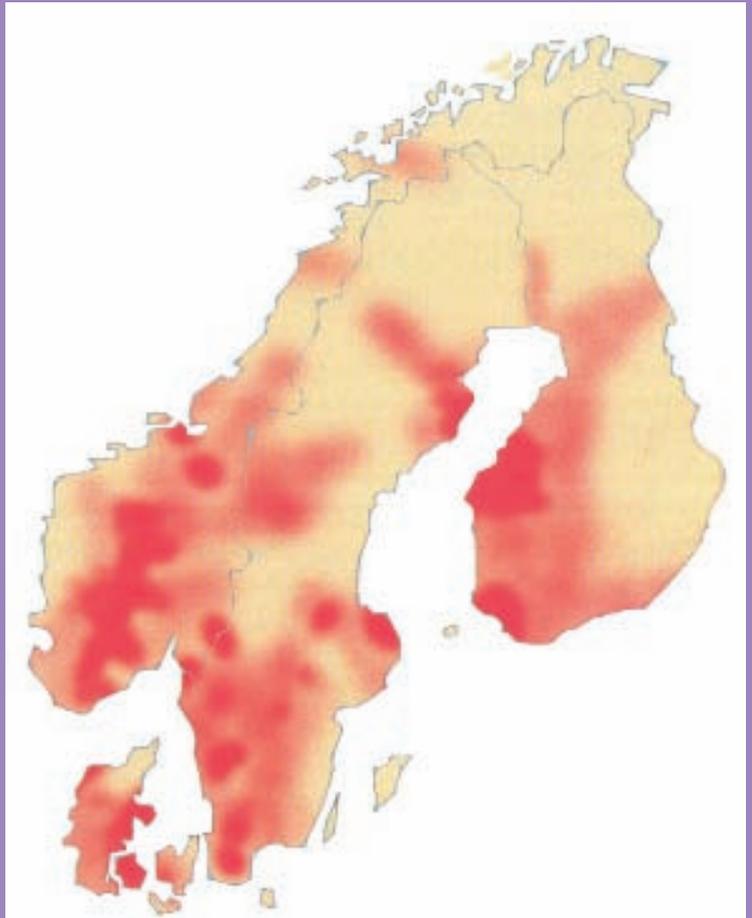




Suvi P. Kallio

Novel Multiple Sclerosis Predisposing Genetic Variants Outside the *HLA* Region



Suvi P. Kallio

Novel Multiple Sclerosis Predisposing Genetic Variants Outside the *HLA* Region

Academic dissertation

*To be presented with the permission of the Medical Faculty,
University of Helsinki, for public examination in the Lecture Hall 2,
Biomedicum Helsinki, on June 26th, at 12 noon.*

National Public Health Institute,
Helsinki, Finland

and

National Institute for Health and Welfare,
Helsinki, Finland

and

Department of Medical Genetics,
University of Helsinki, Finland

RESEARCH 16

Helsinki 2009

© Suvi P. Kallio and National Institute for Health and Welfare

Cover graphic: A heat map illustrating prevalence of MS in Scandinavia. The more red is the area, the higher is the prevalence. Obtained from tutorial "Minulla on MS" with a permission of the copyright holder Dr. Juhani Ruutiainen, The Finnish MS Society. Original data from Kurtzke 1974.

Layout: Christine Strid

ISBN 978-952-245-097-5 (printed)

ISSN 1798-0054 (printed)

ISBN 978-952-245-098-2 (pdf)

ISSN 1798-0062 (pdf)

Helsinki University Print

Helsinki, Finland 2009

Supervised by

Academy Professor Leena Peltonen-Palotie
Finnish Institute for Molecular Medicine and
University of Helsinki
Department of Medical Genetics
Helsinki, Finland and
The Sanger Institute
Cambridge, UK

Adjunct Professor Janna Saarela
Finnish Institute for Molecular Medicine and
University of Helsinki
Department of Medical Genetics
Helsinki, Finland

Reviewed by

Professor Anne Kallioniemi
University of Tampere
Institute of Medical Technology
Tampere, Finland

Professor emeritus Arne Svejgaard
University Hospital of Copenhagen
Department of Clinical Immunology
Copenhagen, Denmark

Opponent

Professor Leif Groop
University of Lund
Department of Endocrinology
Malmö, Sweden

To my family

Abstract

Suvi P. Kallio. Novel Multiple Sclerosis Predisposing Genetic Variants Outside the *HLA* Region. National Institute for Health and Welfare (THL), Research 16/2009. 139 Pages. Helsinki 2009. ISBN 978-952-245-097-5

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS) affecting approximately 0.1 % of Europeans. Similar to what occurs in numerous other complex diseases, both environmental factors and several predisposing genes are required to generate MS. However, despite long-standing and intensive research, these risk factors are largely unknown and the pathogenesis of MS demyelination is still poorly understood. Hence, no accurate diagnostic tools or curative treatment exists.

Both prevalence (0.2 %) and familial occurrence of MS are exceptionally high in a Finnish population subisolate, Southern Ostrobothnia, presumably due to enrichment of predisposing genetic variants within this geographical region. Previous linkage scan on large MS pedigrees from Southern Ostrobothnia detected three main MS loci on chromosomes 5p, 6p (*HLA*) and 17q. Linkage studies in several other populations have also provided independent evidence for the location of MS susceptibility genes in these regions, and further, these loci are syntenic to the experimental autoimmune encephalomyelitis (EAE) susceptibility loci of rodents, supporting their role in predisposition to autoimmune demyelination.

In this thesis work an effort was made to localize MS predisposing alleles of the linked loci outside the *HLA* region and to better understand the molecular mechanisms of MS. Taking into account that MS most probably is not a unitary disorder, but instead may represent an overlapping spectrum of related disorders, we have minimized the genetic and environmental heterogeneity by studying familial MS cases from the Southern Ostrobothnia isolate.

A scan of the 17q locus provided evidence for association with variants of the protein kinase C alpha (*PRKCA*) gene ($p = 0.0001$). Modest evidence for association with *PRKCA* was observed also in MS families from Canada.

Analysis of the 5p locus revealed one region, flanking the complement component 7 (*C7*) and hypothetical protein LOC133558 (*FLJ40243*) genes. The identified relatively rare haplotype seems to have a fairly large effect on genetic susceptibility of MS (frequency 12 % in MS cases and 4 % in controls, $p = 0.000003$, $OR = 2.73$). Evidence for association with alleles of the region and MS was also seen in more heterogeneous populations. Convincingly, plasma *C7* protein levels and total complement activity correlated with the risk haplotype identified.

The fairly strong association with the haplotype flanking *C7* stimulated us to study other complement cascade genes in MS. Previous publications have provided functional evidence for involvement of *C3* in autoimmune demyelination. However,

the data of this work suggests that variation in the complement component coding genes outside 5p is not associated with genetic susceptibility of MS, at least in Finland.

Finally we used a candidate gene approach to identify potential MS loci. Loss-of-function mutations of the *DAP12* and *TREM2* genes cause a recessively inherited CNS white matter disease PLOSL. Interestingly, *DAP12* and *TREM2* are located in MS regions on 6p and 19q, and we tested them as potential candidate genes in the Finnish MS sample. No evidence for association with MS was observed, and the Finnish PLOSL mutation was not over-represented among Finnish MS cases compared to controls (carrier frequency 5/1,000). Thus, the highly conserved *DAP12* and *TREM2* genes unlikely have a role in genetic susceptibility of MS.

This thesis contributes to the existing studies of complex disease genetics by providing an example of how extended families from special populations can be utilized in fine-mapping of the linked loci. A first relatively rare MS variant was here identified utilizing the strength of a Finnish population subisolate. This variant seems to have an effect on activity of the complement system, which has previously been suggested to have an important role in pathogenesis of MS. Thus, according to these results the role of the complement system in MS should be explored more indepth.

Keywords: multiple sclerosis, MS, complex disease, association analysis, linkage, complement cascade

Abstract in Finnish

Suvi P. Kallio. Novel Multiple Sclerosis Predisposing Genetic Variants Outside the *HLA* Region [Multippeliskleroosille altistavien geenivarianttien paikantaminen *HLA*-alueen ulkopuolelta]. Terveystieteiden tutkimuskeskus (THL), Tutkimus 16/2009. 139 sivua. Helsinki 2009. ISBN 978-952-245-097-5

Multippeliskleroosi (MS-tauti) on krooninen tulehduksellinen keskushermostotauti, johon sairastuu noin 0.1 % eurooppalaisista. Kuten monien muidenkin monitekijäisten sairauksien kohdalla, sekä ympäristötekijät että useat alttiusgeenit yhdessä johtavat MS-taudin puhkeamiseen. Pitkäaikaisesta ja intensiivisestä tutkimuksesta huolimatta suurin osa näistä MS-taudin riskitekijöistä on kuitenkin vielä tunnistamatta ja MS-taudissa tapahtuvan myeliinituhon syntymekanismit ovat huonosti ymmärrettyjä. Näin ollen MS-taudille ei ole olemassa täsmällistä diagnostiikkaa tai parantavaa hoitoa.

Sekä MS-taudin esiintyvyys (0.2 %) että suvuittainen esiintyminen ovat poikkeuksellisen korkeat erään Suomen väestösolaatin alueella, Etelä-Pohjanmaalla. Tämä johtuu luultavasti perinnöllisten MS-taudin alttiustekijöiden rikastumisesta tälle maantieteelliselle alueelle. Suomalaisessa kytkentätutkimuksessa on aiemmin Etelä-Pohjanmaan MS-sukuja tutkimalla tunnistettu kolme pääasiallista MS-taudin geenipaikkaa kromosomeissa 5p, 6p (*HLA*) ja 17q. Myös muissa väestöissä tehdyt kytkentäanalyytit tukevat löydöstä. Lisäksi nämä geenipaikat vastaavat jyrskiöiden kokeellisen autoimmuunienkefalomyeliitin (EAE) alttiusalueita, vahvistaen näiden kromosomien merkitystä autoimmuuni-demyelinaation kehittämisessä.

Tässä väitöskirjatutkimuksessa on pyritty paikantamaan MS-taudille altistavia geenivariantteja *HLA*-alueen ulkopuolisilta kytkentäalueilta ja siten ymmärtämään paremmin MS-taudin syntymekanismeja. Koska MS-tauti ei luultavasti ole vain yksi sairaus vaan ennemminkin kirjo samankaltaisia sairauksia, pyrimme määrittämään perinnöllisten alttiustekijöiden ja ympäristötekijöiden vaihtelun tutkimalla MS-sukuja Etelä-Pohjanmaan isolaattialueelta.

Kromosomin 17q alttiusalueen analyysissä havaittiin proteiinikinaasi C alfa-geenin (*PRKCA*) varianttien liittyvän MS-alttiuteen ($p = 0.0001$). Myös kanadalaisia MS-perheitä tutkimalla todettiin assosiaatio *PRKCA*-geeniin.

Kromosomin 5p alttiusalueen analyysissä havaittiin yhdellä suhteellisen harvinaisella haplotyypillä olevan suurehko vaikutus MS-alttiuteen (yleisyys 12 % MS-potilailla ja 4 % kontrolleilla, $p = 0.000003$, $OR = 2.73$). Tämä haplotyyppi sivuaa komplementti komponentti 7 (*C7*) ja hypoteettinen proteiini LOC133558 (*FLJ40243*) geenejä. Alueen alleelien havaittiin liittyvän MS-alttiuteen myös heterogeenisemmissä väestöissä. Lisäksi plasman *C7*-proteiinitasojen sekä komplementtisysteemin aktiivisuuden havaittiin korreloivan tunnistetun riskihaplotyyppin kantajuuteen.

Suhteellisen vahva assosiaatio *C7*-geeniin osuvaan haplotyyppiin kannusti tutkimaan myös muiden komplementtikomponentteja koodaavien geenien osuutta MS-taudissa. Lukuisissa aiemmissä tutkimuksissa on havaittu yhteys *C3*-proteiinitasojen ja immunologisen myeliinituhon välillä. Tässä tutkimuksessa kromosomin 5p ulkopuolisten komplementtigeenien ei kuitenkaan voitu todeta liittyvän MS-alttiuteen suomalaisilla.

Väitöskirjan viimeisessä osatyössä tutkittiin MS-taudin ehdokasgeenejä. *DAP12* ja *TREM2* geenien toimimattomuuteen johtavat mutaatiot aiheuttavat peityvästi periytyvän keskushermoston valkean aineen sairauden, PLOSL:n. Mielenkiintoista on, että kyseiset geenit sijaitsevat kromosomien 6p ja 19q alueilla, joiden on havaittu kytkeytyvän myös MS-tautiin. Osatyössä selvitetttiinkin, altistavatko *DAP12* ja *TREM2* myös MS-taudille. Näiden geenien ei todettu assosioituvan MS-alttiuteen, eikä suomalaisen PLOSL-mutaation havaittu rikastuneen Suomen MS-potilaille verrattuna verrokkeihin (kantajuus 5/1000).

Tämä väitöskirjatyö tukee monitekijäisten tautien geneettistä tutkimusta kuvaamalla, miten erityisväestöjen sukuaineistoja voidaan hyödyntää alttiusvarianttien paikantamiseksi kytkentäalueilta. Tutkimuksessa tunnistettiin suomalaisista väestöisolaattia aineistona hyödyntäen ensimmäinen suhteellisen harvinainen MS-taudille altistava geenivariantti. Tämä variantti vaikuttaisi säätelevän komplementtiaktiivisuutta, jolla on jo aiemmin todettu olevan tärkeä rooli MS-taudin patogeneesissa. Havaitut tulokset kannustavatkin tutkimaan tarkemmin komplementtisysteemin roolia MS-taudin kehittämisessä.

Avainsanat: multipeliskleroosi, MS-tauti, monitekijäinen sairaus, assosiaatioanalyysi, kytkentä, komplementtikaskadi

Contents

Abstract

Abstract in Finnish

Abbreviations	13
List of original publications	15

1 INTRODUCTION	17
2 REVIEW OF THE LITERATURE	18
2.1 Multiple sclerosis	18
2.1.1 Epidemiology and diagnosis	18
2.1.2 Pathological hallmarks and theories of pathogenesis	22
2.2 Genetics of complex diseases	25
2.2.1 The human genome	25
2.2.2 Sequence variation	26
2.2.3 Strategies to identify genes underlying complex traits	28
2.2.4 The Finnish population and its subisolates	31
2.3 Genetics of multiple sclerosis	34
2.3.1 Genome-wide linkage screens of MS	34
2.3.2 MS candidate gene studies	37
2.3.3 The first genome-wide association scan of MS	40
2.3.4 Other strategies to map MS predisposing genetic variants	42
3 AIMS OF THE STUDY	44
4 MATERIALS AND METHODS	45
4.1 Study sample	45
4.1.1 Finnish MS sample	45
4.1.2 Selection of cases and controls for the Finnish GWA study	46
4.1.3 Study samples from more heterogeneous populations	48
4.2 Laboratory methods and statistical analyses	49
5 RESULTS AND DISCUSSION	52
5.1 Fine-mapping of the linked loci on chromosomes 17 and 5 (I, II)	52
5.1.1 MS locus on chromosome 17q (I)	52
5.1.2 MS locus on chromosome 5p (II)	58
5.2 A follow-up study: variation in other complement cascade genes in MS (III)	66
5.3 Candidate genes for immune-mediated demyelination on MS linked loci (IV)	68
6 CONCLUDING REMARKS	72
7 ACKNOWLEDGEMENTS	74
References	76

Original publications

Abbreviations

<i>2H2D2A</i>	SH2 domain protein 2A
AITD	Autoimmune thyroid disease
BBB	Blood-brain barrier
C3	Complement component 3
C5	Complement component 5
C7	Complement component 7
C9	Complement component 9
CD-CV	Common disease-common variant
CEU	Caucasians of European origin
CNS	Central nervous system
CNV	Copy number variation
CSF	Cerebrospinal fluid
<i>CTLA-4</i>	Cytotoxic T lymphocyte antigen 4
DNA	Deoxyribonucleic acid
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
<i>FLJ40243</i>	Hypothetical protein LOC133558
<i>FYB</i>	Fyn binding protein
GWA	Genome-wide association
HGP	Human genome project
HHRR	Haplotype-based haplotype relative risk
HHV-6	Human herpesvirus-6
<i>HLA</i>	Human leukocyte antigen
IBD	Identical by descent
IBS	Identical by state
<i>ICAM2</i>	Intercellular adhesion molecule 2
<i>IRF5</i>	Interferon regulatory factor 5
<i>IL2RA</i>	Interleukin 2 receptor, alpha chain
<i>IL7R</i>	Interleukin 7 receptor
LD	Linkage disequilibrium
LOD	Logarithm of odds
MAC	Membrane attack complex
MAF	Minor allele frequency
MBP	Myelin basic protein
MHC	Major histocompatibility complex
<i>MHC2TA</i>	MHC class II transactivator
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
NK	Natural killer cell

OR	Odds ratio
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
<i>PECAM1</i>	Platelet/endothelial cell adhesion molecule 1
PLOSL	Polycystic lipomembranous osteodysplasia with sclerosing leucoencephalopathy
<i>PRKCA</i>	Protein kinase C alpha
PP	Primary progressive
<i>PTPN22</i>	Protein tyrosine phosphatase, non-receptor type
RA	Rheumatoid arthritis
RNA	Ribonucleic acid
RR	Relapsing remitting
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SO	Southern Ostrobothnia
SP	Secondary progressive
T1D	Type 1 diabetes
TCC	Terminal complement complex
TDT	Transmission disequilibrium test
<i>TREM2</i>	Triggering receptor expressed on myeloid cells 2
<i>TYROBP</i>	TYRO protein tyrosine kinase binding protein

List of original publications

This thesis is based on the following original articles referred to in the text by their Roman numerals:

- I Saarela J, **Kallio SP**, Chen D, Montpetit A, Jokiaho A, Choi E, Asselta R, Bronnikov D, Lincoln MR, Sadovnick AD, Tienari PJ, Koivisto K, Palotie A, Ebers GC, Hudson TJ, Peltonen L (2006) *PRKCA* and multiple sclerosis: association in two independent populations. *PLoS Genet* 2(3):e42.
- II **Kallio SP**, Jakkula E, Purcell S, Suvela M, Koivisto K, Tienari PJ, Elovaara I, Pirttilä T, Reunanen M, Bronnikov D, Viander M, Meri S, Hillert J, Lundmark F, Harbo HF, Lorentzen ÅR, De Jager PL, Daly MJ, Hafler DA, Palotie A, Peltonen L and Saarela J (2009) Use of a genetic isolate to identify rare disease variants: *C7* on 5p associated with MS. *Hum Mol Genet* 18(9):1670-1683.
- III Leppä V, **Kallio SP**, Kempainen A, Koivisto K, Tienari PJ, Elovaara I, Pirttilä T, Reunanen M, Meri S, Palotie A, Peltonen L and Saarela J. Variation in complement cascade genes in MS patients and population controls. *Submitted*.
- IV Sulonen A-M*, **Kallio SP***, Ellonen P, Suvela M, Elovaara I, Koivisto K, Pirttilä T, Reunanen M, Tienari PJ, Palotie A, Peltonen L and Saarela J (2009) No evidence for shared etiology in two demyelinating disorders, MS and PLOSL. *J Neuroimmunol* 206(1-2):86-90.

* These authors contributed equally to the study.

These articles are reproduced with the kind permission of their copyright holders.

1 INTRODUCTION

Multiple sclerosis (MS), affecting approximately two million people worldwide, is the most common cause of neurological disability in young adults in the developed world (Oksenberg and Barcellos 2005). MS is a clinically heterogeneous demyelinating disease of the human central nervous system (CNS) with a putative autoimmune pathogenesis and complex inheritance. However, despite long-standing and intensive research, both genetic and environmental risk factors are still largely unknown. Thus, pathogenesis of MS is poorly understood and no curative treatment exists.

Genetics provides tools to dissect the molecular background of MS, of which the target tissue is challenging to study using conventional cell biological approaches. During this thesis study, the emphasis in genetic mapping of complex diseases has largely shifted from linkage studies in families to genome-wide association (GWA) studies in unrelated cases and controls. However, it also has become obvious that these large GWA studies primarily expose common variants contributing to disease pathogenesis with modest effects, and alternative strategies are needed to identify relatively rare, more penetrant alleles, which most probably give rise to a familial concentration of cases.

Previous linkage scan on large Finnish MS pedigrees detected three main MS loci on chromosomes 5p, 6p (HLA) and 17q, and linkage studies in several other populations have provided independent evidence for the location of MS susceptibility loci in these regions. In this thesis study current knowledge of human genetics as well as strengths of Finnish population history have been utilized to localize MS predisposing genetic variants of the linked loci outside the *HLA* region and, further, to better understand the molecular mechanisms of MS susceptibility.

2 REVIEW OF THE LITERATURE

2.1 Multiple sclerosis

2.1.1 Epidemiology and diagnosis

The onset of MS happens typically in early adulthood. Despite the fact that most patients die due to unrelated reasons, MS causes significant neurological disability and no curative treatment is available. Based on twin and population studies MS has a complex inheritance: both environmental factors and several predisposing genes are required to generate the disease. The sibling relative risk (λ_s) of 20-40, and higher concordance rate of ~20% for monozygotic twins compared to ~5% for dizygotic twins demonstrate a moderate genetic component for MS (Sadovnick 1993, Dymment et al. 2004). The role of childhood environmental factors on MS susceptibility is supported by migration studies: if migration occurs during childhood, the migrant acquires the new region's susceptibility to MS, whereas migration occurring later in life has little effect on MS susceptibility (Gale and Martyn 1995). Notably, dizygotic twins of MS patients seem to have a higher risk to get the disease than siblings of patients, even though there is no difference in genetic sharing (Ebers 2008) (Table 1). This may reflect the importance of the environmental factors of prenatal period and early childhood on MS susceptibility. On the other hand, adoption studies have provided convincing evidence that the aggregation of MS within families is largely explained by shared genes rather than a shared environment (Ebers et al. 1995) (Table 1).

TABLE 1. Population-based prevalence rates of MS in relatives of patients.

Genetic sharing	Relationship	Prevalence ¹
0%	General population (European origin)	1/1000
0%	Adoptive sibling	1/1000
12.5%	Cousin	7/1000
25%	Half sibling	20/1000
50%	Child	30/1000
50%	Parent	30/1000
50%	Full sibling	30/1000
50%	Dizygotic twin	40–50/1000
100%	Monozygotic twin	200–300/1000

¹ Estimates from Ebers 2008 and Dymment et al. 2004.

Similar to most diseases with a putative autoimmune etiology, MS affects twice as many females as it does males. The prevalence of MS is highest (1/1,000) in populations of Northern European descent living in temperate climate (Compston 1997), especially in the coastlines of Scandinavia, Iceland, the British Isles and in countries settled by their inhabitants. Thus, it has been proposed that the Vikings may have disseminated the risk alleles of MS, and the alleles may have entered Finland via them along the rivers in the Southwestern Finland (Tienari 2004). Notably, even though the prevalence of MS is quite uniform in Europe, both its incidence and prevalence are two times higher in Southern Ostrobothnia in Western Finland (Sumelahti et al. 2000, Sumelahti et al. 2001, Wikström and Palo 1975), most probably due to enrichment of risk alleles within this population subisolate (Figure 1). Further, there are also more familial MS cases in this isolated region than elsewhere in Europe (Wikström 1975b). MS is extremely rare in certain ethnic groups, including sub-Saharan Africans and Maori of New Zealand (Pugliatti et al. 2002). The uneven geographical distribution of MS is currently considered to be due to both regional variation in frequency of genetic risk factors and ubiquitously distributed environmental risk factors (Ebers 2008).

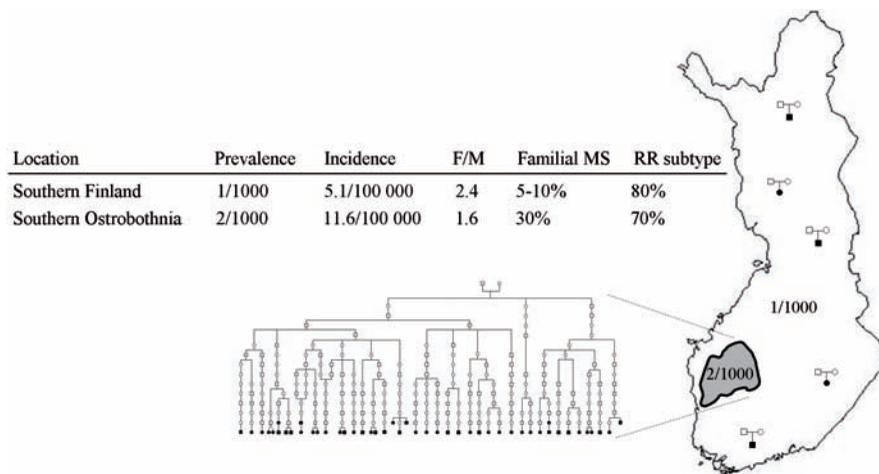


FIGURE 1. Special features of MS in distinct parts of Finland.

The prevalence of MS is two times higher in Southern Ostrobothnia than in other parts of Finland, and there are also more familial MS cases (large MS pedigrees) in this isolated region. The Southern Ostrobothnia subisolate region is marked with grey color. Values are obtained from Sumelahti et al. 2000, Sumelahti et al. 2001, Sumelahti et al. 2003, Wikström and Palo 1975. F/M = female to male ratio. RR = relapsing remitting MS.

