From the DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

CONQUERING COMPLEXITY:

SUCCESSFUL STRATEGIES FOR FINDING DISEASE GENES IN MULTIPLE SCLEROSIS

Kerstin Imrell

Karolinska Institutet

Stockholm 2009

Published by Karolinska Institutet. Printed by [name of printer]

ISBN 978-91-7409-515-9

“Before you know, you must imagine.”

Nobel Laureate Richard Axel

ABSTRACT

Multiple sclerosis (MS) is a complex disease; at an intra-individual level several contributory factors cause the disease; at an inter-individual level different factors contribute to the disease. The aim of the thesis project was to reduce this complexity by focusing of subsets of patients that were likely to share to a higher extent contributory causes and to identify in these subsets genes conferring susceptibility.

In paper I we wanted to know the *HLA-DRB1* allele associated to the small fraction of MS patients lacking signs of immunoglobulin synthesis within the central nervous system (OCB-negative MS). In papers III and IV we performed whole-genome single-nucleotide-polymorphism (SNP) scans to identify genetic susceptibility regions in distantly related patients from an MS cluster in a parish in Värmland (paper III) and in a consanguineous family with several affected family members (paper IV). We also asked the question if differences in etiology are reflected in clinical parameters such as the severity of the disease, which we looked at in both papers I and II in relation to OCB status (I and II) and *HLA-DRB1* alleles (II).

Some of the main results in this thesis project and their congruence with previous reports of genetic susceptibility in MS are the following: the association between the OCB-negative entity and *HLA-DRB1*04* seen both in our population and in Japan; the potential importance of the *ACCN1* gene, identified in our distantly related MS patients and in an isolated population in Sardinia; the role of mutations in the *PLP1* gene on the X chromosome reported in two MS case reports, thus indicating the plausibility of monogenic X-linked MS.

The results derived from this thesis project merit follow-up.

LIST OF PUBLICATIONS

I. **Kerstin Imrell**, Anne-Marie Landtblom, Jan Hillert, Thomas Masterman.

Multiple sclerosis with and without CSF bands: Clinically indistinguishable but immunogenetically distinct. *Neurology 2006;67:1062–1064*

II. **Kerstin Imrell**, Eva Greiner, Jan Hillert, Thomas Masterman.

*HLA-DRB1*15* and cerebrospinal-fluid specific oligoclonal immunoglobulin G bands lower age at attainment of important disease milestones in multiple sclerosis. *Journal of Neuroimmunology 2009*

III. **Kerstin Imrell**, Thomas Masterman, Boel Brynedal, Izaura Lima Bomfim, Jerome Wojcik, Jan Hillert, Ingrid Kockum, Anne-Marie Landtblom.

Disease genes uncovered in 11 distantly related individuals affected with multiple sclerosis through SNP-based identical-by-descent heterozygosity mapping. *Manuscript*

IV. **Kerstin Imrell**, Izaura Lima Bomfim, Homayoun Roshanisefat, Marco Zucchelli, Benjamin Lamb, Jan Hillert, Ingrid Kockum, Thomas Masterman.

SNP-based gene mapping in a consanguineous multiple sclerosis family. *Manuscript*

CONTENTS

LIST OF ABBREVIATIONS

Ab	Antibody
BBB	Blood-brain barrier
Chr	Chromosome
CI	Confidence interval
CSF	Cerebrospinal fluid
EAE	Experimental animal encephalitis
ECTRIMS	European Committee for Treatment and Research in MS
EDSS	Kurtzke's Expanded Disability Status Scale
GWAS	Genome wide association study
HLA	Human leukocyte antigen
HR	Hazard ratio
IBD	Identical-by-descent
IBS	Identical-by-state
ISNI	International Society of Neuroimmunology
kb	Kilobases
LD	Linkage disequilibrium
LOD	Logarithm-of-the-odds
Mb	Megabases
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
NMO	Neuromyelitits optica
NPL	Non-parametric linkage
OCB	Oligoclonal immunoglobulin-G bands
OR	Odds ratio
PCR	Polymerase chain reaction
PLP	Proteo-lipid-protein
PMD	Pelizaeus-Merzbacher disease
SNP	Single-nucleotide polymorphism
SPG2	Spastic paraplegia 2

1 PREFACE

The title: I hoped for some readers apart from the most inveterate researchers, so I compromised with the scientific correctedness. A more scientifically appropriate title would have been: "Using reduced heterogeneity strategies to identify genetic susceptibility in multiple sclerosis." Hello, are you still awake?

The background: The background is directed to those not so "inveterate-researcher-in-this-narrow-little-field" readers. I have tried as simply as I can to take you through some rather complicated things (it took me half a decade to get the pieces together, so if you don't agree that these things are complicated –don't let me know!)

The main section: The main section is directed to other researchers in the field, who actually can start right in with the main section; thus, I chose to have this as the first chapter. If this book was a DVD box and the articles the four movies included, then the part in the synopsis about each paper could be described as "the story behind it and extra material".

Pleasant reading!

Yours sincerely,

Kerstin

2 MAIN SECTION

2.1 CONQUERING COMPLEXITY

2.1.1 Imagining complexity

A complex disease is, according to the definition I use, a disease which is both multifactorial and etiologically heterogeneous. I visualize these terms by the pie model. During my first year as a PhD student, my picture of complexity (although I was not yet familiar with the terminology) was extremely complex, involving differently sized spheres overlapping with each other in a system that required more than three dimensions. It was a picture hopeless to convey. But then, during my first course in epidemiology, I was introduced to the pie model of Rothman[1]: "Wow, this is a two-dimensional way of describing what I'm thinking off", I thought. I immediately adopted it and made it mine. Since then, the pies have been my trademark in our research group.

2.1.2 The pie model

To be affected by a complex disease one needs to have been exposed to several risk factors; these can be both genetic and environmental. For example, not everyone that smokes gets lung cancer, but let's say that if you smoke and have a certain genetic susceptibility and achieve a certain age you will be affected. To develop a disease you need a number of contributing causes, which by themselves do not cause the disease, but together make up a sufficient cause for developing the disease. The sufficient cause can be seen as a pie with many slices, where the slices are the contributing causes. In a strictly monogenic disease, you need only one slice in the pie; if a few slices are needed then it would be called a polyfactorial disease; however, most diseases are multifactorial.

Not all patients sharing a diagnosis will share the same set of contributing causes making up the sufficient cause. For a complex disease there exist several sufficient causes – or several pies. The number of slices in these pies may vary. One particular contributing cause may exist in several pies. Depending on the level of resolution (from

smallest genetic variations to signaling pathways), one may regard the number of pies for a very complex disease as being close to infinite, or at least close to the number of affected individuals. That makes complex diseases heterogeneous.

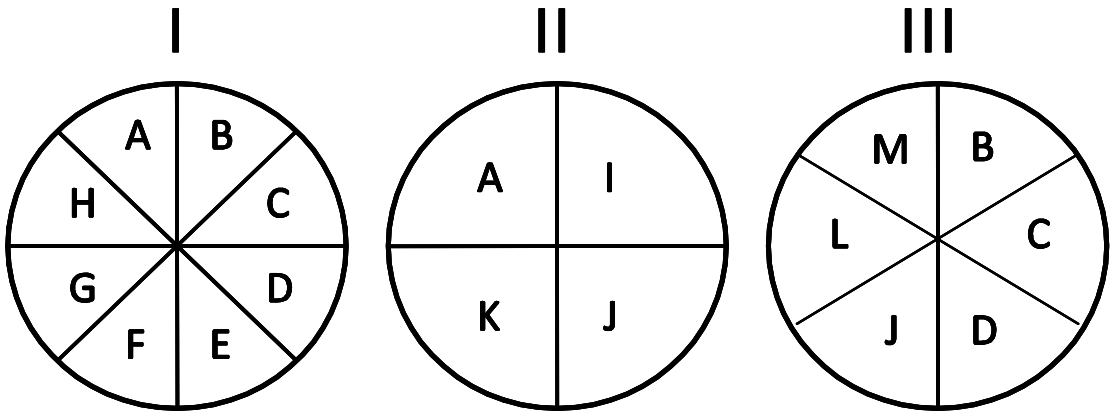


Figure 1. The pies I to III are all sufficient causes for developing a phenotype (disease). Each pie consists of a number of contributory causes: A to M. The number of contributory causes may vary between pies. The same contributory cause (pie slice) may be part of different pies: A,B,C,D and J. Unaffected individuals may carry several contributing causes as long as they do not complete the pie.

2.1.3 Finding the needle(s) in the haystack

"Disease is a lousy phenotype. …A disease is like the Mississippi river."

Joe Terwilliger, HGM Helsinki 2006

A consequence of the fact that several contributing causes are needed for developing a disease is that most individuals exposed for these contributing causes – genetic or environmental – will stay healthy. In other words, each contributing cause has only a small effect. One can also say that the penetrance of a risk factor is low. If an individual is exposed to a certain risk factor it may not even double his or her risk of developing disease. The approaches of the past few years in complex genetics, derived from the insight that researchers need to collaborate to achieve the large datasets of patients and controls needed to detect these effect sizes and the biotechnological advances in genetics, have made it possible to thoroughly investigate the whole genome for common variants associated with complex diseases. Such efforts have indeed shown some successes. But they have also demonstrated that in heterogeneous diseases such as MS there are no common variants with individual large effects. The common

variants associated with MS will change the individual risk of developing the disease from a background risk for a Swede of 1 in 500 (life time prevalence in Sweden) to 1 per 350 (based on a OR of 1.42 as seen for the best IL7R SNP[2]). So how does one find the needles in the haystack?

2.1.4 Increasing the signal-to-noise ratio

Influenced by my supervisors Thomas and Anne-Marie, I was early on convinced of the advantages of using small but more homogeneous sets of patients to find disease genes. Let us return to the pie model: looking at the whole heterogeneous MS population identifies separate pie slices. In this heterogeneous group of patients, some share more risk factors with each other than others, i.e. patients aggregate into different pies. The term endophenotypes are used in psychiatry. We started to use that concept for patients sharing etiology (belonging to the same pie or similar pies). By focusing on endophenotypes of a disease one concentrates one's efforts to finding slices in a small set of more similar pies on the assumption that a greater fraction of the patients in that endophenotype will share slices, i.e. the signal-to-noise ratio will increase. Looking at endophenotypes could potentially have two advantages: finding disease genes of large effect in that particular etiological fraction; such genes would not even necessarily have an effect above the detectable threshold in the whole patient population. The second advantage is that in a certain subset of patients we may directly detect more than one slice acting in the same pie.

This reasoning is theoretical, and includes a catch 22: we would like to use endophenotypes to identify susceptibility genes in an etiological fraction of patients, but we cannot really know if a group of patients belongs to the same endophenotype until we know whether they share etiology, which we will know once we have identified the risk factors.

2.1.5 Obtaining homogeneous groups in MS

Some group of patients can be assumed to share more of their etiology.

