by the International HapMap Consortium [85] (<http://www.hapmap.org/>) were used as comparison to the Swedish data.

RESULTS AND INTERPRETATIONS

STUDY I

The first study was a genome-wide screen for linkage in 136 Nordic sib-pairs with multiple sclerosis. Although the results showed no regions of genome-wide significance, 17 regions were identified where the lod score exceeded 0.7 (the 5% nominal significance level). These regions include 1q11-24, 2q24-32, 3p26.3, 3q21.1, 4q12, 6p25.3, 6p21-22, 6q21, 9q34.3, 10p15, 10p12-13, 11p15.5, 12q21.3, 16p13.3, 17q25.3, 22q12-13 and Xp22.3. In a genome wide screen of this density, only ten such regions would be expected by chance alone. The fact that our results exceed expectation suggests that at least some of these regions are likely to be genuine.

The results show both overlap and differences with previous screens in multiple sclerosis and other autoimmune complex traits. The differences with respect to other linkage screens in multiple sclerosis may reflect type I error (false positive findings), or result from differences in the ethnic genetic background of the populations studied.

In the beginning of 2000, when the laboratory-work for “Study I” was started, four genome-wide screens for linkage in multiple sclerosis had been published [127-130]. Meanwhile, the “Nordic MS Genetics Group” had since its inception in 1994, on a collaborative basis collected DNA from families with several members affected by multiple sclerosis. The high prevalence of multiple sclerosis in the Nordic countries together with the relative genetic homogeneity of the Nordic population encouraged us to perform a genome-wide screen for linkage in these families. The number of sib-pairs included in the study (136 sib-pairs) was according to power estimations at the time of the start of the study regarded to be sufficiently high.

Disappointingly however, this screen, similar to other screens of different sizes in other populations did not show any regions of genome-wide significance [94]. With the increased knowledge we have today, we now know that a study size of 136 sib-pairs is not enough to be able to find peaks of genome-wide significance in multiple sclerosis. The results from the Nordic screen has however contributed to the general scientific knowledge in the field of genetic analysis of multiple sclerosis and should be viewed as one of several pieces in a large jig-saw puzzle.

For stage 1 of the screen, we had decided not to genotype parents or unaffected sibs. This decision was based on suggestions by Hauser [93] and Holmans [89]. As a second stage we typed the available parents for the 31 best markers from the first stage. Genotyping the available parents and unaffected sibs resulted in only a minor increase in information extraction (3%). This very low level of information increase, could probably largely be explained by the fact that the number of unaffected relatives in stage 2 was low (see Table 2) and the increase in information might have been slightly higher if there had been more unaffected relatives available.

When planning the screen, we decided not to include families with more than two affected siblings. Our hypothesis was that these families (with > 2 affected sibs) are

more likely to include parents who are homozygous for potential susceptibility alleles, and therefore not informative for linkage. This effect was suspected in the Inflammatory Bowel Disease meta-analysis [181], where excess sharing on chromosome 16 only was seen only in the group of families with a single affected sib-pair and not in dense families with clustering of Crohn disease, although in fact this was subsequently shown not to be the case.

Although the results were somewhat disappointing in that there were no regions of genome-wide significance, the number of regions reaching nominal significance is higher than expected by chance alone, suggesting that some of the regions are likely to be true peaks containing disease genes. There is however no easy way of telling the true peaks from the false positive ones. Basically there are two main strategies for moving forward based on these results. One is to put together the raw-data from all screens and perform a meta-analysis. This has been done and the data from the Nordic screen was included in the latest meta-analysis [140]. Another way forward is to compare the results from the Nordic screen with the screens from other populations and look for similarities indicating overlapping results (Table 1 & Fig. 4). This is what was performed in Study II.

STUDY II

One area with partially overlapping peaks appearing in several of the genome-wide screens including the Nordic, Italian, Sardinian, the second screen from United Kingdom, as well as the second meta-analysis of genome-wide screens in multiple sclerosis, is the short arm and centromere of chromosome 10 (Fig. 5a). This region had not previously been pinpointed as a candidate region harbouring genes of interest in the genetic susceptibility for multiple sclerosis.

Study II therefore aimed at exploring these partially overlapping regions on chromosome 10. The aim of the study was to better delineate the peak(s) in the region by increasing the marker density and thereby get increased information extraction. This study was not confined to the Nordic population, but also included the families from United Kingdom, Sardinia and Italy that had been used in each original screen (Table 3). All sib-pairs (n=449) were genotyped for a high-density set of 13 microsatellite markers (average distance between markers 5.3 cM) in an area covering 64 cM on the short arm and centromere of chromosome 10. The first step was to analyse all the original data using the same genetic map for each of the individual populations thereby making it easier to compare the data. We thereafter added the additional genotyping and re-analysed the data. In order to look for both domestic and ubiquitous genes, the data was analysed in two ways: a) each population separately and b) all populations together.

The additional genotyping led to an increase in the average information extraction in the region from 52% to 79% (Fig. 5b). This narrowed the fairly wide peak somewhat to be focused on the more p-telomeric end of the chromosome and revealed increased support for linkage (MLS 2.5), with maximum at 10p15 (Fig. 5c).

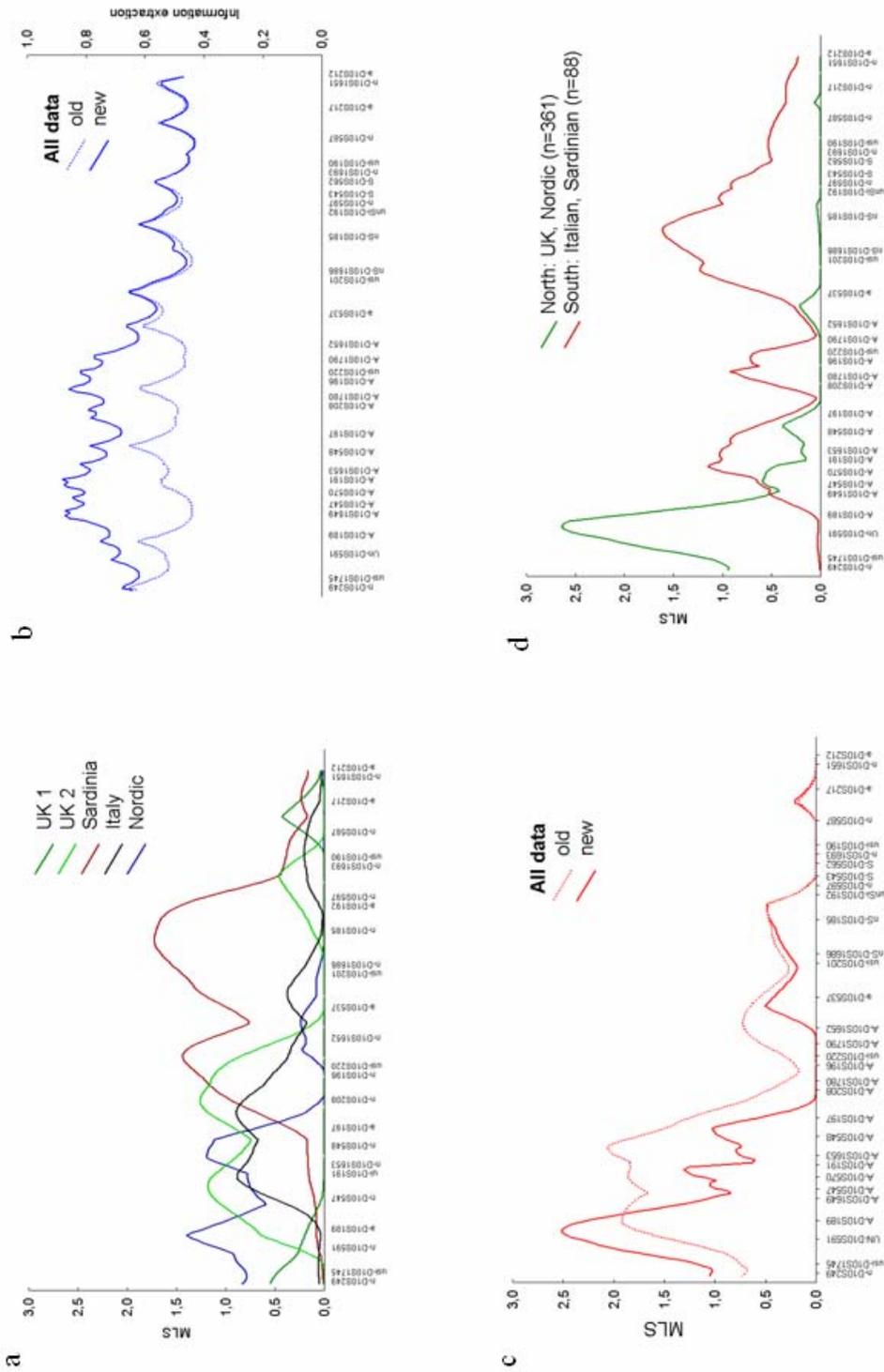


Fig 5 a) Results on chromosome 10 in genome-wide screens for linkage (original screens) in UK1[121], UK2[131], Sardinia[124], Italy[125] and the Nordic countries[126]. **b**) Mean information extraction for all populations included in study II. The dotted line depicts the original screens and the solid line the information extraction obtained in study II. **c**) MLS (maximum lod score) for all populations in study II, based on data from the original screens (dotted line) and the additional genotyping in study II (solid line). **d**) Results from study II divided into “Northern European populations”: UK and Nordic (green line) and “Southern European populations”: Sardinia and Italy (red line).