Due to largely unknown pathogenesis, no specific diagnostic test for MS exists. The diagnosis is mainly based on symptoms and clinical findings, supported by laboratory and radiologic data. MS is a spectrum of various neurological

symptoms and findings, which are caused by inflammatory demyelinating lesions, axonal damage and inflammatory burst. Characteristic to MS is that lesions, which develop in the brain, the optical nerve and the spinal cord, affect different sites, separated in time. In consequence, the symptoms of the disease are protean. The most typical symptoms are unilateral painful loss of vision (optical neuritis), diplopia, problems with senses of feeling and spastic pareses of the limbs, ataxia and vertigo, tremor, dysarthria, bladder dysfunction, constipation, erectile impotence and fatigue (Compston and Coles 2002). Revised McDonald's diagnostic criteria (McDonald et al. 2001, Polman et al. 2005), representing an update for previously used Poser's diagnostic criteria (Poser et al. 1983), are currently used for MS diagnosis (Table 2).

TABLE 2. McDonald's diagnostic criteria for MS

	Clinical Presentation	Additional Evidence Needed for Diagnosis
RR	≥ 2 attacks and ≥ 2 objective clinical lesions	-
	1 attack and ≥ 2 objective clinical lesions	Dissemination in time (MRI)
	≥ 2 attacks and 1 objective clinical lesion	Dissemination in space (MRI / CSF+ and ≥ 2 MRI lesions)
	1 attack and 1 objective clinical lesion	Dissemination in time and space (see above)
PP	Insidious neurological progression suggestive of MS (≥ 1 year)	Two of the following: * brain MRI + * spinal cord MRI + * CSF +

RR = relapsing remitting; PP = primary progressive; MRI = magnetic resonance imaging; CSF = cerebrospinal fluid

Most MS patients (80-85%) experience a relapsing-remitting (RR) disease characterized by attacks followed by periods with no new signs of disease activity (Compston and Coles 2002, Inglese 2006) (Figure 2). However, the majority of patients with this clinical subtype will evolve into a secondary progressive (SP) MS within three decades, leading to chronic increases of symptoms and disability. Approximately 10-20% of MS patients experience a primary progressive (PP) form of the disease, characterized by chronic increase of symptoms and disability straight from the onset of the disease (Compston and Coles 2002, Inglese 2006) (Figure 2). A very rare clinical subtype is a malignant MS, leading to severe neurological symptoms or death within a short period after the disease onset.

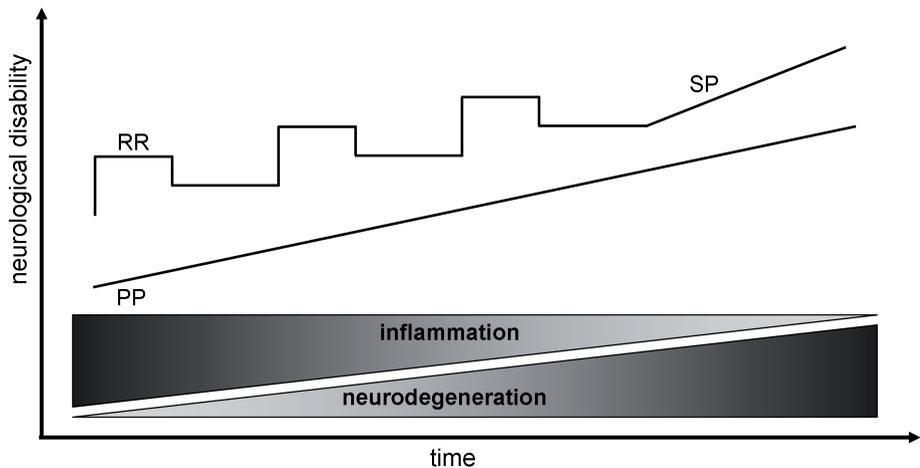


FIGURE 2. Clinical subtypes and progression of MS.

A relapsing-remitting (RR) MS is characterized by attacks followed by periods with no new signs of disease activity. The majority of patients with this clinical subtype will evolve into a secondary progressive (SP) MS, leading to chronic increase of disability. A primary progressive (PP) MS is characterized by chronic increase of symptoms and disability straight from the onset of the disease. Modified from Oksenberg and Barcellos 2005 by permission from Macmillan Publishers Ltd: [Genes and Immunity], copyright (2005).

MS subtypes are important for both prognosis and therapeutic decisions, since current MS therapies are immunomodulatory, partially protecting against relapses, but being ineffective against progressive symptoms (Inglese 2006). However, in a recent study it was shown, that the clinical course of MS is typically similar between affected siblings but not between affected parents and their children (Hensiek et al. 2007) (Figure 3). Thus, familial factors influence the clinical course of MS, but different clinical subtypes of MS can exist in one pedigree. No effect of family concordance for disease severity was observed (Figure 3). The authors concluded that different clinical subtypes of MS most probably represent a continuous spectrum of inflammation and neurodegeneration, balance of which is determined by both genetic and environmental risk factors; Individuals with a low threshold for neurodegeneration would more probably manifest disease progression, whereas individuals with a lower risk of this component but more inflammatory activity would tend to the relapsing remitting phenotype. Further, they suggest that the analysis of genetic studies should be stratified according to clinical course rather than disease severity (Hensiek et al. 2007). However, taking into account that different clinical subtypes of MS exist in one pedigree and most probably represent a continuous spectrum of inflammation and neurodegeneration, stratification of the study sample according to the clinical course in order to identify genetic risk factors for MS may be unnecessary and even lead to loss of statistical power.

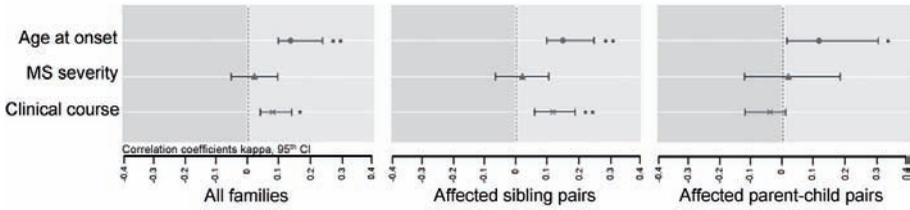


FIGURE 3. Intrafamilial correlation for various clinical features in multiplex MS families (n = 1,083). Affected parent-child pairs and affected sibling pairs also analysed separately. Modified from Hensiek et al. 2007.

*The clinical course of MS is typically similar between affected siblings but not between affected parents and their children. Thus, familial factors influence the clinical course of MS, but different clinical subtypes of MS can exist in one pedigree. No effect of family concordance for disease severity was observed. The authors suggested that the analysis of genetic studies should be stratified according to clinical course rather than disease severity. Concordance for age of onset was observed for all family members studied. *p<0.05, **p<0.001. Reproduced with a permission of the copyright holder.*

2.1.2 Pathological hallmarks and theories of pathogenesis

The pathological hallmarks of MS are inflammatory demyelinating lesions (plaques) within the CNS white matter. Pathogenesis leading to development of these lesions is currently unknown, but several theories exist.

According to data obtained from experimental autoimmune encephalomyelitis (EAE), an animal model of MS caused by immunizing rodents with myelin peptides, autoreactive peripherally activated CD4+ T lymphocytes enter the CNS by penetrating the blood-brain barrier (BBB), recognize myelin as foreign and attack it, and simultaneously activate the complement cascade, stimulate other immune cells like macrophages/microglia and B lymphocytes, and trigger expression of cytokines and antibodies (Gold et al. 2006). In the CNS, more epitopes of destructed myelin are represented to CD4+ T cells in the context of human leukocyte antigen (HLA) class II molecules. This process called epitope spreading further accelerates the inflammatory reaction and leads to a vicious cycle of myelin destruction and CNS inflammation. However, active MS lesions have been shown to be colonised also by CD8+ T lymphocytes, their pathogenetic relevance still being unclear (Babbe et al. 2000).

There is still a need for many explanations for tissue pathogenesis. The key questions are: why do the T lymphocytes get activated against myelin and how do they reach the CNS via BBB, which is normally unpermeable to lymphocytes. The most apparent explanation is that the trigger to both autoreactivity and decreased BBB integrity is an infection, which must be fairly ubiquitous within the

population and infect people before early adulthood. According to the molecular mimicry hypothesis, lymphocytes have a sensitization to myelin proteins because of homologous sequences found on antigenic viral proteins (Libbey et al. 2007). One of the candidate viruses is Epstein-Barr virus (EBV), a human herpes virus infecting 80-90% of the general population (Ascherio and Munch 2000, Ebers 2008). A systematic review of eight case-control studies comparing EBV seropositivity in MS cases and unaffected controls found increased odds of MS among EBV seropositive individuals (Ascherio and Munch 2000). Further, in a recent study, evidence for EBV infection in over half of brain-infiltrating B cells and plasma cells was observed in the post-mortem brain tissue of MS patients, potentially indicating an important role for EBV reactivation in MS pathogenesis (Serafini et al. 2007). However, it remains unclear whether homing of EBV infected B cells into the CNS is a primary event in MS or just a consequence of an unknown disease process.

Another hypothesis called the neural hypothesis suggests that a latent viral infection within the CNS leads to chronic infection of neurons, which in turn causes release of tissue antigens, increases permeability of the BBB and activates autoreactivity against myelin (Prat and Antel 2005). Among the candidate pathogens are herpes viruses and especially the human herpesvirus-6 (HHV-6), which is the causative agent in a common febrile rash (exanthema subitum) of children (Christensen 2007). Interestingly, the cellular receptor for HHV-6 is a complement regulatory protein MCP (Santoro et al. 1999). It was recently shown that ~10% of the CSF samples of MS patients are positive for HHV-6 and the total prevalence of human herpesviruses in the CSF of patients is around 15% compared to 2.3% for controls (Alvarez-Lafuente et al. 2008). However, it is worth noting that a reactivation of herpes viruses, including EBV and HHV-6, is typical for immune-mediated diseases and is possibly just a consequence of rather than a reason for MS.

Despite numerous attempts, no causal pathogen has been unequivocally linked to MS and the role of infectious diseases on MS predisposition still remains unproven. After all, even if a particular virus is not involved in MS pathogenesis, virus infections generally can predispose to MS by increasing production of cytokines, activating the complement system, altering the permeability of the BBB and boosting the autoreactive response against the CNS. On the other hand, MS patients may have defects in the BBB itself, facilitating penetration of lymphocytes into the CNS, and failure of regulatory T lymphocytes, which normally keep autoreactive T cells in control, might lead to proliferation and activation of autoreactive T lymphocytes and enhance the chronic CNS inflammation (Zozulya and Wiendl 2008).

Can something be learned about the MS etiology by studying the pathology of demyelinating lesions, the hallmarks of MS? Recently, four different immunopathological patterns of MS were characterized based on the composition of early active plaques in a large set of biopsies and necropsy samples (Luchinetti et

al. 2000). Based on this study, demyelination can be induced by T cells, macrophages and their toxic products (pattern I), by antibodies and the complement system (pattern II), by distal oligodendrogliopathy and apoptosis (pattern III) or by primary degeneration of oligodendrocytes in periplaque white matter (pattern IV) (Figure 4). Interestingly, these pathological patterns differ between patients but remain similar through the disease course, suggesting that MS is more heterogeneous disorder than expected and that most probably genetic heterogeneity also exists (Luchinnetti et al. 2000). The clinical significance of the pathological patterns was verified in a recent publication, in which pathological patterns of 19 MS patients treated with plasma exchange for an attack of fulminant demyelination were retrospectively studied (Keegan et al. 2005). All patients with pattern II (n=10), but none with other patterns, achieved functional neurological improvement after the plasma exchange, which depletes antibodies and complement. Notably, no clear correlation between the immunopathological patterns of demyelination and the classical clinical subtypes of MS (RR, PP) have been observed (Pittock et al. 2005). It is still unknown whether there is concordance in the immunopathological patterns of MS between members of the same pedigree.

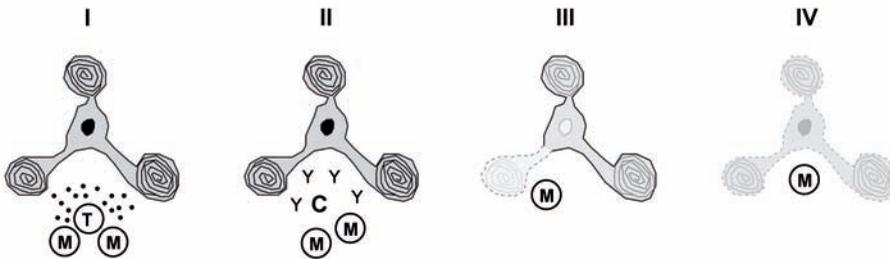


FIGURE 4. Pathological mechanisms behind the four distinct demyelination patterns of MS.

Four different immunopathological patterns of MS according to Luchinnetti et al. 2000. Demyelination is induced by T cells (T), macrophages (M) and their toxic products (pattern I), by antibodies (Y) and complement (C) (pattern II), by distal oligodendrogliopathy and apoptosis (dashed line) (pattern III) or by primary degeneration of oligodendrocytes in periplaque white matter (pattern IV).

Pattern II, characterized by immunoglobulins and complement activation, is the most common pathological subtype of MS, accounting for over 50% of patients (Luchinnetti et al. 2000). The complement system is a biochemical cascade of the innate immune system that helps to clear pathogens and cellular debris from an organism. Activation of C3 by the classical, lectin or alternative pathway leads to activation of the terminal components (C5-C9), which then form the TCC (C5b-9) (Figure 5). The TCC makes a transmembrane channel (membrane attack complex, MAC) in the cell membrane of the target cell, resulting in osmotic lysis of the target. Complement system has been shown to have an important role in pathogenesis of

many immunological diseases like hereditary angioedema, membranoproliferative glomerulonephritis, hemolytic uremic syndrome and SLE (Meri 2007), and based on several publications it may play an important role also in the etiology of MS (see Results and Discussion). Interestingly, oligodendrocytes, myelin forming cells of the CNS, are especially sensitive to complement mediated injury due to relative deficiency of regulatory proteins, which normally protect host cells from complement-mediated lysis (Scolding et al. 1998).

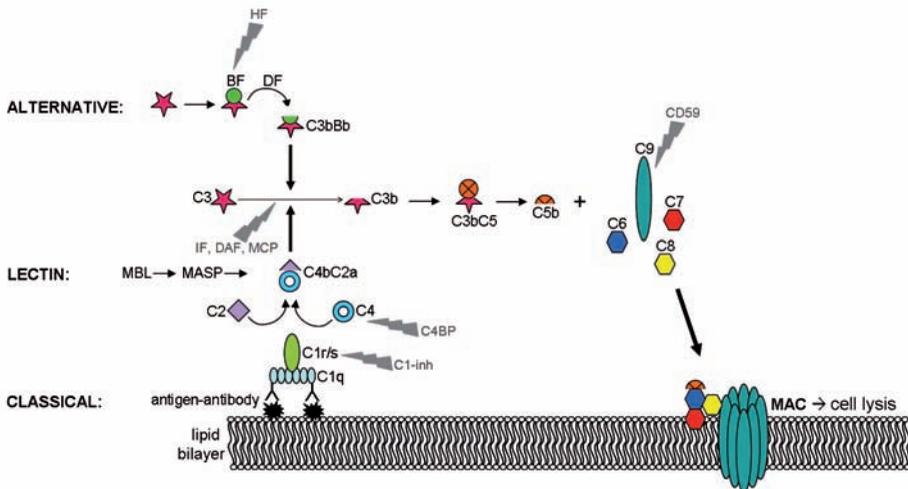


FIGURE 5. Complement cascade.

Activation of C3 by the classical, lectin or alternative pathway leads to activation of the terminal components (C5-C9), which then form a transmembrane channel (C5b-9 = membrane attack complex, MAC) in the cell membrane of the target cell, resulting in osmotic lysis of the target. Host cells are protected from complement-mediated lysis by regulatory proteins (marked as grey thunderbolts).

2.2 Genetics of complex diseases

2.2.1 The human genome

James Watson and Francis Crick discovered the structure of deoxyribonucleic acid (DNA) already 55 years ago (Watson and Crick 1953) (Figure 6), but the number of human genes and their exact order in the map of human chromosomes was not known until the beginning of this century, when the human genome was sequenced by the Human Genome Project (HGP) (Lander et al. 2001). The project revealed the 3.2 billion base pairs long human genome to contain about 20,000-25,000 protein coding genes (HGP website, http://www.ornl.gov/sci/techresources/Human_

Genome/home.shtm). This is much less than expected. Indeed, less than 2% of the human genome encodes proteins, while the rest of the sequence is composed of introns, promoters, other regulatory regions, non-translated RNA and so called “junk DNA”, for which the function is still largely unknown. It has become clear, that the human genome is much more than the sum of its genes.

2.2.2 Sequence variation

Humans are genetically approximately 99.5% the same (Levy et al. 2007). The rest 0.5% makes individuals genetically different and mainly explains why one individual is more susceptible to a certain hereditary disease than another. This genetic difference is caused by sequence variation (Table 3).

Variation in the human genome arises from mutations. Mutations occurring in the germline are further transmitted to offspring. The variants can be silent, modify protein products of genes or alter gene expression (Table 3). Single-nucleotide polymorphisms (SNPs) constitute the great majority of sequence variations: at least 6.6 million polymorphic SNPs are known to exist in the human genome (dbSNP build 129, validated SNPs, <http://www.ncbi.nlm.nih.gov/projects/SNP/>). Since our knowledge of the human genomic sequence is currently based on only a few sequenced individuals, the magnitude of genetic variation is still largely unknown. To identify majority of interindividual sequence variation and to better understand the role of this variation in human diseases, a large sequencing effort called The 1,000 Genomes Project was launched year 2008 (<http://www.1000genomes.org>).

TABLE 3. The main types of sequence variation in the human genome.

Variation	Description	Potential role in diseases
Single nucleotide polymorphism (SNP)	A single nucleotide of DNA sequence differs compared to the reference sequence.	Altered protein product or gene expression.
Microsatellite, minisatellite	Tandem arrays of repeat units.	Altered protein product or gene expression.
Structural variation, copy number variation (CNV)	Insertions, deletions, translocations, inversions, (segmental) duplications. Gains or losses of DNA segments compared to the reference sequence.	Altered protein product, gene dosage or gene expression. Rearrangement of the genome (non-allelic homologous recombination)

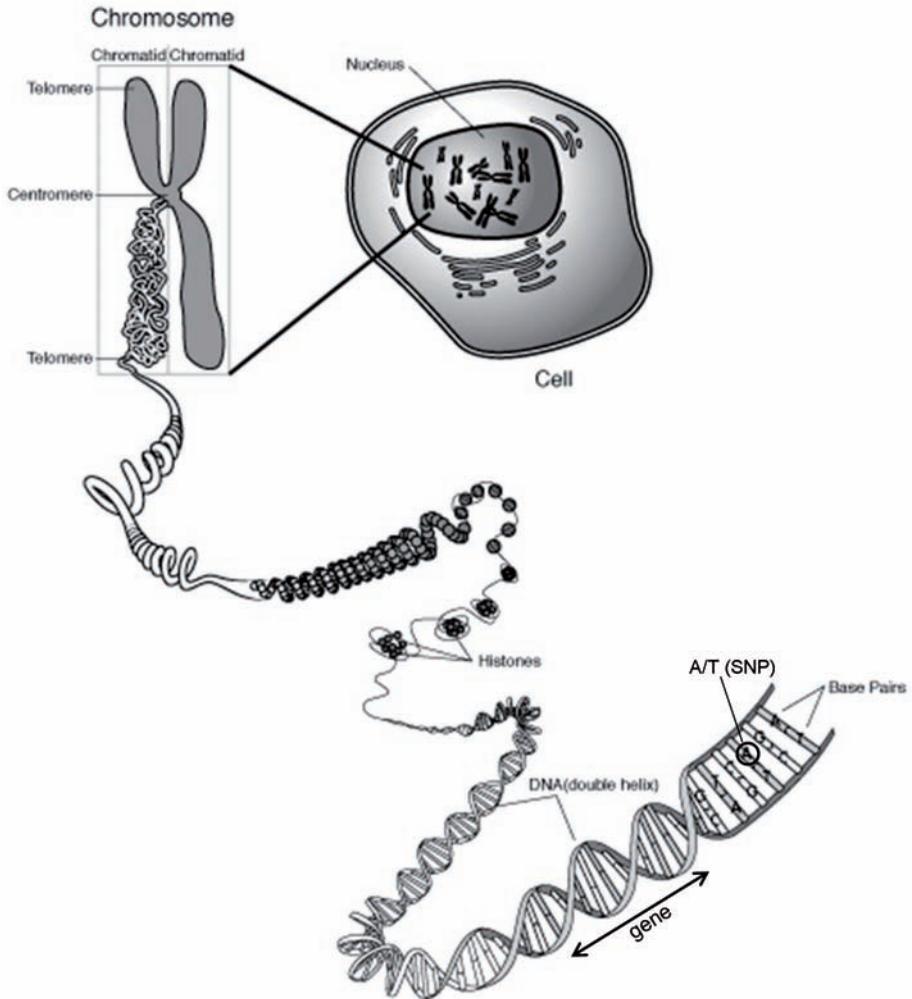


FIGURE 6. The Human Genome.

The 3.2 billion base pairs long human genome contains about 20,000-25,000 protein coding genes and numerous sequence variations, including SNPs, which appear roughly 1/1000 nucleotide sites. DNA is packed into 23 pairs of chromosomes preserved in the nucleus of a cell. Modified from National Human Genome Research Institute (<http://www.genome.gov>).

2.2.3 Strategies to identify genes underlying complex traits

To identify genetic variants predisposing to hereditary diseases, sequence variation is studied in affected cases, their family members and healthy controls. When the molecular mechanism leading to a disease is known, particular candidate genes can be selected based on their biological function. However, for many complex diseases, including MS, the etiology is largely unknown and hypothesis-free mapping methods might be more productive. One strategy to identify susceptibility genes of common diseases is the linkage approach, which aims to identify loci that co-segregate with the disease within multiplex families (multiple affected individuals in one pedigree) or affected sib pairs (Botstein and Risch 2003). After a genome-wide linkage scan, typically using a set of less than 1,000 polymorphic markers, has identified candidate loci, fine-mapping of these regions is usually needed to identify the causative variants. The subsequent association analysis aims to identify a marker allele, which co-segregates with the disease across individuals or families (Botstein and Risch 2003). Further, findings should usually be repeatable in independent study samples, and functional studies are needed to confirm biological significance of the finding. Notably, positional cloning, although being successful in gene identification of Mendelian diseases, have had substantially more limited success in complex diseases, which are influenced by multiple genes with modest effects, gene-gene interactions as well as allelic and locus heterogeneity (Hirschhorn and Daly 2005). However, a few success stories in complex polygenic disease genetics have already been reported, including the discovery of the *NOD2/CARD15* gene as an inflammatory bowel disease susceptibility gene. Linkage to chromosome 16 was identified already in 1996 (Hugot et al. 1996), and fine mapping of this locus finally led to the identification of *NOD2/CARD15* (Hugot et al. 2001), rare variants of which are associated with inflammatory bowel disease.

Recently, progress in high-throughput genotyping technologies and a better understanding of the human genome have enabled genome-wide association (GWA) scans, which have become popular and have largely replaced the conventional whole-genome linkage studies. In this new, more sensitive mapping method the genome is covered with 300,000-1,000,000 SNPs, each of which is then tested for association, typically in large study samples of several thousands or even tens of thousands of individuals (McCarthy et al. 2008). However, like other mapping methods, the GWA approach also has its problems. Perhaps most importantly, the huge number of tests performed unavoidably leads to false positive results, highlighting the importance of replication of the findings (Pearson and Manolio 2008).

Every variation in the genome still can not be studied before sequencing technologies get more cost-efficient. Instead, information produced by the HapMap project is utilized to “tag” the common variation of the human genome (The International HapMap Consortium 2003). The fundamental idea behind

tagging is that when a new mutation arises in the genomic sequence, where specific SNP allele already exists, the combination of SNP alleles, a haplotype, is further passed down to descendants as a unit unless a rare recombination event breaks the link. Due to this non-random association of alleles at closely linked chromosomal loci, linkage disequilibrium (LD), so called haplotype-tagging SNPs provide information not only about themselves but also about several other SNPs located nearby (Hirschhorn and Daly 2005). Hence, even when the actual causative variant is not tested, the genetic variants in proximity and in LD will be co-inherited more often in the same haplotype with the disease variant than expected under independent assortment and will show association with the trait. The amount of LD between two variants varies, to some extent, between different populations due to genetic drift, natural selection, mutations, recombinations, ancestral population demographics and mating patterns (Varilo and Peltonen 2004). Due to population history of Finland, the genome of Finns exhibits an increase in LD compared to mainland Europe and Africa, thus making gene mapping especially advantageous in Finland (Jakkula et al. 2008).