During the diagnostic work up a paraclinical test is routinely performed, which, if positive, is a sign of intrathecal immunoglobulinG synthesis. The test result is viewed as bands on an electrophoresis gel, and since the immunoglobins are derived from a small number of B-cell clones, a positive patient is said to have "the presence of

oligoclonal bands (OCB) in cerebrospinal fluid (CSF)". In Northern Europe, almost 95% of the patients are positive for OCB, compared to 10% of healthy controls[3]. In Japan, the MS prevalence is much lower than in Sweden, and less than 60% of the patients are OCB-positive[4; 5]. Assuming they are not simply the result of discrepant laboratory techniques, these differences in frequencies of OCB positivity between countries could reflect variations in the distribution of both genetic and environmental risk factors. The MS associated *HLA-DRB1*15* allele is less frequent among the Japanese patients than in most Western countries. However, Fukazawa et al.[4] published a study in 1998 showing that the frequency of the *HLA-DRB1*15* allele was the same in Japanese OCB-positive MS patients as in MS-patients in Western countries. They also showed that there was no association between OCB-negative western-type MS and *HLA-DRB1*15*, but that OCB-negative MS showed an association to the *HLA-DRB1*04* allele. Although the Japanese study was rather small, we were convinced that OCB-negative patients may be an endophenotype of MS and worth investigating in a Swedish dataset, which is what we did in paper I.

Some pies contain *HLA-DRB1*15* and cause OCB-positive MS; other pies contain *HLA-DRB1*04* and cause OCB-negative MS. The OCB status may be a marker for pathological events with different etiological backgrounds. Do patients belonging to different pies have the same prognosis? In paper II, we looked at the effect of *HLA-DRB1*15* and the effect of OCB status on age at onset and age at EDSS 6.0, which is the disability score when patients no longer can walk 100 meters without a cane.

In paper III, the included patients share a common ancestor and were brought up in the same parish. The study includes only 11 individuals, and they are connected through a common ancestral couple 10 generations back. The prerequisite for performing such a study is the assumption of homogeneity. We had no presumption about what the genetic background for the disease in these patients might be, and therefore the idea was to scan the whole genome and look for segments shared identical-by-descent. The low degree of relatedness between the patients increases the probability that a shared segment is due to the shared MS phenotype.

Paper IV also includes only 11 individuals, but the presupposition is completely different from that in paper III since they are all belong to the same family and the pedigree contains two loops of first cousin marriages. The study included 7 affected

individuals and 4 unaffected. This family may display a monogenic form of MS. Our research group previously performed a linkage study on this family using around 800 microsatellites distributed over the whole genome. This generated a promising peak on chromosome 9 and a modest peak on the X chromosome[6]. The reason for reinvestigating this family is that new cases have been identified and the new techniques allow us to scan the whole genome with high-throughput microarray-based genotyping of densely spaced SNP markers while at the same time fine-mapping the previously reported peaks.

2.2 THESIS OBJECTIVE

By studying less heterogeneous groups of patients, and thereby reducing the complexity of disease, I aimed at a better understanding of genetic susceptibility to MS.

2.3 PAPER I

Multiple sclerosis with and without CSF bands: Clinically indistinguishable but immunogenetically distinct

Our objective was to determine whether, in Sweden, patients with OCB-positive and OCB-negative MS constitute distinct subpopulations, clinically and immunogenetically.

The thesis by Thomas Masterman included the extensive work of going through the clinical records for all patients included in studies of MS genetics in our group. This work opened up new possibilities for studies, including this one. This was my very first real scientific work, and my main supervisor Thomas was devoted to teaching me everything about good scientific policy.

As a first step, we used Thomas's database to identify patients of Scandinavian ethnicity, fulfilling the consensus diagnostic criteria[7], and for whom information about CSF status was available. From 1505 MS patients we identified 83 OCB-negative patients; thus, we attained a clinic-based frequency of OCB-negative patients of 5.5%. Our first question was whether the MS patients with and without OCB shared the same clinical features, so we compared the sex ratio, the frequency of primary-progressive MS, the frequency of patients fulfilling the MRI criteria included in the McDonald diagnosis[7-9] and the severity. The last parameter was measured using the MS severity score (MSSS)[10], an algorithm which allows cross-sectional comparisons of disability.

Inspired by the study of Fukazawa et al.[4], and the follow-up paper by Kikuchi et al.[5] we wanted to explore the *HLA-DRB1* allele carrier frequencies in these two subgroups. Most low-resolution HLA genotypes were already available in our lab but I did some of them myself, using the same methodology (Olerup SSP AB, Saltsjöbaden, Sweden). This is a method where each individual is genotyped with 24 polymerase-chain-reactions (PCR) to reveal their serological specificity. The first results showed the same thing as those in the Japanese studies: no association between OCB-negative MS and *HLA-DRB1*15*, but instead an association between this subgroup and *HLA-DRB1*04*. We therefore went further and I genotyped the

*HLA-DRB1*04* positive patients and controls with higher resolution to reveal their genomic alleles (Olerup SSP AB, Saltsjöbaden, Sweden).

The result of the clinical characterization is presented in table 1. No differences reach a significant level of $p<0.05$. One question regarding the group of OCB-negative patients is the possibility that it is diluted by cases that are not truly MS. The clinical similarities between these two groups in this study may indicate that our use of strict criteria for inclusion has prevented the dilution by false diagnosis. This shows that although patients (presumable) belong to different etiological pies they share clinical features; thus, different etiological pies may possibly lead to the same pathological pathway. In paper II we reinvestigate the clinical parameter with the greatest impact for the patient – the prognosis. But instead of using cross-sectional data, we used longitudinal data for disability, data which were not yet available at the time of the first study.

Table1. Clinical characteristics of patients with and without oligoclonal bands (OCB).

	OCB-pos MS n=1422[*]	**OCB-neg MS** n=83[†]	*p*
Men/women	1:2.6	1:2.0	0.26
Mean age at onset, *y*	32.1	33.6	0.17
Primary progressive %	7.5	8.4	0.67
Positive MRI %	77.3	74.1	0.62
Mean MS Severity Score	4.80	4.89	0.75

[*] MRI, n=917; MSSS, n=1404; [†]MRI, n=54; MSSS, n=82

Table 2 shows the result of both the *HLA-DRB1* low resolution genotyping as well as the suballele genotyping of *HLA-DRB1*04*-positive individuals. In the first line, it is the absence of association between OCB-negative MS and *HLA-DRB1*15* that is the interesting result. The association between *HLA-DRB1*04* and the OCB-negative entity is due to the sub-allele DRB1*0404 which among carriers increased the risk of being affected by OCB-negative MS four-fold.

Table 2. Odds ratios (OR) for HLA-DRB1 risk alleles in patients with multiple sclerosis with and without OCBs.

	OCB-pos MS		**OCB-neg MS**	
DRB1	OR	p	OR	p
*15	3.5 (2.3-5.4)	0.0001	1.7 (0.9-3.0)	0.09
*04	1.1 (0.7-1.7)	0.7	2.1 (1.2-3.8)	0.01
*0401	0.8 (0.5-1.4)	0.5	1.1 (0.5-2.3)	0.8
*0404	1.2(0.6-2.6)	0.7	4.3 (1.9-9.7)	0.0008

Carriage frequencies are compared to ethnically matched blood-donor controls.

Figure 2 shows the carrier frequencies of *HLA-DRB1*15* and *HLA-DRB1*04* in Swedish and Japanese OCB-positive MS patients, OCB-negative MS-patients and controls. For the *HLA-DRB1*15* alleles the pattern is strikingly similar. *HLA-DRB1*04* seems to be more frequent in Japan but the distribution among the groups correlates with the distribution in our dataset.

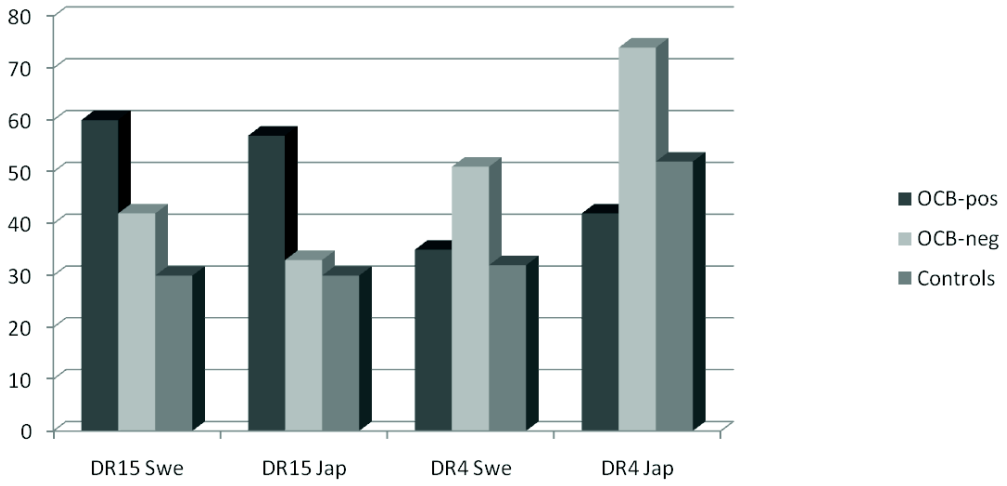


*Figure . Carrier frequencies of HLA-DRB1*15 and HLA-DRB1*04 in Swedish and Japanese OCB-positive MS-patients, OCB-negative MS patients and controls.*

The Japanese follow-up study[5] also included results regarding the *HLA-DRB1*04* suballeles. The association between *HLA-DRB1*04* and OCB-negative MS are, in the Japanese study, explained by the suballele *HLA-DRB1*04*05, while it is explained by *HLA-DRB1*04*04 in the Swedish patients. *HLA-DRB1*04*05 is uncommon among ethnically Swedes and the opposite holds for *HLA-DRB1*04*04 in Japanese, but the two alleles are closely related in terms of genetic sequence.

Although *HLA-DRB1*04* has not previously been shown to be associated to MS in Sweden or other Nordic countries, it seems to have some effect on the disease risk in Mediterranean countries[11-13]. In the autumn of 2004, I attended my two first conferences ever. My abstract regarding my preliminary findings in this study had been selected for oral presentation both at ISNI in Venice and ECTRIMS in Vienna. I was eager to meet and discuss my data with Professor Marrosu from University of Cagliari in Sardinia, one of the chairs of the ISNI session. After the session I went back on stage again and confronted Professor Marrosu with my question: "*HLA-DRB1*04* is associated to MS in Sardinia, *HLA-DRB1*15* is very rare in the Sardinian population. It would therefore be logical if the proportion of OCB-negative MS is relatively high in the Sardinian patients." A few months later, Professor Marrosu's postdoc Elenora Cocco provided us with preliminary data answering my question: around one fourth of Sardinian MS patients lack oligoclonal bands.