Hypothesizing that there may be regional differences between the populations; we also performed an analysis with the results divided into “Northern European populations” – UK and the Nordic countries and “Southern European populations” – Italy and Sardinia. In this analysis (Fig. 5d), the peaks tend to split into one peak at the p-telomeric end, only present in the “Northern” group and one peak on the long arm (10q) only present in the “Southern” group together with some minor peaks in the area in between. This could indicate the presence of population specific genes. Sardinia for instance has a different HLA-association than Northern Europe. This result however should be considered with caution as the number of sib-pairs especially in the “Southern group” is small.

STUDY III

In complex disease genetics it is not possible to define a critical linkage region that is *certain* to contain a candidate gene, but only to give a confidence interval with a *certain probability* of containing a potential candidate gene [182]. Even if the peak on the short arm of chromosome 10 now is more well-defined, it still covers an area which is too wide to comprehensively screen by studying all potential candidate genes in the region. We therefore chose to focus on one candidate gene located in the region of the peak on 10p15. We decided to study the gene coding for the alpha receptor of IL-15 (IL-15RA). This gene can be regarded as both a positional and functional candidate (thereby increasing the prior probability).

IL-15 is a cytokine expressed at the mRNA level in numerous human tissues in a broad range of cell types, including activated monocytes, dendritic cells, macrophages, epithelial cells, activated astrocytes and microglia [183]. At high doses, IL-15 has been regarded as a general T-cell growth factor. The IL-15 receptor is heterodimeric and contains the IL-2 β - and γ chain, together with a unique α -chain (IL-15RA). Previous studies by colleagues in our laboratory have shown increased levels of IL-15 mRNA in patients compared to healthy subjects [184, 185]. Another interesting point is that human herpesvirus-6 (HHV-6), a virus implicated in the pathogenesis of multiple sclerosis, has been shown to be a potent stimulator of IL-15 expression [186, 187].

Study III was a case-control association study in which we genotyped 553 patients with multiple sclerosis and 530 controls (healthy blood donors). All patients and controls were Caucasians residing in the Stockholm area (Sweden), who were born and raised in either Sweden or one of our neighbouring Nordic countries (Denmark, Finland or Norway).

Six out of eleven SNPs initially selected, met our criteria for continued genotyping. (The SNPs should be polymorphic in this population and have a high success rate >85% in the genotyping process.) These six SNPs were genotyped in the whole study material. We performed both single-point and haplotype analysis of the data. Neither genotype frequencies (Table 1 in the manuscript) nor allele frequencies (data not shown) showed any indication of association with multiple sclerosis. The four most common haplotypes together accounted for more than 90% of the variation (Table 3 in

the manuscript). No haplotype was found to be associated with the disease. The conclusion is that our data do not support the alpha-receptor of IL15 as a gene conferring susceptibility for multiple sclerosis, with the reservation of the possibility of a genetic contribution too small to prove with our study-design or other variations not tested.

Despite the fact that the gene we studied had both positional and potential functional support, we failed to find a disease association. This illustrates the difficulties in choosing genes for candidate gene analysis.

Deloukas et al. at the Sanger Institute recently published an article [188] on the sequencing of chromosome 10. They conclude that the finished sequence of chromosome 10 comprises a total of 131,666,441 base pairs. The sequence annotation revealed 1,357 genes of which 816 are protein coding. Excluding the pseudogenes, human chromosome 10 has an average gene density of 10.4 genes / Mb. A total of 4,204 transcripts were annotated for the 1,357 genes emphasizing the abundance of alternative splicing. It is estimated that 73% of the protein-coding genes have more than one transcript (average: 5.8). Deloukas et al. mapped 143,364 SNPs (dbSNP release 115) to the chromosome 10 sequence. These figures highlight the complexity of the genome, here represented by chromosome 10.

STUDY IV

In parallel with the published genome-wide screens for linkage in multiple sclerosis, a genome-wide linkage screen was performed in a cohort of Icelandic families with multiple sclerosis, by deCode Genetics, Reykjavik, Iceland (unpublished data). This screen showed a haplotype on chromosome 10p12 potentially associated with multiple sclerosis in these families. As this region overlaps with one of the peaks in Study I & II, we found it interesting to investigate this chromosomal area in our cohort of Swedish patients with multiple sclerosis and controls. This study was performed as an academic collaboration with researchers at deCode Genetics in Iceland.

The approach in study IV was different from study III. Instead of choosing a *gene* from a peak as in study III, this study focuses on typing a relatively dense set of markers in a chromosomal *region*. The region was in this case chosen based on positional data from Icelandic families.

A total of 48 markers were genotyped in the Swedish patients and controls. Thirty-nine of these (33 microsatellites and 6 SNPs) were located in the region of haplotype-sharing that had been found to be possibly associated with MS in the Icelandic genome-wide screen for linkage (unpublished data). We performed single-point and haplotype analysis based on the EM-algorithm and randomization testing.

An LD-map was established based on the result from our analysis (Fig 1b in the manuscript). The estimations of LD in this figure were based on data from all Swedish individuals that were genotyped (both patients and controls). In order to compare the distribution and location of LD-blocks with that of a dense map of SNPs, we analysed

252 SNPs from same the same chromosomal region that had been genotyped in 90 CEPH individuals by the International HapMap Consortium [85], illustrated in Fig 1a in the manuscript. Although the three LD-blocks found in the Swedish data, with some difficulty can be recognized in the LD-map based on HapMap data, the latter map contains more and shorter blocks. The main difference between the two analyses is that the HapMap-data has a density of 6.7 kb between the SNPs compared to 46 kb in the Swedish data. Other differences include type of markers: the HapMap-set only contained SNPs, while the Swedish data mainly was based on microsatellites. The analyses are also based on different study-populations (Swedish versus CEPH-individuals, which are Utah residents with ancestry from northern and western Europe).

The analysis of our Swedish data showed several areas of D' values indicating intermediate LD. As mentioned earlier in this thesis, the interpretation of intermediate values of D' is more complicated than with high LD. Therefore it is difficult to give clear unambiguous haplotype-block boundaries.

Although the study did not show any definite significantly associated haplotypes, we detected a marginal significant core haplotype in the Swedish multiple sclerosis cohort, overlapping a haplotype shared in four affected sibs in one Icelandic family (labeled “Swe hap” and “Ice 1 hap” respectively in Table 3 in the manuscript), giving some support for continuing the genetic research in the region as one of the candidate regions on the short arm of chromosome 10 for susceptibility genes for multiple sclerosis.

Figure 6 gives an illustration of the regions on chromosome 10 that were investigated in this thesis.

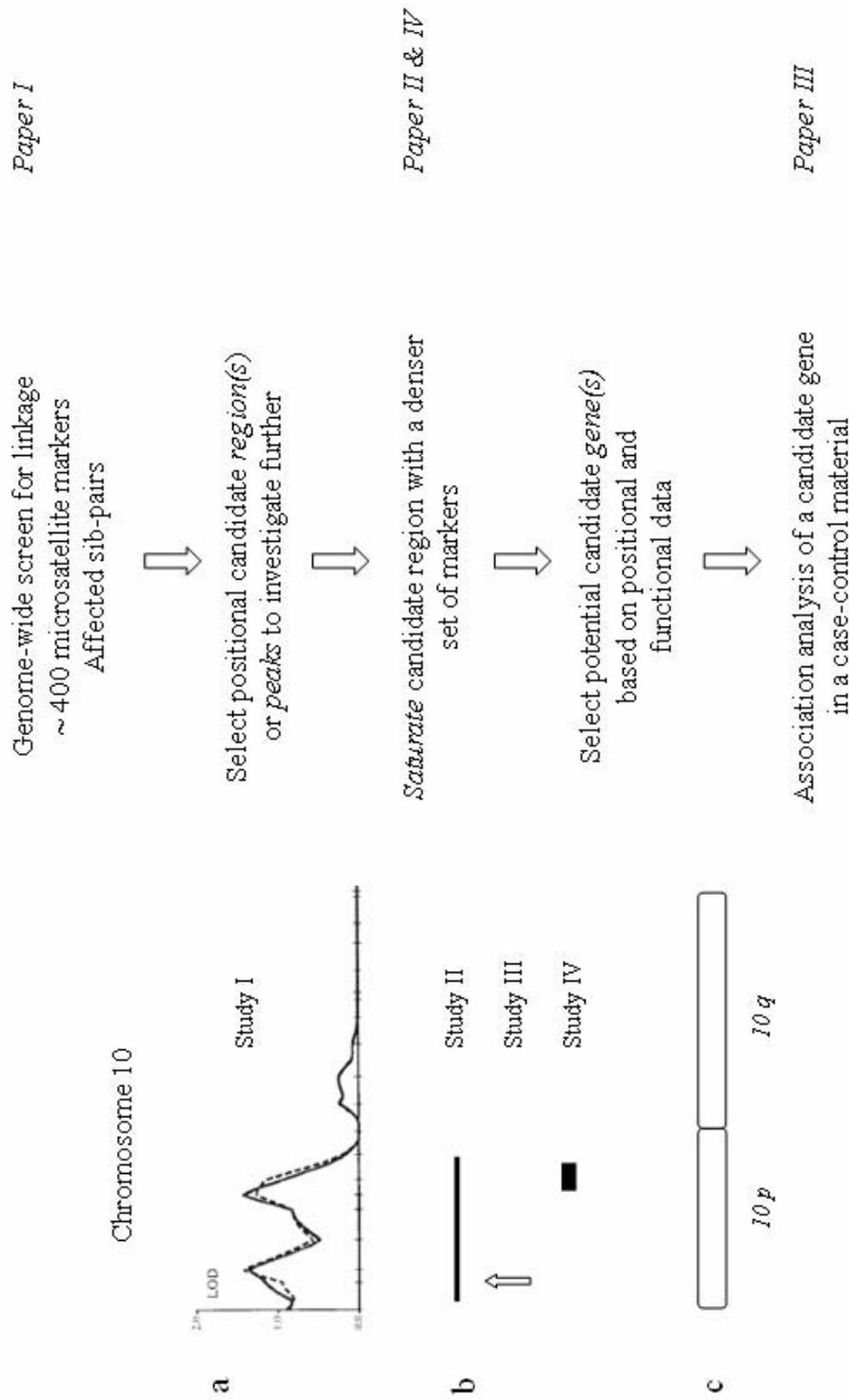


Fig 6 Flowchart summarizing the studies included in this thesis. **a)** chromosome 10 in the Nordic screen (Study I), **b)** location of regions investigated in study II-IV in relation to the peaks in the Nordic screen, **c)** a schematic picture of chromosome 10.