It is worth noting that the HapMap project focuses only on common variation (minor allele frequency is at least 1-5% on the population level). For any disease allele frequency, the power of an association study is greatest when the marker and disease allele frequencies match. Thus the GWA panels based on HapMap data are not optimized to detect genomic variants or rare genetic variants, which may also have a significant role on susceptibility of many common complex diseases (Hirschhorn and Daly 2005). According to the common disease-common variant (CD-CV) theory, common complex traits are mainly caused by genetic variants, which have a relatively high frequency and are found in all human populations (Reich and Lander 2001). However, these kinds of variants must have only small effects on the disease phenotype, since common variants with high penetrance would already have been detected in genome scans (McCarthy et al. 2008) (Figure 7). Common variants of complex diseases can be detected by studying large international study samples. For example a variant of the interleukin 7 receptor (*IL7R*) gene has a frequency of approximately 70% in the general population and it is further slightly enriched among MS patients from several populations (Lundmark et al. 2007, Gregory et al. 2007). An alternative opinion also exist, postulating that common diseases are caused by relatively rare mutations, each having a moderate or high effect on disease phenotype (Bodmer and Bonilla 2008). Further, there may be different mutations in different populations and the mutations may be even family specific. This seems to hold true, for example, for autism (Abrahams and Geschwind 2008). Notably, the HapMap tagging SNPs, detecting mainly common haplotypes, may not capture the relatively rare disease alleles very well, making mapping of these variants challenging. Only studies in special populations with unusual histories or exceptional pedigrees at high risk might provide sufficient number of cases to explore the association (Weiss and Terwilliger 2000) (Figure 7). It has also been suggested that, since environmental factors have so important role in

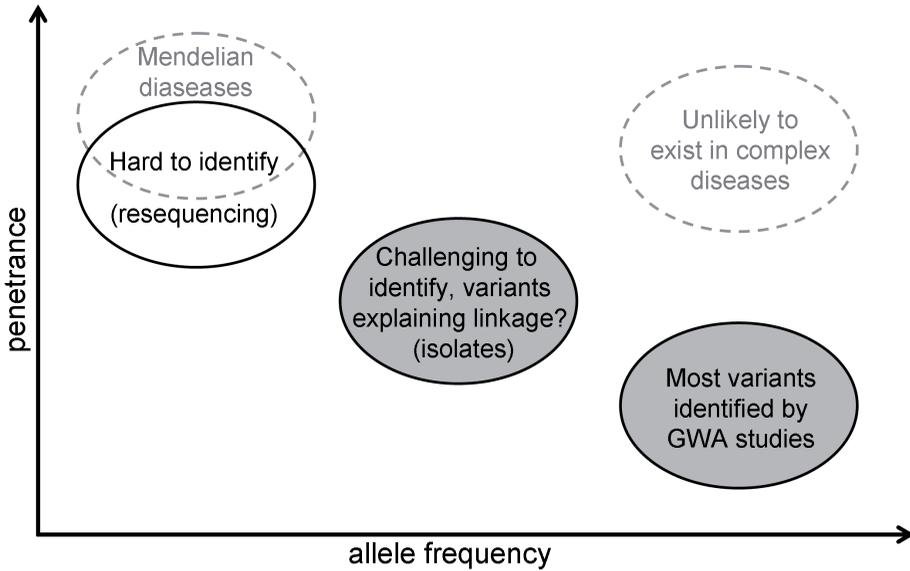


FIGURE 7. Identification of low and high frequency variants of complex diseases.

Current GWA studies are designed to identify common small effect variants of common diseases, and the rare variation has remained largely uncharacterized, mainly due to challenges in identifying such variants using the current methods. Modified from McCarthy et al. 2008. Adapted by permission from Macmillan Publishers Ltd: [Nature reviews genetics], copyright (2008).

the development of complex diseases, individual genetic variants must be rare and have a small effect on disease trait (Weiss and Terwilliger 2000), making detection of this kind of variants extremely challenging. For most complex diseases the truth probably lies between these extremities, and both rare and common variants, together with environmental factors, have a role in disease predisposition.

The current GWA studies aim to identify common variants of complex diseases, and the rare variation has remained largely uncharacterized, mainly due to challenges in identifying such variants with current platforms favouring common variants (Hirschhorn and Daly 2005, McCarthy et al. 2008, Frazer et al. 2009). Is there any point to try to map the rare variation when common variants more likely have an effect on disease phenotype globally? Firstly, the relatively rare variants with at least moderate penetrance most probably give rise to a familial concentration of cases and therefore explain the linkage identified in large pedigrees and provide opportunities for family specific diagnostic testing (Frazer et al. 2009). Secondly, common variants are common also among healthy population and typically increase the probability to get the disease only marginally. Thus, these variants do not necessarily much improve either diagnostics or understanding of the pathogenesis of the disease. For example, studies of several immunological diseases have revealed associations with interleukin genes (a catalog of published GWA studies at <http://>

www.genome.gov/gwastudies), as expected. To better understand the molecular mechanisms behind the diseases of unknown pathogenesis, relatively rare variants with higher impact on disease phenotype should also be looked for, since they provide information of defective metabolic pathways (Frazer et al. 2009).

2.2.4 The Finnish population and its subisolates

The human species originated in Africa around 150,000 years ago, of which the first waves of migration occurred approximately 100,000 years ago (Cavalli-Sforza 2007). After that several waves of migration have occurred. Finland was inhabited mainly from two immigration waves, occurring about 4,000 years ago from East and 2,000 years ago from South (Kittles et al. 1998). In the 16th century the internal migrations within Finland created regional subisolates, which were established typically by only a few founders (Figure 8A) (Varilo and Peltonen 2004). Since then, multiple bottlenecks like famines, wars and infectious diseases have temporarily reduced the population size, causing loss of genetic variation, and the subsequent rapid population expansion characterized by relative inbreeding and isolation has remarkably reduced allelic diversity (Varilo and Peltonen 2004) (Figure 8B).

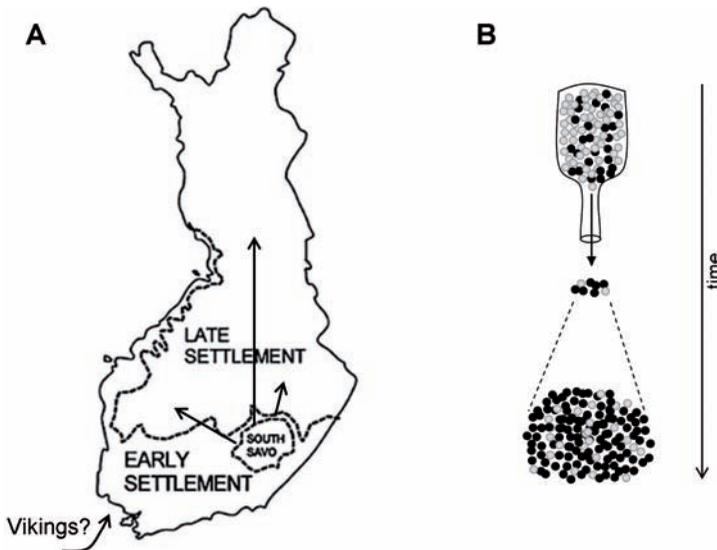


FIGURE 8. Characteristics of the population history of Finland.

A. First only the coastal region of Finland was inhabited (early settlement). An internal migration movement originated mainly from south Savo in the 16th century, resulting in genetically isolated subpopulations established by only few founders and isolated by distance (late settlement). The Vikings might also have disseminated their genome into the Southwestern Finland. The map is modified from Peltonen et al. 1999. Reproduced with a permission of the copyright holder. B. Multiple bottlenecks, temporarily reducing the population size and causing loss of genetic variation, and the subsequent rapid population expansion have remarkably reduced allelic diversity.

Due to this unusual population history of Finland, the Finnish genome, especially in the young subisolates, shows a decrease of genetic diversity and an increase in LD compared to other parts of Europe and especially to Africa (Varilo and Peltonen 2004). This has been a key to success in positional cloning of monogenic diseases in Finland and can further be beneficial in the identification of genetic variants of complex diseases, as the common HapMap markers of the GWA panels capture more variation through haplotype blocks in Finland than in more heterogeneous populations (Service 2006, Jakkula et al. 2008). On the other hand, high LD can complicate identification of actual causative variants. Hence, in some cases fine-mapping might be more meaningful in more outbred populations (Varilo and Peltonen 2004).

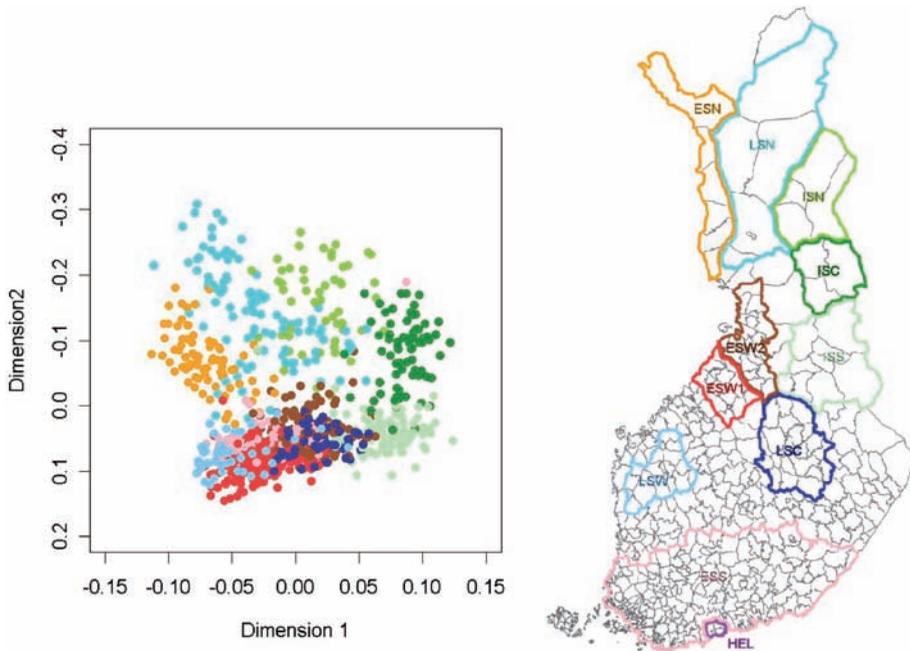


FIGURE 9. The population substructures within Finland.

Pairwise IBS sharing data of samples from ten distinct early- and late-settlement subpopulations is visualized with multidimensional scaling. The coloured dots indicate samples from the corresponding coloured areas of the map. Modified from Jakkula et al. 2008. Reproduced with a permission of the copyright holder.

The population history of Finland has led to uneven geographical distribution of disease alleles. Thus, the prevalence of several traits varies significantly between different subisolates, and typically birthplaces of the patients' grandparents represent regional clustering (Norio 2003). Similar population substructure can still be detected at a very high resolution by studying the "genetic fingerprint" of Finnish individuals (Jakkula et al. 2008, Salmela et al. 2008) (Figure 9). Sample

sizes of the population subisolates are often too small to detect common disease alleles with modest effects, but the subisolate populations can be especially valuable for identification of rare, high-impact variants of the isolate-enriched diseases. The majority of affecteds are identical by descent (IBD), meaning that chromosomes descending from a common ancestral chromosome carry the same disease allele in similar haplotypes, making the genetic background of complex diseases resemble that of monogenic disease (Varilo and Peltonen 2004) (Figure 10). However, use of isolates in gene mapping can expose to false positive associations due to population stratification (cases and controls originate from genetically distinct population subsets having distinct allele frequencies due to population history) (Hirschhorn and Daly 2005), and attention should be paid to selection of study sample, even when studying a seemingly homogeneous population like Finns (Jakkula et al. 2008). Another way of avoiding stratification is to use family-based study samples.

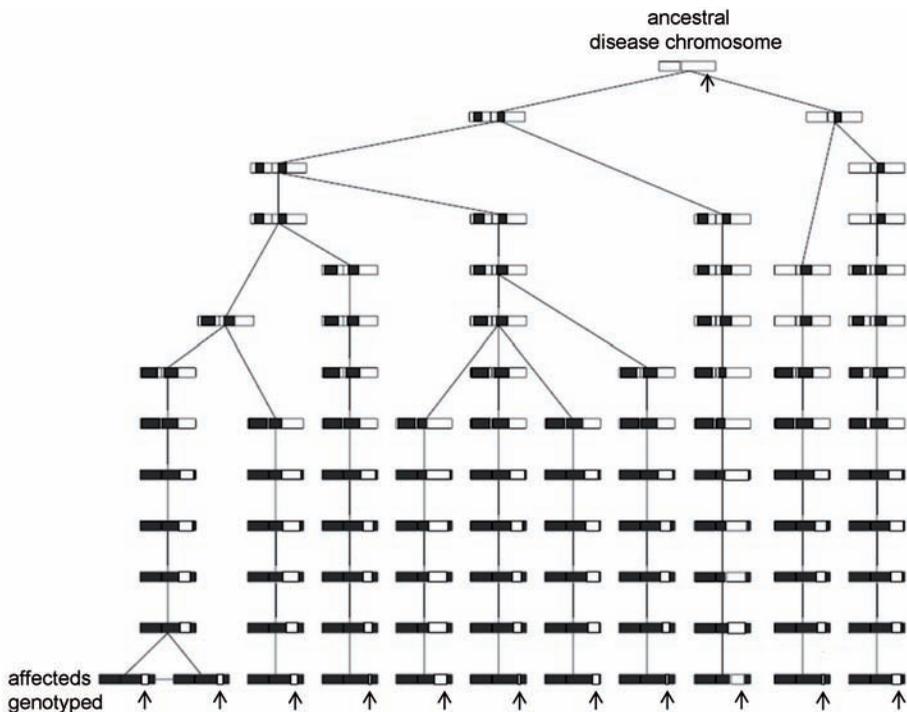


FIGURE 10. The use of LD and the founder effect in identification of the disease alleles.

Chromosomes descending from a common ancestral chromosome carry the same disease allele (arrow) in similar haplotypic background. However, recombinations have limited the shared haplotype (white). Modified from Varilo and Peltonen 2004. Reproduced with a permission of the copyright holder.

Incidence, prevalence and familial occurrence of MS are exceptionally high in Southern Ostrobothnia (SO) (Sumelahti et al. 2000, Sumelahti et al. 2001,

Wikström and Palo 1975), which is approximately 2,000 years old subisolate in Western Finland (Figure 1). Notably, SO is a relatively old subisolate, and the extent of LD and the length of homozygous segments are not as substantial as in younger subisolates of Finland (Jakkula et al. 2008).

2.3 Genetics of multiple sclerosis

2.3.1 Genome-wide linkage screens of MS

Linkage approach has been successful in gene identification of Mendelian diseases, but it has had substantially more limited success in genetic mapping of complex diseases (Hirschorn and Daly 2005). In MS, several genome-wide linkage scans, usually performed using sparse marker maps and small study samples, have resulted in identification of numerous potential disease loci. However, only a handful of them have been repeatable. The first genome-wide scans were published in 1996–1997 and studied affected sibling pairs and MS families from UK, US (Americans of European descent), Canada and Finland (Sawcer et al. 1996, Haines et al. 2006, Ebers et al. 1996, Kuokkanen et al. 1996 and 1997). Evidence for a shared MS locus was observed in all four studies for 6p21 (MHC region). Other potential regions of consensus were 2q24-33, 3q21-24, 5p14-tel, 5q13-23, 7q21-22, 10q21-22, 17q22-24, 18p11 and 19q13 (bolded in table 4).

To increase power to detect linkage, the data of the first four genome scans and five additional genome-wide screens was combined in a meta-analysis, resulting in 719 MS families from US, Australia, Canada, Finland, Italy, Scandinavia, Sardinia, Turkey and UK (GAMES 2003). The *HLA* locus provided strongest evidence for linkage and the loci on chromosomes 17q21 and 22q13 were the next strongest findings. Further, the International Multiple Sclerosis Genetics Consortium (IMSGC) performed a high-density linkage screen utilizing 4,500 SNPs in 730 multiplex MS families of Northern European descent (IMSGC 2005). Again, the peak logarithm of odds (LOD) score 11.7 was found in the *HLA* locus, and no other locus reached genome-wide significance. Promisingly, the second most significant LOD score was again detected on chromosome 17q.

The *HLA* gene cluster on chromosome 6p21 (Figure 11) is no doubt the strongest susceptibility locus for MS. As a matter of fact, immunologist discovered HLA-DR2 as a risk factor for MS already before molecular geneticists using serological methods (Jersild et al. 1973). HLA molecules are heterodimeric cell surface glycoproteins presenting antigens to T lymphocytes. Association with *HLA* locus and MS has been observed in most populations studied and with different clinical subtypes of MS (McDonnell 1999). The association signal primarily arises from the *HLA-DRB1*1501-DQB1*0602* haplotype (recognized by

TABLE 4. Suggestive linkage regions according to the first MS genome scans

Chromosome	Region
1	p35 ^C , p21 ^U
2	p23 ^A , p21 ^C , q24-33 ^{FC}
3	p25 ^C , p14 ^C , q21-24 ^{ACF} , q26 ^C
4	p16 ^C , q26-28 ^C , q31-35 ^A
5	p14-12(-tel) ^{FC} , q11-13 ^U , q13-23 ^{AC}
6	p21 ^{FACU} , q14 ^C
7	p21 ^C , p15 ^U , p14 ^C , q11 ^A , q21-22 ^{AC} , q32-35 ^A
9	p24-22 ^A , q34 ^A
10	q21-22 ^{AF} , q26 ^C
11	p15 ^A , q22 ^C
12	p13 ^U , q23-24 ^A ,
13	q33-34 ^A
14	q32 ^C
15	q21 ^C
16	p13 ^A , q12 ^C
17	q22-24 ^{FU}
18	p11 ^{FCA} , q21 ^C
19	q13 ^{FAC}
22	q12-13 ^U
X	p21 ^C , p11 ^C , p22 ^C , q26 ^C

U = UK (Sawcer et al. 1996); A = American (Haines et al. 2006); C = Canadian (Ebers et al. 1996); F = Finnish (Kuokkanen et al. 1996 and 1997).

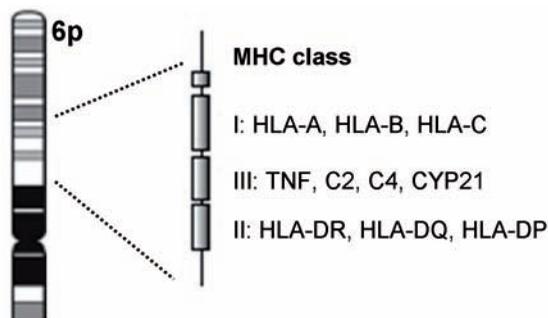


FIGURE 11. The MHC region on chromosome 6p21-23.

MCH class I and II code for HLA molecules involved in antigen presentation, whereas class III includes genes for complement proteins (C2 and C4), tumor necrosis factor (TNF) and cytochrome P450 (CYP21). Modified from Oksenberg and Barcellos 2005. Adapted by permission from Macmillan Publishers Ltd: [Genes and Immunity], copyright (2005).

the HLA-DR2 serotype) and especially from the *HLA-DRB1* gene located on the major histocompatibility complex (MHC) class II segment (Lincoln et al. 2005). However, identification of the actual MS predisposing variant within the *HLA* locus has been extremely complex due to extensive polymorphism and LD across the region. The allele frequency of the MS risk haplotype is approximately 35% in Finnish MS patients and 15% in unaffected population controls (unpublished result). Notably, this *HLA-DRB1*1501-DQB1*0602* haplotype has a relatively low frequency in Africa (frequency of 4% in African Americans, 6% in African American MS patients; Oksenberg et al. 2004), suggesting that positive selection of the haplotype has occurred in Europeans, potentially due to some infectious pathogen specific to Europe (Oksenberg and Barcellos 2005). The low frequency of the *HLA-DRB1*1501-DQB1*0602* haplotype may mostly explain the low MS prevalence in Africa. Even though African Americans are known to carry other MS risk haplotypes (for example *HLA-DRB1*0301-DQB1*0200*), the effect size of these haplotypes seems to be substantially smaller than that of the *HLA-DRB1*1501-DQB1*0602* haplotype in Caucasians (Oksenberg et al. 2004).

Roles for *HLA* class I loci in MS have also been suggested in several studies (Fogdell-Hahn et al. 2000, Harbo et al. 2004, Yeo et al. 2007, Brynedal et al. 2007). However, many of the reports suggesting association between *HLA* class I and MS are likely to be secondary to LD with class II loci. This was supported by the recent study, in which class I and II interactions were analysed in 1,260 individuals from almost 300 MS families (Chao et al. 2007). Overtransmission of the *HLA-A* and *HLA-B* alleles could be detected only in *HLA-DRB1*15* positive but not in *HLA-DRB1*15* negative MS families.

The *HLA* locus has been estimated to explain 14-50% of the genetic susceptibility of MS (Hafler et al. 2005), and the population attributable risk of MS has been estimated to be 48% for *HLA-DRB1*1501* positive individuals (Svejgaard 2008). Hence, a substantial fraction of the genetic component still remains to be explained. In addition to *HLA*, whole genome scans and other more targeted linkage approaches in several populations have provided evidence for the location of susceptibility genes on chromosomes 5p and 17q (Ebers et al. 1996, Oturai et al. 1999, Eraksoy et al. 2003, Sawcer et al. 1996, Larsen et al. 2000, Dyment et al. 2001, IMSCG 2005) (Figure 12). Similarly, the Finnish genome-wide linkage analysis in multicase MS families from Southern Otsrobothnia, together with other mapping methods, has revealed *HLA* and two wide regions on chromosomes 5p15-q11 and 17q22-q24 as the main MS susceptibility loci (Kuokkanen et al. 1996, Kuokkanen et al. 1997) (Figure 12). Interestingly, these regions are syntenic to the EAE susceptibility loci of rodents, supporting their role in predisposition to autoimmune demyelination (Butterfield et al. 1998, Sundvall et al. 1995, Jagodic et al. 2001). In addition, some evidence for linkage and association to the myelin basic protein (*MBP*) locus on 18q and a region on 19q has also been observed in Finnish MS families (Tienari et al. 1992, Reunanen et al. 2002).

Importantly, only loci with moderate to high effect on disease outcome can be detected using linkage approach. If the CD-CV hypothesis holds true, very many these kinds of predisposing variants do not exist for MS. Secondly, linkage analyses utilizing exceptional MS families as a study sample can detect only variants with at least moderate penetrance, and this kind of variants must be relatively rare (Figure 7). As mentioned above, the most effective mapping of relatively rare variants is carried out in populations with unusual histories, like Southern Ostrobothnia in Finland. By contrast, common, low effect variants of MS might be more meaningful to map using the sensitive GWA method and large study samples.

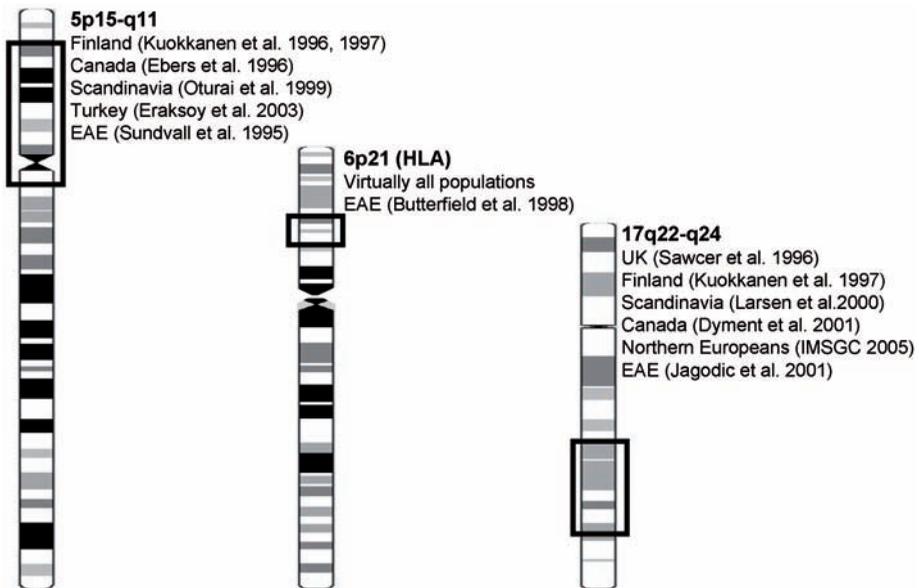


FIGURE 12. The main MS susceptibility loci in Finnish pedigrees.

Evidence for linkage to these regions has been observed also in other populations, and the loci are syntenic to EAE susceptibility loci of rodents.

2.3.2 MS candidate gene studies

Numerous candidate genes for MS (genes coding for proteins with meaningful biological function regarding MS) have been studied in several populations, but the findings have mainly been inconsistent between different sample sets. Notably, most of these studies have been performed using small or modest sized study samples and sparse marker maps, potentially leading to either false positive or false negative results. Variants outside the linked loci tend to have small effect sizes, and the smaller the effect is, the larger should the sample size be to detect association. Further, the pathogenesis of MS is largely unknown and thus, selection of particular

candidate genes based on their biological function can be misleading. Some of the numerous MS candidate studies are described below.

There is commonly an overlap in families with different autoimmune diseases like rheumatoid arthritis (RA), autoimmune thyroid disease (AITD) and type 1 diabetes (T1D) (Maier and Hafler 2008). Similarly, a number of genetic loci show association with several autoimmune diseases, suggesting that these immune-mediated diseases may, at least partially, share a common molecular background. Increased prevalence rates of various autoimmune diseases, including AITD and T1D, have been reported also among relatives of MS probands (Barcellos et al. 2006, Midgard et al. 1996, Broadley et al. 2000), and MS seems to occur more frequently, for example, in families with systemic lupus erythematosus (SLE) than in the general population (Corporaal et al. 2002). However, based on a recent study, no excess of common autoimmune diseases could be identified in MS patients or their families when the data was adjusted for sex (Ramagopalan 2007 A). Neither does MS explicitly fit into the “genetic cluster of autoimmunity”. For example, the cytotoxic T lymphocyte antigen 4 (*CTLA-4*) gene, involved in the regulation of T cell proliferation, show association with T1D and AITD (Kavvoura and Ioannidis 2005, Kavvoura et al. 2007), but discrepant results have been reported in MS. Nominal evidence for association has been observed in two Scandinavian studies (Harbo et al. 1999, Ligers et al. 1999) whereas no linkage or association to the region have been detected in several other studies (Dyment et al. 2002, Lorentzen et al. 2005, Bonetti et al. 2004, Greve et al. 2008, Wray et al. 2008). Similarly, no association between the protein tyrosine phosphatase (*PTPN22*) gene, a regulator of T cell receptor signalling associated with SLE, RA and T1D (Vang et al. 2007), and MS has been observed (Begovich et al. 2005). *IL2RA*, being strongly associated both with MS and T1D, as well as the *STAT3* gene, associated with MS, T1D and Crohn’s disease, are obviously exceptions (see 2.3.3 and 5.1.1). Moreover, some evidence for association with MS have been observed with the interferon regulatory factor 5 (*IRF5*) gene (Kristjansdottir et al. 2008), which is associated also with SLE, RA and inflammatory bowel disease (Dideberg et al. 2007, Sigurdsson et al. 2007, Demirci et al. 2007), and the MHC class II transactivator (*MHC2TA*) gene, which is associated also with RA (Swanberg et al. 2005). However, replication of these findings is still needed to validate the role of *IRF5* and *MHC2TA* in genetic susceptibility of MS.