I would like to believe, based on the collected evidence from my study, the Japanese studies and the Sardinian data that OCB-positive and OCB-negative entities are the same across different populations; but their distributions differ between populations, reflecting the genetic, and perhaps also the environmental, national background. For me it does not seem unlikely that these entities can be broken down to even smaller fractions, whose distribution would also depend on the genetic backgrounds of the population.

In conclusion, we here showed that MS patients with and without OCBs are indistinguishable regarding the investigated clinical features. However, they constitute two different immunogenetic entities. Even Jan was convinced.

2.4 PAPER II

HLA-DRB1*15 and cerebrospinal-fluid-specific oligoclonal immunoglobulin G bands lower age at attainment of important disease milestones in multiple sclerosis

Our objective was to determine the extent to which carriage of HLA-DRB1*15 and the presence of OCB influence the age at attainment of two important MS milestones: the advent of clinical symptoms; and, as an index of the longitudinal accumulation of disability, an EDSS score of 6.0 (corresponding to the inability to walk 100 m without the assistance of unilateral support).

The first task in this project was to obtain survival data and other clinical parameters from the Swedish MS Registry (SMSreg). In SMSreg, patients' EDSS scores are entered at each clinical visit and that was our basis for a transformation to survival data. Instead of going through all the patients in the registry, we wanted to extract data it in an automatic way, which forced us to set up strict rules and follow them. The Swedish MS Registry started a decade ago and its use has accelerated. It is the clinicians who enter the information into the database. Some patients have been in the database from onset, and others enter the database at an advanced stage of disease. Thus it took us a while to figure out the rules. A number of patients overlapped with the clinical database that Thomas made during his thesis project, which was based on the survey of clinical records for all patients who had donated blood for genetic studies. We went through all the discrepancies between these two databases to find explanations. Combining the two databases we could finally, after many discussions and a lot of work, proudly present a database with survival data regarding duration and age at EDSS 6.0 for 2094 patients fulfilling the diagnostic criteria and for which we also had CSF status. Of these, we had genotyping results for the *HLA-DRB1* locus for 1488 patients.

Thereafter the discussion started regarding what possibly could confound and potentially interact with our variables of interest: *HLA-DRB1*15*, *HLA-DRB1*04* and OCB, with regard to their effect on age at attainment of EDSS 6.0.0 and onset. We used Mantel-Cox to assess these issues and defined interaction as a test of equality ≤ 0.1. In the absence of interaction we looked for confounding, which we defined as the presence of a variable in the model that changed the HR ratio of the variable of interest

with more than 10%. Since the two databases had different time frames, we needed to control for source of information. The main effects of the variables in the analysis are presented in table 3.

Table 3.The main effects of the investigated variables on age at attainment of EDSS 6.0

Cox regression age at EDSS6 main effects
Adjusted for source of information

	HR	95% CI	p	Comment
The presence of OCB	2.08	1.30-3.32	0.002	Main variable of interest
HLA-DR15 carrier	1.39	1.16-1.67	0.000	Variable of interest
HLA-DR4 carrier	1.05	0.86-1.27	ns	Variable of interest, interacts with OCB
Male sex	1.08	0.90-1.31	ns	interacts with ocb
Residency south Stockholm	1.39	1.15-1.68	0.001	interacts with ocb in males
Scandinavian etnicity	0.64	0.49-0.84	0.001	confound ocb

The statement of confounding and interaction are based on analysis with Mantel -Cox.
Interaction was consider if test of equality p=<0.1
Confounding was consider in absent of interaction if a strata or
the combined HR differed more than 10% from OCB crude HR

If you read the paper, you might notice that this table says another thing than we are stating in the paper. That is because the assessment of interaction and confounding was performed in the maximum dataset obtainable for the variable of interest and the variable under investigation, meaning sometimes the dataset of 2094 patients and sometimes the dataset of 1488 patients. When we realized that HLA and OCB interacted we had huge discussion regarding whether we then needed to exclude the patients for which we did not have *HLA-DRB1* genotypes. Otherwise it would have been impossible to control for the impact of *HLA-DRB1* risk alleles of the effect of OCB status. Actually the genotyped and the ungenotyped patients differed in several ways, one being that ungenotyped patients were more frequently OCB-negative. A probable explanation was the fact that patients who more frequently visited the clinic were both more likely to have undergone a second or third lumbar puncture and more likely to have donated a blood sample for genotyping. We also suspected that the ungenotyped patient group could be diluted with misdiagnosed patients. Thus, in the end, we excluded patients without *HLA-DRB1* genotypes.

After excluding the ungenotyped patients the interaction effects presented in table 3 disappeared and thus the model presented in table 4 includes interaction terms that no

longer fulfilled the criteria for being allowed in the model (which is possibly a debatable issue). However, I've chosen anyway to show you this table since it provides some data that may be interesting to have a look at, although it should not be given too much credibility. It is based on the dataset of 1488 patients.

Table 4. The hazard ratios for age of attainment of EDSS 6.0 in male OCB-positive patients differ dependent of area of residency in Stockholm.
Cox regression model adjusted for all in the table included variables and source of information.

	HR	95% CI	p	Interpretation
The presence of OCB	1.78	0.98-3.25	0.061	The effect of ocb in females
HLA-DR15 carrier	1.40	1.17-1.68	0.000	
HLA-DR4 carrier	1.01	0.83-1.23	ns	
Male sex	0.79	0.31-2.00	ns	The effect of male sex in ocb negatives
Residency south Stockholm	1.18	0.95-1.47	ns	The effect of residency above the effect of male sex and OCB
Scandinavian etnicity	0.65	0.50-0.86	0.002	
*OCB*sex*	0.93	0.34-2.54	ns	HR OCB posive male in north Stockhom: 1.78*0.79*0.93=1.31
*OCB*sex*residency*	1.80	1.16-2.78	0.008	HR OCB posive male in south Stockhom: 1.31*1.80=2.36

One very plausible explanation for the effects seen in table 4 is that the OCB patients constitute a small fraction and that male patients are also underrepresented; thus the strata comprised of OCB-negative males is very small. On the other hand, regarding the effect of residency, which in table 3 is shown to have a main effect of the same size as *HLA-DRB1*15* carriage, this may be a true effect since the southern part of Stockholm consists of more areas of low socioeconomic status than the wealthier northern part of town. And socioeconomic status is a well recognized variable affecting health. A study in Läkartidningen[14] investigated the capacity of working until the retirement age of 65 in women in Stockholm. In that study, the authors made a rough classification of the areas south of Slussen (which is located in the middle of Stockholm) as being of low socioeconomic class and areas north of Slussen as being of high socioeconomic class. They went on to show that among the highly-educated women residency had a great impact on sick leave before the age of 65 (OR 3.2; 95% CI 1.3-7.8). Our effect with residency interacts with being a male patient. Living conditions in most families are − due to unequal salaries in the sexes − more commonly dependent of the male income, thus potentially the conditions are different for male and females with MS in these

areas with the causality going either way; i.e., the change of income in a family due to sickness may have an impact of the choice of living area. In the end, these results should be viewed with great caution since the study was not designed to answer the question of the effect of residency. Therefore, we chose to exclude this result from the paper, although residency among other confounders is a thing we are controlling for in the models in the publication.

Let us now focus the discussion on the questions we hoped the answer by the study design. Our main results, expressed as ages at onset and at attainment of EDSS 6.0, are shown in table 5 and 6. The use of age instead of duration allowed us to also investigate the effect of the variables of interest on age at onset.

Table 5.

Median age at onset			
	Onset age	95% CI	subject (n)
Absence of both OCB and *HLA-DRB1*15*	32	28-37	55
Absence of OCB, presence of *HLA-DRB1*15*	32	29-36	29
Presence of OCB, absence of *HLA-DRB1*15*	32	32-34	580
Presence of both OCB and *HLA-DRB1*15*	30	29-30	824

Table 6.

Median age at attainment of EDSS 6.0	
	age
Presence of *HLA-DRB1*15*	57
Presence of OCB	58
Presence of both OCB and *HLA-DRB1*15*	57
Absence of *HLA-DRB1*15*	61
Absence of both OCB and *HLA-DRB1*15*	67
Absence of both OCB and *HLA-DRB1*04*	67
Absence of OCB, presence of *HLA-DRB1*04*	58

In the Cox regression model including all variables of interest, the effect of carriage of *HLA-DRB1*15* has a significant effect on age at onset (presented in paper II). The

effect of OCB positivity on age at onset, over and above the effect of *HLA-DRB1*15*, is zero (HR=1.00).

Regarding age at attainment of EDSS 6.0, OCB positivity has a greater impact than *HLA-DRB1*15* carriage, however the hazard at each age of attaining EDSS 6.0 for OCB negative *HLA-DRB1*04* carriers is the same as that of OCB positive patients.

During the work with this project I've acquired a deeper understanding of the features of our patient dataset. Looking at parameters such as progression makes it much more important to remember that these patients are ascertained through the neurology clinic. The most severe cases will receive care elsewhere and the most benign cases will not been seen at the clinic very often. Connecting back to the pie concepts – do patients not available in our dataset belong to the same pies as the ascertained ones? Does using a clinic-based dataset cause a distribution of subentities that has an impact not only on progression and clinical characteristics but potentially also on which genetic susceptibility variants we detect?

2.5 PAPER III

Disease genes uncovered in 11 distantly related individuals affected with multiple sclerosis through SNP-based identical-by-descent heterozygosity mapping

The ultimate objective of this study was to locate disease genes in MS. A corollary objective, however, was to achieve this goal by combining a cutting-edge technique−high-throughput microarray-based genotyping of densely spaced SNP markers−with a bold epidemiological idea: that it should be possible to find susceptibility genes in a complex disease by analyzing a small number of patients of common, but distant, ancestry. Together, the technique and the idea would, first, deliver a dataset of subjects in whom MS was less heterogeneous than in the background population; and, second, allow identification of segments shared identical-by-descent between subjects homozygous or subjects heterozygous for each segment.

Lysvik was previously identified as an area with a high prevalence of MS[15-17]. Genealogical research has enabled us to link 22 patients to a common ancestral couple that settled in this parish in the beginning of the 17th century. This parish has a dramatic demographic history with periods of rapid increases in population size and rapid decreases as a consequence of emigration.

The manuscript of this paper is quite extensive and contains all the details of the analysis we did concerning the MS cluster in Lysvik. I will instead use this section to present the story of this project and then a simplified overview of the analytical steps. The flowchart of the work behind this study may give you the idea that this was a straightforward, easily performed study. It was the opposite. This is by far the most complex study in this thesis project and I spend at least half of my time as a PhD student on this project. There were many occasion when I thought, this is not working, this is a complete flop. At the same time I had to convince myself as well as my supervisors that I should continue with this project. I owed it to the patients and to all people engaged in this project, nurses and clinicians in Värmland, our fantastic genealogists, especially Arne Linnarud and his wife Moira whom I have visited several

times in Karlstad, and not least Anne-Marie who with her great enthusiasm has spent so much research time on this MS-cluster.