GENERAL DISCUSSION

As mentioned previously, extensive and thorough studies of the genetic epidemiology of multiple sclerosis have provided supporting evidence for multiple sclerosis being a complex genetic trait. This means that there are genes to be found. However, as in many other complex diseases, the research-field of genetics in multiple sclerosis has been faced with problems. Some of the important issues are discussed here.

GENOME-WIDE SCREENS

What reliable conclusions can we draw from the genome-wide screens for linkage performed to date in multiple sclerosis (including study I)?

A review of the genome-wide screens for linkage performed to date in multiple sclerosis, shows a wide variety of genomic regions with different evidence of potential or suggestive linkage. Figure 4 illustrates the many overlapping as well as non-overlapping regions of peaks on all the chromosomes. How can we decide which of these regions are linkage peaks containing a true susceptibility gene to multiple sclerosis? The screens are based on people with multiple sclerosis from different populations, the screens used somewhat different ascertainment criteria and have employed different statistical analysis methods, mainly non-parametric, but some groups with a higher number of multiple affected people also applied parametric methods.

One way to get guidance is to look at the results from the latest (and so far the largest) meta-analysis [140]. This analysis was based on the raw-data from 719 families (944 sib-pairs). It is the only screen that has shown a region of genome-wide significance. That the region showing genome-wide significance is the MHC-region on the short arm of chromosome 6 probably comes as no surprise to anyone working in this field. The association between HLA and multiple sclerosis was described already in the early 1970s' and is still the only unquestionable genetic association in multiple sclerosis. The question is rather; why haven't any of the individual screens managed to show unequivocal linkage for this region? And why have the screens not shown significant linkage to any "non-MHC" region?

It is well-known that linkage is less powerful than association analysis in finding modest genetic effects. But is that the sole answer to the question? How much power did the screens actually have? What factors could have influenced the power of the studies?

One important aspect partly discussed with regard to study II is the *information extraction*. The results from the second meta-analysis [140] showed that the mean information extraction from the 719 families was 44% (highest level 68% in the MHC-region, where a higher density of markers had been typed).

Recent reports [189, 190] point to low information extraction as one important contributing reason to the lack of success for genome-wide screens for linkage in

complex diseases, despite the use of standard “state-of-the-art” sets of highly polymorphic microsatellite markers.

Evans and Cardon have recently published suggested guidelines [191] for genotyping in genome-wide screens for linkage. They performed simulations whereby they tested marker maps with different densities for both microsatellites and SNPs. They showed that traditional maps of microsatellite markers with an average density of 1 marker / 10 cM, are associated with a significantly lower information content than a dense map of SNPs or microsatellites. The information content associated with a traditional map of microsatellites (1/10cM) could be as low as about 30% in regions between markers! The results of Evans’ and Cardon’s study further pointed out that the information extraction was highly sensitive to lack of parental data. When parental data is missing, the loss of information is even higher.

A study by “The International Multiple Sclerosis Genetics Consortium” [190] comparing the genotypes from the original UK genome-wide screen in multiple sclerosis from 1996 with modern linkage mapping sets showed a marked increase in information extraction with the new methods. The information extraction in the original UK screen ranged from an average of 32.4% (without parents) to 61.7% (with parents). These figures were increased to 54.4-58.4% (without parents) and 90.7-91.5% (with parents) with the use of high-density SNP-sets.

Referring the issues of information extraction to the linkage studies in this thesis (Study I and II), gives the following figures: The average information extraction from the Nordic screen where no parental data was available (except for a few unaffected relatives typed for a few markers in stage 2) was 57%, ranging from 15% on the p-telomeric region of chromosome 19 to 86% on a region on chromosome 22. The region on chromosome 10 that was investigated in Study II had an initial information content of 52% in the screen, which increased to 79% by typing of additional microsatellites.

Another problem, which has previously not received much attention, is the impact of *genotyping error* on the power of detecting linkage. Abecasis et al. [192] used Monte-Carlo simulations to study the influence of genotyping error on the analysis. The results of that study indicated that genotyping error can have devastating effects on the possibility of detecting an existing true linkage peak. They showed that a rate of 5% genotyping error eliminated virtually all evidence for a true linkage for a modest effect size. At a λ_s of 2.0, a rate of 2% error reduced the expected lod score by almost half. As an example, they also described that a study of 1000 ASPs with a locus of relatively small effect (e.g. $\lambda_s = 1.25$) which in simulations yielded an average peak lod score of 2.90 in the absence of genotyping error, would be reduced to 1.40 by the introduction of 1% genotyping error. For affected sib-pair studies, genotyping errors have a greater impact on multipoint than two-point analysis.

In the region on chromosome 10 investigated in Study II, two markers (D10S189 and D10S197) had been genotyped both in the respective original screens as well as in Study II. We therefore had the opportunity of comparing the concordance of genotyping results for these two markers. By comparing the original genotypes with the

results from Study II, we could get a measurement of the genotyping accuracy. With 2565 and 2357 alleles called for respective marker, we found a mean error of allele calling of 1.8%. In this case all electrophoreses in the original screens had been performed on a 373A DNA sequencing machine (Applied Biosystems) using acrylamide slab gels, while in Study II a capillary based 3700 Genetic Analyser (Applied Biosystems) was used. This fact might have contributed to a difference in the quality of the traces. It is difficult to draw any generalized conclusions from this error rate of 1.8% as it merely describes the error rates for these markers. Nevertheless it gives a hint of the possible level of genotyping error.

With these problems in mind, what can be done to increase the likelihood of finding the true peaks? With the development of high-density genome-wide SNP assays during the last few years, many researchers advocate that screens which have been performed with traditional low/medium density microsatellite 1/10cM marker sets should be re-analysed with new SNP-assays in order to give a more reliable and true result [189, 190]. An advantage of SNP-assays is that SNPs are more abundant in the genome than microsatellites. They are however less polymorphic, but studies have estimated that 2-3 SNPs could be used to replace one microsatellite [193]. SNPs are associated with a lower genotyping error rate and are better amenable to high-throughput genotyping. If possible, one should aim to include parents or unaffected sibs, both in order to increase the information extraction and to reduce the genotyping error rate by being able to check for Mendelian inconsistencies.

One important thing however is that, if an increase in marker density with the new techniques also increases the number of genotyping errors, the outcome might be a decrease in the power to detect true linkage. It is thus important to keep genotyping errors at a minimum.

HAPLOTYPE-ANALYSIS

How can we make the best use of linkage disequilibrium and haplotype analysis in future studies?

As mentioned earlier, the last few years have shown a “hype” in studies using haplotype-analysis. What evidence do we have for using haplotypes in analysis? Why should we use it, and how? What can we gain from the HapMap project and what are its limitations?

There are several reasons for using haplotype-analysis. In a review [194], Clark points out three main reasons why haplotype analysis confers an improvement compared to ordinary single-point analysis. The first, he says, is that the haplotype structure corresponds directly to the biologically functional unit (the protein sequence). Secondly, studies of population genetics have shown that population variation is structured into haplotypes. Finally he points out statistical advantages as haplotype analysis reduces the dimension of statistical tests involved.

However, the basis for using haplotype analysis in mapping genes with complex genetic background, assumes that the risk-enhancing SNP is a “common SNP”. This is the basis in the Common Disease Common Variant (CDCV) hypothesis. The CDCV hypothesis predicts that complex genetic diseases will be due to disease-predisposing alleles with relatively high frequencies, i.e. each of the contributing disease loci will only contain one or a few predominating alleles. Examples supporting this hypothesis include APOE ϵ 4 in Alzheimer’s disease [122, 123, 195] and PPAR γ in type II diabetes [196]. If the common disease – common variant is true, association mapping with the help of the data acquired through the HapMap project [85] has the prospect of being successful.

Some scientists doubt whether the CDCV-hypothesis is generally applicable to complex genetic diseases [197]. It may be that there are *rare* interacting factors underlying the complex traits, and if that is the case, haplotype analysis and the use of “common” SNPs will not be able to locate the disease predisposing genes/variants. One could also argue that our knowledge of the genome-wide distribution of genetic variation still is too weak, with a very variable LD pattern throughout the genome. Whether or not the CDCV is true, there are still many factors regarding the genomic pattern that are unknown, which complicates the study-designs.

With future genotyping studies to an increasing extent being based on large-scale genotyping, one important factor for designing a study is how to cut the genotyping cost without losing power. Many scientists advocate the use of haplotype-tagging SNPs “htSNPs” as one way of cutting the costs of large scale genotyping. Haplotype-tagging means that only a few “tagging” SNPs need to be genotyped in order to cover the majority of the haplotype information from a region. There are many algorithms by which these “ht-SNPs” can be chosen. With the emerging work of the HapMap project, the hope is that the publicly available information will help in choosing the correct “tagging”-SNPs.

Another issue is the definition of a “haplotype block”. As described in a review by Wall et al. [198], different methods have been proposed for defining haplotype blocks. These methods can be classified into two main groups depending on the way they define the blocks. The first group of methods defines a block as a region with limited haplotype diversity [199-202], while the other group uses pair-wise disequilibrium to identify regions with evidence for extensive historical recombination [102, 203-205]. A different approach was used by Jeffreys et al. [206], who performed a single-sperm study to show that very much of the recombination in the MHC class II region was restricted to narrow recombination “hot-spots”.