Most of the MS candidate genes studied possess immunological functions. One of them is *IL7R*, which codes for a receptor of regulator of lymphopoiesis, IL7. A Swedish research group selected 66 candidate genes based on their immunological functions and/or location in linked regions, *IL7R* located on the linked region of chromosome 5p being one of them (Zhang et al. 2005). In a small Swedish sample set of 670 cases and 670 controls nominal evidence for association with *IL7R* was observed ($p=0.004$). The *IL7R* association was further validated in two simultaneous publications studying rather large study samples from Scandinavia, US (European

descent), UK and Belgium (Lundmark et al. 2007, Gregory et al. 2007). Moreover, the nonsynonymous SNP rs6897932 (T244I) in the alternatively spliced exon 6 of *IL7R* was reported to most likely be the causative variant. The MS-associated variation T244I is located in a transmembrane domain of the *IL7R* protein. The associated allele C of SNP rs6897932 is the major allele (allele frequency ~70%) and has only a modest effect on MS susceptibility (OR 1.2-1.3) (Lundmark et al. 2007, Gregory et al. 2007). Thus, it is very unlikely that the risk variant of *IL7R* alone explains the linkage observed to chromosome 5p in MS families. Some functional data for the MS-associated variant of *IL7R* already exists. The SNP rs6897932 has been shown to influence the amount of soluble (non-functional) and membrane-bound (functional) isoforms of the *IL7R* protein by putatively disrupting an exonic splicing silencer, individuals carrying the C allele having higher levels of circulating soluble receptor (Gregory et al. 2007). Further, levels of *IL7R* and *IL7* transcripts have been reported to be higher in the CSF of MS patients than in that of unaffected controls (Lundmark et al. 2007). On the other hand, this is not necessarily due to the genetic variant of *IL7R*, but can just be a consequence of the active inflammatory process of MS CNS.

Numerous other immunological candidate genes have also been studied as candidate genes for MS, including complement components 6 and 7, complement-like perforin and regulators of T lymphocytes, to mention but a few. In a small study by Chataway et al., only three SNPs of *C7* and one SNPs of *C6* were genotyped (Chataway 1999). Suggestive evidence for linkage and association with *C7* and MS was observed, but correction for multiple testing diluted the signals and the authors concluded that *C6* and *C7* do not confer susceptibility to MS. However, the study was underpowered.

Perforin is involved in CD8+ T cell and natural killer (NK) cell mediated cytotoxicity. After perforin and granzymes are released from these cells on a target cell upon its recognition, perforin forms MAC resembling pores on the target cell membrane allowing entry of apoptosis triggering granzymes into the target. Homozygous mutations of its gene *PRF1* cause a rare immune deficiency syndrome due to decreased capacity of the immune system to clear viral infections, and it has been suggested that some heterozygous variations may also favor development of several autoimmune diseases (Cappellano et al. 2008). Hence, *PRF1* has been studied also as a candidate gene for MS. Recently the entire coding region of *PRF1* was sequenced in 190 MS patients and 268 controls, and frequency of the exonic variations of the perforin gene was observed to be higher in patients than in controls (17% vs 9%) (Cappellano et al. 2008). The finding was replicated in a larger independent study sample. The authors suggested that the variations of *PRF1* may be important for MS development by altering perforin activity and thus by delaying virus clearance, potentially favoring development of molecular mimicry (Cappellano et al. 2008).

In a Norwegian study sample nominal evidence for association with the SH2 domain protein 2A (*2H2D2A*) gene and MS has been observed (Dai et al. 2001, Lorentzen et al. 2008). *2H2D2A* is a good candidate gene for MS since it encodes a T cell specific adaptor protein, which is important for normal differentiation and activation of T cells. However, replication of the association in other populations has been problematic (Lorentzen et al. 2008).

Myelin basic protein (MBP) is a key player in myelin maintenance and repair and is a potential target for immune-mediated demyelination, the *MBP* gene being thus one of the MS candidates. Further, the golli form of MBP has been shown to negatively regulate signal transduction in T lymphocytes (Feng et al. 2004). Evidence for linkage and association between MS and *MBP* has been observed in MS samples from Southern Ostrobothnia in Finland (Tienari 1992, Tienari 1998, Pihlaja et al. 2003). However, the findings mainly have not been confirmed in other populations studied (Pihlaja et al. 2003).

2.3.3 The first genome-wide association scan of MS

It has been suggested that association studies in complex diseases should involve at least 2,000 cases and 2,000 controls to achieve significance level where p-values $<5 \times 10^{-7}$ would more commonly be true positives than false positives (Wellcome Trust Case Control Consortium (WTCCC) 2007). Even larger, international study samples are needed to identify common MS variants with very small effects. On the other hand, allelic and locus heterogeneity most probably exists in clinically heterogeneous MS, and large study samples combining cases from several populations can be disadvantageous in identification of relatively rare susceptibility variants, which may be even population specific (Bodmer and Bonilla 2008).

Recently, the first MS GWA study was published, involving 931 trio families and 2,431 controls from UK and US (IMSGC 2007). A total of 70 SNPs were selected for validation based on association signal ($p < 0.0001$ for families (Figure 13), $p < 0.001$ for the case-control set). Further, SNPs showing only very modest association with MS ($p < 0.01$) but located in proximity to autoimmune loci were also selected for validation, resulting in 40 SNPs. Thus, the study design was actually partially candidate gene-based. Despite the fairly small study sample of the first stage of the study, validation of the findings was performed using a largish study sample of 2,322 MS cases, 5,418 unaffected controls and 1,540 trio families. Three non-*HLA* markers in two genes exceeded the p-value threshold suggested by the Wellcome Trust (WTCCC 2007): rs6897932 in the *IL7R* gene ($p = 2.94 \times 10^{-7}$) and rs12722489 and rs2104286 in the *IL2RA* gene ($p = 2.96 \times 10^{-8}$ and $p = 2.16 \times 10^{-7}$, respectively). However, it is worth noting that *IL7R* showed only trivial level of significance in the first phase of the study and was actually selected for the study based on its immunological function and the previous publications (Zhang et al.

2005, Lundmark et al. 2007, Gregory et al. 2007). Weak evidence for association in the replication sample was observed also with some of the SNPs selected for validation based on association observed in the first phase (SNPs located on the following genes: *RPL5*, *CD58*, *FAM69A*, *ANKRD15* and *CBLB*) (IMSGC 2007), but none of these SNPs exceeded the p-value threshold suggested by the Wellcome Trust (WTCCC 2007).

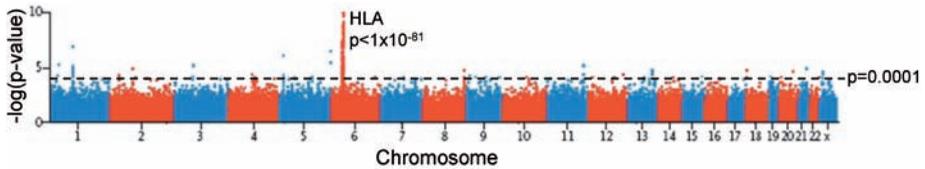


FIGURE 13. Association results of the first stage of the MS GWA scan.

P values (shown as $-\log_{10}$ values) for results of transmission disequilibrium testing (TDT) in 931 MS trio families are plotted across the genome. SNP with $p < 0.0001$ (dashed line) in this analysis were selected for validation. Modified from IMSGC 2007 with a permission of the copyright holder.

IL2RA was selected as a candidate gene for MS due to its association with another immune-mediated disease, type 1 diabetes (WTCCC 2007). The *IL2RA* gene, located on chromosome 10p15, codes for interleukin 2 receptor alpha chain. IL2R in turn mediates the action of the T-cell growth factor IL2. Originally, two SNPs (rs12722489, rs2104286), being in LD ($r^2=0.5$) with each other, showed association with MS (IMSGC 2007). Later, SNP rs12722489 was showed to provide the primary association (IMSGC 2008). Notably, even the modest LD between the SNPs rs12722489 and rs2104286 was enough to reveal the primary association of the common risk allele when the study sample was large enough. Like the risk allele of *IL7R*, the common susceptibility variant of *IL2RA* has only a modest effect on MS susceptibility, with an OR of ~ 1.2 (IMSGC 2007 and 2008). The risk variants of *IL7R* and *IL2RA* have further been genotyped in a large set of over 20,000 individuals from Australia, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Norway, Sardinia, Spain, Sweden and UK (IMSGC 2008) and in 600 multiplex MS families from Canada (Ramagopalan et al. 2007b) to refine understanding of the findings. Association was replicated in all but three populations. In the large combined study sample impressing p-values of even 1×10^{23} were obtained, even though the variants are common and have only modest effect on disease outcome. Associations could not be replicated in three populations, namely Australia and Ireland for *IL7R* and Holland for *IL2RA*. Notably, the small case-control sample from Holland was significantly underpowered (power to detect an OR of 1.2 was only $< 10\%$).

All the 17 SNPs showing even weak evidence for association in the first MS GWA scan were later genotyped in a large set of 1,134 MS cases and 1,265 controls

from Australia (Rubio et al. 2008). Again, no association could be detected with *IL7R* and MS, even though the statistical power was estimated to be almost 90%; the risk allele of *IL7R* was actually less frequent in Australian MS cases than in controls. In this study associations with *KIAA0350*, *IL2RA*, *RPL5* and *CD58* could be replicated.

At least the risk allele of *IL7R* seems to be independent of the *HLA* (Lundmark et al. 2007). Interestingly, on contrary to the prevalence of MS, the risk alleles of both *IL7R* and *IL2RA* are more common among non-white populations than in populations of European origin (IMSGC 2008). These variants explain very small proportion of the genetic risk of MS, which has been estimated to be only 0.2% (IMSGC 2007). However, it is still not clear whether the variants identified are the actual causative ones, and fine-mapping and functional studies are required to fully understand the role of *IL7R* and *IL2RA* in pathogenesis of MS.

Whereas this first GWA study of MS highlighted the power of collaboration in identification of genetic risk variants, it also made clear that identification of other MS risk genes with even smaller effects of disease outcome can only be revealed by studying much larger samples. Hence, the extended consortium, supported by the Wellcome Trust, has started even larger MS GWA scan (http://www.neurodiscovery.harvard.edu/research_initiatives/imsgc.html). In total this study will examine approximately 20,000 patients and 20,000 non-MS controls. Hopefully this large international collaborative project, of which massive data will be analysed during the next couple of years, will reveal novel MS genes and pathways and results in better understanding of MS pathogenesis.

2.3.4 Other strategies to map MS predisposing genetic variants

One method of association analysis is admixture mapping, which can be used when two populations have different prevalence of a disease and there exists a third population admixed of these first two populations (Zhu et al. 2008). Prevalence of MS is extremely low in Africans and much higher in European-Americans, whereas the prevalence in African-Americans is in between, suggesting that MS predisposing genetic variants most probably exist in the European genome but not in the African genome. In an admixture study for MS, genomes of African-American MS-patients were analysed to find regions that have an increased proportion of European ancestry due to potential risk alleles of European origin (Reich et al. 2005). Strongest evidence for association was found on the centromeric region of chromosome 1. Later, modest association with the *CD58* gene of the admixture locus was detected in the MS GWA scan (IMSGC 2007). The *CD58* gene encodes the CD58 antigen which, together with its counterreceptor CD2, optimizes immune

recognition and, on the other hand, promotes differentiation of regulatory T cells (Arthur et al. 2008), thus being a good candidate for MS.

Another way to try to dissect the molecular background of MS is to compare transcriptional profiles of MS patients and unaffected controls using microarray technology. In MS lesions, overexpression of inflammation-related genes and underexpression of myelin component coding genes have been observed, reflecting an important role for the immune system in MS and suggesting that ineffective remyelination may predispose to chronic demyelination and neuronal damage (Lock et al. 2002).

Finally, it is worth noting that the role of structural variation and epigenetics in MS still remains mainly uncharacterized. For example, rare and even family specific high penetrance CNVs (both de novo and inherited variation) seem to have an important role in genetic susceptibility to autism (Abrahams and Geschwind 2008). The role of this kind of more complex variation should also be dissected in more detail in future studies of MS.

§ AIMS OF THE STUDY

The aim of this thesis was to better understand the genetic architecture of MS, pathogenesis of which is largely unknown. The following specific aims were addressed:

1. To identify MS predisposing risk alleles within the two wide MS linkage loci on chromosomes 17q and 5p utilizing the strength of the Finnish population history (**I, II**).
2. To study relevant biological pathways for MS based on the findings (**III**).
3. To test if allelic variation of the *DAPI2* and *TREM2* genes, mutations of which cause a recessively inherited white matter disease PLOSL, would have an impact on another immune-mediated demyelinating disease, MS (**IV**).

4 MATERIALS AND METHODS

4.1 Study sample

4.1.1 Finnish MS sample

The Finnish MS samples have been collected from the hospital districts of Helsinki, Kuopio, Tampere, Oulu and Seinäjoki (Southern Ostrobothnia). The diagnosis of MS has strictly followed Poser's diagnostic criteria (Poser et al. 1983). All individuals have given their informed consent and the study has been approved by the Ethics Committee for Ophthalmology, Otorhinolaryngology, Neurology and Neurosurgery in the Hospital District of Helsinki and Uusimaa (Decision 46/2002, Dnro 192/E9/02). Unfortunately, only little clinical data for the MS patients was available during this thesis study.

The Finnish study sample is described in Figure 14. The family-based study material from Southern Ostrobothnia MS high-risk region consists of 22 Finnish multiplex MS families with two to six affected cases per pedigree, and ~140 MS patients with their parents and/or unaffected siblings. The multiplex families have previously been utilized in the Finnish genome-wide linkage scan (Kuokkanen et al. 1997). For case-control analyses, ~390 regional population controls have also been collected from Southern Ostrobothnia. In addition, ~730 unrelated MS cases and ~960 population controls have been collected from other parts of Finland.

At the time of the first publication (**I**), no population control samples were available and only MS families were studied. Case-control samples have since been utilized to increase statistical power (**II-IV**). However, to verify the findings and to avoid false positive association signals induced by population stratification, the families have been studied as well (**II**). The number of Southern Ostrobothnian MS families has slightly diminished during this thesis study due to revised genealogical data.

A small proportion of the Finnish genotyping sample was utilized for further studies. The genes of interest were sequenced in approximately ten Southern Ostrobothnian MS cases and population controls (**I, II**), and expression of the genes were tested in lymphocyte and mononuclear cell samples of ~10-20 cases and controls from Southern Ostrobothnia (**I, II**). Further, serum and plasma samples, as well as some clinical data, were collected for 20 Southern ostrobothnian MS cases (**II**). Finally, 174 Finnish MS cases and 172 population controls, of which half originated from Southern Ostrobothnia, were utilized in the CNV analysis (**IV**).

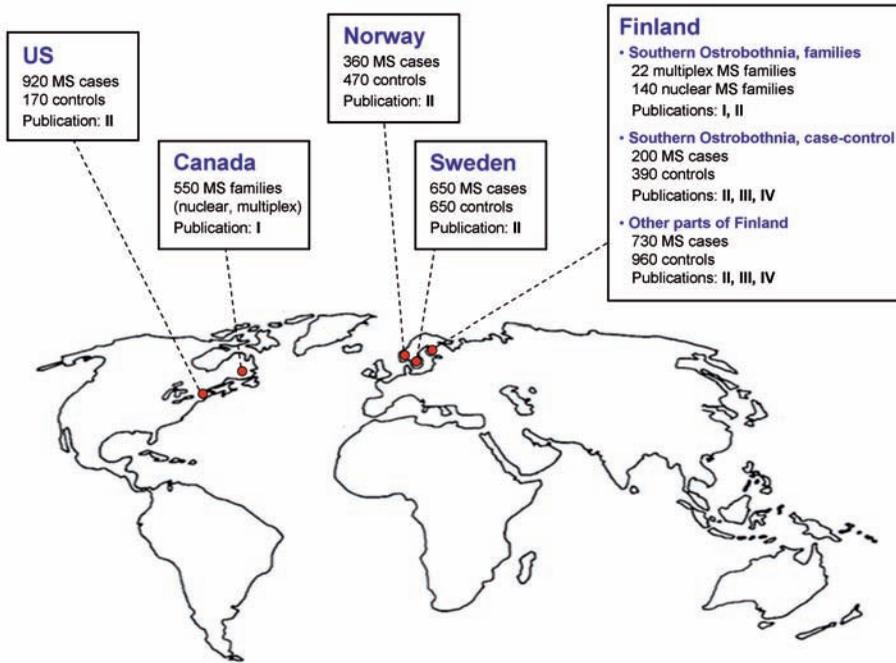


FIGURE 14. Study samples used in this thesis.

The rounded number of MS families, MS cases and population controls from different countries used in original publications (I-IV).

4.1.2 Selection of cases and controls for the Finnish GWA study

We utilized the data of the Finnish GWA study (Jakkula et al., manuscript in preparation) to screen the chromosome 5p linked region (II). Specifically, 72 MS cases, having either one parent born in the high-risk region of Southern Ostrobothnia and a family history of MS ($n=8$) or both parents born in Southern Ostrobothnia ($n=64$), have been genotyped for the GWA. Of these 72 MS cases, 41 belonged to either one ($n=14$) or both ($n=27$) of the two large interconnected mega-pedigrees, which we were able to construct via genealogical studies (Figure 15). However, no 1st degree relatives were included. 68 identity-by-state (IBS) matched population controls from Finnish genome-wide studies (from a total of 227 control individuals with GWA data) were used as the control set. We used genome-wide SNP data and IBS, identity-by-descent (IBD) and multidimensional scaling analyses to select these controls so that their genetic background would be similar with the cases, as parental birthplace information was not available for all the controls (Figure 16). Fourteen of the controls were known to have both parents born in Southern

Ostrobothnia (Figure 16, SOB ctrls) while 13 of the controls were known to live in Southern Ostrobothnia (Figure 16, SOB living ctrls) and 41 controls were part of the Health 2000 project (Figure 16, H2000 selected ctrls).

The genomic inflation factor (a comparison of unassociated genetic markers with those of control subjects for potential differences in allele frequency related to imperfect matching between case subjects and control subjects) was 1.0758 for our GWA data set, which suggests that cases and controls are well-matched (no difference over the majority of markers tested) and thus, there is no large-scale population stratification within our final study set. Importantly, no 1st degree relative pairs were found in any pair combination (case-case, case-control, control-control) according to the IBD sharing estimates.

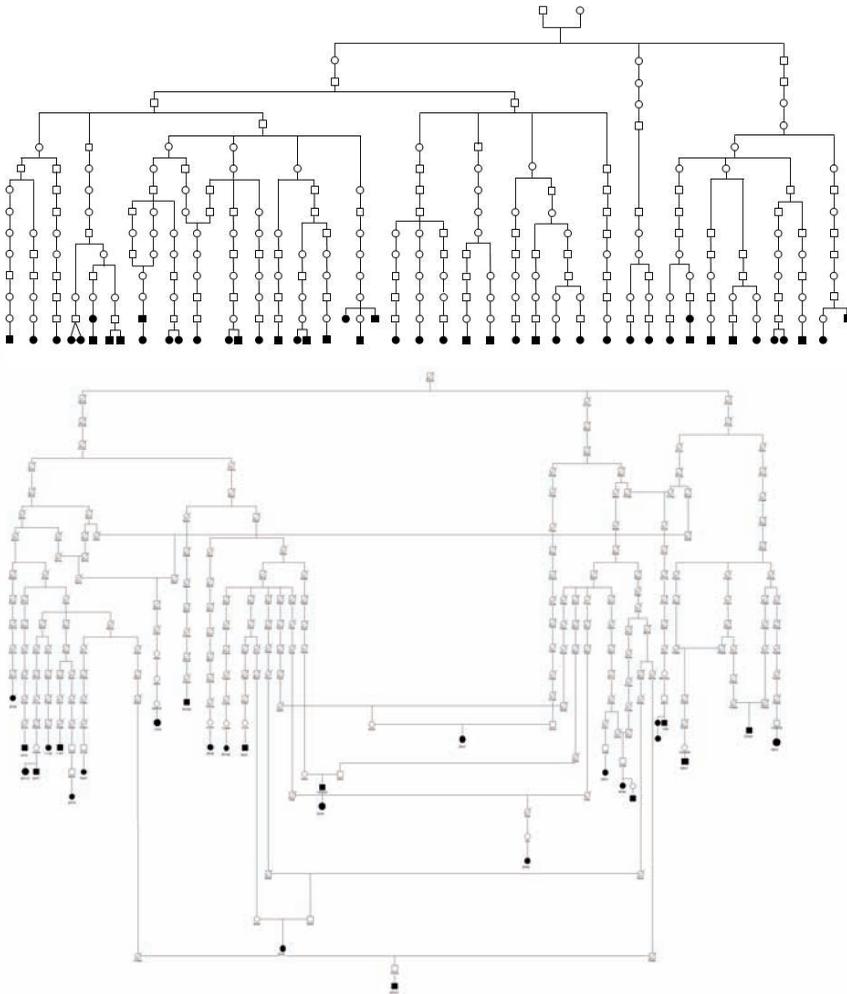


FIGURE 15. Finnish mega-pedigrees.

Majority of the MS cases of the Finnish GWA study belonged to either one or both of the two large interconnected mega-pedigrees constructed via genealogical studies.

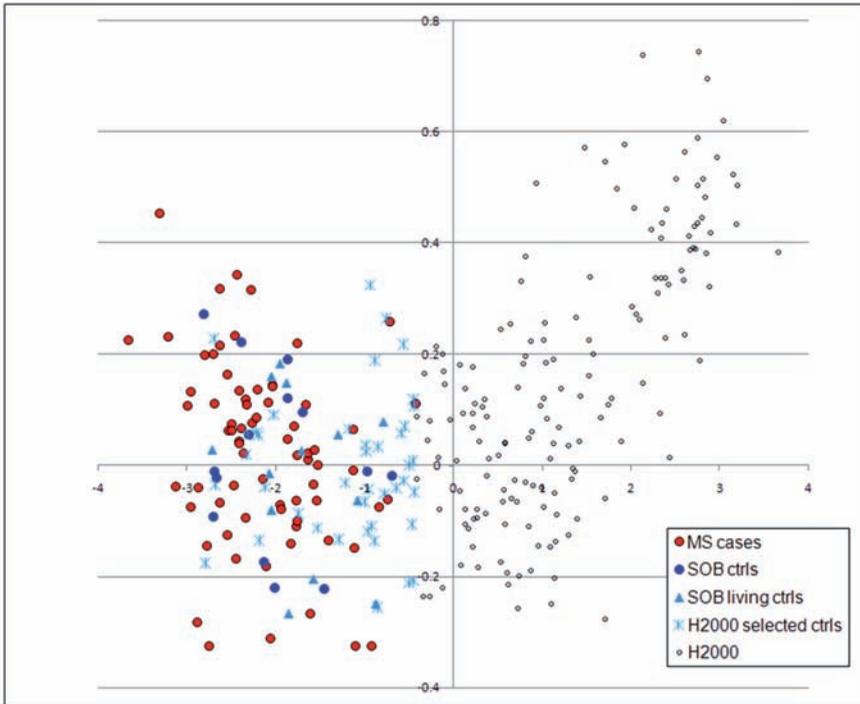


FIGURE 16. Selection of the controls for the Finnish GWA study.
Multidimensional scaling of pairwise IBS sharing data.

4.1.3 Study samples from more heterogeneous populations

In this thesis study a Finnish population subisolate has been utilized in fine-mapping. To study the role of the findings in more heterogeneous populations, MS families, cases and controls have been obtained from the collaborators of the Canadian Collaborative Project on the Genetic Susceptibility to MS (Figure 14, Canada), the Nordic MS Genetics Network (Figure 14, Sweden and Norway) and the Partners Multiple Sclerosis Center in Boston, Massachusetts (Figure 14, US). The Canadian cohort consisted simplex, extended, affected sibling pair and affected parent-child pair families. All the samples used were of Northern European descent, and the diagnosis of MS has strictly followed Poser's or McDonald's diagnostic criteria (Poser et al. 1983, Polman et al. 2005). All individuals have given their informed consent and the study has been approved by the ethics committees of the institutions involved.