It seemed very straightforward at first. My supervisors and I consulted the famous Swedish/Finnish genetic researcher Juha Kere and some of his group-members on account of their expertise before we had made any decisions about the design. I knew already that I strongly believed that unaffected individuals in this pedigree also would carry risk alleles. An affected-only design has been applied with success in recessive disorders[18-20] using microsatellites. But my hypothesis was that these patients had more than one contributing cause for MS. The conclusion from this initial meeting was to take the most distantly related patients in the pedigree, genotype their whole genome with densely spaced SNP markers (which could be, and was, performed using cheap microarray-chips) and then look at segmental sharing. Easy? This meeting was held on March 29, 2006.

My frustration in this project peaked in January 2008. We had a whole-day genetic meeting within the research group, at which I presented a mindmap of all problems I was confronting. Although I don't think the problems would have been solved using microsatellites, most of the problems were related to the SNP markers. The reason for not being better off with microsatellites is their high error rate and their high mutation frequency which may have caused some serious problems since the study individuals are related to one another through a common ancestor 10 generations back. In addition, microsatellites are expensive to genotype. The good thing with microsatellites though is that they are very polymorphic and the same marker contains multiple variants, making them informative. SNPs are not very informative, since they most commonly are biallelic. On the 250k chip we used, a high proportion of SNPs are not polymorphic at all in the Swedish population (at that time I could not know the extent of this problem, just suspect it). Obviously, this would give me regions that were shared homozygously among all affected individuals, but without having anything to do with shared ancestry. Another problem is that the genome is transmitted through the generations in chunks containing several SNPs; the physical distance but also other important features of the genome makes certain loci more prone to be inherited together than other. The measurement of this is called *linkage disequilibrium* (LD). The extent of LD between the SNP markers on the chip will make some areas more prone than others to contain long segments of consecutive SNPs to be shared among the patients. However, we

were not interested in regions caused by the background LD pattern in the genome but in regions that are truly (presumably) shared from the common ancestor. Inspired by a recent publication about runs of homozygosity in schizophrenia patients[21], I made a naïve attempt to look for homozygous sharing using Excel. I also wanted to look for heterozygous sharing, but with the unphased chromosomal data I had it seemed undoable – at least in Excel.

Two meetings shortly after I presented the "mindmap of problems" had a decisive impact on the completion of this project. My naïve attempt had generated some loci of interest, and coincidently one of these corresponded exactly to the result of the search for MS susceptibility regions in a Finnish subisolate; the Finnish data were presented at the Nordic collaborative meeting. I should add that the Finnish subisolate was settled by emigrants from the same Finnish region as the one our ancestral couple emigrated from. After the formal meeting I sat down with Janna Saarela and Eveliina Jakkula from the Finnish group and we discussed our respective studies. Their approach was completely different from mine; they even used another genotyping platform. It was striking how we both could get exactly the same region, not larger than half a megabase (the refined analysis in both our and the Finnish study has however exluded this region as being of interest). I also took the opportunity to discuss the issues I was confronting in my project and this was how I was introduced to the "Segmental sharing" analysis in Plink[22; 23] (a feature that still is in its beta development stage).

A couple of weeks later I was in Paris with Jan and Thomas and presented my naïve analysis for a group led by Professor Françoise Clerget-Darpoux. Their expertise is statistical genetics with a focus on complex neurological diseases. After my presentation, Françoise, Catherine Bourgain and Anne-Louise Leutenegger stayed and discussed my project. It was a luxury to get the chance to discuss this with such competent and intelligent women. They were rather skeptical, pointing at the same weaknesses that I was very aware of myself and they strongly questioned the use of an affected-only design. I had started to think of that I might make use of some data we had available in our group: 664 sporadic cases and controls genotyped with the same microarray chip for the purpose of another study. These genotypes could be used to get rid of some of the problems related to the SNPset on the chip. I tested this idea on them, and received some positive responses.

In a very short time I had got the solution for many critical issues. It still was not straightforward to just do it. To handle almost a quarter of millions of data-points each for 664 individuals by far exceeds the limits of Excel. I had to learn to use a Linux and the program *R*, writing codes in the terminal window, etc. But I enjoyed it!

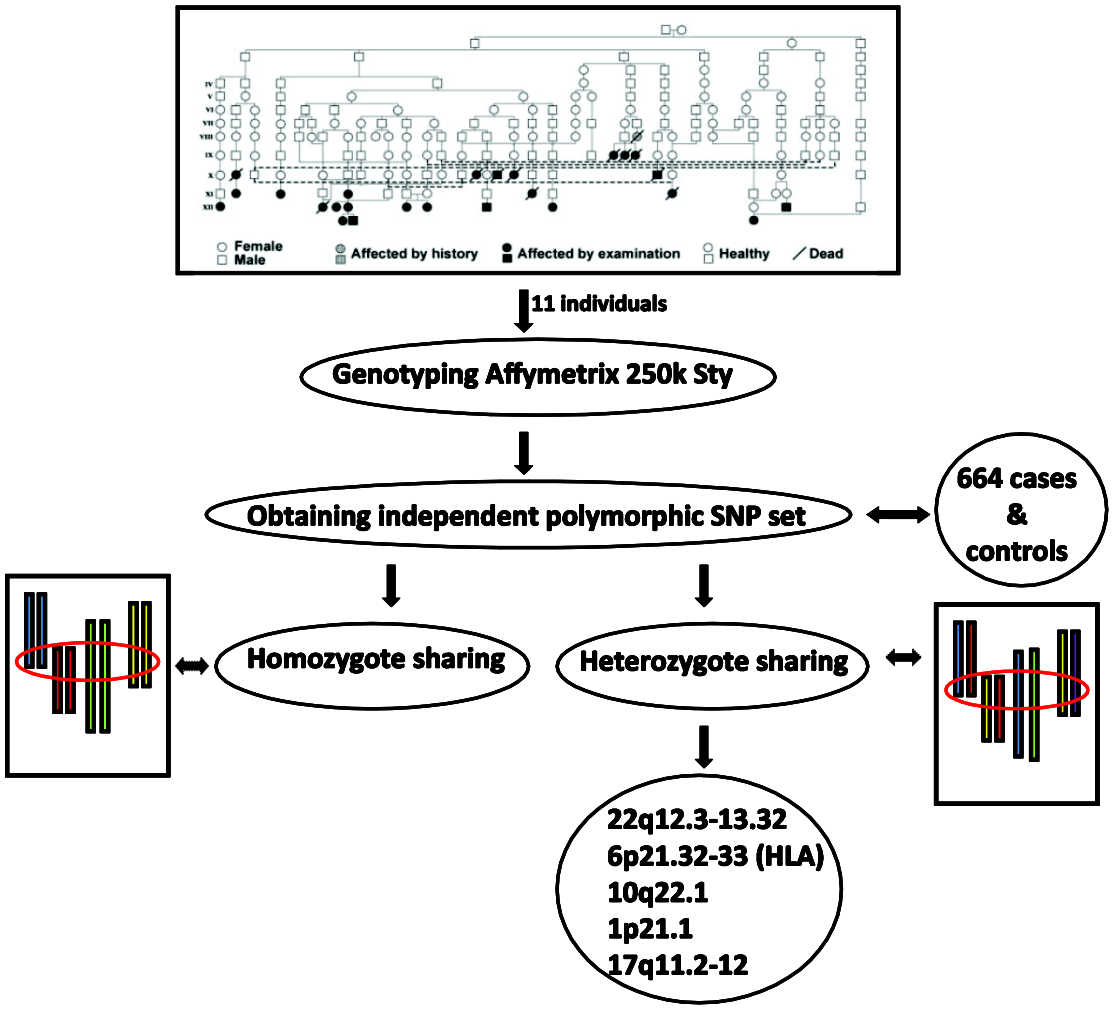


Figure 3. Flowchart of study III. The pedigree is published with the courtesy of Acta Neurologica Scandinavica[17].

Figure 3 shows an overview flowchart of the steps in this project. We used an affected-only design, selecting 9 distantly related individuals to be included in the study as well as the affected children to one of them. The individuals were genotyped on the Affymetrix platform using a microarray chip which generated almost 250,000 genotypes per individual. In our group we had genotypes from the same microarray-chip available for 664 sporadic cases and controls, which we used to obtain a SNPset of polymorphic independent SNPs. The LD pruning step was performed in Plink 1.05[22; 23] and excluded SNPs in pairs with an r^2 greater than 0.5. The marker map generated after

these procedures was then used for detection of segmental sharing in the Lysvik patients. The concept of segmental sharing is the identification of chromosomal regions that are shared identical-by-descent (IBD; derived from the same ancestor). Since we cannot know for sure if a region is shared identical-by-descent, we have to make assumption about that from the identity-by-state (IBS; sharing of an allele). A paper about runs of homozygosity in European populations[24] from last year indicated that segmental sharing of a length of 1500 kb indicates parental relatedness; this length may be too conservative for our purposes but we adopted it.

The analysis of homozygous sharing was performed in all eleven individuals. First, runs of homozygosity within individuals were identified and overlaps of these runs between individuals were noted. The analysis was performed in Plink 1.05 and so were also the heterozygosity mapping.

In the heterozygositye mapping, the algorithm first identifies segmental sharing by comparing all individuals pairwise, and then overlaps of pairwise runs of sharing are detected between the pairs. We included only the nine distantly related individuals since the inclusion of a trio would have skewed the results, although we also performed a second run to see if the affected sibling pair shared with the others, the same regions as reported from the first run. For the probability calculation of sharing we used a formula[25] which took into account the relatedness of the patients and the fact that the whole genome was searched. We also looked for sharing in the case-control dataset for the purpose of detecting false positive results due to specific features of that genomic region.

The homozygosity sharing revealed no regions that overlapped in more than two individuals. The heterozygosity sharing on the other hand revealed 5 chromosomal regions of interest which were shared among at least 4 of the 9 distantly related individuals: chromosome (chr) 22q12.3-13.32, chr 6p21.32-33 (the HLA region), chr 10q22.1, chr 1p21.1 and chr 17q11.2-12 (inside the *ACCN1* gene). The sharing of the HLA region in the case-control dataset was at the same magnitude as among the Lysvik pairs, thus the sharing of this region in the Lysvik patients may have nothing to do with shared ancestry. An evaluation of the methodology is the congruence of the results with what we know about MS pathogenesis and previously reports of MS susceptibility.

This is thoroughly discussed in the discussion part of the manuscript; here I will just mention what I regard as the most promising findings from this study.