Cardon and Abecasis discuss this issue and the problem with comparing blocks between studies [207]. Their opinion is that: *“It is clear that “observed” haplotype blocks summarize a mixture of the underlying haplotype structure and the whims of the investigator reporting them. Presently, methods for detecting blocks remain subjective and no single approach is appropriate for all datasets”*.

LACK OF REPLICATION AND PUBLICATION BIAS

How to deal with the issue of significance levels, lack of replication and publication bias in association studies?

One main issue that has become an important concern regarding the results from genetic association studies is that many of the initially reported positive significant associations fail to be replicated, despite several independent follow-up studies. A related problem is that of *publication bias*, i.e. that fact that scientific journals preferentially publish studies with a positive finding regardless of study design. Therefore there is a risk that the articles published give a “skewed” picture of true genetic associations. The tendency to only publish “positive” findings is not restricted to the scientific journals, but is also present with the investigator themselves, who often have less or no interest in submitting results that are not positive. As a result of this, there is also a risk that scientists, when not encountering a positive finding in the first round, continue to perform “post-hoc” sub-analysis, by stratifying the results in many different ways, until a result with a p-value of <0.05 appears.

What are the reasons for the failure to replicate positive findings? In a review by Colhoun et al. [208], the authors suggest that the most important factors behind the inability to replicate previous positive associations are: “publication bias, failure to attribute results to chance and inadequate sample sizes”.

How stringent criteria should be used to be able to exclude “chance” as an explanation for failure to replicate initial positive studies? It is crucial to find an appropriate method of correcting for multiple testing. As mentioned in the “association” chapter, the Bonferroni correction is often applied to give the correct *α -value* when multiple tests have been performed. There are however disadvantages with the Bonferroni method. There are different opinions about which number of tests to correct for: Is it the number of tests that were performed and presented in the present study? Is it the number of all tests, the scientist *intended* to perform? Or, is it the number of tests that *ALL scientists intend* to perform??? [209]. The Bonferroni method has been criticized for being too conservative, leading to a risk of missing true associations (type II errors). It is important not only to avoid false positive results, but also false negative results, as negative results are less likely to be reconsidered. As the Bonferroni method assumes that the tests performed are independent, it is not really well suited for haplotype analysis. For haplotype analysis, permutation tests are therefore commonly used instead. These however, are often computationally intense but lately new methods, not based on permutation tests have been developed (for instance the SNPSpD approach by Nyholt) [210].

Another important issue is the power of the study-design. In a recent article, Zondervan and Cordon [211] discuss important factors influencing the power of an association study. They point out the importance of taking into account the following factors: the effect size of the susceptibility locus, the frequencies of the marker alleles as well as the disease alleles and finally, the extent and distribution of linkage disequilibrium in the region.

Many scientific journals have posted guidelines for the study-design of genetic analysis studies [212, 213]. Some of the points made are: to use a sufficient number of subjects, a carefully chosen control group to avoid population stratification, adjustments for multiple testing, provide power calculation and effect size estimates and that there should be a plausible biological role for the gene in question. Finally, before an association is proclaimed, the association should be confirmed in an independent sample with appropriate power.

In summary, it is important to design a study in an accurate way to be able to draw the correct conclusions from the outcome.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

It is very likely that complex genetics will be of increasing importance in many areas in the near future. Future studies are likely to involve analysis of possible interactions, including both gene-gene interaction as well as and gene-environment interaction, requiring methods to analyse and handle large dataset to an ever increasing extent.

Already now, the need for large samples in order to study complex genetic traits, have given rise to collaborations between research groups. With a growing demand for increasing sample sizes and large number of markers to be genotyped, work on a collaborative level is likely to play an even more important role.

The article “Prospects for whole-genome linkage disequilibrium mapping of common disease genes”, by Kruglyak was published in 1999 [214]. In this article it was estimated that around 500,000 SNPs would be required in order to perform whole-genome association studies. At that time, the possibility of genotyping half a million SNPs sounded like a utopia. Today, five years later, this is close to reality. In a talk titled: “The road to 500k” at the 54th annual meeting of The American Society of Human Genetics, October 2004, a speaker from Affymetrix described the next generation of SNP-chips containing assays for 500,000 SNPs, which should be available within the next year.

One can only speculate what progress in the genetic analysis of complex traits the next five years will reveal, but it is worthwhile to remember that what might sound like a utopia today might be a reality in a not too far future...

ACKNOWLEDGEMENTS

The work for this thesis was performed at the Division of Neurology, Neurotec Department at Karolinska Institutet. Part of the work, including study I & II, was performed at the Neurology Unit, Addenbrooke's Hospital, University of Cambridge, UK, where I had the opportunity to work for one and a half year during 2000 and 2001. I have also had the opportunity to work with the MS-group at deCode in Reykjavik, Iceland, where I performed the laboratory work for study IV.

I would like to express my warmest gratitude to all patients with multiple sclerosis and their families as well as the blood-donors who donated blood for these studies, without whom this work would not have been possible.

There are many people who have contributed to this work and to whom I want to express my warmest appreciation. This includes not only the scientific and academic work, but also colleagues and friends who have shared my experiences during the years of this work.

First of all I would like to thank my supervisor Jan Hillert for encouraging me to embark on this work, by inviting me to do a PhD in your group and for giving me the opportunity to spend time abroad during my thesis as well. You have built up an excellent research group and I am proud of being part of your team!

My co-supervisor Stephen Sawcer in Cambridge for sharing your wide knowledge of genetics. For excellent supervision including tutorials, as well as supervision in the lab. And together with the rest of the Sawcer family (Jane, Frances, Becky and Jack) you have always shown a warm hospitality during my visits, both in Sawston and in Oslo.

Professor Alastair Compston at the Neurology Unit in Cambridge, for inviting me to work with your "Gene-team". I have always felt welcome, from the very beginning. Also for letting me take part in the Tuesday meetings, Calman days and journal-club. Your enthusiasm for neurology has made a great impression on me.

I want to thank all past and present PhD-students and post-docs/researchers at the Division of Neurology. In particular, I would like to thank the MS genetics group for being such a great group. Helena Modin, for so much fun together, from the time as two complete novices in the lab to conferences and courses. Thomas Masterman for sharing your wisdom with the rest of us in the group. Kristina Duvefelt, my office-mate, for endless encouraging discussions and for valuable comments on this thesis. Iza Roos for inspiring discussions regarding every aspect of genetic analysis. Frida Svensson for introducing the excellent idea of celebrating an accepted manuscript with champagne. Kerstin Imrell, soon our most experienced conference-speaker. Boel Brynedal, our future large-scale data expert. Virginija Kerrenbauer for your constant enthusiasm.

Sanjaya Adikari, Jenny Ahlqvist, Tjerk Bueters, Mabel Cruz, Yaimei Dai, Ulrika Einarsson, Gabriel von Euler, Anna Fogdell-Hahn, Vilmantas Giedraitis, Andreia Gomes, Kristina Gottberg, Wen-Xin Huang, Yassir Hussein, Sverker Johansson, Lena von Koch, Kosta Kostulas, Nikos Kostulas, Matilde Kouwenhoven, Eva Laudon-Meyer, Arturs Ligers, Susanna Mjörnheim, Prabhakar Putheti, Rayomand Press, Uros Rot, Christina Sjöstrand, Ajith Sominanda, Leszek Stawiarz, Anna Steinberg, Dan Svensson, Mathula Thangarajh, Marina Vita, Katarzyna Wecliewicz, Lotta Widén Holmqvist, Chun Xu, Volkan Özenci. Our "summer-workers": Malin Andersson, Pernilla Andersson, Karin Moraitis, Wiveka Olsson.

Gunnel Larsson for excellent help with all practical matters. Cecilia Svarén-Quiding, Anna Ljungberg, Merja Kanerva and Faezeh Vejdani for help with the work in the lab. Helena, Anny and all nurses at the outpatient-clinic for help with sampling from the patients.

I want to thank all my friends and colleagues from the time in Cambridge for making those years such a great experience. Especially Francesca Coraddu and Anke Hensiek, my friends and co-authors, for collaboration in the lab as well as endless fun together. Everyone in the lab at Neurology Unit for making me feel at home there; Steve Sawcer, Alastair Compston, Julia Gray, Katrina Dedman, Richard Roxburgh, Mel Maranian, Belinda Smillie, Jackie Deans, Alasdair Coles + co-workers, Sabine, Peter and everyone else in the "journal-club group". GAMES-players and/or house-mates in "the safe-house": Sandra, Hanne, Murat, Mehdi, Maria, Giulia, Monica, Patricia, Pamel, Maria, Ann, Alexandra, Bartosz and others.

All members of the "Nordic MS Genetics Group" for an excellent collaboration: Copenhagen: Pamel Datta, Annette Oturai, Per Solberg-Sorensen, Arne Svejgaard, Lars Ryder; Oslo: Hanne Flinstad Harbo, Anne Spurkland, Frode Vartdal, Elisabeth G Celius; Bergen: Kjell-Morten Myhr, Harald Nyland; Lund: Magnhild Sandberg-Wollheim; Göteborg: Oluf Andersen; Åbo: Jorma Ilonen, Mikko Laaksonen; Helsinki: Lena Peltonen, Janna Saarela. A special thanks to Hanne, Anne, Pamel and Annette for inspiring collaborations and friendship including the time in Cambridge (Hanne) and Reykjavik (Pamel).

All people at deDode, Reykjavik involved in study IV including Ragnheidur Fossdal, Jeff Gulcher, Kari Stefansson, Hjörvar Petursson, Tinna Thorlacius, Valla Gudnadottir. A special warm thanks to Ragnheidur who together with the whole CNS-department & the "MS-plus"-group made my time in Reykjavik a great experience.