4.2 Laboratory methods and statistical analyses

The methods used in this study are described in Table 5. Most of the SNPs have been genotyped using the Illumina 317K and the Illumina Golden gate assays and the Sequenom's MassArray system (Sequenom, San Diego, CA, USA), either with the hME or the iPLEX reaction. The Sequenom method utilizes chip arrays and matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Gabriel and Ziaugra 2004, Gabriel et al. 2009). Genotype counts between MS cases and unaffected controls were compared using Pearson's chi-square statistics (II-IV). The standard measure of effect in the case-control study is the odds ratio (OR), defined as the odds of exposure among cases divided by the odds of exposure among controls. Transmission disequilibrium test (TDT) (Spielman et al. 1993), haplotype-based haplotype relative risk (HHRR) (Terwilliger and Ott 1992) analysis as well as the Gamete Competition option of the Mendel 7.0.0 program (Lange et al. 2001) were used to monitor for association in MS families (I, II). In TDT analysis the transmission of alleles from heterozygous parents to affected children is compared to the expected 1:1 ratio. In HHRR, the two parental alleles, which have not been transmitted to the affected child, are combined to form the marker genotype of the "control individual", overcoming the stratification problem of the case-control design (Terwilliger and Ott 1992). TDT utilizes only genotype data of full trios. This is worth noting since only ~30% of the Finnish and Canadian MS families studied had both parents available. HHRR uses genotype data of cases also when data for both parents is not available. Further, HHRR is able to make use of families in which both parents are not heterozygotes for a given marker (Terwilliger and Ott 1992). The TDT-based Gamete Competition analysis utilizes the genotype data of the whole pedigree by treating transmission to normal children as complementary to transmission to affected children and is better adapted to missing data than the classical TDT test (Lange et al. 2001). Notably, if more than one affected child per family is used, all the family-based methods can confound linkage and association. Statistical significance of an association analysis is usually defined with p-value, which is the probability of obtaining by chance a result at least as extreme as that observed, even when the null hypothesis (no association) is true and no real difference exists. The smaller the p-value, the more strongly the test rejects the null hypothesis and the more unlikely the result is explained by chance alone. A p-value of 0.05 (corresponding to a 5% probability) or less is commonly used to reject the null hypothesis.

Illumina HumanHap300 SNP chip (Illumina, San Diego, CA, USA) was used to genotype samples of the Finnish GWA study. High quality SNPs ($n=3,981$) mapping to 11.1-56.0 Mb of chromosome 5 were used in the sliding window 5 SNP haplotype analyses (II), which were performed using the PLINK program (Purcell et al. 2007).

This study utilized two methods for multiple testing correction. In SNPSpD method (Nyholt 2004) the effective number of variants was estimated and the significance threshold was adjusted according to estimated number of independent SNPs (**I**). In permutation method (Chuchill and Doerge 1994) a false positive rate was obtained by generating several data sets by breaking the link between genotype and phenotype data (**II**).

TABLE 5. Methods used in this thesis study.

Laboratory method	Original publication
DNA and RNA extraction, quality controls	I-IV
Polymerase chain reaction (PCR)	I-IV
Agarose gel electrophoresis	I-IV
Genotyping	
Allele-specific primer extension on microarray	I
Pyrosequencing	I
Sequenom MassArray	I-IV
Illumina BeadArray	I
Illumina HumanHap300 SNP chip	II
ABI Prism DNA Sequencer	IV
Quantitative real time (RT) PCR using TaqMan	I, II, III
Sequencing using ABI Prism DNA Sequencer	I, II
Complement level quantification	
Radial immunodiffusion (Mancini technique)	II
Enzyme-linked immunosorbent assay (ELISA)	II
Statistical method or analysis program	Original publication
SNP and primer selection	
dbSNP database	I-IV
HapMap tagger	II, III
Primer3	I, II
Sequenom AssayDesigner	I-IV
Sequenom MassArray Typer	I-IV
Linkage and association analyses	
PLINK	II, III
Pedcheck	I, II
Chi-square test	I-IV
Merlin	I
MLINK	IV
TDT, Analyze package	I
HHRR, Analyze package	I
Mendel Gamete Competition	II
Haploview	I-IV
Phase	I, II
Genepop	I
GeneMapper	IV
Correction for multiple testing	
Permutation using Haploview	I, II
SNPSpD	I
ABI Prism 7900HT Sequence Detection System	I, II
Sequencher	I, II
Mann-Whitney U test	II
Copy number variation analysis using SeLoFit	III

5 RESULTS AND DISCUSSION

5.1 Fine-mapping of the linked loci on chromosomes 17 and 5 (I, II)

In addition to *HLA*, whole genome linkage scans in several populations, including Finns, have provided evidence for the location of MS susceptibility genes on chromosomes 5p and 17q (Figure 12). Further, these regions are syntenic to the EAE susceptibility loci of rodents, supporting their role on predisposition of autoimmune demyelination. An effort has previously been made to narrow these wide linked loci by haplotype analysis in Finnish MS families (Saarela et al. 2002, Riise-Stensland et al. 2005). The 17q ~20Mb linked locus has been restricted to a ~3.4Mb region (Saarela et al. 2002), which was further fine-mapped in this thesis study (Figure 17A). For chromosome 5p no gene in the restricted region was found to explain the linkage observed (unpublished), thus, in this thesis study the complete linked region was further scanned through to pinpoint a MS predisposing gene (Figure 18).

Notably, most probably MS is not a unitary disorder, but may represent an overlapping spectrum of related disorders (Lucchinetti 2000, Barcellos et al. 2002). To minimize the genetic and environmental heterogeneity, MS patients and families from the homogeneous Southern Ostrobothnian MS high-risk isolate have been utilized as a study sample in these fine-mapping efforts (Figure 14).

5.1.1 MS locus on chromosome 17q (I)

Fine-mapping of the chromosome 17q linked locus was started at the time when no extensive marker maps, HapMap or high-throughput genotyping techniques existed. In the first stage of the study, the previously restricted 3.4Mb region was flanked and covered with a sparse marker map of 67 unevenly distributed SNPs (Figure 17A), which were genotyped in a small set of 63 Finnish MS families. Nine of the SNPs studied showed some evidence for association ($p < 0.05$) in TDT or HHRR tests, four of them mapping to the protein kinase C alpha (*PRKCA*) gene (Figure 17A). The second stage of the study focused on *PRKCA* and the 1Mb flanking sequence. Over 200 SNPs were selected from the dbSNP database to somewhat evenly cover the target region (Figure 17A). These markers were genotyped in the whole set of Finnish nuclear MS families from Southern Ostrobothnia and in a large set of Canadian MS families. The strongest p-valuewise evidence for association was observed with SNP rs887797 ($p = 0.0001$), which is located in

intron 3 of the *PRKCA* gene (Figure 17A). This association observed seemed to be independent of the *HLA*. Although the same SNP of *PRKCA* failed to reveal association in Canadian families, two SNPs next to it provided suggestive evidence for association in this more heterogeneous study sample (Figure 17A). However, some evidence for association with MS was observed also with SNPs mapping to other genes within the critical region (Figure 17A), and not all of these SNPs were in high LD with *PRKCA*.

Further, all the exons, the promoter as well as parts of the intron 3 of the *PRKCA* gene were sequenced in ten MS patients and eight unaffected population controls, but no causative variant was identified. To dissect the allelic background in more detail, haplotypes over the critical region were constructed. A haplotype, flanking introns 3 and 8 of *PRKCA*, was observed to be over-represented in Finnish MS cases compared to their healthy family members (OR 1.34, 95th CI 1.07-1.68), whereas another haplotype of the same region showed association in Canadian MS pedigrees (OR 1.64, 95th CI 1.39-1.94). In a small sample of MS cases and their unaffected family members (n=20) a slight correlation, although not statistically significant, with *PRKCA* expression in CD4- blood mononuclear cells and the putative risk allele was observed, *PRKCA* expression being lower in individuals with two copies of the risk allele compared to carriers of only one copy of the allele (Figure 17B).

The study represented a large-scale fine-mapping effort of that time. When this fine-mapping study was started only few SNPs were known to exist in chromosome 17q according to public databases. No tagging SNP information was available even when the second stage of the study was started, and not all the common variation was captured. Further, at the time there was a debate in science community on how to construct and analyse haplotypes. Thus, the study really emphasizes the rapid progress in the field of human genetics.

The first stage of the study was conducted in a small set of MS families, of which only a fraction could be utilized in the TDT-based analyses. However, association with a SNP in the *PRKCA* gene was observed, and this association was validated in a larger set of Finnish MS families. Thus, SNPs within *PRKCA* seem to be in LD with some yet unidentified MS variant, at least in the Finnish population subsolate. At the time we published our results, modest evidence for association with *PRKCA* (p=0.001) was observed also in UK population (Barton et al. 2004). Specifically, in that study 35 SNPs of *PRKCA* were genotyped in a small sample of 184 MS cases and 340 controls from UK. The associated variants were located in the 5' and 3' ends of the large 0.5Mb *PRKCA* gene, and no evidence for association was detected with SNPs of the intron 3. Notably, the 5' end association was later on suggested to be a false positive finding; no evidence for association with the same variants was observed in 947 MS families from UK (Ban et al. 2005).

PRKCA (located at 17q24.2) encodes a protein kinase C type alpha, which is a calcium-activated, phospholipid-dependent serine- and threonine-specific

enzyme. When activated by diacylglycerol, PKC phosphorylates a range of cellular proteins. *PRKCA* is fairly ubiquitously expressed.

Interestingly, it has been shown that PKC alpha plays a critical role in signal transduction pathway via which the cytokine CCL2 induces permeability of the blood-brain barrier (BBB) (Stamatovic et al. 2006), making *PRKCA* a good functional candidate for MS predisposition. However, based on our data MS cases carrying two *PRKCA* risk alleles seem to express less PKC alpha compared to individuals with one copy of the risk allele. Obviously, this is in discordance with the hypothesis that BBB of MS patients is anomalously permeable, facilitating penetration of lymphocytes into the CNS.

Like potential causative variants of other linked regions, the putative MS variant of the 17q locus is most probably relatively rare (Figure 7). Thus, it is not totally surprising that no very strong evidence for association with any of the SNPs within the 17q locus was observed in the first international MS GWA, which has been optimized to detect common variation (IMSGC 2007). By contrast, such relatively rare variant most probably should have been revealed by the Finnish GWA study (Jakkula et al., manuscript in preparation). However, none of the variants within the restricted 3.4Mb region provided strong evidence for association in the Finnish scan. Allele frequency of the most promising SNP of the *PRKCA* gene (rs887797) remained to be different between the MS cases and the unaffected population controls studied (MS 0.80 versus controls 0.75, $p > 0.05$), but this variant unlikely explains the linkage observed to 17q.

The wide linked region of chromosome 17q has previously been restricted by selecting a region that was shared by all affecteds from each of the 20 multiplex families studied (Saarela et al. 2002). In fact, the affecteds shared the whole linked region in all but three families. Thus the MS variant may be located also outside the 3.4Mb region, which was shared by all affected individuals of each family. Hence, like in question of chromosome 5p linked locus, we later decided to scan through the complete region under the original wide 17q linkage peak utilizing the Finnish GWA data. Indeed, variants of the *STAT3* (signal transducer and activator of transcription) gene, located outside the restricted 3.4Mb region, provided strong evidence for association with MS ($p \sim 5 \times 10^{-5}$) (Jakkula et al., manuscript in preparation). The finding was further replicated in case-control samples from six populations of Northern European origin (a combined CMH analysis $p = 2.65 \times 10^{-10}$) (Jakkula et al., manuscript in preparation). SNPs within *STAT3* have provided nominal evidence for association also in the international MS GWA scan ($p = 0.002$ in the first phase, $p = 0.03$ in the validation phase), and these SNPs were observed to be in LD with the associated variants of the Finnish scan ($r^2 > 0.7$ according to the HapMap CEU data). Interestingly, *STAT3* has been reported to show association also with Crohn's disease, ulcerative colitis and T1D (WTCCC 2007, Barrett et al. 2008, Franke et al. 2008, Fung et al. 2008), the risk allele of MS being protective for Crohn's disease and ulcerative colitis. The protein encoded by *STAT3* acts as a

transcription activator when activated in response to certain cytokines and growth factors (including IL5, IL6, interferons, HGF, LIF and BMP2). Interestingly, in a recent study it was shown that mice with targeted deletion of *STAT3* in CD4⁺ T-cells do not develop EAE (Liu et al. 2008). The authors hypothesize that *STAT3* may have a critical role in shaping T-cell repertoire: activation of *STAT3* seems to be required for generation of Th17 lineage and restriction of the Th1 lineage. The role of *STAT3* in MS predisposition will certainly be tested in future studies.

Other biologically highly relevant candidate genes within the 17q linked region are *ICAM2* and *PECAM1* (both located at 17q23.3), which code for intracellular adhesion molecules involved in, for example, transendothelial migration of lymphocytes. The Illumina SNP chip used in the Finnish GWA scan includes only five SNPs in *PECAM1* and no SNPs in *ICAM2*. None of the studied SNPs provided evidence for association with MS in the Finnish GWA scan (Jakkula et al., manuscript in preparation). Neither have previous candidate gene based studies revealed evidence for association between *PECAM1* and MS (Nelissen et al. 2000, Nelissen et al. 2002, Sciacca et al. 2000), but notably, all of these studies analysed only one microsatellite polymorphism in a fairly small study sample. Thus, the role of these genes in MS predisposition can not be definitely excluded.

Interestingly, the restricted 3.4Mb region on 17q24 is flanked by palindromic segments and highly homologous duplicated sequences. These can predispose to large chromosomal rearrangements by nonallelic homologous recombination (Chen et al. 2004). Further, this region is inverted in the chimp and human with respect to the order in the mouse genome (Chen et al. 2004) (Figure 17A). Considering the complex structure of this chromosomal region, the SNPs showing association with MS could be in LD with a yet unidentified CNV or structural variant, affecting potentially even several genes within the critical region. Unfortunately, this could not be tested in Finnish MS, since the marker map of the GWA panel was relatively sparse at the duplicated regions, which could potentially predispose to the rearrangements.

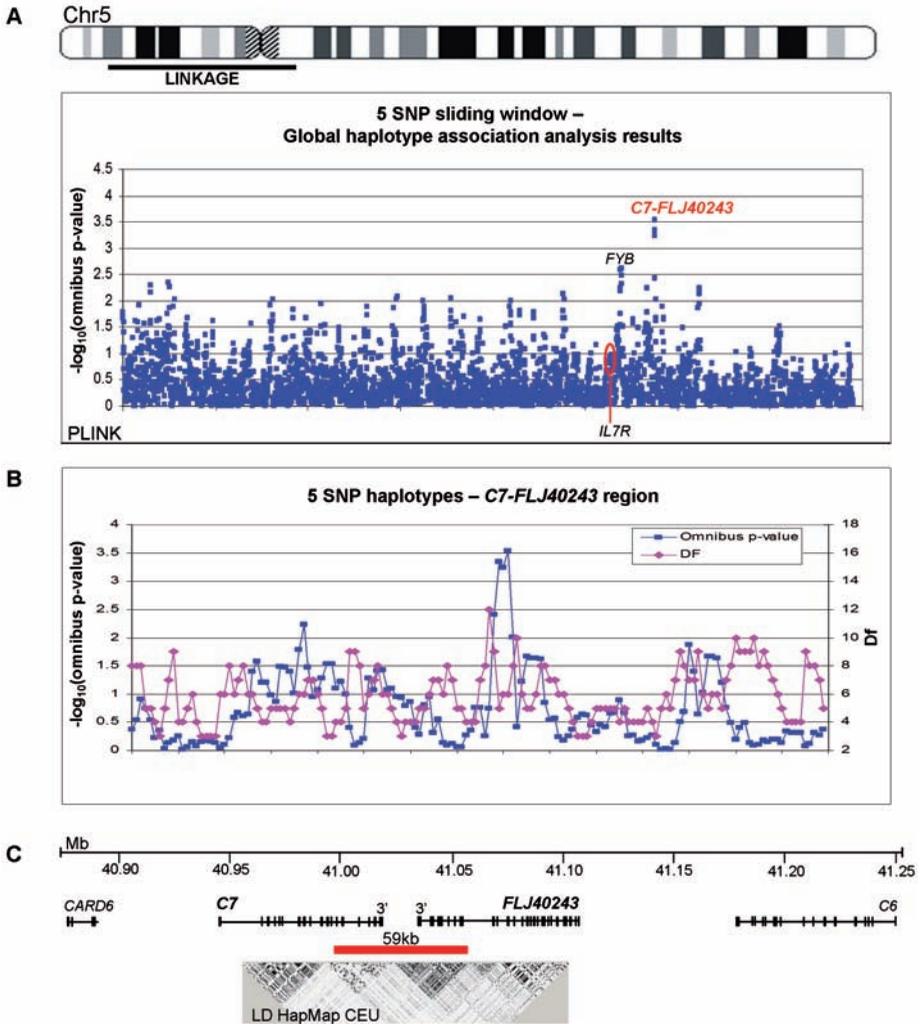


FIGURE 18. Fine-mapping of the MS locus on chromosome 5p.

A. The linked region was screened with a five SNP haplotype association analysis from the Illumina HumanHap 300 chip genotypes. The $-\log(p\text{-values})$ for the haplotypes are shown as blue dots. The analysis revealed one region with omnibus $p\text{-values}$ 10^{-4} (C7-FLJ40243). B. The haplotype showing strongest evidence for association was extended with PLINK to both orientations until reaching sites with increased number of degrees of freedom (DF) in recombination spots. C. The identified 59kb risk region, extending from 3' end of C7 to 3' end of FLJ40243, can be divided into a low LD region and a more clear block structure in heterogeneous populations according to the HapMap CEU data.

5.1.2 MS locus on chromosome 5p (II)

The first MS GWA scan (IMSGC 2007), together with other studies (Zhang et al. 2005, Lundmark et al. 2007, Gregory et al. 2007), revealed *IL7R* associated with increased risk of MS. Notably, *IL7R* is located in chromosome 5p MS linked region (Figure 12). However, the associated allele C of the likely causative variant rs6897932 of *IL7R* is very common also among healthy population, and it has been estimated to explain only a tiny fraction of the variance in the risk of development of MS. As speculated above, common variants with low penetrance, like *IL7R*, unlikely give rise to a familial concentration of cases. Thus, it is likely that other MS predisposing variant(s) also exists in the 5p linked region.

We first studied whether the variants of *IL7R* contribute to MS susceptibility in Finland. We genotyped SNP rs6897932 as well as three additional SNPs from previous association studies in the whole Finnish study sample of over 900 MS cases, of which 200 originated from Southern Ostrobothnia isolate, and 1,300 population controls (Figure 14). Only modest association was observed (rs6897932 $p=0.002$, OR 1.24, 95th CI 1.09-1.41), the odds ratio corresponding to that observed in other studies.

To study whether other loci on 5p could be identified as susceptibility loci for MS in Finland, we scanned through the complete ~45 Mb linked region utilizing the SNP data of the Finnish GWA study (Figure 18A). Specifically, the Finnish GWA has been aimed to enrich relatively rare, penetrant variants, which most probably give rise to a familial concentration of MS cases. The 72 MS patients studied had parents born in the Southern Ostrobothnia MS high-risk region, characterized with isolation and a founder effect, and majority of the patients were noticed to belong to either one or both of the two large interconnected mega-pedigrees, which we were able to construct via genealogical studies (Figure 15). Even though most of the MS patients were distantly related, none of them were first-degree relatives with each other. We hypothesized that the relatively short history, with common ancestors only 14-16 generations ago, might expose shared haplotypes between the distantly related MS cases. To avoid large-scale population stratification, we used 68 IBS-matched Finns as controls.

The haplotype analysis over the 5p linked locus revealed one region, located over 5 Mb centromeric from *IL7R* (Figure 18A). This haplotype, covering the 3' ends of the *C7* (complement component 7) and *FLJ40243* (hypothetical protein LOC133558) genes, had a frequency of 0.18 in MS cases compared to 0.04 in controls in the GWA sample ($p=0.0001$). Due to the small sample size of the GWA scan, we could not correct the result for multiple testing. Thus, to validate the finding, we genotyped the haplotype in an independent set of 125 Southern Ostrobothnian MS cases and over 350 population controls from the same geographical region and were able to replicate the association ($p=0.0004$). To estimate the effect size of the identified 59kb risk haplotype, the two study sets from the isolate were

combined. Frequency of the *C7-FLJ40243* risk allele was 0.12 among MS cases and 0.04 among population controls, resulting to a p-value of 3×10^{-6} ($p_{\text{perm}} = 5 \times 10^{-5}$) and a fairly high odds ratio of 2.73 (95th CI 1.67-4.47). Convincingly, also family-based association between the critical region and MS was observed ($p=0.006$), suggesting that the association observed is not just a false positive finding caused by population stratification. Due to a low frequency of the *C7-FLJ40243* risk allele and the relatively small sample size we could not test whether the association was independent of the *HLA* and the *PRKCA*.

We sequenced the coding regions of the *C7* and *FLJ40243* genes in 8 MS cases and 8 controls from the isolate. Four of the MS cases were known to carry two copies, four MS cases one copy, and eight controls no copies of the *C7-FLJ40243* risk haplotype. We identified altogether nine SNPs. Four nonsynonymous SNPs were located within *C7*, one of them being a novel variant. Two nonsynonymous and three synonymous SNPs were located within *FLJ40243*. However, none of the variants was in tight LD with the risk allele, thus, these SNPs are not likely candidates for the causative variant.

In addition, the *C7* promoter and most of the non-coding sequence covering the 3' end of *C7* were sequenced through. The sequenced region reached from the beginning of the intron 12 of *C7* to a recombination hotspot between the *C7* and *FLJ40243* genes. Several polymorphisms were found, but again, no definite causative variant was identified (unpublished result). As evolution conserves function, we carefully looked at the conserved motives within the sequenced region. One such sequence motive was noticed to be located in the middle of the *C7* risk haplotype in intron 14 (~41,003,460–41,003,640 Mb according to UCSC, hg18 assembly, Mar2006). The function of this motive is unknown. Interestingly, the same conserved sequence is found in several locations of the human genome. However, there was no variation in the sequence of the conserved motive between the individuals studied.

There is a potential micro-RNA binding site within the 3' end of *C7*, allele A of the SNP rs1061429 (A/C) enabling binding of miR-591. However, the major allele C was present in the identified risk haplotype as well as in most of the other haplotypes. Instead, a SNP rs3805226 could be a good candidate since it is in complete LD with the risk haplotype of the isolate and is located in a conserved element within intron 15 of *C7*. According to the transcription binding site predicting softwares SNP rs3805226 could potentially alter binding of a brain specific transcription factor Brn2. However, to verify the functionality of the SNP, changes in DNA binding capability should be tested for example using the electrophoretic mobility shift assay (EMSA). Interestingly, of the vertebrates with conservation information available for this SNP rs3805226 in the USCS Genome Browser (<http://genome.ucsc.edu/>), almost all carry an allele A (Opossum, Elephant, Armadillo, Dog, Cow, Horse, Rabbit, Mouse, Rhesus, Macaque, Chimp), except Tenrec (C), whereas the genomes of Cat and Rat contain the rare allele G. However, eventhough the isolate

enriched haplotype contains the rare allele G, it is not present in risk haplotypes of the more heterogeneous populations. Thus, this SNP is not a likely candidate for the causative variant, at least globally. Sequencing of the rest of the non-coding regions and the intergenic sequence is still warranted.

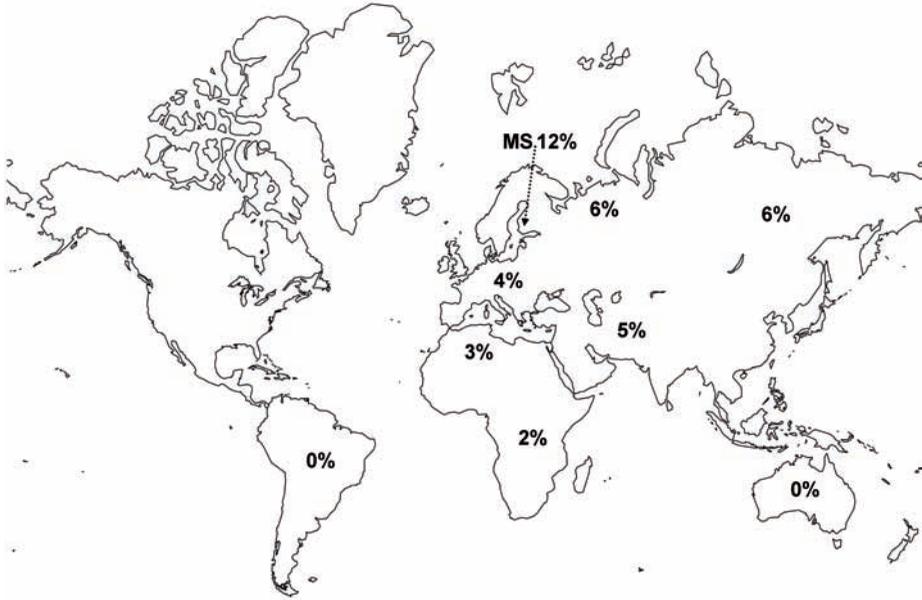


FIGURE 19. Frequency of the identified C7-FLJ40243 risk haplotype in various populations according to the Human Diversity Panel data.

The haplotype was observed to be relatively rare globally, having the highest frequency in Eastern populations. It has further been enriched in MS pedigrees of Southern Ostrobothnia, having a frequency of 12% among the MS cases of this population subisolate.

The identified C7-FLJ40243 risk haplotype seems to have a fairly large effect on genetic susceptibility of MS, at least in the Finnish MS isolate where it has got enriched due to the founder effect and isolation. We monitored the frequency of this haplotype in various populations utilizing the Human Diversity Panel SNP data (<http://www.cephb.fr/en/hgdp/diversity.php/>). Interestingly, the haplotype was observed to be relatively rare globally, found at the ~4% of alleles in the general European population, being almost absent in the Africans, Southern Americans and Oceanians, and having the highest frequency of ~6% in Eastern populations (Figure 19).