When I went through the genes within the derived regions of interest, I obviously thought that the ACCN1 gene, which encompassed the chr 17 hit region, seemed interesting. I was astonished when I did a Pubmed search and found that it two years earlier had been described as an MS-associated gene in a Sardinian isolate[26], and that mutations in this gene could be involved in neurodegeneration[27; 28]. I continued to read about the protein, acid-sensing ion channel 2 (ASIC2), which is encoded by the gene and came across two publications from a French group[29; 30] , the first one describing the ASIC partner protein PICK-I and the second showing that PICK-I is stimulated by protein kinase C (PKC), presumably the α version. Figure 4 is adapted from one of these papers. PKCα is encoded by the *PRKCA* gene which was previously shown to be associated with MS[31] in a Finnish and a Canadian population; this candidate gene was selected on the basis of findings in a Finnish isolate. I was surprised to find that the gene encoding PICK-I, *PICK1* was located in the chromosomal region that came up on chr 22. All the genomic regions containing these genes have also been reported in MS linkage studies[16; 32-34]. The ASICs are pH-sensitive sodium-channels, and when certain neuropeptides bind their G-coupled receptors a signal cascade modulates these channels and changes the neuronal excitability. ASIC2 is expressed abundantly in CNS. Thus this pathway not only shows congruence with previous reports of MS susceptibility but also makes biological sense.

Another gene making biological sense is to be found in the consensus region of chromosome 10. This hit segment contained 7 genes and actually three of them have previously been described as associated with MS[35-37]. One of these, the *PRF1* gene, caught my attention both since it was associated in a combined cohort summing to almost 3000 cases and controls and in addition the authors identified an association between mutations in this gene and MS. On top of that, the protein encoded by this gene, perforin, has been shown to be involved in MS pathogenesis in a number of studies[38-40].

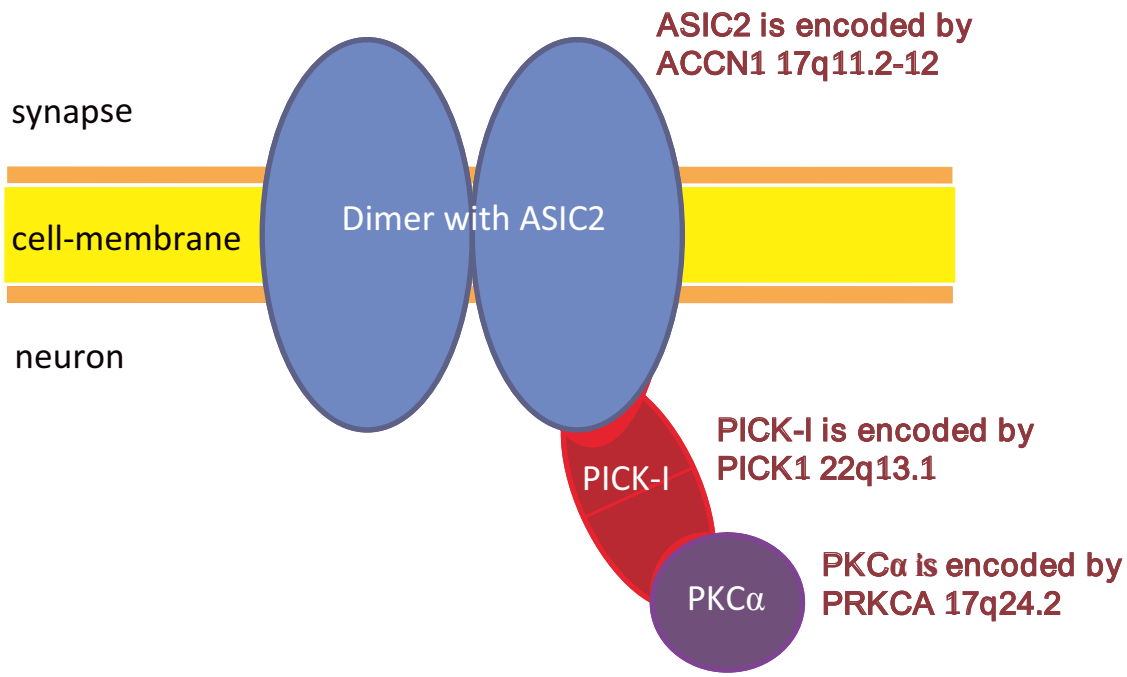


Figure 4. ASIC2 forms dimers with other ASICs. The ion channel is modulated by PICK-I, which interacts with PKCα. The genes encoding ASIC2 and PKCα, ACCN1 and PRKCA have previously shown association toMS[26; 31]*. The genomic regions containing ACCN1,PRKCA and* PICK1 *have all shown some evidences of linkage in MS*[16; 32-34]*. A study in nematodes shows that mutations in ACCN1 potentially could cause neurodegeneration*[27; 28]*. Adapted figure*[30].

Gene-environment interactions are thought to underlie some of the disease mechanisms of MS. Interestingly, the parish of Lysvik appears on the Swedish Environmental Protection Agency's list of 38 areas plagued by hazardous pollution; the sawmill in Lysvik closed in 1967, but earlier this decade, it was discovered that it was contaminating the municipal drinking water with pentachlorophenol. The degree of contamination during previous decades remains unknown. The neurotoxic effect of heavy metals can be mediated through ASIC channels[41-46], and phenols have been found to modulate synaptic transmission in the CNS[47].

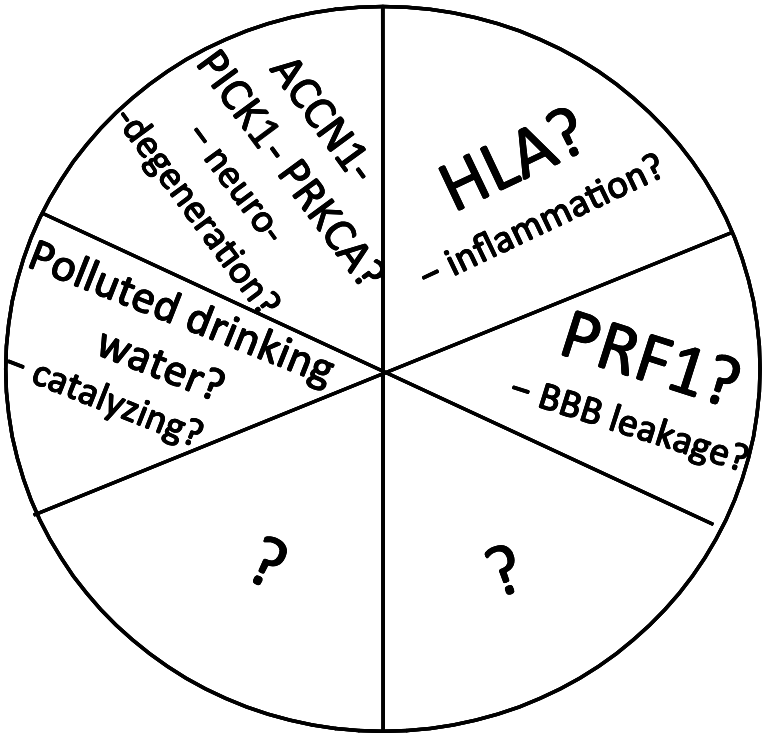


Figure 5. Potentially, we have identified several contributing causes of disease in this small but presumable homogeneous set of MS patients. These candidate genes make sense both with regard to previous studies of genetic susceptibility in MS and with regard to functional studies of the MS pathogenesis and neurodegeneration. These patients may also display other, as yet unidentified, contributing causes.

Certainly there may be more genes to explore under the consensus regions. It is a dangerous, but understandable, thing to always connect your findings to what is previously known; but by doing that one may really miss the real breakthrough. Many studies in complex genetics have employed far-fetched ways of trying to fit their candidate gene into a picture making biological sense. I do not think the involvement of the *ACCN1* gene and the *PRF1* gene in the pathogenesis of MS is far-fetched. However, I'm not convinced that we have uncovered the complete genetic effect in this set of patients, but we may well have identified some important pieces of information, which are summarized in figure 5.

In conclusion, bold ideas have their price and their rewards.

2.6 PAPER IV

SNP-based gene mapping in a consanguineous multiple sclerosis family

Taking advantage of the tremendous technical advances in genotyping that have occurred since the last report on this family[6] we chose to use a dense single-nucleotide-polymorphism (SNP) map to reinvestigate the whole genome, while including both previously investigated family members as well as two additional cases. At the same time, we were able to fine-map the previously reported linkage peaks in a cost-effective way.

This study investigates a seemingly Mendelian MS family with two first-cousin marriages and seven affected family members (pedigree in flowchart in figure 6). In 2003 our group published a whole-genome microsatellite scan on which a non-parametric linkage analysis was performed. The analysis included 9 family members, five of whom had an MS diagnosis, and revealed a peak on chromosome 9 with a LOD score of 2.29 (p=0.0009) and a modest peak on the X chromosome with a LOD score of 1.7[6]. We wanted to reinvestigate this family after two additional affected members were brought to our attention, with new genotyping techniques allowing for both fine-mapping of previous peaks and a whole-genome search at once. We also wanted to perform not only non-parametric linkage analysis but also parametric analysis testing for different models, an autosomal-recessive model being the a priori hypothesis in regard of the pedigree structure. We also wanted to include the complete pedigree with the two loops in the analysis. Figure 6 shows a flowchart of the steps included in the study.

Our study did not support the previously reported peak on chromosome 9. The only chromosome showing good evidence of linkage in this study was the X chromosome; however, the affection status of one of the sisters in the pedigree is uncertain since she has neurological symptoms but failed to show up for a MRI scan and lumbar puncture, which could have confirmed an MS diagnosis. The affection status of this sister has a great impact on which model of inheritance and which locus at the X chromosome to believe in (figure 7). Figure 8 shows the estimated haplotypes in each individual. Table 7 shows the LOD score obtained with the different models under the two peaks.