All other co-authors: Peter Holmans, Maria Giovanna Marrosu, Luca Massacesi, Maria Trojano, Patricia Momigliano-Richiardi, Eleonora Cocco, Raffaele Murru

Juha Kere, MAF, Novum, KI for collaborations on SNP-genotyping and Cecilia Lindgren for nice and supportive discussions.

Gudrun Jonasdottir, Maria Grünewald and Juni Palmgren at MEB, KI for valuable collaboration and discussions regarding biostatistics.

Colleagues at the Solna-site, KI for interesting on-going collaboration; Tomas Olsson, Fredrik Piehl, Maria Swanberg, Olle Lidman, Lars Klareskog, Leonid Padyukov, Anders Hamsten, Per Eriksson.

All members of MS Centrum at Huddinge, past and present; including nurses, doctors, physiotherapists, occupational therapists, secretaries, psychologist, welfare officers, including Kristina Gottberg, Birgitta Holm, Anna-Carin Oscarsson, Anna Österlund, Marita Ingemarsson, Ewa Roos-Waldebäck, Eva Johansson, Anna Aronsson, Jan Hillert, Sten Fredrikson, Claes Martin, Magnus Andersson, Anders Svenningsson, Hans Link, Pia Kivisäkk, Virginija Kerrenbauer, Johan Berg, Houmayoun Roshanisefat.

All colleagues at Neurologkliniken Huddinge. Even if I during the last few years mainly have been away on either research or “randningar”, I have not forgotten you. “Neurologen” is still the workplace where I feel at home. A special thanks to Bo Höjeberg and Vasilios Kostulas for encouraging me to go into neurology and research. Rayomand Press for valuable discussions. Göran Solders who was head of the clinical department during almost all my PhD-time and Lars-Olof Ronnevi, present head of the newly formed department for giving me the opportunity to do this research.

While working on a PhD it is easy to become absorbed by the work. Friends play a crucial role in reminding you of the real values of life. I want to give my warmest thanks to my friends, who might not have seen much of me lately because of all the work, but you all mean a lot to me: Sanne & Niklas, Anki & Robert + family, Carolina & Jonas + family, Lisa, Sanna, Cilla, Anna, Susanna, Agneta, Ulrika + families and Lindströms in Örebro among others.

Finally, my special warm thanks to my family, Torsten and Ulla for being supportive and helpful through-out the years. Erik and Jenny for friendship and fun together. Jörgen, for always finding the right words of encouragement, for motivating me during the work, but also constantly reminding me of the world outside of work, and for just being who you are.

The studies included in these thesis were supported by the European Commission (project number CT97-2422), the Multiple Sclerosis Society of Great Britain and Northern Ireland (grant 591/00), The Wellcome Trust (grant 057097), the Swedish Medical Research Council (project number 11023 and 11220), and the Swedish Association of Neurologically Disabled, the Bibbi och Niels Jensen Foundation, the Danish Multiple Sclerosis Society, the Danish Medical Research Council, a stipend from Biogen Idec and Karolinska Institutet.

REFERENCES

1. Compston, A. and A. Coles, *Multiple sclerosis*. Lancet, 2002. **359**(9313): p. 1221-31.
2. Noseworthy, J.H., et al., *Multiple sclerosis*. N Engl J Med, 2000. **343**(13): p. 938-52.
3. Olerup, O., et al., *Primarily chronic progressive and relapsing/remitting multiple sclerosis: two immunogenetically distinct disease entities*. Proc Natl Acad Sci U S A, 1989. **86**(18): p. 7113-7.
4. Lucchinetti, C. and W. Bruck, *The pathology of primary progressive multiple sclerosis*. Mult Scler, 2004. **10 Suppl 1**: p. S23-30.
5. Rudick, R., *Mechanisms of disability progression in primary progressive multiple sclerosis: are they different from secondary progressive multiple sclerosis?* Mult Scler, 2003. **9**(2): p. 210-2.
6. Vukusic, S. and C. Confavreux, *Primary and secondary progressive multiple sclerosis*. J Neurol Sci, 2003. **206**(2): p. 153-5.
7. Weinshenker, B.G., et al., *The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course*. Brain, 1989. **112 (Pt 6)**: p. 1419-28.
8. Weinshenker, B.G., et al., *The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability*. Brain, 1989. **112 (Pt 1)**: p. 133-46.
9. Barkhof, F., et al., *Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis*. Brain, 1997. **120 (Pt 11)**: p. 2059-69.
10. Tintore, M., et al., *Isolated demyelinating syndromes: comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis*. AJNR Am J Neuroradiol, 2000. **21**(4): p. 702-6.
11. Andersson, M., et al., *Cerebrospinal fluid in the diagnosis of multiple sclerosis: a consensus report*. J Neurol Neurosurg Psychiatry, 1994. **57**(8): p. 897-902.
12. Link, H. and G. Tibbling, *Principles of albumin and IgG analyses in neurological disorders. III. Evaluation of IgG synthesis within the central nervous system in multiple sclerosis*. Scand J Clin Lab Invest, 1977. **37**(5): p. 397-401.
13. Poser, C.M., et al., *New diagnostic criteria for multiple sclerosis: guidelines for research protocols*. Ann Neurol, 1983. **13**(3): p. 227-31.
14. McDonald, W.I., et al., *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis*. Ann Neurol, 2001. **50**(1): p. 121-7.
15. Jacobs, L.D., et al., *Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG)*. Ann Neurol, 1996. **39**(3): p. 285-94.
16. *Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group*. Lancet, 1998. **352**(9139): p. 1498-504.
17. *Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group*. Neurology, 1995. **45**(7): p. 1277-85.
18. Johnson, K.P., et al., *Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. The Copolymer 1 Multiple Sclerosis Study Group*. Neurology, 1995. **45**(7): p. 1268-76.
19. Comi, G., M. Filippi, and J.S. Wolinsky, *European/Canadian multicenter, double-blind, randomized, placebo-controlled study of the effects of glatiramer*

- acetate on magnetic resonance imaging--measured disease activity and burden in patients with relapsing multiple sclerosis. *European/Canadian Glatiramer Acetate Study Group*. *Ann Neurol*, 2001. **49**(3): p. 290-7.
20. Miller, D.H., et al., *A controlled trial of natalizumab for relapsing multiple sclerosis*. *N Engl J Med*, 2003. **348**(1): p. 15-23.
 21. Pettinelli, C.B. and D.E. McFarlin, *Adoptive transfer of experimental allergic encephalomyelitis in SJL/J mice after in vitro activation of lymph node cells by myelin basic protein: requirement for Lyt 1+ 2- T lymphocytes*. *J Immunol*, 1981. **127**(4): p. 1420-3.
 22. O'Connor, K.C., A. Bar-Or, and D.A. Hafler, *The neuroimmunology of multiple sclerosis: possible roles of T and B lymphocytes in immunopathogenesis*. *J Clin Immunol*, 2001. **21**(2): p. 81-92.
 23. Persidsky, Y., *Model systems for studies of leukocyte migration across the blood - brain barrier*. *J Neurovirol*, 1999. **5**(6): p. 579-90.
 24. Hickey, W.F., *Migration of hematogenous cells through the blood-brain barrier and the initiation of CNS inflammation*. *Brain Pathol*, 1991. **1**(2): p. 97-105.
 25. Mosmann, T.R., et al., *Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins*. *J Immunol*, 1986. **136**(7): p. 2348-57.
 26. Mosmann, T.R. and R.L. Coffman, *TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties*. *Annu Rev Immunol*, 1989. **7**: p. 145-73.
 27. Aloisi, F., F. Ria, and L. Adorini, *Regulation of T-cell responses by CNS antigen-presenting cells: different roles for microglia and astrocytes*. *Immunol Today*, 2000. **21**(3): p. 141-7.
 28. Weiner, H.L., *Multiple sclerosis is an inflammatory T-cell-mediated autoimmune disease*. *Arch Neurol*, 2004. **61**(10): p. 1613-5.
 29. van Oosten, B.W., et al., *Increased MRI activity and immune activation in two multiple sclerosis patients treated with the monoclonal anti-tumor necrosis factor antibody cA2*. *Neurology*, 1996. **47**(6): p. 1531-4.
 30. Cross, A.H., J.L. Trotter, and J. Lyons, *B cells and antibodies in CNS demyelinating disease*. *J Neuroimmunol*, 2001. **112**(1-2): p. 1-14.
 31. Archelos, J.J., M.K. Storch, and H.P. Hartung, *The role of B cells and autoantibodies in multiple sclerosis*. *Ann Neurol*, 2000. **47**(6): p. 694-706.
 32. Kornek, B. and H. Lassmann, *Axonal pathology in multiple sclerosis. A historical note*. *Brain Pathol*, 1999. **9**(4): p. 651-6.
 33. Brex, P.A., et al., *Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis*. *J Neurol Neurosurg Psychiatry*, 2001. **70**(4): p. 544-7.
 34. Lassmann, H., W. Bruck, and C. Lucchinetti, *Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy*. *Trends Mol Med*, 2001. **7**(3): p. 115-21.
 35. Lucchinetti, C., et al., *Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination*. *Ann Neurol*, 2000. **47**(6): p. 707-17.
 36. Kurtzke, J.F., *Epidemiologic evidence for multiple sclerosis as an infection*. *Clin Microbiol Rev*, 1993. **6**(4): p. 382-427.
 37. Dahl, O.P., et al., *Multiple sclerosis in Nord-Trondelag County, Norway: a prevalence and incidence study*. *Acta Neurol Scand*, 2004. **109**(6): p. 378-84.
 38. Celius, E.G. and B. Vandvik, *Multiple sclerosis in Oslo, Norway: prevalence on 1 January 1995 and incidence over a 25-year period*. *Eur J Neurol*, 2001. **8**(5): p. 463-9.
 39. Svenningsson, A., et al., *Incidence of MS during two fifteen-year periods in the Gothenburg region of Sweden*. *Acta Neurol Scand*, 1990. **82**(3): p. 161-8.
 40. Sundstrom, P., L. Nystrom, and L. Forsgren, *Incidence (1988-97) and prevalence (1997) of multiple sclerosis in Vasterbotten County in northern Sweden*. *J Neurol Neurosurg Psychiatry*, 2003. **74**(1): p. 29-32.