The advantages of the Finnish population isolates are that most of the affected individuals typically share the same major risk allele and that the relatively rare variants can be exposed by the common HapMap markers due to the wide LD intervals. However, the relatively rare variants identified using an isolated population are much more challenging to detect in more heterogeneous populations with

distinct LD patterns. As expected, the LD pattern of the 59kb risk region identified here was noticed to differ between the isolate and the general European population: based on the GWA data, the haplotype seems to cover only one haplotype block in the isolate whereas the critical region can be divided into a low LD region and a more clear block structure according to the HapMap CEU data (Figure 18C).

We studied the identified *C7-FLJ40243* risk region also in case-control samples from more heterogeneous populations, namely Finland outside the MS high-risk region, Sweden, Norway and US (Figure 14). The risk allele of the isolate had a comparable frequency in control samples of the isolate (0.04) and the other populations of Northern European origin (0.03-0.06) but it was not over-represented among the MS cases from these more heterogeneous populations. However, the Finnish MS cases were observed to carry another risk allele (freq MS 0.20, controls 0.16, $p=0.003$), which was also marginally over-represented in the MS cases from Sweden and US (Sweden: MS 0.18, Ctrl 0.15; US: MS 0.17, Ctrl 0.14), and a third allele was over-represented among the Norwegian cases (MS 0.05, Ctrl 0.03).

Use of population isolates is a double-edged sword. Our strategy was to study familial MS cases having both parents born in Southern Ostrobothnia, exposing the most extreme load of similar genetic background. As expected, the LD pattern of the identified risk region differed between the isolate and other populations, thus, it was not totally surprising that unequivocal association could not be detected. There seems to be more allelic heterogeneity within the region in more heterogeneous populations, and the potential risk variant may be carried in diverse allelic backgrounds. Only hard functional evidence or identification of the actual causative variant may enable uniform replication of the finding. Moreover, it is also possible that, like in many monogeneous diseases, there are various causative mutations within the same gene in different populations and in that case replication of the original allelic association in a more heterogeneous population might not be even feasible.

Notably, it is likely that MS is a more heterogeneous disorder than expected, and most probably genetic heterogeneity also exists. The MS cases from the Southern Ostrobothnia isolate presumably carry the same predisposing genes and have been exposed to same environmental factors, thus they are likely to experience the same immunopathological pattern of MS. Taking into account our results, it is tempting to speculate that majority of the MS cases of the Finnish MS subisolate experience the most common pathological subtype of MS, pattern II, characterized by immunoglobulins and complement activation and that the *C7-FLJ40243* risk allele identified predisposes especially to this type of disease. Hence, sorting the MS cases of the more heterogeneous populations according to the immunopathological pattern type could potentially reduce the genetic heterogeneity and unmask the association. Further, taking into account the frequencies of the identified MS risk haplotype in different populations, it possibly has drifted into Europe from East, and it could be worth the effort to study the *C7-FLJ40243* in Asian MS.

The *C7* gene is an excellent candidate for MS and there already exists strong evidence for involvement of the complement system in MS and EAE pathology (see discussion of study III). The seventh component of the complement system is a component of the terminal complement complex (TCC, C5b-9) which, when assembled on a cell membrane, forms the cytolytic MAC complex. Interestingly, *C7* is a critical limiting factor of complement activation: only when the local expression of *C7* is sufficient, *C7* binds to preformed C5b6 and the resulting C5b-C7 complex is able to insert into the phospholipid membrane to start the formation of the MAC (Thompson and Lachmann 1970). *C7* has been reported to be synthesized at least by endothelial cells, polymorphonuclear cells, macrophages, platelets, fibroblasts, synovial tissue and even in the CNS by astrocytes and oligodendrocytes (Gasque et al. 1995, Hogasen et al. 1995, Langeeggen et al. 2000, Hosokawa et al. 2003, Morgan and Gasque 1997). However, we observed no expression of *C7* in peripheral blood mononuclear cells (PBMCs), of which we had RNA available.

To study whether the identified haplotype has an effect on *C7* protein levels, we collected plasma samples of 20 MS patients and 32 unaffected controls. Eleven of the cases and thirteen of the controls were known to carry the MS risk allele. Eventhough most of the *C7* levels were within the reference range, a correlation between the protein level and the risk allele carriership was observed, carriers of the risk allele having slightly more circulating *C7* protein in plasma compared to non-carriers. Convincingly, this correlation was seen both in MS cases and in unaffected controls (Figure 20A).

We hypothesized that the complement system of MS cases could overall be more active than that of unaffected controls and that the observed increase in *C7* levels could potentially affect the three complement activation pathways: classical (CP), alternative (AP) and lectin (MBL). Therefore, we studied the total complement activity in same individuals by measuring the number of TCC formed as a consequence of activation of each pathway. As expected, the complement system was significantly more active in MS cases than in controls (Figure 20B). Importantly, the cascade was most active in MS cases carrying the identified *C7FLJ40243* risk allele, suggesting that this allele further boosts the complement system when it gets activated, which happens for example in chronic inflammatory diseases, including MS.

In the future, it would be interesting to study protein levels of the three TCC components located on 5p, namely *C7*, *C6* and *C9*, also in CSF samples of MS cases and unaffected controls. It would also be interesting to test whether the myelin forming oligodendrocytes are injured more easily when being in contact with the serum of the carriers of the identified *C7-FLJ40243* risk allele. The hypothesis would be that the carriers of this risk allele have more active complement system compared to noncarriers and that excess of functional TCCs following the complement activation would lead to more efficient destruction of the oligodendrocytes, which are fairly defenceless against the complement mediated lysis (Scolding et al. 1998).

Notably, the potential causative variant of the *C7-FLJ40243* region is still unidentified. Both *C7* and *FLJ40243* are present at the corresponding location of the mouse genome 15 (Mouse Genome Informatics database). The function of the other gene of the risk haplotype region, *FLJ40243*, is still unresolved but it is known to encode for a protein. Based on the human GNF Expression Atlas 2 Data (<http://genome.ucsc.edu/>) *FLJ40243* is expressed at extremely low levels fairly ubiquitously. To test whether *FLJ40243* also is a good biological candidate for MS, we monitored its expression in several human tissues. In concordance with the GNF Expression Atlas data, *FLJ40243* was observed to be expressed at extremely low levels at least in spleen, lymph node, fetal liver and fetal skeletal muscle. No expression was observed in PBMCs, of which we had RNA available, thus, we could not test whether the identified *C7-FLJ40243* risk allele has an effect on *FLJ40243* expression in MS.

The *C7-FLJ40243* region and *IL7R* are 5 Mb apart from each other and there is no LD between these two genes. To test whether these two variants are independent risk factors for MS susceptibility, we calculated how large proportion of all the studied Finnish and Swedish individuals carrying a certain number of the *IL7R* and *C7-FLJ40243* risk alleles were affected. Figure 21 shows how these two MS risk factors contribute to the probability of developing the disease in an additive way: individuals carrying the *C7-FLJ40243* risk allele are in higher risk to get the disease compared to non-carriers, and the risk is even higher when an additional *IL7R* risk allele is present (Figure 21A). However, at least *C7* and *IL7R* are expressed in different tissues and cell types, unlikely acting in same cellular pathways in MS.

Interestingly, in addition to the identified *C7-FLJ40243* region, two other haplotypes within the MS linked locus also provided a p-value <0.001 in the original scan: an intergenic region in 5p15.2 and the *FYB* region in 5p13.1 at 39.2 Mb (Figure 18a). *FYB* encodes a FYN-binding protein isoform 2, which acts as an adapter protein of the FYN and LCP2 signaling cascades in T cells and modulates the expression of interleukin-2 (IL-2). Interestingly, *Fyb*-deficient T lymphocytes of mice are defective in adhering to mouse *Icam1*, to human *ICAM2* (the gene is located on chromosome 17q MS linked region), and to other substrates mediated by integrins (Griffiths et al. 2001). Thus, *FYB* is also a good candidate gene for MS.

Besides *IL7R*, *C7* and *FYB*, the chromosome 5p MS linked locus encompasses also several other immunological genes, which potentially might also have a role in MS susceptibility. Importantly, this thesis study demonstrates the complexity of the genome regions initially identified as potential loci for common diseases and suggests that several independent genetic risk factors may exist in a single locus showing evidence for linkage in many populations.

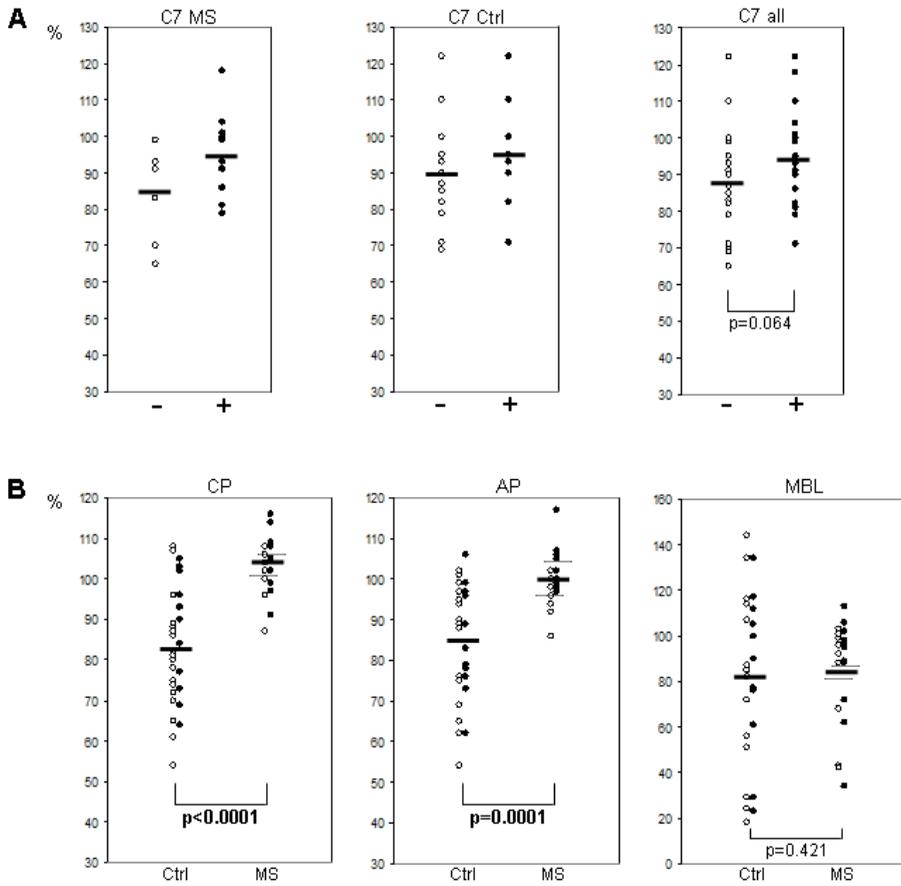


FIGURE 20. C7 protein levels and the complement activity in 20 Finnish MS cases and 32 unaffected controls.

Carriers of the risk allele are indicated with black dots and the non-carriers with open dots. A. Plasma complement component 7 protein levels (C7) were noticed to be higher in carriers of the risk haplotype (+) compared to non-carriers (-). Reference range: 80-120%. B. Amount of serum terminal complement complexes (TCC), formed as a consequence of the activation of classical (CP) and alternative (AP) pathways, was observed to be significantly higher in MS patients compared to controls. Interestingly, the complement system was most active in MS cases carrying the identified C7-FLJ40243 risk allele. No statistically significant difference was observed in the lectin pathway (MBL). Dashed lines indicate the mean values of carriers and noncarriers of the risk haplotype among MS cases. Reference values: CP>60%, AP>40%, MBL>10%.

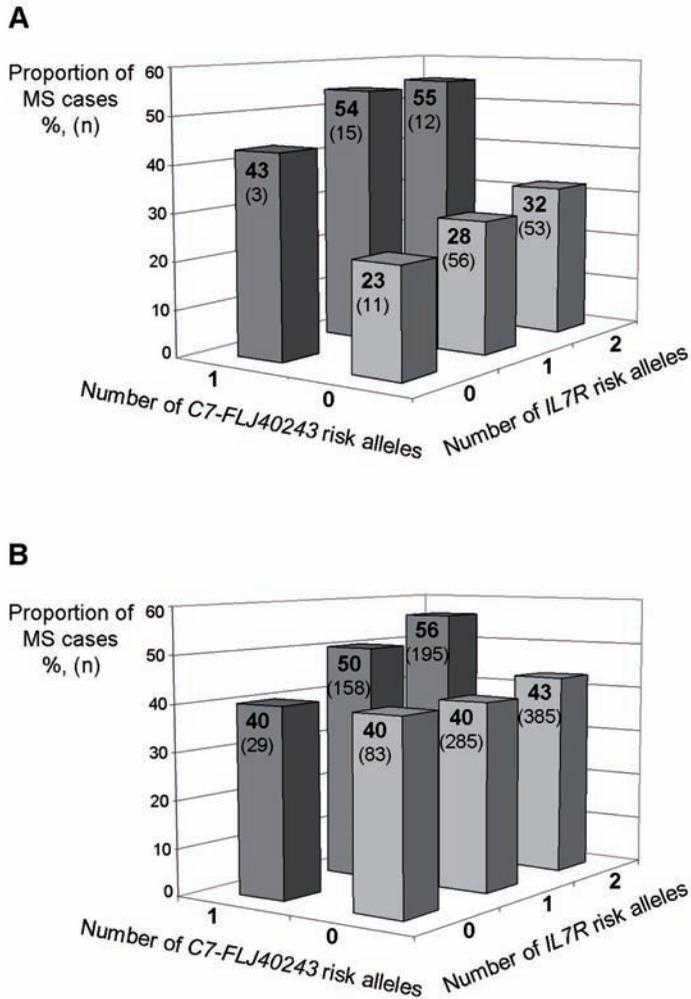


FIGURE 21. Co-effect of the *C7-FLJ40243* and *IL7R* risk alleles on MS predisposition.

Y-axis refers to the proportion of the MS cases among all the individuals of the study sets carrying a certain number of the IL7R (0, 1 or 2 copies) and C7-FLJ40243 (0 or 1 copies) risk alleles. The absolute number of the MS cases in each risk group is shown within the bars in parenthesis. Groups with two copies of the C7-FLJ40243 risk alleles were excluded from the analysis due to the small number of individuals (2 C7-FLJ40243 risk alleles and 0 IL7R risk alleles: Southern Ostrobothnia MS n=0; other parts of Finland and Sweden MS n=4). A. Southern Ostrobothnia MS high-risk region. B. Two study sets from more heterogeneous population, Finland outside the high-risk region and Sweden, combined.

5.2 A follow-up study: variation in other complement cascade genes in MS (III)

After we had observed association with a haplotype flanking the *C7* gene, we wanted to study both genetic and genomic variation of the other complement cascade genes in MS. Interestingly both functional studies in humans as well as genetic studies in EAE rodents have previously provided evidence for involvement of *C3* and TCCs in autoimmune demyelination. In a recently published study, CSF samples of patients with MS and patients with other neurological diseases were compared in order to identify genes that are differentially expressed between the groups and potentially indicate relevant biological pathways for MS. Two important factors of complement-mediated inflammation were identified: clusterin, a regulator of complement activity, and *C3* (Stoop et al. 2008). Further, depositions of complement components have been detected in active MS lesions (Barnett and Prineas 2004, Compston et al. 1989), and increased levels of TCC in the CSF of MS patients have been shown to correlate with neurological disability (Sellebjerg et al. 1998). Moreover, *C6*-deficient rats (unable to form the TCC) as well as *C3*-deficient mice have attenuated EAE, the animal model of MS, and show little to no demyelination compared to wild-type littermates (Nataf 2000, Mead 2002, Szalai 2007). In contrast, deletion of *C5* has no significant effect on the course of EAE (Weerth 2003).

In addition to *C7*, the third and the fifth components of the cascade are critical limiting factors of the complement activation (Figure 5). Thus, we first used HapMap tagger to capture most of the common variation in the *C3* and *C5* genes and genotyped the tagging SNPs in the whole Finnish case-control sample (Figure 14). The SNPs did not show any evidence for association in the well-matched case-control sample from the Southern Ostrobothnia MS high-risk isolate. Suggestive evidence for association was observed with three SNPs of the *C3* gene ($p=0.003$) when the whole Finnish sample set was analysed, but this association was then noticed to be due to population stratification, thus being a false positive finding and highlighting importance of properly matched controls, even in the relatively non-admixed population like Finns.

Next, we monitored for potential MS associated variation in other complement cascade related genes utilizing the SNP data of the two MS GWA studies (IMSGC 2007; Jakkula et al., manuscript in preparation). SNPs showing nominal evidence for association in these studies (at least two SNPs within the gene or the surrounding 10kb region having $>5\%$ difference in case-control allele frequencies in the Finnish GWA or $p<0.03$ in the international GWA) were selected for validation. The 23 potentially interesting SNPs were then genotyped in the whole Finnish case-control sample. However, none of these SNPs provided significant evidence for association in this expanded Finnish sample.

Interindividual variation in the copy-number of the fourth component of the complement cascade is known to exist, the number of total *C4* genes (*C4A* and *C4B*) varying between two and six and four copies being the most common count (Yang 2007). We hypothesized that high copy-numbers of *C4* might predispose to MS by potentially leading to higher complement activity. Thus, we examined the CNV for total *C4* in 174 MS cases and 172 population controls. The variation showed a pattern close to normal distribution both in the cases and in the controls, the majority having four copies of *C4*, as expected (Figure 22). Case-control copy-number frequencies observed in this study were comparable to known copy number frequencies of healthy European Americans (Yang 2007). Importantly, no evidence for association with *C4* CNV and MS was observed ($p>0.05$).

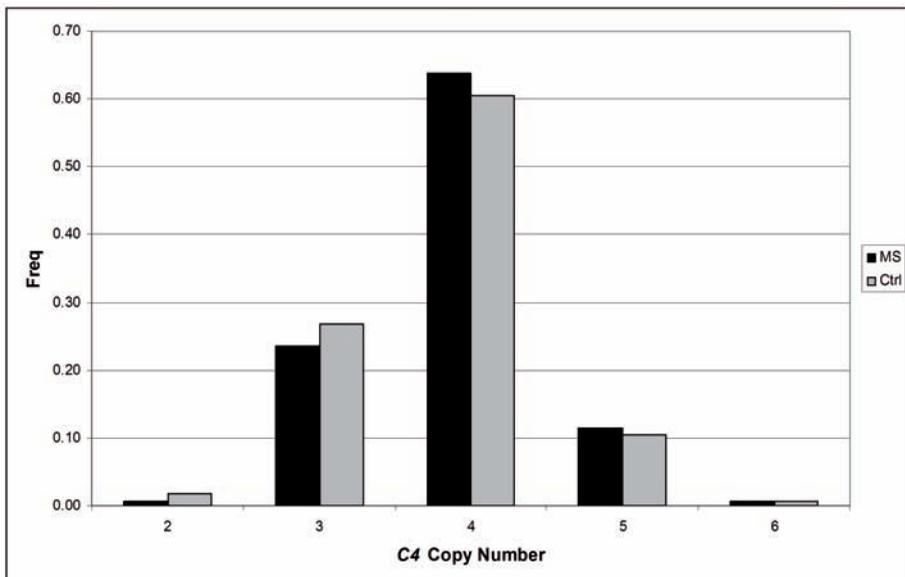


FIGURE 22. *C4* copy number variation.

Frequency of different copy numbers (2-6) in Finnish MS cases (black columns; n=174) and population controls (grey columns; n=172).

The data of this work suggests that variation in the complement component coding genes outside 5p is not associated with genetic susceptibility of MS, at least in Finland. However, since previous publications have provided functional evidence for involvement of both TCCs and C3 in autoimmune demyelination, it would still be interesting to study also C3 protein levels in CSF samples of Finnish MS cases and unaffected controls.

5.3 Candidate genes for immune-mediated demyelination on MS linked loci (IV)

Polycystic lipomembranous osteodysplasia with sclerosing leucoencephalopathy (PLOSL), also known as Nasu-Hakola disease, is a recessively inherited rare disease of the bone and the white matter of the brain. The estimated population prevalence of PLOSL is $1-2 \times 10^{-6}$ (Hakola 1990). The first symptoms of the disease are typically pain and fractures in wrists and ankles at early adulthood, followed by neuropsychiatric symptoms, dementia and premature death. Notably, the most prominent feature of PLOSL is myelin loss in the CNS (Paloneva et al. 2001).

PLOSL is caused by mutations either in the *DAP12* (*TYROBP*, TYRO protein tyrosine kinase binding protein) or in the *TREM2* (triggering receptor expressed on myeloid cells 2) gene (Paloneva et al. 2000, Paloneva et al. 2002). All Finnish patients carry a homozygous 5.3 kb PLOSL_{Fin}-deletion of *DAP12*, whereas inactivating point mutations of *TREM2* have been found in PLOSL patients of other populations (Paloneva et al. 2002). Together *DAP12* and *TREM2* form a signalling receptor complex (Figure 23), which is expressed in various cell types of the myeloid lineage and has quite recently been discovered to be an important regulator of the innate immune system (Klesney-Tait et al. 2006). However, the ligands for *TREM2* as well as the downstream effects of the *DAP12*-*TREM2* mediated signal transduction are still largely unrecognized.

The fairly small *DAP12* (3.9kb) and *TREM2* (4.7kb) genes are located on chromosomes 19q13.12 and 6p21.1, respectively (Figure 24). Interestingly, linkage to these loci has been reported also in families affected by another, more common immune-mediated demyelinating disease, MS (Ebers et al. 1996, Haines et al. 1996, Sawcer et al. 1996, Kuokkanen et al. 1997, Reunanen et al. 2002). Even though no MS cases are known to exist in the Finnish PLOSL pedigrees, the number of Finnish PLOSL families is too small to make any final conclusions. Thus, we wanted to test if allelic variation in *DAP12* or *TREM2* predisposes also MS.

Since homozygous PLOSL_{Fin}-deletion of *DAP12* results in severe white matter changes and premature death in Finnish PLOSL, we hypothesized that the same mutation as heterozygous form could lead to a milder, relapsing phenotype like MS. According to the prevalence of PLOSL, carrier frequency of this 5.3 kb deletion has previously been estimated to be 2.4/1000 in Finland (Hakola 1990), but since high throughput genotyping of the deletion has been difficult, its prevalence in the healthy Finnish population has never been established. Neither has carrier frequency of this mutation been previously checked in MS.

To test our hypothesis, we genotyped the *DAP12* deletion in 744 unrelated Finnish MS cases (randomly selected from the 900 Finnish MS cases to equally represent our full study set) and in the whole set of 1,350 Finnish controls (Figure 14) using an in-house developed high throughput method. Two carriers of the

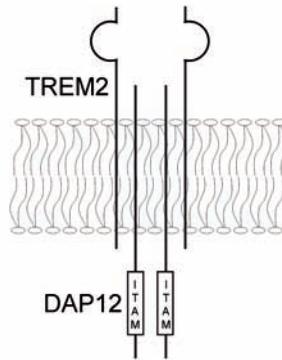


FIGURE 23. The DAP12-TREM2 receptor complex.

The DAP12 and TREM2 proteins form a signalling receptor complex, which is expressed in various cell types of the myeloid lineage.

deletion were identified among the MS cases, corresponding to carrier frequency of 2.7/1,000. Importantly, the clinical picture of these MS patients did not differ from that of the non-carriers. DNA was available for the first degree relatives of one of the carriers, and they were also observed to be carriers. Of the studied MS cases 138 originated from the Southern Ostrobothnia MS high-risk isolate. None of these Ostrobothnian MS cases carried the mutation. Six carriers of the deletion were identified among the controls, corresponding to carrier frequency of 4.9/1,000. Thus, the previous estimation of the $PLOSL_{Fin}$ carrier frequency in Finland was now shown to be slightly underestimated. Importantly, the *DAP12* deletion was not observed to be over-represented among the MS cases.

We further made an effort to study the role of allelic variation of the highly conserved *DAP12* and *TREM2* in MS by linkage and association analyses. All putative SNPs ($n=24$) mapping to *DAP12* and *TREM2* were initially selected from public databases. However, only one of these SNPs was found to be polymorphic (minor allele frequency (MAF) ≥ 0.05 in Europeans) and was selected for genotyping. To find more polymorphisms, we re-sequenced parts of *DAP12*. We found four novel non-coding SNPs, but again, only one of these was observed to have a MAF ≥ 0.05 in Finns, and this SNP was included in the genotyping panel. Further, 15 polymorphic SNPs flanking the *TREM2* gene were selected for genotyping to capture possible variation in this highly conserved locus. The final set of 17 SNPs was genotyped in the whole Finnish case-control sample (Figure 14). No evidence for association was observed with any of the SNPs. Due to a low number of polymorphic SNPs in *TREM2* and *DAP12*, two STS markers nearby these genes were also genotyped in Finnish multiplex MS families (Figure 14), but no evidence for linkage was observed.

Both microglia, the resident immune cells of the CNS, and oligodendrocytes, myelin forming cells of the CNS, express *DAP12* and *TREM2* (Kaifu et al. 2003,

Roumier et al. 2004, Kiialainen et al. 2005, Takahashi et al. 2007). However, no expression of *TREM2* and only modest expression of *DAP12* was detected in PBMCs, of which we had RNA available. Further, there was no difference in expression levels of *DAP12* between the 15 MS patients and 6 unaffected controls studied (unpublished data).

As has been demonstrated previously, the low number of variants in *DAP12* and *TREM2* most probably indicates the crucial role of this receptor complex in immune response modulation. Fenoglio et al. analysed the known polymorphisms of the *TREM2* coding regions in 100 patients with Alzheimer’s disease, 56 patients with frontotemporal lobe degeneration, 78 patients with MS and 140 population controls (Fenoglio et al. 2007). None of the SNPs were polymorphic in this Italian study sample. Further, they sequenced the coding regions of *TREM2* in Alzheimer’s patients and healthy controls but no new mutations were found.