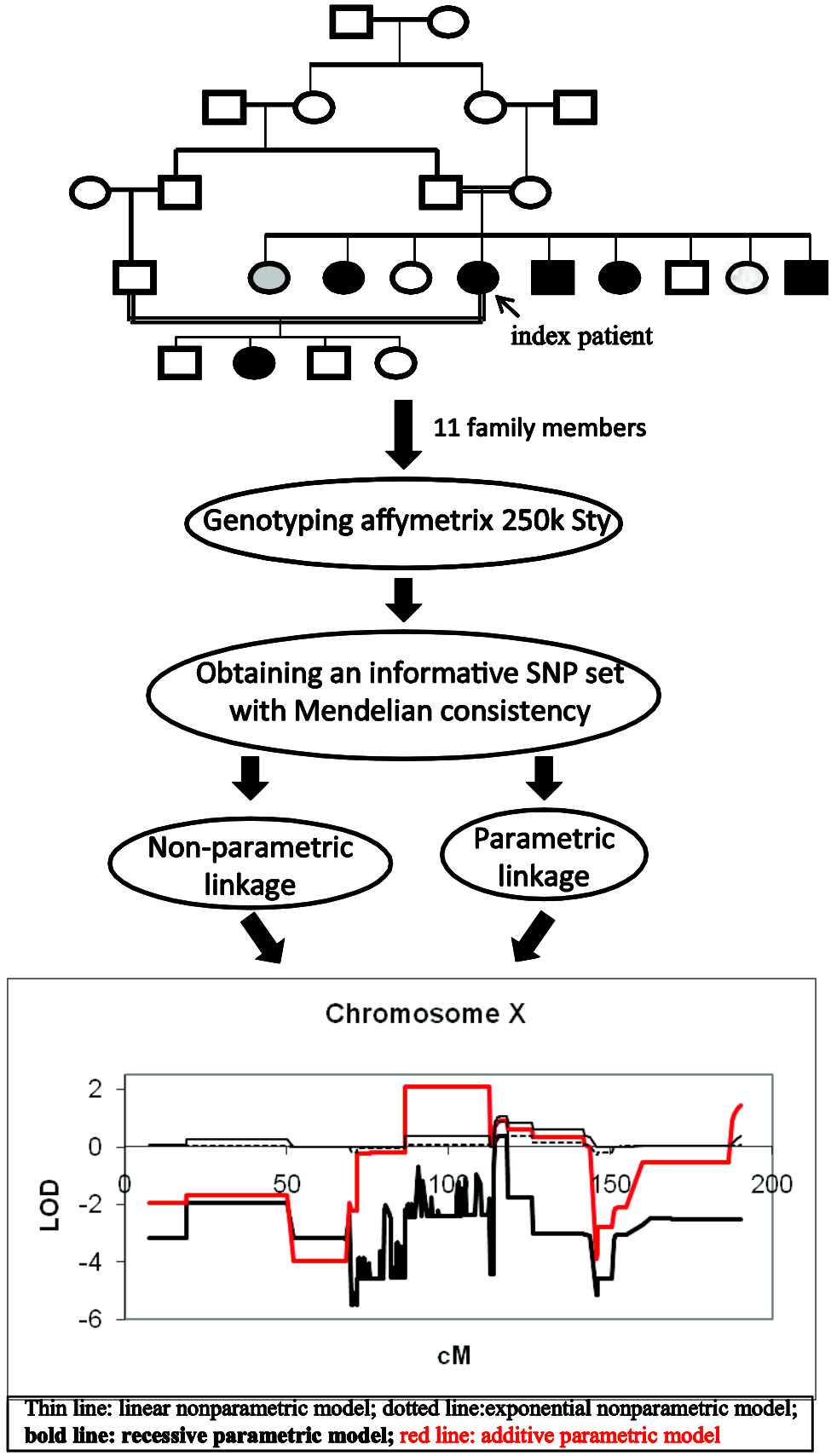
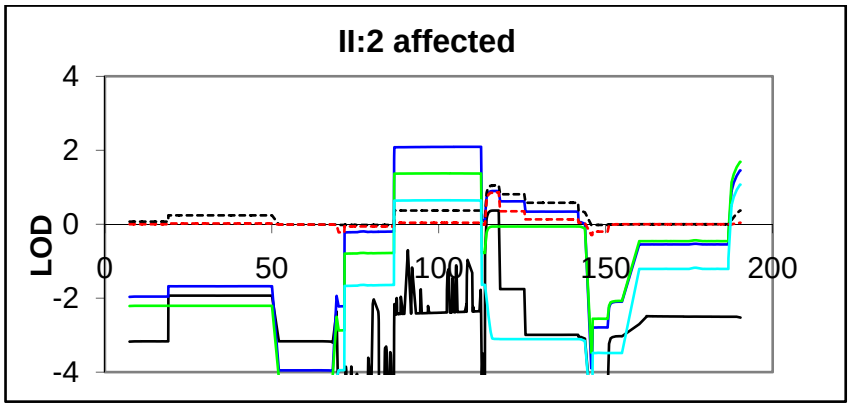
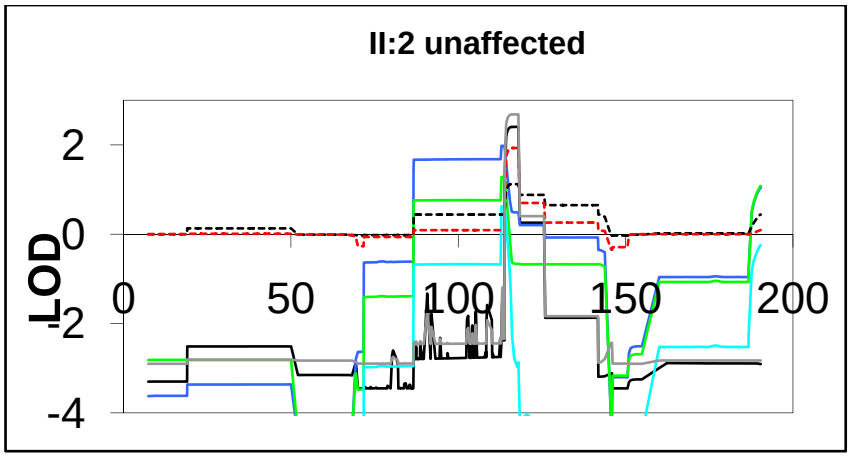
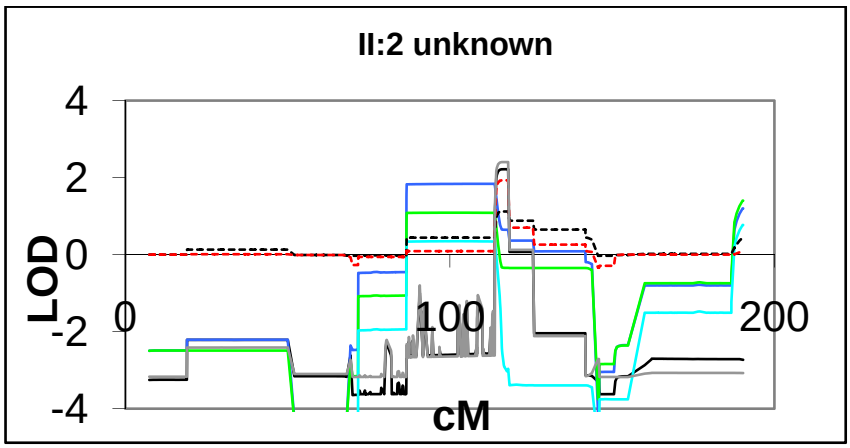


Figure 6. Flowchart of study III. The graph shows the LOD-score for the kinship when II:2 is coded as affected.







		Penetrance n alleles:		
Line color	**LOD**	*0*	*1*	*2*
black	Recessive 1:	0.005	0.005	0.7
grey	Recessive 2:	0.005	0.005	0.95
blue	Additive:	0.005	0.5	0.95
green	Dominant 1:	0.005	0.7	0.7
turquoise	Dominant 2:	0.005	0.95	0.95
black dotted	Linear	-	-	-
red dotted	Exponential	-	-	-

Figure 7. Linkage results for the X chromosome*. The additive model gives the highest LOD score if II:2 is coded as affected, the peak includes the PLP1 gene. The recessive model with high penetrance gives the highest LOD score with II:2 as unaffected at a position including the* IL13A1 *gene.*

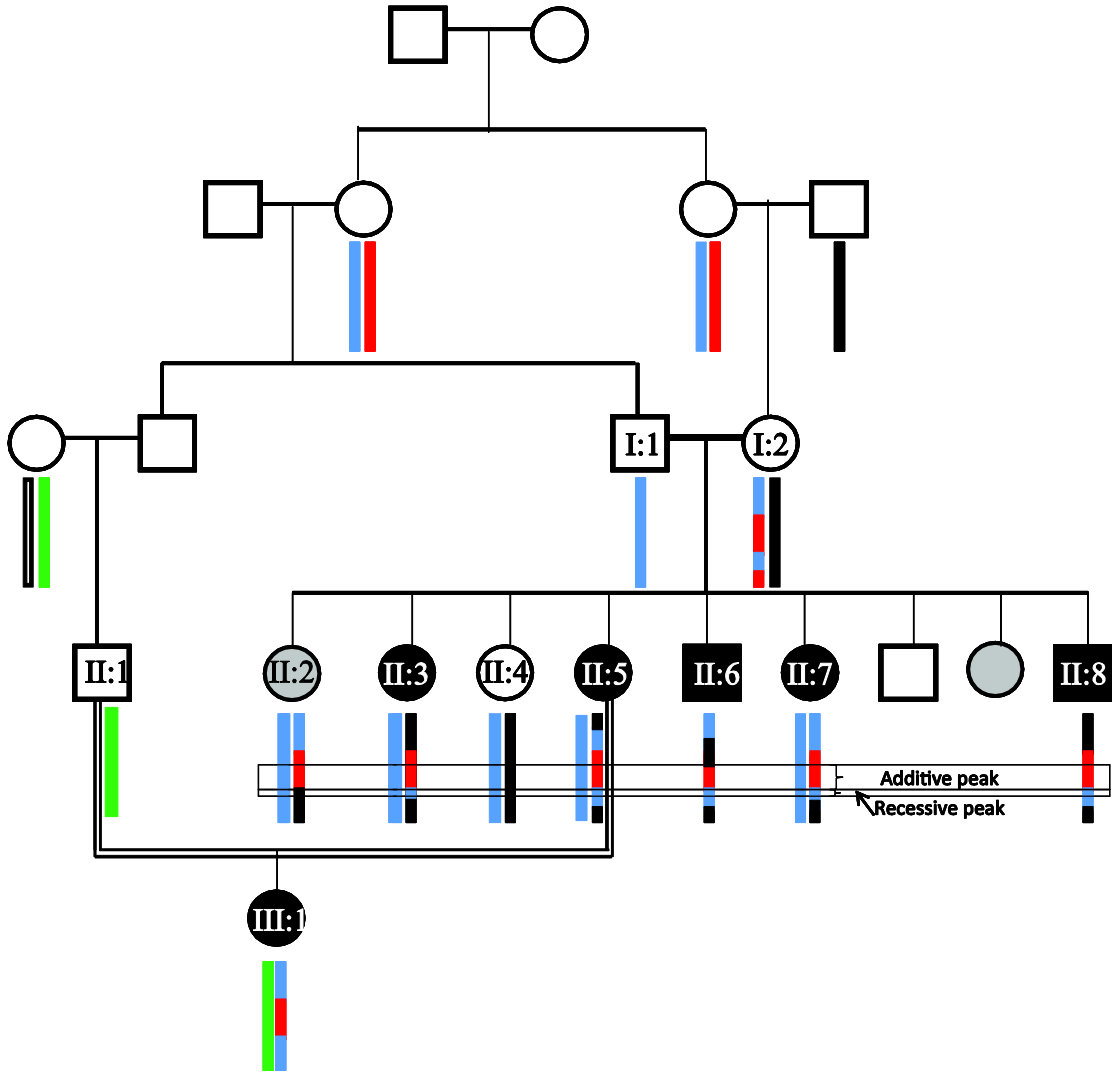


Figure 8. The X-chromosomal haplotypes in the kinship as estimated by the algorithm used in the Merlin programme[48; 49].