41. Sarasoja, T., et al., *Occurrence of multiple sclerosis in central Finland: a regional and temporal comparison during 30 years*. Acta Neurol Scand, 2004. **110**(5): p. 331-6.
42. Sumelahti, M.L., et al., *Increasing prevalence of multiple sclerosis in Finland*. Acta Neurol Scand, 2001. **103**(3): p. 153-8.
43. Benedikz, J., et al., *The natural history of untreated multiple sclerosis in Iceland. A total population-based 50 year prospective study*. Clin Neurol Neurosurg, 2002. **104**(3): p. 208-10.
44. Koch-Henriksen, N., *The Danish Multiple Sclerosis Registry: a 50-year follow-up*. Mult Scler, 1999. **5**(4): p. 293-6.
45. Fox, C.M., et al., *The epidemiology of multiple sclerosis in Devon: a comparison of the new and old classification criteria*. J Neurol Neurosurg Psychiatry, 2004. **75**(1): p. 56-60.
46. Robertson, N., et al., *Multiple sclerosis in the north Cambridgeshire districts of East Anglia*. J Neurol Neurosurg Psychiatry, 1995. **59**(1): p. 71-6.
47. Pugliatti, M., S. Sotgiu, and G. Rosati, *The worldwide prevalence of multiple sclerosis*. Clin Neurol Neurosurg, 2002. **104**(3): p. 182-91.
48. Pugliatti, M., et al., *Multiple sclerosis distribution in northern Sardinia: spatial cluster analysis of prevalence*. Neurology, 2002. **58**(2): p. 277-82.
49. Poser, C.M., *The dissemination of multiple sclerosis: a Viking saga? A historical essay*. Ann Neurol, 1994. **36 Suppl 2**: p. S231-43.
50. Poser, C.M., *Viking voyages: the origin of multiple sclerosis? An essay in medical history*. Acta Neurol Scand Suppl, 1995. **161**: p. 11-22.
51. Kurtzke, J.F., K.R. Gudmundsson, and S. Bergmann, *Multiple sclerosis in Iceland: 1. Evidence of a postwar epidemic*. Neurology, 1982. **32**(2): p. 143-50.
52. Cook, S.D., et al., *Multiple sclerosis and distemper in Iceland 1966-1978*. Acta Neurol Scand, 1980. **61**(4): p. 244-51.
53. Cook, S.D., et al., *Declining incidence of multiple sclerosis in the Orkney Islands*. Neurology, 1985. **35**(4): p. 545-51.
54. Sheremata, W.A., et al., *Unusual occurrence on a tropical island of multiple sclerosis*. Lancet, 1985. **2**(8455): p. 618.
55. Sadovnick, A.D., D. Dyment, and G.C. Ebers, *Genetic epidemiology of multiple sclerosis*. Epidemiol Rev, 1997. **19**(1): p. 99-106.
56. Benedikz, J., H. Magnusson, and G. Guthmundsson, *Multiple sclerosis in Iceland, with observations on the alleged epidemic in the Faroe Islands*. Ann Neurol, 1994. **36 Suppl 2**: p. S175-9.
57. Gronlie, S.A., et al., *Multiple sclerosis in North Norway, and first appearance in an indigenous population*. J Neurol, 2000. **247**(2): p. 129-33.
58. Skegg, D.C., et al., *Occurrence of multiple sclerosis in the north and south of New Zealand*. J Neurol Neurosurg Psychiatry, 1987. **50**(2): p. 134-9.
59. Compston, A., *Genetic epidemiology of multiple sclerosis*. J Neurol Neurosurg Psychiatry, 1997. **62**(6): p. 553-61.
60. Gale, C.R. and C.N. Martyn, *Migrant studies in multiple sclerosis*. Prog Neurobiol, 1995. **47**(4-5): p. 425-48.
61. Kurtzke, J.F. and W.F. Page, *Epidemiology of multiple sclerosis in US veterans: VII. Risk factors for MS*. Neurology, 1997. **48**(1): p. 204-13.
62. Hernan, M.A., M.J. Olek, and A. Ascherio, *Geographic variation of MS incidence in two prospective studies of US women*. Neurology, 1999. **53**(8): p. 1711-8.
63. Hammond, S.R., et al., *The epidemiology of multiple sclerosis in Queensland, Australia*. J Neurol Sci, 1987. **80**(2-3): p. 185-204.
64. Hammond, S.R., et al., *The contribution of mortality statistics to the study of multiple sclerosis in Australia*. J Neurol Neurosurg Psychiatry, 1989. **52**(1): p. 1-7.
65. Hammond, S.R., et al., *The epidemiology of multiple sclerosis in three Australian cities: Perth, Newcastle and Hobart*. Brain, 1988. **111 (Pt 1)**: p. 1-25.

66. Poser, C.M., *The epidemiology of multiple sclerosis: a general overview*. Ann Neurol, 1994. **36 Suppl 2**: p. S180-93.
67. Sola, P., et al., *Human herpesvirus 6 and multiple sclerosis: survey of anti-HHV-6 antibodies by immunofluorescence analysis and of viral sequences by polymerase chain reaction*. J Neurol Neurosurg Psychiatry, 1993. **56**(8): p. 917-9.
68. Moore, F.G. and C. Wolfson, *Human herpes virus 6 and multiple sclerosis*. Acta Neurol Scand, 2002. **106**(2): p. 63-83.
69. Levin, L.I., et al., *Multiple sclerosis and Epstein-Barr virus*. Jama, 2003. **289**(12): p. 1533-6.
70. Lily, O., *Epstein-Barr virus and risk of multiple sclerosis*. Jama, 2003. **290**(2): p. 192; author reply 193.
71. Sundstrom, P., et al., *An altered immune response to Epstein-Barr virus in multiple sclerosis: a prospective study*. Neurology, 2004. **62**(12): p. 2277-82.
72. Sriram, S., et al., *Chlamydia pneumoniae infection of the central nervous system in multiple sclerosis*. Ann Neurol, 1999. **46**(1): p. 6-14.
73. Swanborg, R.H., J.A. Whittum-Hudson, and A.P. Hudson, *Human herpesvirus 6 and Chlamydia pneumoniae as etiologic agents in multiple sclerosis - a critical review*. Microbes Infect, 2002. **4**(13): p. 1327-33.
74. Willer, C.J., et al., *Timing of birth and risk of multiple sclerosis: population based study*. Bmj, 2004.
75. Sadovnick, A.D., P.A. Baird, and R.H. Ward, *Multiple sclerosis: updated risks for relatives*. Am J Med Genet, 1988. **29**(3): p. 533-41.
76. Robertson, N.P., et al., *Age-adjusted recurrence risks for relatives of patients with multiple sclerosis*. Brain, 1996. **119 (Pt 2)**: p. 449-55.
77. Risch, N., *Linkage strategies for genetically complex traits. I. Multilocus models*. Am J Hum Genet, 1990. **46**(2): p. 222-8.
78. Ebers, G.C., et al., *A population-based study of multiple sclerosis in twins*. N Engl J Med, 1986. **315**(26): p. 1638-42.
79. Mumford, C.J., et al., *The British Isles survey of multiple sclerosis in twins*. Neurology, 1994. **44**(1): p. 11-5.
80. Ebers, G.C., A.D. Sadovnick, and N.J. Risch, *A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group*. Nature, 1995. **377**(6545): p. 150-1.
81. Sadovnick, A.D., et al., *Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group*. Lancet, 1996. **347**(9017): p. 1728-30.
82. Robertson, N.P., et al., *Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis*. Lancet, 1997. **349**(9065): p. 1587-90.
83. Ebers, G.C., et al., *Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. Canadian Collaborative Study Group*. Ann Neurol, 2000. **48**(6): p. 927-31.
84. Ellegren, H., *Microsatellites: simple sequences with complex evolution*. Nat Rev Genet, 2004. **5**(6): p. 435-45.
85. *The International HapMap Project*. Nature, 2003. **426**(6968): p. 789-96.
86. Holmans, P., *Affected sib-pair methods for detecting linkage to dichotomous traits: review of the methodology*. Hum Biol, 1998. **70**(6): p. 1025-40.
87. Risch, N., *Linkage strategies for genetically complex traits. II. The power of affected relative pairs*. Am J Hum Genet, 1990. **46**(2): p. 229-41.
88. Risch, N., *Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs*. Am J Hum Genet, 1990. **46**(2): p. 242-53.
89. Holmans, P., *Asymptotic properties of affected-sib-pair linkage analysis*. Am J Hum Genet, 1993. **52**(2): p. 362-74.
90. Holmans, P. and D. Clayton, *Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers*. Am J Hum Genet, 1995. **57**(5): p. 1221-32.
91. Kruglyak, L. and E.S. Lander, *Complete multipoint sib-pair analysis of qualitative and quantitative traits*. Am J Hum Genet, 1995. **57**(2): p. 439-54.

92. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)*. *Am J Hum Genet*, 1993. **52**(3): p. 506-16.
93. Hauser, E.R., et al., *Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations*. *Genet Epidemiol*, 1996. **13**(2): p. 117-37.
94. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results*. *Nat Genet*, 1995. **11**(3): p. 241-7.
95. Sawcer, S., et al., *Empirical genomewide significance levels established by whole genome simulations*. *Genet Epidemiol*, 1997. **14**(3): p. 223-9.
96. Zavattari, P., et al., *Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection*. *Hum Mol Genet*, 2000. **9**(20): p. 2947-57.
97. Labuda, M., et al., *Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians*. *Am J Hum Genet*, 1996. **59**(3): p. 633-43.
98. Peltonen, L., A. Palotie, and K. Lange, *Use of population isolates for mapping complex traits*. *Nat Rev Genet*, 2000. **1**(3): p. 182-90.
99. Varilo, T. and L. Peltonen, *Isolates and their potential use in complex gene mapping efforts*. *Curr Opin Genet Dev*, 2004. **14**(3): p. 316-23.
100. Lewontin, R.C., *On measures of gametic disequilibrium*. *Genetics*, 1988. **120**(3): p. 849-52.
101. Reich, D.E., et al., *Linkage disequilibrium in the human genome*. *Nature*, 2001. **411**(6834): p. 199-204.
102. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome*. *Science*, 2002. **296**(5576): p. 2225-9.
103. Tonin, P.N., *Genes implicated in hereditary breast cancer syndromes*. *Semin Surg Oncol*, 2000. **18**(4): p. 281-6.
104. Compston, A. and S. Sawcer, *Genetic analysis of multiple sclerosis*. *Curr Neurol Neurosci Rep*, 2002. **2**(3): p. 259-66.
105. Giordano, M., S. D'Alfonso, and P. Momigliano-Richiardi, *Genetics of multiple sclerosis: linkage and association studies*. *Am J Pharmacogenomics*, 2002. **2**(1): p. 37-58.
106. Kenealy, S.J., M.A. Pericak-Vance, and J.L. Haines, *The genetic epidemiology of multiple sclerosis*. *J Neuroimmunol*, 2003. **143**(1-2): p. 7-12.
107. Dyment, D.A., G.C. Ebers, and A.D. Sadovnick, *Genetics of multiple sclerosis*. *Lancet Neurol*, 2004. **3**(2): p. 104-10.
108. Jersild, C., A. Svejgaard, and T. Fog, *HLA-A antigens and multiple sclerosis*. *Lancet*, 1972. **1**(7762): p. 1240-1.
109. Jersild, C., et al., *Histocompatibility-linked immune-response determinants in multiple sclerosis*. *Transplant Proc*, 1973. **5**(4): p. 1791-6.
110. Hillert, J., *Human leukocyte antigen studies in multiple sclerosis*. *Ann Neurol*, 1994. **36 Suppl**: p. S15-7.
111. Marrosu, M.G., et al., *Multiple sclerosis in Sardinia is associated and in linkage disequilibrium with HLA-DR3 and -DR4 alleles*. *Am J Hum Genet*, 1997. **61**(2): p. 454-7.
112. Marrosu, M.G., et al., *DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population*. *Hum Mol Genet*, 1998. **7**(8): p. 1235-7.
113. Marrosu, M.G., et al., *Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia*. *Hum Mol Genet*, 2001. **10**(25): p. 2907-16.
114. Masterman, T., et al., *HLA-DR15 is associated with lower age at onset in multiple sclerosis*. *Ann Neurol*, 2000. **48**(2): p. 211-9.
115. Hensiek, A.E., et al., *HLA-DR 15 is associated with female sex and younger age at diagnosis in multiple sclerosis*. *J Neurol Neurosurg Psychiatry*, 2002. **72**(2): p. 184-7.

116. Celius, E.G., et al., *Sex and age at diagnosis are correlated with the HLA-DR2, DQ6 haplotype in multiple sclerosis*. J Neurol Sci, 2000. **178**(2): p. 132-5.
117. Barcellos, L.F., et al., *HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course*. Am J Hum Genet, 2003. **72**(3): p. 710-6.
118. Modin, H., et al., *Modes of action of HLA-DR susceptibility specificities in multiple sclerosis*. Am J Hum Genet, 2004. **74**(6): p. 1321-2.
119. Fogdell-Hahn, A., et al., *Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease*. Tissue Antigens, 2000. **55**(2): p. 140-8.
120. Harbo, H.F., et al., *Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis*. Tissue Antigens, 2004. **63**(3): p. 237-47.
121. Rubio, J.P., et al., *Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis*. Am J Hum Genet, 2002. **70**(5): p. 1125-37.
122. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families*. Science, 1993. **261**(5123): p. 921-3.
123. Farrer, L.A., et al., *Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium*. Jama, 1997. **278**(16): p. 1349-56.
124. De Stefano, N., et al., *Influence of apolipoprotein E epsilon4 genotype on brain tissue integrity in relapsing-remitting multiple sclerosis*. Arch Neurol, 2004. **61**(4): p. 536-40.
125. Enzinger, C., et al., *Accelerated evolution of brain atrophy and "black holes" in MS patients with APOE-epsilon 4*. Ann Neurol, 2004. **55**(4): p. 563-9.
126. Savettieri, G., et al., *Apolipoprotein E genotype does not influence the progression of multiple sclerosis*. J Neurol, 2003. **250**(9): p. 1094-8.
127. Haines, J.L., et al., *A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. The Multiple Sclerosis Genetics Group*. Nat Genet, 1996. **13**(4): p. 469-71.
128. Sawcer, S., et al., *A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22*. Nat Genet, 1996. **13**(4): p. 464-8.
129. Ebers, G.C., et al., *A full genome search in multiple sclerosis*. Nat Genet, 1996. **13**(4): p. 472-6.
130. Kuokkanen, S., et al., *Genomewide scan of multiple sclerosis in Finnish multiplex families*. Am J Hum Genet, 1997. **61**(6): p. 1379-87.
131. Coraddu, F., et al., *A genome screen for multiple sclerosis in Sardinian multiplex families*. Eur J Hum Genet, 2001. **9**(8): p. 621-6.
132. Broadley, S., et al., *A genome screen for multiple sclerosis in Italian families*. Genes Immun, 2001. **2**(4): p. 205-10.
133. Akesson, E., et al., *A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis*. Genes Immun, 2002. **3**(5): p. 279-85.
134. Eraksoy, M., et al., *A whole genome screen for linkage in Turkish multiple sclerosis*. J Neuroimmunol, 2003. **143**(1-2): p. 17-24.
135. Ban, M., et al., *A genome screen for linkage in Australian sibling-pairs with multiple sclerosis*. Genes Immun, 2002. **3**(8): p. 464-9.
136. Kenealy, S.J., et al., *A second-generation genomic screen for multiple sclerosis*. Am J Hum Genet, 2004. **75**(6): p. 1070-8.
137. Dyment, D.A., et al., *An extended genome scan in 442 Canadian multiple sclerosis-affected sibships: a report from the Canadian Collaborative Study Group*. Hum Mol Genet, 2004. **13**(10): p. 1005-15.
138. Hensiek, A.E., et al., *Updated results of the United Kingdom linkage-based genome screen in multiple sclerosis*. J Neuroimmunol, 2003. **143**(1-2): p. 25-30.
139. *A meta-analysis of genomic screens in multiple sclerosis. The Transatlantic Multiple Sclerosis Genetics Cooperative*. Mult Scler, 2001. **7**(1): p. 3-11.

140. *A meta-analysis of whole genome linkage screens in multiple sclerosis.* J Neuroimmunol, 2003. **143**(1-2): p. 39-46.
141. Vitale, E., et al., *Linkage analysis conditional on HLA status in a large North American pedigree supports the presence of a multiple sclerosis susceptibility locus on chromosome 12p12.* Hum Mol Genet, 2002. **11**(3): p. 295-300.
142. Modin, H., et al., *Genome-wide linkage screen of a consanguineous multiple sclerosis kinship.* Mult Scler, 2003. **9**(2): p. 128-34.
143. Barcellos, L.F. and G. Thomson, *Genetic analysis of multiple sclerosis in Europeans.* J Neuroimmunol, 2003. **143**(1-2): p. 1-6.
144. Sawcer, S. and A. Compston, *The genetic analysis of multiple sclerosis in Europeans: concepts and design.* J Neuroimmunol, 2003. **143**(1-2): p. 13-6.
145. Barcellos, L.F., et al., *Association mapping of disease loci, by use of a pooled DNA genomic screen.* Am J Hum Genet, 1997. **61**(3): p. 734-47.
146. Sawcer, S., et al., *A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility.* Brain, 2002. **125**(Pt 6): p. 1337-47.
147. Goedde, R., et al., *A genome screen for linkage disequilibrium in HLA-DRB1*15-positive Germans with multiple sclerosis based on 4666 microsatellite markers.* Hum Genet, 2002. **111**(3): p. 270-7.
148. Alizadeh, M., et al., *Genetic analysis of multiple sclerosis in Europeans: French data.* J Neuroimmunol, 2003. **143**(1-2): p. 74-8.
149. Ban, M., et al., *A genome-wide screen for linkage disequilibrium in Australian HLA-DRB1*1501 positive multiple sclerosis patients.* J Neuroimmunol, 2003. **143**(1-2): p. 60-4.
150. Bielecki, B., et al., *A whole genome screen for association in Polish multiple sclerosis patients.* J Neuroimmunol, 2003. **143**(1-2): p. 107-11.
151. Coraddu, F., et al., *A genome-wide screen for linkage disequilibrium in Sardinian multiple sclerosis.* J Neuroimmunol, 2003. **143**(1-2): p. 120-3.
152. Eraksoy, M., et al., *A genome screen for linkage disequilibrium in Turkish multiple sclerosis.* J Neuroimmunol, 2003. **143**(1-2): p. 129-32.
153. Goertsches, R., et al., *A genomic screen of Spanish multiple sclerosis patients reveals multiple loci associated with the disease.* J Neuroimmunol, 2003. **143**(1-2): p. 124-8.
154. Goris, A., et al., *New candidate loci for multiple sclerosis susceptibility revealed by a whole genome association screen in a Belgian population.* J Neuroimmunol, 2003. **143**(1-2): p. 65-9.
155. Harbo, H.F., et al., *Two genome-wide linkage disequilibrium screens in Scandinavian multiple sclerosis patients.* J Neuroimmunol, 2003. **143**(1-2): p. 101-6.
156. Heggarty, S., et al., *A genome wide scan for association with multiple sclerosis in a N. Irish case control population.* J Neuroimmunol, 2003. **143**(1-2): p. 93-6.
157. Jonasdottir, A., et al., *A whole genome association study in Icelandic multiple sclerosis patients with 4804 markers.* J Neuroimmunol, 2003. **143**(1-2): p. 88-92.
158. Laaksonen, M., et al., *A whole genome association study in Finnish multiple sclerosis patients with 3669 markers.* J Neuroimmunol, 2003. **143**(1-2): p. 70-3.
159. Liguori, M., et al., *A whole genome screen for linkage disequilibrium in multiple sclerosis performed in a continental Italian population.* J Neuroimmunol, 2003. **143**(1-2): p. 97-100.
160. Martins Silva, B., et al., *A whole genome association study in multiple sclerosis patients from north Portugal.* J Neuroimmunol, 2003. **143**(1-2): p. 116-9.
161. Rajda, C., et al., *A genome-wide screen for association in Hungarian multiple sclerosis.* J Neuroimmunol, 2003. **143**(1-2): p. 84-7.
162. Santos, M., et al., *A whole genome screen for association with multiple sclerosis in Portuguese patients.* J Neuroimmunol, 2003. **143**(1-2): p. 112-5.
163. Weber, A., et al., *A genome-wide German screen for linkage disequilibrium in multiple sclerosis.* J Neuroimmunol, 2003. **143**(1-2): p. 79-83.

164. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells*. Nucleic Acids Res, 1988. **16**(3): p. 1215.
165. Don, R.H., et al., *'Touchdown' PCR to circumvent spurious priming during gene amplification*. Nucleic Acids Res, 1991. **19**(14): p. 4008.
166. Rithidech, K.N., J.J. Dunn, and C.R. Gordon, *Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms*. Biotechniques, 1997. **23**(1): p. 36, 40, 42, 44.
167. Palsson, B., et al., *Using quality measures to facilitate allele calling in high-throughput genotyping*. Genome Res, 1999. **9**(10): p. 1002-12.
168. Jurinke, C., et al., *Automated genotyping using the DNA MassArray technology*. Methods Mol Biol, 2002. **187**: p. 179-92.
169. Buetow, K.H., et al., *High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Proc Natl Acad Sci U S A, 2001. **98**(2): p. 581-4.
170. Gyapay, G., et al., *The 1993-94 Genethon human genetic linkage map*. Nat Genet, 1994. **7**(2 Spec No): p. 246-339.
171. Collins, A., et al., *A metric map of humans: 23,500 loci in 850 bands*. Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14771-5.
172. Davies, J.L., et al., *A genome-wide search for human type 1 diabetes susceptibility genes*. Nature, 1994. **371**(6493): p. 130-6.
173. Broman, K.W., et al., *Comprehensive human genetic maps: individual and sex-specific variation in recombination*. Am J Hum Genet, 1998. **63**(3): p. 861-9.
174. Kong, A., et al., *A high-resolution recombination map of the human genome*. Nat Genet, 2002. **31**(3): p. 241-7.
175. Ehm, M. and M. Wagner, *A test statistic to detect errors in sib-pair relationships*. Am J Hum Genet, 1998. **62**(1): p. 181-8.
176. Holmans, P., *Detecting gene-gene interactions using affected sib pair analysis with covariates*. Hum Hered, 2002. **53**(2): p. 92-102.
177. Clayton, D., *A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission*. Am J Hum Genet, 1999. **65**(4): p. 1170-7.
178. Excoffier, L. and M. Slatkin, *Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population*. Mol Biol Evol, 1995. **12**(5): p. 921-7.
179. Gretarsdottir, S., et al., *The gene encoding phosphodiesterase 4D confers risk of ischemic stroke*. Nat Genet, 2003. **35**(2): p. 131-8.
180. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps*. Bioinformatics, 2004.
181. Cavanaugh, J., *International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16*. Am J Hum Genet, 2001. **68**(5): p. 1165-71.
182. Kruglyak, L. and E.S. Lander, *High-resolution genetic mapping of complex traits*. Am J Hum Genet, 1995. **56**(5): p. 1212-23.
183. Liew, F.Y., *The role of innate cytokines in inflammatory response*. Immunol Lett, 2003. **85**(2): p. 131-4.
184. Pashenkov, M., et al., *Levels of interleukin-15-expressing blood mononuclear cells are elevated in multiple sclerosis*. Scand J Immunol, 1999. **50**(3): p. 302-8.
185. Kivisakk, P., et al., *IL-15 mRNA expression is up-regulated in blood and cerebrospinal fluid mononuclear cells in multiple sclerosis (MS)*. Clin Exp Immunol, 1998. **111**(1): p. 193-7.
186. Arena, A., et al., *Role of IL-15 on monocytic resistance to human herpesvirus 6 infection*. New Microbiol, 2000. **23**(2): p. 105-12.
187. Flamand, L., I. Stefanescu, and J. Menezes, *Human herpesvirus-6 enhances natural killer cell cytotoxicity via IL-15*. J Clin Invest, 1996. **97**(6): p. 1373-81.
188. Deloukas, P., et al., *The DNA sequence and comparative analysis of human chromosome 10*. Nature, 2004. **429**(6990): p. 375-81.

189. John, S., et al., *Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites*. *Am J Hum Genet*, 2004. **75**(1): p. 54-64.
190. Sawcer, S.J., et al., *Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping*. *Hum Mol Genet*, 2004. **13**(17): p. 1943-9.
191. Evans, D.M. and L.R. Cardon, *Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps*. *Am J Hum Genet*, 2004. **75**(4): p. 687-92.
192. Abecasis, G.R., S.S. Cherny, and L.R. Cardon, *The impact of genotyping error on family-based analysis of quantitative traits*. *Eur J Hum Genet*, 2001. **9**(2): p. 130-4.
193. Kruglyak, L., *The use of a genetic map of biallelic markers in linkage studies*. *Nat Genet*, 1997. **17**(1): p. 21-4.
194. Clark, A.G., *The role of haplotypes in candidate gene studies*. *Genet Epidemiol*, 2004. **27**(4): p. 321-33.
195. Pericak-Vance, M.A., et al., *Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage*. *Am J Hum Genet*, 1991. **48**(6): p. 1034-50.
196. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes*. *Nat Genet*, 2000. **26**(1): p. 76-80.
197. Evans, D.M., L.R. Cardon, and A.P. Morris, *Genotype prediction using a dense map of SNPs*. *Genet Epidemiol*, 2004. **27**(4): p. 375-84.
198. Wall, J.D. and J.K. Pritchard, *Haplotype blocks and linkage disequilibrium in the human genome*. *Nat Rev Genet*, 2003. **4**(8): p. 587-97.
199. Daly, M.J., et al., *High-resolution haplotype structure in the human genome*. *Nat Genet*, 2001. **29**(2): p. 229-32.
200. Patil, N., et al., *Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21*. *Science*, 2001. **294**(5547): p. 1719-23.
201. Dawson, E., et al., *A first-generation linkage disequilibrium map of human chromosome 22*. *Nature*, 2002. **418**(6897): p. 544-8.
202. Zhang, K., et al., *A dynamic programming algorithm for haplotype block partitioning*. *Proc Natl Acad Sci U S A*, 2002. **99**(11): p. 7335-9.
203. Phillips, M.S., et al., *Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots*. *Nat Genet*, 2003. **33**(3): p. 382-7.
204. Wang, N., et al., *Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation*. *Am J Hum Genet*, 2002. **71**(5): p. 1227-34.
205. Stumpf, M.P. and D.B. Goldstein, *Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium*. *Curr Biol*, 2003. **13**(1): p. 1-8.
206. Jeffreys, A.J., L. Kauppi, and R. Neumann, *Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex*. *Nat Genet*, 2001. **29**(2): p. 217-22.
207. Cardon, L.R. and G.R. Abecasis, *Using haplotype blocks to map human complex trait loci*. *Trends Genet*, 2003. **19**(3): p. 135-40.
208. Colhoun, H.M., P.M. McKeigue, and G. Davey Smith, *Problems of reporting genetic associations with complex outcomes*. *Lancet*, 2003. **361**(9360): p. 865-72.
209. Thomas, D.C. and D.G. Clayton, *Betting odds and genetic associations*. *J Natl Cancer Inst*, 2004. **96**(6): p. 421-3.
210. Nyholt, D.R., *A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other*. *Am J Hum Genet*, 2004. **74**(4): p. 765-9.
211. Zondervan, K.T. and L.R. Cardon, *The complex interplay among factors that influence allelic association*. *Nat Rev Genet*, 2004. **5**(2): p. 89-100.

212. Bird, T.D., G.P. Jarvik, and N.W. Wood, *Genetic association studies: genes in search of diseases*. *Neurology*, 2001. **57**(7): p. 1153-4.
213. Cooper, D.N., R.L. Nussbaum, and M. Krawczak, *Proposed guidelines for papers describing DNA polymorphism-disease associations*. *Hum Genet*, 2002. **110**(3): p. 207-8.
214. Kruglyak, L., *Prospects for whole-genome linkage disequilibrium mapping of common disease genes*. *Nat Genet*, 1999. **22**(2): p. 139-44.