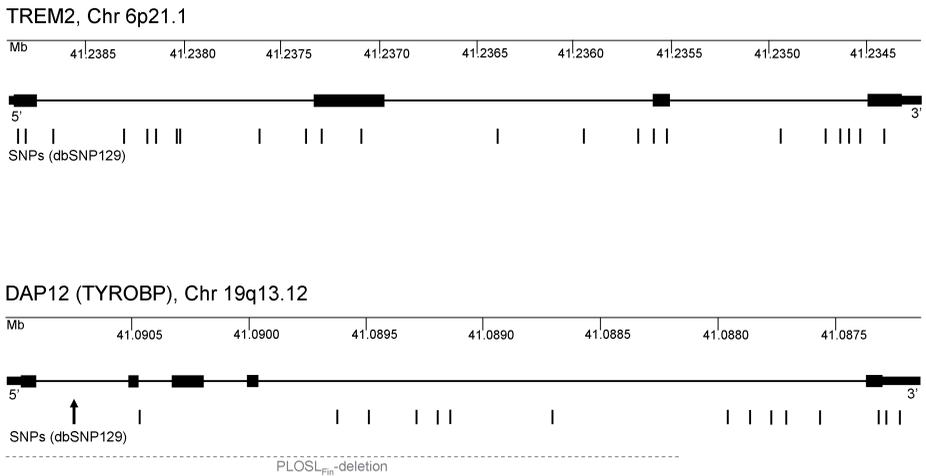


FIGURE 24. The *DAP12* and *TREM2* genes and their chromosomal locations. *The only SNP with MAF>5% in the general European population is marked with an arrow.*

The lack of polymorphisms in *DAP12* and *TREM2* is intriguing but makes these genes very difficult to study. The Affymetrix 550K SNP panel used in the first MS GWA study included only one SNP in *TREM2* (MAF 0.02 in Finland) and no SNPs in *DAP12*. Likewise, the Illumina HumanHap300 SNP panel used in the Finnish GWA study included only one SNP in *DAP12* (genotyped also in this study) and no SNPs in *TREM2*. Neither do the latest GWA panels cover these genes properly. Thus, the future GWA studies most probably do not bring more enlightenment for potential involvement of genetic variation of *DAP12* and *TREM2* in MS.

None of the SNPs in the *TREM2* coding regions was polymorphic in Italian MS patients (Fenoglio et al. 2007). Furthermore, we have sequenced 80% of the

DAP12 gene in Finnish MS patients and found no new mutations. We also studied the non-coding polymorphisms of *DAP12* and *TREM2* and found no evidence for association with MS. However, we did not study the rarest variants ($MAF < 5\%$ in general European population) in MS, and to finish up the study, those variants should also be tested. Another way to try to dissect the role of the DAP12-TREM2 receptor complex in MS is more extensive deep sequencing, which could potentially reveal some rare sporadic mutations. However, these mutations most probably would not explain the linkage observed to the *DAP12* and *TREM2* loci. Further, the future genome wide association studies hopefully will bring some enlightenment for potential involvement of copy number variations in these genes in MS pathogenesis.

The pathogenesis of PLOSL, like that of MS, is mostly unknown. MS and PLOSL share some common features. However, both the clinical picture and the pathology of these two diseases are also dissimilar in many respects. Firstly, PLOSL patients, lacking functional DAP12 or TREM2, suffer from a dramatic and progressive loss of CNS white matter in the deep frontal and temporal white matter (Paloneva et al. 2001), whereas in MS demyelination occurs both in the brain and in the spinal cord, usually in the relapsing-remitting way. Secondly, colonization of activated T lymphocytes is characteristic to MS CNS but not to PLOSL. Moreover, *DAP12* and *TREM2* are expressed also in osteoclasts, and thus, PLOSL patients with inactivating point mutations of the receptor complex develop osteoporosis, but such bone abnormalities have not been observed to be over-represented among MS patients.

We have confirmed that the Finnish PLOSL mutation is not enriched among Finnish MS patients. To conclude, the DAP12-TREM2 receptor complex unlikely has any role in genetic susceptibility to MS in Finland, and the strong linkage of 6p is most probably explained by the *HLA* region, whereas the linkage of chromosome 19 is most probably due to variation in some other locus than *DAP12*. On the other hand, 19q linkage has not been unequivocally replicated in multiple populations and it may also exemplify a false positive finding.

6 CONCLUDING REMARKS

Despite long-standing and intensive research the etiology and pathogenesis of MS are still poorly understood and few predisposing genetic variants have been identified. In this thesis study an effort was made to better understand the molecular background of MS taking advantage of genetics. Traditionally, linkage approach has been used to map the susceptibility loci of genetic diseases. However, the importance of linkage studies in complex diseases has been under debate during this thesis study, and the emphasis in genetic mapping has largely shifted from genomewide linkage studies in families to genome-wide association studies in unrelated cases and controls. These kinds of large analyses in mixed populations are optimized to detect common variants of complex diseases, but will not bring out the relatively rare, penetrant variants, which give rise to a familial concentration of cases.

Here an effort was made to identify MS predisposing genetic variants within the most promising non-*HLA* loci showing evidence for linkage to MS. Taking into account that MS most probably is not a unitary disorder, but instead may represent an overlapping spectrum of related disorders, we have minimized the genetic and environmental heterogeneity by studying familial MS cases from the Southern Ostrobothnian MS high-risk isolate. Presumably, most of these MS patients share the same risk alleles, which can be exposed by the common HapMap markers due to the wide LD intervals. In reference to the aims listed for this thesis, the following findings were presented:

1. In the scan of the 5p linked locus, strongest evidence for association was detected with a haplotype flanking the complement component 7 (*C7*) gene. The identified haplotype is relatively rare, has become enriched in Finland and especially in Southern Ostrobothnian MS pedigrees and seems to have a fairly large effect on genetic susceptibility of MS. Interestingly, there are already multiple lines of evidence to suggest the involvement of the complement system in MS. Plasma *C7* levels and complement activity were here observed to correlate with the risk haplotype identified, the complement system being most active in MS cases carrying the risk allele. The identified risk variant may predispose especially to the most common pathological subtype of MS, pattern II.

The scan of the 17q linked locus showed evidence for association with variants of the protein kinase C alpha (*PRKCA*) gene. Thus, these variants are likely to be in LD with the putative causative variant of the locus, at least in the Finnish isolate. Another variants of *PRKCA* provided nominal evidence for association with MS also in Canadian MS families. We would conclude that the MS risk locus on 17q is more complex than previously assumed and might

- contain multiple genes, different genes potentially playing role in different families.
2. The strong association with the *C7* region stimulated us to study other complement cascade genes in the Finnish case-control sample. No evidence for association could be observed with the complement component coding genes outside 5p and MS.
 3. The highly conserved *DAP12* and *TREM2* genes, located on the MS linked regions of chromosomes 6p and 19q, unlikely have a role in genetic susceptibility of MS. Most importantly, the Finnish PLOSL-mutation is not over-represented among the Finnish MS cases.

This thesis work provides an example of how extended families from special populations can be utilized in fine-mapping of the linked loci, even in this new era of complex genetics. The study also suggests that the genome regions initially identified as potential loci for common diseases most probably are more complex than assumed; It appears that there exists at least two independent risk variants within the chromosome 5p MS locus, and the same may hold true with other loci showing evidence for linkage with MS in several populations. Moreover, the commonly accepted conception that the same allele of the same genetic variant should be repeatable globally to verify the significance of the finding is here challenged. Such may be feasible in question of the common genetic variants, but a new mindset is needed to define less common and more penetrant variants of complex diseases. Finally, the study highlights the rapid process of both knowledge and the technologies of Human genetics during the past few years.

In future, classification of MS patients according to the immunopathological patterns of demyelination could be relevant, since distinct patterns might have different molecular background which, in turn, might require different therapeutic strategies. Such classification could also enable more straightforward replication of the findings by reducing the genetic heterogeneity of mixed populations and thereby unmasking the association.

7 ACKNOWLEDGEMENTS

This study was carried out at the Department of Molecular Medicine, National Public Health Institute, and the Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki Finland during the years 2002-2009. The former and present Director General of the Institutes, Jussi Huttunen and Pekka Puska, as well as the Head of the Department of Molecular Medicine, Anu Jalanko, are acknowledged for providing excellent research facilities.

Professor Leif Groop is acknowledged for accepting the role as an Opponent in my thesis defense. Professor emeritus Arne Svejgaard and Professor Anne Kallioniemi are warmly thanked for thorough revision of this thesis. Their constructive comments led to some very essential improvements. I would also like to thank Päivi Saavalainen and the members of my thesis committee, Aarno Palotie, Pentti Tienari, Per-Henrik Groop and Petteri Arstila, for both critical comments and support.

I wish to thank all the participating MS patients and their families. Without their voluntary assistance none of these conceptions could have been made.

Financial support for this thesis was provided by the National Institutes of Health (NIH) (grant RO1 NS 43559), the Center of Excellence for Disease Genetics of the Academy of Finland, the Paulo Foundation, the Sigrid Juselius Foundation, the Finnish Cultural Foundation, the Biocentrum Helsinki Foundation, the Research Foundation of the Helsinki University Central Hospital, the Neuropromise EU project (grant LSHM-CT-2005-018637), the Multiple Sclerosis Foundation of US, the Canadian Institutes of Health Research (CIHR), the European Molecular Biology Organization (EMBO), Genome Canada and Genome Quebec, the US National Center for Research Resources (grant U54 RR020278), the SGENE EU project (grant LSHM-CT-2006-037761), Simons Foundation (R01MH71425-01A1) and a Harry Weaver Neuroscience Scholar Award from US National Multiple Sclerosis Society, which all are gratefully acknowledged.

I wish to express my deepest gratitude to my supervisors, Professor Leena Peltonen-Palotie and Adjunct Professor Janna Saarela. I have had a privilege to be supervised by two brilliant female scientists who have managed to get me to do my best. In the enthusiastic and inspiring atmosphere created by them I have had a unique chance to grow both as a scientist and as a person.

My sincere thanks go to our collaborators, Pentti Tienari, Keijo Koivisto, Irina Elovaara, Tuula Pirttilä, Mauri Reunanen, the Nordic MS genetics Network, the collaborators from US and Canada, Joe Terwilliger, Mark J Daly, Markku Viander and Seppo Meri, for providing their study samples and expertise for these projects.

I am also deeply indebted for the senior researchers of our laboratory, Marjo Kestilä, Samuli Ripatti, Kaisa Silander, Ismo Ulmanen, Teppo Varilo, Markus Perola, Anu Loukola, who have helped me in many ways during these years. Special thanks also go to the wonderful secretaries Sari Kivikko, Tuija Svanbäck and Mika Kivimäki, the DNA extraction team of Minttu Jussila and Outi Törnwall as well as the sequencing team of Pekka Ellonen. Anne Nyberg, Sisko Lietola, Anne Vikman, Elli Kempas, Siv Knaappila, Lea Puhakka and especially Minna Suvela are acknowledged for skilful technical assistance. Jari Raikko, Juha Saharinen, Juha Knuuttila, Juri Ahokas and Hannu Turunen, thank you for the IT-support and all the help. Päivi Hauhia and Christine Strid are thanked for the assistance in the publication process.

“Missukat” Anu, Ansku, Virpi, Eveliina, your company has made this work joyful – we have had great time, great discussions (and great coffee too). Many thanks also to all who have helped me with my project and/or made the lab such a nice place to work at: Roxy, Denis, Anna, Kati, PP, Annu, Jussi, Emma, Marika, Olli, Nora, Karola, Sampo, Joni, Mervi, Kirsi, Annika, Jarkko, Johannes, Pia, Tintti, Juho, Annina, Tero, Markus, and anyone I failed to mention here. Mari, Tämer, Ulla and Petteri from ”family Kankkunen”, Anne and Mari from ”Pachamamas”, Salla, Laura, Varpu, Henri, Liisa, Markku, Petra, Hanna-Mari and all friends from the Medical School, my dearest friends, thank you for lending me an ear and maintaining the balance between work and relationships.

Last but not least, I warmly thank my family. Father, you have been my role model; I admire your determination and resilience. Antti, thank you for your unconditional love, faith and support.

Without you all I could not fulfill my dreams today.

Helsinki, May 13th, 2009

Suvi Kallio

References

- Abrahams BS, Geschwind DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 9(5):341-55.
- Alvarez-Lafuente R, García-Montojo M, De Las Heras V, Domínguez-Mozo MI, Bartolome M, Benito-Martin MS, Arroyo R (2008) Herpesviruses and human endogenous retroviral sequences in the cerebrospinal fluid of multiple sclerosis patients. *Mult Scler* 14(5):595-601.
- Arthur AT, Armati PJ, Bye C; Southern MS Genetics Consortium, Heard RN, Stewart GJ, Pollard JD, Booth DR (2008) Genes implicated in multiple sclerosis pathogenesis from consilience of genotyping and expression profiles in relapse and remission. *BMC Med Genet* 9:17.
- Ascherio A, Munch M (2000) Epstein-Barr virus and multiple sclerosis. *Epidemiology* 11(2):220-4.
- Babbe H, Roers A, Waisman A, Lassmann H, Goebels N, Hohlfeld R, Friese M, Schröder R, Deckert M, Schmidt S, Ravid R, Rajewsky K (2000) Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. *J Exp Med* 192(3):393-404.
- Ban M, Maranian M, Yeo TW, Gray J, Compston A, Sawcer S (2005) No evidence for association of the protein kinase C alpha gene with multiple sclerosis. *J Neurol* 252(5):619-20.
- Barcellos LF, Oksenberg JR, Green AJ, Bucher P, Rimmler JB, Schmidt S, Garcia ME, Lincoln RR, Pericak-Vance MA, Haines JL, Hauser SL; Multiple Sclerosis Genetics Group (2002) Genetic basis for clinical expression in multiple sclerosis. *Brain* 125(Pt 1):150-8.
- Barcellos LF, Kamdar BB, Ramsay PP, DeLoa C, Lincoln RR, Caillier S, Schmidt S, Haines JL, Pericak-Vance MA, Oksenberg JR, Hauser SL (2006) Clustering of autoimmune diseases in families with a high-risk for multiple sclerosis: a descriptive study. *Lancet Neurol* 5(11):924-31.
- Barnett MH, Prineas JW (2004) Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. *Ann Neurol* 55(4):458-68.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhardt AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossom A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghorri J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 40(8):955-62.
- Barton A, Woolmore JA, Ward D, Eyre S, Hinks A, Ollier WE, Strange RC, Fryer AA, John S, Hawkins CP, Worthington J (2004) Association of protein kinase C alpha (PRKCA) gene with multiple sclerosis in a UK population. *Brain* 127(Pt 8):1717-22.
- Begovich AB, Caillier SJ, Alexander HC, Penko JM, Hauser SL, Barcellos LF, Oksenberg JR (2005) The R620W polymorphism of the protein tyrosine phosphatase PTPN22 is not associated with multiple sclerosis. *Am J Hum Genet* 76(1):184-7.
- Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40(6):695-701.
- Bonetti A, Reunanen K, Finnilä S, Koivisto K, Wikström J, Sumelahti ML, Pirttilä T, Elovaara I, Reunanen M, Saarela J, Peltonen L, Rantamäki T, Tienari PJ (2004) A two-stage study on multiple sclerosis susceptibility and chromosome 2q33. *Genes Immun* 5(2):142-6.
- Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes:

- past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33 Suppl:228-37.
- Broadley SA, Deans J, Sawcer SJ, Clayton D, Compston DA (2000) Autoimmune disease in first-degree relatives of patients with multiple sclerosis. A UK survey. *Brain* 123 (Pt 6):1102-11.
- Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson E, Palmgren J, Hillert J (2007) HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. *PLoS ONE* 2(7):e664.
- Butterfield RJ, Sudweeks JD, Blankenhorn EP, Korngold R, Marini JC, Todd JA, Roper RJ, Teuscher C (1998) New genetic loci that control susceptibility and symptoms of experimental allergic encephalomyelitis in inbred mice. *J Immunol* 161(4):1860-7.
- Cappellano G, Orilieri E, Comi C, Chiocchetti A, Bocca S, Boggio E, Bernardone IS, Cometa A, Clementi R, Barizzone N, D'Alfonso S, Corrado L, Galimberti D, Scarpini E, Guerini FR, Caputo D, Paolicelli D, Trojano M, Figà-Talamanca L, Salvetti M, Perla F, Leone M, Monaco F, Dianzani U (2008) Variations of the perforin gene in patients with multiple sclerosis. *Genes Immun* 9(5):438-44.
- Cavalli-Sforza LL (2007) Human evolution and its relevance for genetic epidemiology. *Annu Rev Genomics Hum Genet* 8:1-15.
- Chao MJ, Barnardo MC, Lui GZ, Lincoln MR, Ramagopalan SV, Herrera BM, Dymont DA, Sadovnick AD, Ebers GC (2007) Transmission of class I/II multi-locus MHC haplotypes and multiple sclerosis susceptibility: accounting for linkage disequilibrium. *Hum Mol Genet* 16(16):1951-8.
- Chen DC, Saarela J, Clark RA, Miettinen T, Chi A, Eichler EE, Peltonen L, Palotie A (2004) Segmental duplications flank the multiple sclerosis locus on chromosome 17q. *Genome Res* 14(8):1483-92.
- Christensen T (2007) Human herpesviruses in MS. *Int MS J* 14(2):41-7.
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138(3):963-71.
- Compston DA, Morgan BP, Campbell AK, Wilkins P, Cole G, Thomas ND, Jasani B (1989) Immunocytochemical localization of the terminal complement complex in multiple sclerosis. *Neuropathol Appl Neurobiol* 15(4):307-16.
- Compston A (1997) Genetic epidemiology of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 62(6):553-61.
- Compston A, Coles A (2002) Multiple sclerosis. *Lancet* 359(9313):1221-31.
- Corporaal S, Bijl M, Kallenberg CG (2002) Familial occurrence of autoimmune diseases and autoantibodies in a Caucasian population of patients with systemic lupus erythematosus. *Clin Rheumatol* 21(2):108-13.
- Dai KZ, Harbo HF, Celius EG, Oturai A, Sørensen PS, Ryder LP, Datta P, Svejgaard A, Hillert J, Fredrikson S, Sandberg-Wollheim M, Laaksonen M, Myhr KM, Nyland H, Vartdal F, Spurkland A (2001) The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis. *Genes Immun* 2(5):263-8.
- Demirci FY, Manzi S, Ramsey-Goldman R, Minster RL, Kenney M, Shaw PS, Dunlop-Thomas CM, Kao AH, Rhew E, Bontempo F, Kammerer C, Kamboh MI (2007) Association of a common interferon regulatory factor 5 (IRF5) variant with increased risk of systemic lupus erythematosus (SLE). *Ann Hum Genet* 71(Pt 3):308-11.
- Dideberg V, Kristjansdottir G, Milani L, Libioulle C, Sigurdsson S, Louis E, Wiman AC, Vermeire S, Rutgeerts P, Belaiche J, Franchimont D, Van Gossum A, Bours V, Sävänen AC (2007) An insertion-deletion polymorphism in the interferon regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. *Hum Mol Genet* 16(24):3008-16.
- Dymont DA, Willer CJ, Scott B, Armstrong H, Ligers A, Hillert J, Paty DW, Hashimoto S, Devonshire V, Hooge J, Kastrukoff L, Oger J, Metz L, Warren S, Hader W, Power C, Auty A, Nath A, Nelson R, Freedman M, Brunet D, Paulseth JE, Rice G, O'Connor P, Duquette P, Lapierre Y, Francis G, Bouchard JP, Murray TJ, Bhan V, Maxner C, Pryse-Phillips W, Stefanelli M, Sadovnick AD, Risch N, Ebers GC (2001) Genetic susceptibility to MS: a second stage analysis in Canadian MS families. *Neurogenetics* 3(3):145-51.
- Dymont DA, Steckley JL, Willer CJ, Armstrong H, Sadovnick AD, Risch N, Ebers GC (2002) No evidence to support CTLA-4

- as a susceptibility gene in MS families: the Canadian Collaborative Study. *J Neuroimmunol* 123(1-2):193-8.
- Dyment DA, Ebers GC, Sadovnick AD (2004) Genetics of multiple sclerosis. *Lancet Neurol* 3(2):104-10.
- Ebers GC, Sadovnick AD, Risch NJ (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature* 377(6545):150-1.
- Ebers GC, Kukay K, Bulman DE, Sadovnick AD, Rice G, Anderson C, Armstrong H, Cousin K, Bell RB, Hader W, Paty DW, Hashimoto S, Oger J, Duquette P, Warren S, Gray T, O'Connor P, Nath A, Auty A, Metz L, Francis G, Paulseth JE, Murray TJ, Pryse-Phillips W, Nelson R, Freedman M, Brunet D, Bouchard JP, Hinds D, Risch N (1996) A full genome search in multiple sclerosis. *Nat Genet* 13(4):472-6.
- Ebers GC (2008) Environmental factors and multiple sclerosis. *Lancet Neurol* 7(3):268-77.
- Eraksoy M, Kurtuncu M, Akman-Demir G, Kilinc M, Gedizlioglu M, Mirza M, Anlar O, Kutlu C, Demirkiran M, Idrisoglu HA, Compston A, Sawcer S; Turkish Multiple Sclerosis Genetics Study Group (2003) A whole genome screen for linkage in Turkish multiple sclerosis. *J Neuroimmunol* 143(1-2):17-24.
- Feng JM, Fernandes AO, Campagnoni CW, Hu YH, Campagnoni AT (2004) The golli-myelin basic protein negatively regulates signal transduction in T lymphocytes. *J Neuroimmunol* 152(1-2):57-66.
- Fenoglio C, Galimberti D, Piccio L, Scalabrini D, Panina P, Buonsanti C, Venturelli E, Lovati C, Forloni G, Mariani C, Bresolin N, Scarpini E (2007) Absence of TREM2 polymorphisms in patients with Alzheimer's disease and Frontotemporal Lobar Degeneration. *Neurosci Lett* 411(2):133-7.
- Fogdell-Hahn A, Ligens A, Grønning M, Hillert J, Olerup O (2000) Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* 55(2):140-8.
- Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M, Schreiber S (2008) Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet* 40(6):713-5.
- Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10(4):241-51.
- Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA (2009) Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. *Genes Immun* 10(2):188-91.
- Gabriel S, Ziaugra L (2004) SNP genotyping using Sequenom MassARRAY 7K platform. *Curr Protoc Hum Genet* Chapter 2:Unit 2.12.
- Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet* Chapter 2:Unit 2.12.
- Gale CR, Martyn CN (1995) Migrant studies in multiple sclerosis. *Prog Neurobiol* 47(4-5):425-48.
- GAMES; Transatlantic Multiple Sclerosis Genetics Cooperative (2003) A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* 143(1-2):39-46.
- Gasque P, Fontaine M, Morgan BP (1995) Complement expression in human brain. Biosynthesis of terminal pathway components and regulators in human glial cells and cell lines. *J Immunol* 154:4726-4733.
- Gold R, Lington C, Lassmann H (2006) Understanding pathogenesis and therapy of multiple sclerosis via animal models: 70 years of merits and culprits in experimental autoimmune encephalomyelitis research. *Brain* 129(Pt 8):1953-71.
- Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, Caillier SJ, Ban M, Goris A, Barcellos LF, Lincoln R, McCauley JL, Sawcer SJ, Compston DA, Dubois B, Hauser SL, Garcia-Blanco MA, Pericak-Vance MA, Haines JL; Multiple Sclerosis Genetics Group (2007) Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet* 39(9):1083-91.
- Greve B, Simonenko R, Illes Z, Peterfalvi A, Hamdi N, Mycko MP, Selmaj KW, Rozsa C, Rajczyk K, Bauer P, Berger K, Weissert R (2008) Multiple sclerosis and the