Table 7. The highest LOD scores observed under the recessive and the additive/dominant models map to different locations on the X chromosome. The affection status of II:2 changes not only which inheritance model to prefer but also which location to favor.

	Recessive peak		Additive peak		
II:2 status	**Rec 1**	**Rec 2**	**Add**	**Dom 1**	**Dom 2**
affected	0.37	0.42	2.09	1.67	1.07
unaffected	2.40	2.68	1.98	1.28	0.62
unknown	2.22	2.40	1.84	1.40	0.77

There are two additional features (apart from the consanguinity) with this family that one should take into account regarding the cause of disease; they have emigrated from the Middle East to Sweden before their onset of disease. Another intriguing thing is that the affected daughter of the index patient developed her MS at a much earlier age than the others (age 14) and she is also affected by breast cancer. Known breast-cancer genes are located in the vicinity and within the additive peak on the X chromosome. Does this seemingly autosomal recessive family in fact harbor three contributory causes of the disease: one environmental factor that they became exposed to upon settling in Sweden, one genetic component under the additive model and peak and a second genetic component under the recessive model and peak? The environmental factor can only be speculated about; however patients in this family have suffered bout of, and suspected bout of viral meningitis. Other environmental causes could be change of dietary-habits or decreased exposure to sunlight. The two linkage peaks on the X chromosome harbor two, in my opinion, interesting genes: the *PLP1* gene, mutations in which lead to other non-inflammatory dysmyelinative neurological conditions and the *IL13A1* gene involved in B-cell maturation and HLA class II upregulation (further details in the paper). Thus, these together may contribute to a demyelinating inflammatory disease triggered by an environmental exposure.

2.6.1 An attempt to sequence the *PLP1* gene

Mutations in the *PLP1* gene have been shown to be involved in MS in two case-reports[50; 51], thus we found this gene of sufficient interest to sequence for mutations in this family. The sequencing analysis also included the unaffected brother, from whom we had not yet collected a blood sample at the time of the whole-genome screen. The *PLP1* gene contains 7 exons of which the first 6 are so small that they can be covered by one pair of primers, so we decided to start off with them. We used primers from a publication[52] for all exons except for exon 2, since the exon 2 primers in the publication did not seem to bracket to the entire exon when blasting it in Ensembl (www.ensembl.org). The primers used are seen in table 8.

The polymerase chain reaction (PCR) in the reference protocol[52] needed some optimization for the annealing temperatures for the different primers. We used ExoSAP-IT to clean the PCR product from any excess bases or proteins, BigDye Xterminator Purification Kit and BigDye Terminator v.3.1 Cycle Sequencing Kit for

the sequencing (Applied Biosystems). Variant reporter (Applied Biosystems) was used for the computer analysis of the sequence output.

Table 8. Primers used for PCR amplification and subsequent sequencing of the PLP1 gene.

exon 1a	5'-CAGTGAAAGGCAGAAAGAGA-3'
exon 1b	5'-CTGTGTCCTCTTGAATCTTC-3'
exon 2a	tttgagtggcatgagctacc
exon 2b	cccagtcccctgctagttac
exon 3a	5'-AGATTCCCTGGTCTCGTTTG-3'
exon 3b	5'-TCTTCCTGACCTTCTCGTTC-3'
exon 4a	5'-CATCTGCAGGCTGATGCTGA-3'
exon 4b	5'-AGTGGGTAGGAGAGCCAAAG-3'
exon 5a	5'-TAGAGATGGAAGAAGGGCTC-3'
exon 5b	5'-AGGCACACTTAGCCAACATG-3'
exon 6a	5'-AAAGATATCAACACATTCAG-3'
exon 6b	5'-TCAAGGATGGAAGCAGTCTA-3'

The outcome was not exactly what we expected and I would have considered the outcome purely as artefacts if it was not for two things: first, no one I talked to had ever experience such artefacts and second, I found a publication in *Cell* with a drawing scarily similar to the ones I had made (figure 9) of the problem, concerning this particular gene region[53]. I'll come back to that, but first, what did we see?

For exon 3 the expected target sequence was 455 bases. However, the analysis revealed a much longer high-quality sequence for the sequencing with the reverse primer but not with the forward primer. This was seen in 9 of 12 individuals and a poor quality extended sequence was seen for a tenth family member. The mother (I:2) and the daughter (III:1) of the index patient both displayed very poor quality sequences of about the expected length. I blasted the extended part of the obtained sequence in Ensembl and found that it correlated to an inversion of exon 3, although shorter than the target sequence (figure 9). All family members exhibiting the extended sequence showed at exactly the same position a microhomology between the target sequence and the inverted sequence. The inverted sequence showed mixed bases at some positions, which may be explained by the fact that the reading process from the inversion to the target sequence; the machine will at ones read first the inversion and the target sequence and then continue with only the target sequence that is attached to the inversion. The forward primer cannot bind the inversion and thus not reading it.

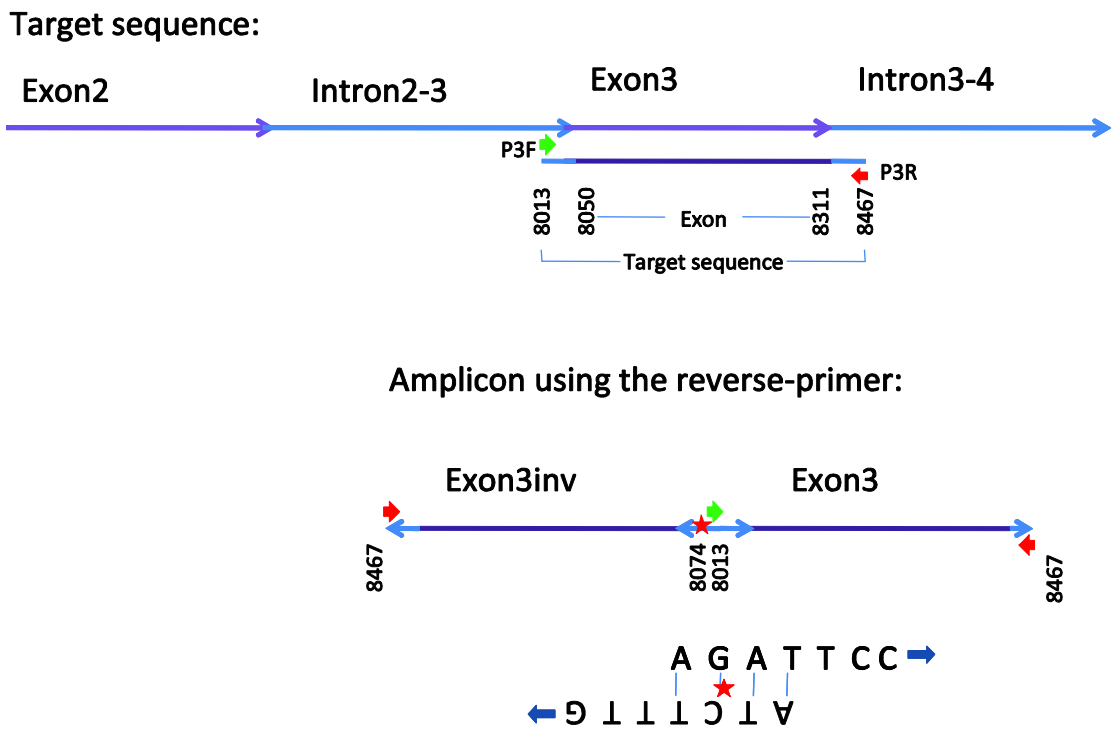


Figure 9. The target sequence of exon 3 continues in an inversion of exon 3. This was seen in several of the family members with the reverse-primer sequence. The location of the junction explains why this couldn't be seen by the forward primer: The inversion lacks the sequence between position ..8013 to ..8074 and the forward primer locates at position ..8013 to ..8033. The star indicates the junction which is a microhomology between position ..8013 to 8016 at the target sequence and position ..8073 to ..8076 in the inversion. Red arrow: reverse primer; green arrow: forward primer.

I had a closer look at the gel pictures from the PCR reactions and for some individuals a very faint band was seen matching with the length of the obtained sequence; thus the sequence may be really there, but had it been created during the laboratory process or does it reflect the genomic sequence of the family members? Does the fact that this was seen in the majority of the family members argue for or against an artefact? Does the low probability of actually placing the primers such that an exon duplication is detected indicate that this is an artefact?

In addition to the strange result for exon 3, the mother of the index patient (I:2, in whom the exon 3 inversion was not detected) showed a similar event on exon 5 (figure

10). In this unaffected mother the exon 5 target sequence was followed by an inversion of exon 5 and the junction was a homology of 17 bases with one mismatch. Three individuals also exhibit inversions at exon 6, although with blurry junctions (which is expected if the primer does not exactly target the junction).

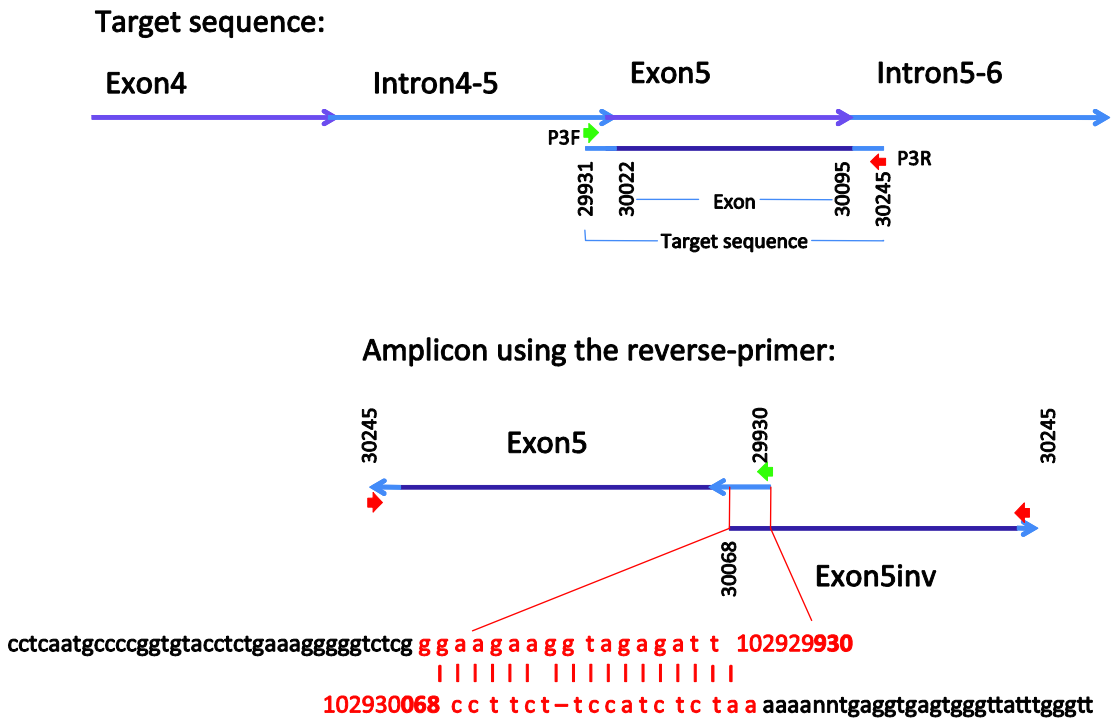


Figure 10. In the unaffected mother of the index patient, an extension of the target sequence with an inversion of exon 5 was found. The target sequence and the inversion adhered to one another with a homology of 17 bases including one mismatch. Again this was only detected by the reverse primer (red arrow).