- CTLA4 autoimmunity polymorphism CT60: no association in patients from Germany, Hungary and Poland. *Mult Scler* 14(2):153-8.
- Griffiths EK, Krawczyk C, Kong YY, Raab M, Hyduk SJ, Bouchard D, Chan VS, Kozieradzki I, Oliveira-Dos-Santos AJ, Wakeham A, Ohashi PS, Cybulsky MI, Rudd CE, Penninger JM (2001) Positive regulation of T cell activation and integrin adhesion by the adapter Fyb/Slap. *Science* 293(5538):2260-3.
- Hafler DA, Slavik JM, Anderson DE, O'Connor KC, De Jager P, Baecher-Allan C (2005) Multiple sclerosis. *Immunol Rev* 204:208-31.
- Haines JL, Ter-Minassian M, Bazyk A, Gusella JF, Kim DJ, Terwedow H, Pericak-Vance MA, Rimmler JB, Haynes CS, Roses AD, Lee A, Shaner B, Menold M, Seboun E, Fitoussi RP, Gartioux C, Reyes C, Ribierre F, Gyapay G, Weissenbach J, Hauser SL, Goodkin DE, Lincoln R, Usuku K, Oksenberg JR, et al (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. The Multiple Sclerosis Genetics Group. *Nat Genet* 13(4):469-71.
- Hakola P (1990) Polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (membranous lipodystrophy): a neuropsychiatric follow-up study. *Monographs of Psychiatria Fennica* 17, 1-114.
- Harbo HF, Celius EG, Vartdal F, Spurkland A (1999) CTLA4 promoter and exon 1 dimorphisms in multiple sclerosis. *Tissue Antigens* 53(1):106-10.
- Harbo HF, Lie BA, Sawcer S, Celius EG, Dai KZ, Oturai A, Hillert J, Lorentzen AR, Laaksonen M, Myhr KM, Ryder LP, Fredrikson S, Nyland H, Sørensen PS, Sandberg-Wollheim M, Andersen O, Svejgaard A, Edland A, Mellgren SI, Compston A, Vartdal F, Spurkland A (2004) Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis. *Tissue Antigens* 63(3):237-47.
- Hensiek AE, Seaman SR, Barcellos LF, Oturai A, Eraksoi M, Cocco E, Vecsei L, Stewart G, Dubois B, Bellman-Strobl J, Leone M, Andersen O, Bencsik K, Booth D, Celius EG, Harbo HF, Hauser SL, Heard R, Hillert J, Myhr KM, Marrosu MG, Oksenberg JR, Rajda C, Sawcer SJ, Sørensen PS, Zipp F, Compston DA (2007) Familial effects on the clinical course of multiple sclerosis. *Neurology* 68(5):376-83.
- Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6(2):95-108.
- Hogasen AK, Wurzner R, Abrahamsen TG, Dierich MP (1995) Human polymorphonuclear leukocytes store large amounts of terminal complement components C7 and C6, which may be released on stimulation. *J Immunol* 154:4734-4740.
- Hosokawa M, Klegeris A, Maguire J, McGeer PL (2003) Expression of complement messenger RNAs and proteins by human oligodendroglial cells. *Glia* 42:417-423.
- Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugierie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JE, Thomas G (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 379(6568):821-3.
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JE, Sahbatou M, Thomas G (2001)
- Inglese M (2006) Multiple sclerosis: new insights and trends. *AJNR Am J Neuroradiol* 27(5):954-7.
- Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411(6837):599-603.
- International Multiple Sclerosis Genetics Consortium (IMSGC) (2005) A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet* 77(3):454-67.
- International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 357(9):851-62

- International Multiple Sclerosis Genetics Consortium (IMSGC) (2008) Refining genetic associations in multiple sclerosis. *Lancet Neurol* 7(7):567-9.
- Jagodic M, Kornek B, Weissert R, Lassmann H, Olsson T, Dahlman I (2001) Congenic mapping confirms a locus on rat chromosome 10 conferring strong protection against myelin oligodendrocyte glycoprotein-induced experimental autoimmune encephalomyelitis. *Immunogenetics* 53(5):410-5.
- Jakkula E, Rehnström K, Varilo T, Pietiläinen OP, Paunio T, Pedersen NL, deFaire U, Järvelin MR, Saharinen J, Freimer N, Ripatti S, Purcell S, Collins A, Daly MJ, Palotie A, Peltonen L (2008) The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet* 83(6):787-94.
- Jersild C, Fog T, Hansen GS, Thomsen M, Svejgaard A, Dupont B (1973) Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. *Lancet* 2(7840):1221-5.
- Kaifu T, Nakahara J, Inui M, Mishima K, Momiyama T, Kaji M, Sugahara A, Koito H, Ujike-Asai A, Nakamura A, Kanazawa K, Tan-Takeuchi K, Iwasaki K, Yokoyama WM, Kudo A, Fujiwara M, Asou H, Takai T (2003) Osteopetrosis and thalamic hypomyelination with synaptic degeneration in DAP12-deficient mice. *J Clin Invest* 111(3):323-32.
- Kavvoura FK, Ioannidis JP (2005) CTLA-4 gene polymorphisms and susceptibility to type 1 diabetes mellitus: a HuGE Review and meta-analysis. *Am J Epidemiol* 162(1):3-16.
- Kavvoura FK, Akamizu T, Awata T, Ban Y, Chistiakov DA, Frydecka I, Ghaderi A, Gough SC, Hiromatsu Y, Ploski R, Wang PW, Ban Y, Bednarczuk T, Chistiakova EI, Chojm M, Heward JM, Hiratani H, Juo SH, Karabon L, Katayama S, Kurihara S, Liu RT, Miyake I, Omrani GH, Pawlak E, Taniyama M, Tozaki T, Ioannidis JP (2007) Cytotoxic T-lymphocyte associated antigen 4 gene polymorphisms and autoimmune thyroid disease: a meta-analysis. *J Clin Endocrinol Metab* 92(8):3162-70.
- Keegan M, König F, McClelland R, Brück W, Morales Y, Bitsch A, Panitch H, Lassmann H, Weinschenker B, Rodriguez M, Parisi J, Lucchinetti CF (2005) Relation between humoral pathological changes in multiple sclerosis and response to therapeutic plasma exchange. *Lancet* 366(9485):579-82.
- Kiialainen A, Hovanec K, Paloneva J, Kopra O, Peltonen L (2005) Dap12 and Trem2, molecules involved in innate immunity and neurodegeneration, are co-expressed in the CNS. *Neurobiol Dis* 18(2):314-22.
- Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. *Am J Hum Genet* 62(5):1171-9.
- Klesney-Tait J, Turnbull IR, Colonna M (2006) The TREM receptor family and signal integration. *Nat Immunol* 7(12):1266-73.
- Kristjansdottir G, Sandling JK, Bonetti A, Roos IM, Milani L, Wang C, Gustafsdottir SM, Sigurdsson S, Lundmark A, Tienari PJ, Koivisto K, Elovaara I, Pirttilä T, Reunanen M, Peltonen L, Saarela J, Hillert J, Olsson T, Landegren U, Alcina A, Fernández O, Leyva L, Guerrero M, Lucas M, Izquierdo G, Matesanz F, Syvänen AC (2008) Interferon regulatory factor 5 (IRF5) gene variants are associated with multiple sclerosis in three distinct populations. *J Med Genet* 45(6):362-9.
- Kuokkanen S, Sundvall M, Terwilliger JD, Tienari PJ, Wikström J, Holmdahl R, Pettersson U, Peltonen L (1996) A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus Eae2. *Nat Genet* 13(4):477-80.
- Kuokkanen S, Gschwend M, Rioux JD, Daly MJ, Terwilliger JD, Tienari PJ, Wikström J, Palo J, Stein LD, Hudson TJ, Lander ES, Peltonen L (1997) Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am J Hum Genet* 61(6):1379-87.
- Kurtzke JF (1974) Further features of the Fennoscandian focus of multiple sclerosis. *Acta Neurol Scand* 50(4):478-502.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. (2001) Initial sequencing and analysis of the

- human genome. *Nature* 409(6822):860-921.
- Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E (2001) Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Am J Hum Genet* 69(supplement):A1886.
- Langegegen H, Pausa M, Johnson E, Casarsa C, Tedesco F (2000) The endothelium is an extrahepatic site of synthesis of the seventh component of the complement system. *Clin Exp Immunol* 121:69-76.
- Larsen F, Oturai A, Ryder LP, Madsen HO, Hillert J, Fredrikson S, Sandberg-Wollheim M, Laaksonen M, Harbo HF, Sawcer S, Fugger L, Sorensen PS, Svejgaard A (2000) Linkage analysis of a candidate region in Scandinavian sib pairs with multiple sclerosis reveals linkage to chromosome 17q. *Genes Immun* 1(7):456-9.
- Lassmann H, Brück W, Lucchinetti C (2001) Heterogeneity of multiple sclerosis pathogenesis: implications for diagnosis and therapy. *Trends Mol Med* 7(3):115-21.
- Libbey JE, McCoy LL, Fujinami RS (2007) Molecular mimicry in multiple sclerosis. *Int Rev Neurobiol* 79:127-47.
- Ligers A, Xu C, Saarinen S, Hillert J, Olerup O (1999) The CTLA-4 gene is associated with multiple sclerosis. *J Neuroimmunol* 97(1-2):182-90.
- Liu X, Lee YS, Yu CR, Egwuagu CE (2008) Loss of STAT3 in CD4+ T cells prevents development of experimental autoimmune diseases. *J Immunol* 180(9):6070-6.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JE, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5(10):e254.
- Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M, Ferretti V, Tienari PJ, Sadovnick AD, Peltonen L, Ebers GC, Hudson TJ (2005) A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nat Genet* 37(10):1108-12.
- Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, Langer-Gould A, Strober S, Cannella B, Allard J, Klonowski P, Austin A, Lad N, Kaminski N, Galli SJ, Oksenberg JR, Raine CS, Heller R, Steinman L (2002) Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. *Nat Med* 8(5):500-8.
- Lorentzen AR, Celius EG, Ekström PO, Wiencke K, Lie BA, Myhr KM, Ling V, Thorsby E, Vartdal F, Spurkland A, Harbo HF (2005) Lack of association with the CD28/CTLA4/ICOS gene region among Norwegian multiple sclerosis patients. *J Neuroimmunol* 166(1-2):197-201.
- Lorentzen AR, Smestad C, Lie BA, Oturai AB, Akesson E, Saarela J, Myhr KM, Vartdal F, Celius EG, Sorensen PS, Hillert J, Spurkland A, Harbo HF (2008) The SH2D2A gene and susceptibility to multiple sclerosis. *J Neuroimmunol* 197(2):152-8.
- Lucchinetti C, Brück W, Parisi J, Scheithauer B, Rodriguez M, Lassmann H (2000) Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Ann Neurol* 47(6):707-17.
- Lundmark F, Duvefelt K, Jacobaeus E, Kockum I, Wallström E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J (2007) Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat Genet* 39(9):1108-13.
- Maier LM, Hafler DA (2008) The developing mosaic of autoimmune disease risk. *Nat Genet* 40(2):131-2.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9(5):356-69.
- McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, Sandberg-Wollheim M, Sibley W, Thompson A, van den Noort S, Weinshenker BY, Wolinsky JS (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis

- of multiple sclerosis. *Ann Neurol* 50(1):121-7.
- McDonnell GV, Mawhinney H, Graham CA, Hawkins SA, Middleton D (1999) A study of the HLA-DR region in clinical subgroups of multiple sclerosis and its influence on prognosis. *J Neurol Sci* 165(1):77-83.
- Mead RJ, Singhrao SK, Neal JW, Lassmann H, Morgan BP (2002) The membrane attack complex of complement causes severe demyelination associated with acute axonal injury. *J Immunol* 168:458-465.
- Meri S (2007) Loss of self-control in the complement system and innate autoreactivity. *Ann N Y Acad Sci* 1109:93-105.
- Midgard R, Grønning M, Riise T, Kvåle G, Nyland H (1996) Multiple sclerosis and chronic inflammatory diseases. A case-control study. *Acta Neurol Scand* 93(5):322-8.
- Morgan BP, Gasque P (1997) Extrahepatic complement biosynthesis: where, when and why? *Clin Exp Immunol* 107:1-7.
- Nataf S, Carroll SL, Wetsel RA, Szalai AJ, Barnum SR (2000) Attenuation of experimental autoimmune demyelination in complement-deficient mice. *J Immunol* 165:5867-73.
- Nelissen I, Fiten P, Vandenbroeck K, Hillert J, Olsson T, Marrosu MG, Opdenakker G (2000) PECAM1, MPO and PRKARIA at chromosome 17q21-q24 and susceptibility for multiple sclerosis in Sweden and Sardinia. *J Neuroimmunol* 108(1-2):153-9.
- Nelissen I, Dubois B, Goris A, Ronsse I, Carton H, Opdenakker G (2002) Gelatinase B, PECAM-1 and MCP-3 gene polymorphisms in Belgian multiple sclerosis. *J Neurol Sci* 200(1-2):43-8.
- Norio R (2003) Finnish Disease Heritage I: characteristics, causes, background. *Hum Genet* 112(5-6):441-56.
- Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 74(4):765-9.
- Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, Lincoln RR, Swerdlin A, Mignot E, Lin L, Goodin D, Erlich HA, Schmidt S, Thomson G, Reich DE, Pericak-Vance MA, Haines JL, Hauser SL (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet* 74(1):160-7.
- Oksenberg JR, Barcellos LF (2005) Multiple sclerosis genetics: leaving no stone unturned. *Genes Immun* 6(5):375-87.
- Oturai A, Larsen F, Ryder LP, Madsen HO, Hillert J, Fredrikson S, Sandberg-Wollheim M, Laaksonen M, Koch-Henriksen N, Sawcer S, Fugger L, Sorensen PS, Svejgaard A (1999) Linkage and association analysis of susceptibility regions on chromosomes 5 and 6 in 106 Scandinavian sibling pair families with multiple sclerosis. *Ann Neurol* 46(4):612-6.
- Paloneva J, Kestilä M, Wu J, Salminen A, Böhling T, Ruotsalainen V, Hakola P, Bakker AB, Phillips JH, Pekkarinen P, Lanier LL, Timonen T, Peltonen L (2000) Loss-of-function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. *Nat Genet* 25(3):357-61.
- Paloneva J, Autti T, Raininko R, Partanen J, Salonen O, Puranen M, Hakola P, Haltia M (2001) CNS manifestations of Nasu-Hakola disease: a frontal dementia with bone cysts. *Neurology* 56(11):1552-8.
- Paloneva J, Manninen T, Christman G, Hovanes K, Mandelin J, Adolfsson R, Bianchin M, Bird T, Miranda R, Salmaggi A, Tranebjærg L, Kontinen Y, Peltonen L (2002) Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *Am J Hum Genet* 71(3):656-62.
- Pearson TA, Manolio TA (2008) How to interpret a genome-wide association study. *JAMA* 299(11):1335-44.
- Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8(10):1913-23.
- Varilo T and Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. *Curr Opin Genet Dev* 14(3):316-23.
- Pihlaja H, Rantamäki T, Wikström J, Sumelahti ML, Laaksonen M, Ilonen J, Ruutiainen J, Pirttilä T, Elovaara I, Reunanen M, Kuokkanen S, Peltonen L, Koivisto K, Tienari PJ (2003) Linkage disequilibrium between the MBP tetranucleotide repeat and multiple sclerosis is restricted to a geographically defined subpopulation in Finland. *Genes Immun* 4(2):138-46.

- Pittock SJ, McClelland RL, Achenbach SJ, König F, Bitsch A, Brück W, Lassmann H, Parisi JE, Scheithauer BW, Rodriguez M, Weinschenker BG, Lucchinetti CF (2005) Clinical course, pathological correlations, and outcome of biopsy proved inflammatory demyelinating disease. *J Neurol Neurosurg Psychiatry* 76(12):1693-7.
- Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, Lublin FD, Metz LM, McFarland HF, O'Connor PW, Sandberg-Wollheim M, Thompson AJ, Weinschenker BG, Wolinsky JS (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann Neurol* 58(6):840-6.
- Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, Johnson KP, Sibley WA, Silberberg DH, Tourtellotte WW (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 13(3):227-31.
- Prat A, Antel J (2005) Pathogenesis of multiple sclerosis. *Curr Opin Neurol* 18(3):225-30.
- Pugliatti M, Sotgiu S, Rosati G (2002) The worldwide prevalence of multiple sclerosis. *Clin Neurol Neurosurg* 104(3):182-91.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559-75.
- Ramagopalan SV, Anderson C, Sadovnick AD, Ebers GC (2007 B) Genomewide study of multiple sclerosis. *N Engl J Med* 357(21):2199-200.
- Ramagopalan SV, Dymnt DA, Valdar W, Herrera BM, Criscuoli M, Yee IM, Sadovnick AD, Ebers GC; Canadian Collaborative Study Group (2007 A) Autoimmune disease in families with multiple sclerosis: a population-based study. *Lancet Neurol* 6(7):604-10.
- Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17(9):502-10.
- Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A, Tandon A, Lincoln RR, DeLoa C, Fruhan SA, Cabre P, Bera O, Semana G, Kelly MA, Francis DA, Ardlie K, Khan O, Cree BA, Hauser SL, Oksenberg JR, Hafler DA (2005) A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 37(10):1113-8.
- Reunanen K, Finnilä S, Laaksonen M, Sumelahti ML, Wikström J, Pastinen T, Kuokkanen S, Saarela J, Uimari P, Ruutiainen J, Ilonen J, Peltonen L, Tienari PJ (2002) Chromosome 19q13 and multiple sclerosis susceptibility in Finland: a linkage and two-stage association study. *J Neuroimmunol* 126(1-2):134-42.
- Riise Stensland HM, Saarela J, Bronnikov DO, Parkkonen M, Jokiaho AJ, Palotie A, Tienari PJ, Sumelahti ML, Elovaara I, Koivisto K, Pirttilä T, Reunanen M, Sobel E, Peltonen L (2005) Fine mapping of the multiple sclerosis susceptibility locus on 5p14-p12. *J Neuroimmunol* 170(1-2):122-33.
- Roumier A, Béchade C, Poncer JC, Smalla KH, Tomasello E, Vivier E, Gundelfinger ED, Triller A, Bessis A (2004) Impaired synaptic function in the microglial KARAP/DAP12-deficient mouse. *J Neurosci* 24(50):11421-8.
- Rubio JP, Stankovich J, Field J, Tubridy N, Marriott M, Chapman C, Bahlo M, Perera D, Johnson LJ, Tait BD, Varney MD, Speed TP, Taylor BV, Foote SJ, Butzkueven H, Kilpatrick TJ (2008) Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes Immun* 9(7):624-30.
- Saarela J, Schoenberg Fejzo M, Chen D, Finnilä S, Parkkonen M, Kuokkanen S, Sobel E, Tienari PJ, Sumelahti ML, Wikström J, Elovaara I, Koivisto K, Pirttilä T, Reunanen M, Palotie A, Peltonen L (2002) Fine mapping of a multiple sclerosis locus to 2.5 Mb on chromosome 17q22-q24. *Hum Mol Genet* 11(19):2257-67.
- Sadovnick AD (1993) Familial recurrence risks and inheritance of multiple sclerosis. *Curr Opin Neurol Neurosurg* 6(2):189-94.
- Salmela E, Lappalainen T, Fransson I, Andersen PM, Dahlman-Wright K, Fiebig A, Sistonen P, Savontaus ML, Schreiber S, Kere J, Lahermo P (2008) Genome-wide analysis of single nucleotide polymorphisms uncovers population structure in northern europe. *PLoS ONE* 3(10):e3519.
- Santoro F, Kennedy PE, Locatelli G, Malnati MS, Berger EA, Lusso P (1999) CD46 is a

- cellular receptor for human herpesvirus 6. *Cell* 99(7):817-27.
- Sawcer S, Jones HB, Feakes R, Gray J, Smaldon N, Chataway J, Robertson N, Clayton D, Goodfellow PN, Compston A (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 13(4):464-8.
- Sciacca FL, Ferri C, D'Alfonso S, Bolognesi E, Martinelli Boneschi F, Cuzzilla B, Colombo B, Comi G, Canal N, Grimaldi LM (2000) Association study of a new polymorphism in the PECAM-1 gene in multiple sclerosis. *J Neuroimmunol* 104(2):174-8.
- Scolding NJ, Morgan BP, Compston DA (1998) The expression of complement regulatory proteins by adult human oligodendrocytes. *J Neuroimmunol* 84(1):69-75.
- Sellebjerg F, Jaliashvili I, Christiansen M, Garred P (1998) Intrathecal activation of the complement system and disability in multiple sclerosis. *J Neurol Sci* 157(2):168-74.
- Serafini B, Rosicarelli B, Franciotta D, Magliozzi R, Reynolds R, Cinque P, Andreoni L, Trivedi P, Salvetti M, Faggioni A, Aloisi F (2007) Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *J Exp Med* 204(12):2899-912.
- Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorius H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, Heutink P, Aulchenko Y, Oostra B, van Duijn C, Jarvelin MR, Varilo T, Peddle L, Rahman P, Piras G, Monne M, Murray S, Galver L, Peltonen L, Sabatti C, Collins A, Freimer N (2006) Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 38(5):556-60.
- Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Rönnelid J, Klareskog L, Huizinga TW, Alm G, Syvänen AC, Rönnblom L (2007) Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis Rheum* 56(7):2202-10.
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52(3):506-16.
- Stamatovic SM, Dimitrijevic OB, Keep RF, Andjelkovic AV (2006) Protein kinase Calpha-RhoA cross-talk in CCL2-induced alterations in brain endothelial permeability. *J Biol Chem* 281(13):8379-88.
- Stoop MP, Dekker LJ, Titulaer MK, Burgers PC, Sillevius Smitt PA, Luiders TM, Hintzen RQ (2008) Multiple sclerosis-related proteins identified in cerebrospinal fluid by advanced mass spectrometry. *Proteomics* 8(8):1576-85.
- Sumelahti ML, Tienari PJ, Wikström J, Palo J, Hakama M (2000) Regional and temporal variation in the incidence of multiple sclerosis in Finland 1979-1993. *Neuroepidemiology* 19(2):67-75.
- Sumelahti ML, Tienari PJ, Wikström J, Palo J, Hakama M (2001) Increasing prevalence of multiple sclerosis in Finland. *Acta Neurol Scand* 103(3):153-8.
- Sundvall M, Jirholt J, Yang HT, Jansson L, Engström A, Pettersson U, Holmdahl R (1995) Identification of murine loci associated with susceptibility to chronic experimental autoimmune encephalomyelitis. *Nat Genet* 10(3):313-7.
- Svejgaard A (2008) The immunogenetics of multiple sclerosis. *Immunogenetics* 60(6):27586.
- Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, Jagodic M, Lobell A, Khademi M, Börjesson O, Lindgren CM, Lundman P, Brookes AJ, Kere J, Luthman H, Alfredsson L, Hillert J, Klareskog L, Hamsten A, Pielh F, Olsson T (2005) MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat Genet* 37(5):486-94.
- Szalai AJ, Hu X, Adams JE, Barnum SR (2007) Complement in experimental autoimmune encephalomyelitis revisited: C3 is required for development of maximal disease. *Mol Immunol* 44:3132-6.
- Takahashi K, Prinz M, Stagi M, Chechneva O, Neumann H (2007) TREM2-transduced myeloid precursors mediate nervous tissue debris clearance and facilitate recovery in an animal model of multiple sclerosis. *PLoS Med* 4(4):e124.
- Terwilliger JD, Ott J (1992) A haplotype-based 'haplotype relative risk' approach

- to detecting allelic associations. *Hum Hered* 42(6):337-46.
- The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426(6968):789-96.
- Thompson RA and Lachmann PJ (1970) Reactive lysis: the complement-mediated lysis of unsensitized cells. I. The characterization of the indicator factor and its identification as C7. *J Exp Med* 131:629-641.
- Tienari PJ, Kuokkanen S, Pastinen T, Wikström J, Sajantila A, Sandberg-Wollheim M, Palo J, Peltonen L (1998) Golli-MBP gene in multiple sclerosis susceptibility. *J Neuroimmunol* 81(1-2):158-67.
- Tienari PJ, Wikström J, Sajantila A, Palo J, Peltonen L (1992) Genetic susceptibility to multiple sclerosis linked to myelin basic protein gene. *Lancet* 340(8826):987-91.
- Tienari PJ, Sumelahti ML, Rantamäki T, Wikström J (2004) Multiple sclerosis in western Finland: evidence for a founder effect. *Clin Neurol Neurosurg* 106(3):175-9.
- Vang T, Miletic AV, Bottini N, Mustelin T (2007) Protein tyrosine phosphatase PTPN22 in human autoimmunity. *Autoimmunity* 40(6):453-61.
- Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts. *Curr Opin Genet Dev* 14(3):316-23.
- Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171(4356):737-8.
- Weerth SH, Rus H, Shin ML, Raine CS (2003) Complement C5 in experimental autoimmune encephalomyelitis (EAE) facilitates remyelination and prevents gliosis. *Am J Pathol* 163:1069-80.
- Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? *Nat Genet* 26(2):151-7.
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661-78.
- Wikström J, Palo J (1975) Studies on the clustering of multiple sclerosis in Finland I: Comparison between the domiciles and places of birth in selected subpopulations. *Acta Neurol Scand* 51(2):85-98.
- Wikström J (1975b) Studies on the clustering of multiple sclerosis in Finland II: microepidemiology in one high-risk county with special reference to familial cases. *Acta Neurol Scand* 51(3):173-83.
- Wray BN, Stankovich J, Whittock L, Dwyer T, Ponsonby AL, van der Mei IA, Taylor B, Dickinson J, Foote S, McMorran BJ (2008) CTLA-4 and multiple sclerosis: the A49G single nucleotide polymorphism shows no association with multiple sclerosis in a Southern Australian population. *J Neuroimmunol* 196(1-2):139-42.
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K, Blanchong CA, McBride KL, Higgins GC, Rennebohm RM, Rice RR, Hackshaw KV, Roubey RA, Grossman JM, Tsao BP, Birmingham DJ, Rovin BH, Hebert LA, Yu CY (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037-54.
- Yeo TW, De Jager PL, Gregory SG, Barcellos LF, Walton A, Goris A, Fenoglio C, Ban M, Taylor CJ, Goodman RS, Walsh E, Wolfish CS, Horton R, Traherne J, Beck S, Trowsdale J, Caillier SJ, Ivinson AJ, Green T, Pobywajlo S, Lander ES, Pericak-Vance MA, Haines JL, Daly MJ, Oksenberg JR, Hauser SL, Compston A, Hafler DA, Rioux JD, Sawcer S (2007) A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Ann Neurol* 61(3):228-36.
- Zhang Z, Duvefelt K, Svensson F, Masterman T, Jonasdottir G, Salter H, Emahazion T, Hellgren D, Falk G, Olsson T, Hillert J, Anvret M (2005) Two genes encoding immune-regulatory molecules (LAG3 and IL7R) confer susceptibility to multiple sclerosis. *Genes Immun* 6(2):145-52.
- Zhu X, Tang H, Risch N (2008) Admixture mapping and the role of population structure for localizing disease genes. *Adv Genet* 60:547-69.
- Zozulya AL, Wiendl H (2008) The role of regulatory T cells in multiple sclerosis. *Nat Clin Pract Neurol* 4(7):384-98.