An easy explanation for the frequent occurrence of these inversions is that they are all just artefacts from the PCR reaction; however, I spoke to several researchers with experience in sequencing and supportdesk at Applied Biosystems and searched the Internet, but nowhere did I find any information suggesting that this kind of artefact exists. Considering the small microhomology seen in exon 3, the problem should be common in sequencing reactions.

An excellent paper in Cell from 2007[53], with Jennifer A Lee as the first author, describes the particularity of the chromosomal region on the X chromosome surrounding the *PLP1* gene. Lee et al. explain a replication-based mechanism through

which microhomologies emerge leading to chromosomal rearrangement causing Pelizaeus-Merzbacher disease (PMD). PMD is a *PLP1*-dosage-sensitive dysmyelinative disease. The study investigates the rearrangements in the surrounding region of *PLP1* in PMD patients and reveals up to 4 junctions for a single patient. The authors manage to pinpoint several of the junctions and reveal duplications adhering to one another with microhomologies as small as two basepairs. Thus, this paper shows that duplications may occur in this chromosomal region through a replication-based mechanism involving adherence of sequences by microhomologies. Were we so incredibly fortunate that we pinpointed such duplications in this family by our sequence analysis?

One of the first things I'll do after my thesis defense is to re-perform the sequencing of the *PLP1* gene. A mistake we did this time was to regard the healthy individuals of the family as adequate controls; when re-doing this I'll include unrelated individuals for the identification of methodology problems.

Mutations in the *PLP1* gene do cause a spectrum of neurological symptoms from benign to so severe that those affected do not survive their childhood. Two diseases are related to these mutations: PMD and spastic paraplegia 2 (SPG2). Two case reports of MS patients indicates that mutations in this gene can cause MS[50; 51]. Thus, it is not far-fetched to speculate in the involvement of *PLP1* in this particular family with evidence of linkage in the X-chromosome region. However, I would not bet a million on the prospect that the duplicated inverted exons are not artefacts; on the other hand, I'm not betting a million on the prospect that they are either.

2.7 SUCCESSFUL STRATEGIES?

2.7.1 Reduced-heterogeneity strategies versus GWAS

When I started in the field of MS genetics in 2003, the climate in the group was quite depressed due to many years of lack of success in the field of MS genetics. The only susceptibility gene that we were certain about was the *HLA-DRB1*15* allele – and that one was discovered in 1972[54]. It was not only in MS genetics though – it was the whole field of complex genetics that despaired.

The completion of the human genome sequence and the dissection of human genetic variation (HapMap), together with great advances in genotyping techniques and the formation of international research consortia for obtaining the large sample sets required, changed the world of complex genetics in just a few years. In the past 2 years, more than 100 genome-wide association studies (GWAS) have been conducted, uncovering more than 250 genetic loci that have been shown to be reproducible, for more than 40 diseases and human phenotypes (reviewed in [55-58]). The optimism reached its climax. We were able to find disease genes in a rather straight-forward hypothesis-free way in the whole patient population! To investigate genetic susceptibility in a clinical subset of patients, or in isolated populations or families was not in fashion.

This very week while I'm writing my thesis, the *New England Journal of Medicine* (*NEJM*) presents a review and three perspective articles about GWAS[55-58]. Without necessarily agreeing with any of the authors, I will try to depict their views. They all agree on one thing, that the results obtained via GWAS have properly accounted for multiple testing, and that most reported associations are thus true. Moreover, they agree that the identified loci explain only a small proportion of the heritability of the investigated traits.

"In pointing at everything, genetics would point at nothing."

Goldstein, *NEJM* 2009[55]

Goldstein provides an extreme interpretation of the small effects on disease that are found for identified risk loci. Pointing out that these effects are the best we can find with common variants, he uses a calculation that shows - under the assumption that

all risk loci have an additive affect – that the population variability of height needs 93,000 SNPs for 80% of its explanation. Goldstein still thinks the initial GWAS were worth doing, since they revealed a sizable fraction of the heritability of certain conditions; but he also states that the un-manageable number of genes for a trait makes it point at nothing concerning the biological cause of diseases.

"...because there are so many risk loci, the probability of detecting at least 1 is good"

Kraft and Hunter, *NEJM* 2009[56]

Kraft and Hunter hihglight the extremely low power to detect a disease locus with such modest effects that we are looking for; however, the large number of risk alleles in each disease makes it highly likely to find one, and the number hits generated it is likely that at least a few will be replicated by chance.

"In response to the skeptics, I offer a new bet. I predict that by the 2012 ASHG meeting, GWAS will have yielded important new biologic insights for at least four common diseases or polygenic traits…."

Hirschhorn, *NEJM* 2009[57]

If the first two perspectives have a pessimistic view, Hirschhorn is definitely more positive about the possibilities of GWAS and the future. He points out that the question asked when conducting these studies has nothing to do with individual risk estimation but rather with identifying biological pathways included in diseases. He points at studies that have indeed showed that associated loci cluster in pathways. Hirschhorn is also optimistic about the prospect that when in the near future, we are looking at variants that are less common than 5%, the explained variation will increase, and that the same thing will happen when taking into account gene-gene interactions and gene-environmental interactions.

How successful has the GWA approach been? In the light of the issues discussed here it seems that the field of complex genetic still has everything to gain by simultaneously applying different approaches. The different approaches will also give us different kinds of knowledge about traits and genetics – in some cases knowledge we didn't

foresee. A lot of money, samples and collective working hours have been spent on GWAS – and a lot of knowledge has been gained. However, a few more years are needed before we can say whether GWAS is an approach which will lead to the ultimate goal of new treatments for the individuals affected by these diseases.

I think GWAS opens up extraordinary possibilities, not least regarding the large number of genotyped individuals. It will be possible to use this data to further explore population genetics, to study at a whole-genome level at homogenous groups of patients, to look for genetic relationships between patients, to perform segmental-sharing analysis that will allow us to identify genetic variants of large effect (although this analysis requires extreme computer capacity) and more.

In contrast to GWAS, the reduced-heterogeneity strategies in this thesis used small sample sets; the studies have been relatively cheap; they have indeed cost some time from everybody involved in the studies, and they have cost me half a decade of my life (it was worth it).

How successful have studies using reduced-heterogeneity strategies been in MS? An already classic example despite its relatively recent publication is the identification of the neuromyelitis optica antibody (NMO-IgG) in 2004[59]. The finding of the NMO-IgG made it possible at an early stage of disease to distinguish NMO patients from MS patients and further to classify Asian optic-spinal MS as NMO. The benefits for the patients were immediate since these two conditions response to different treatments. In 2005, the antigen for NMO-IgG was identified[60], which led to substantial increase of knowledge about the pathogenesis in this disease (reviewed in[61]). However, although the antibody very specifically distinguishes NMO from MS, 30 to 40% of typical NMO patients are negative for the NMO antibody; thus further sub-classification of patients may reveal as yet undiscovered pathogenic mechanisms.

The use of isolated populations makes possible the search for rare variants, and has also been beneficial in MS, not least shown in work in Finnish isolates by the group of Leena Peltonen and Janna Saarela. Their most recent study[62] has shed some light on the genetic effect of chromosome 5. Chromosome 5 harbors the *IL7R* gene, which has been called the first non-HLA gene identified in MS since it was confirmed in 2007[2; 63]. However, the Finnish isolated population study showed only a modest effect for *IL7R*

on risk, but instead a large effect for the complement component 7 (*C7*) gene 5.1 Mb from *IL7R*. The *C7* gene was validated in an independent dataset and in the combined dataset the odds ratio was 2.73 (p=0.000003).

How successful was the work contained in this thesis? Let me take you through how I feel my results correspond to the things we know about MS pathogenesis.

2.7.2 Our findings in correlation to MS pathogenesis

Figure 11 shows pathogenic mechanisms in MS in relation to the findings I present in this thesis. The explanation for the association between the *HLA-DRB1*15* allele and MS (shown for the first time over three decades ago) is still shrouded in mystery. Antigen presenting through HLA class II reactivates CD4+ T cells once they have crossed the blood-brain barrier, which initiates a cascade of events resulting in the recruitment of other inflammatory cells such as B lymphocytes. Immunoglobulin G synthesis by B cells of oligoclonal origin within the brain is detectable in cerebrospinal fluid (CSF). The reason for the clonal expansion of B cells is unknown, as are the antigens against which the produced antibodies (Ab) are directed. Antibodies directed against myelin components have been studied extensively and results indicate that anti-myelin Ab are present in MS patients (reviewed in [64]). The majority of these studies concern Ab directed against myelin oligodendrocyte glycoprotein (MOG), although these Ab are also present in unaffected individuals in some studies[64]. Ab directed toward proteolipid protein, encoded by our candidate gene in study IV, have been found in the CSF of MS patients[64]. The presence of Ab in MS could be primarily pathogenic, or it could be a secondary humoral response or even part of the self-repair (remyelination or axonal protection) as reviewed in[64]. Thus, OCB are not necessarily directed against the disease-causing antigen.

In study I, we showed that OCB-positive MS is associated with *HLA-DRB1*15* and OCB-negative MS with *HLA-DRB1*04*. This suggests that antigen presenting through the molecule encoded by *HLA-DRB1*15* produces an inflammatory response including B-cell clonal expansion, while antigen presenting through the molecule encoded by *HLA-DRB1*04*04 may have a more neurodegenerative effect, perhaps through activated cytotoxic CD8+ T-cells. Unpublished preliminary data from colleagues of mine (Malin Lundkvist, Eva Greiner and Anna Fogdell-Hahn, manuscript) shows that OCB-negative patients also are less prone to develop antibodies (Ab) against

therapeutically used interferon beta (drug-neutralizing Ab are a therapeutic nuisance which is reported in almost one third of patients). Thus, OCB-negative patients may have a maladaptive incapacity to produce antibodies −or, alternatively, an adaptive capacity to produce antibodies in appropriate quantities, which OCB-positive patients lack.

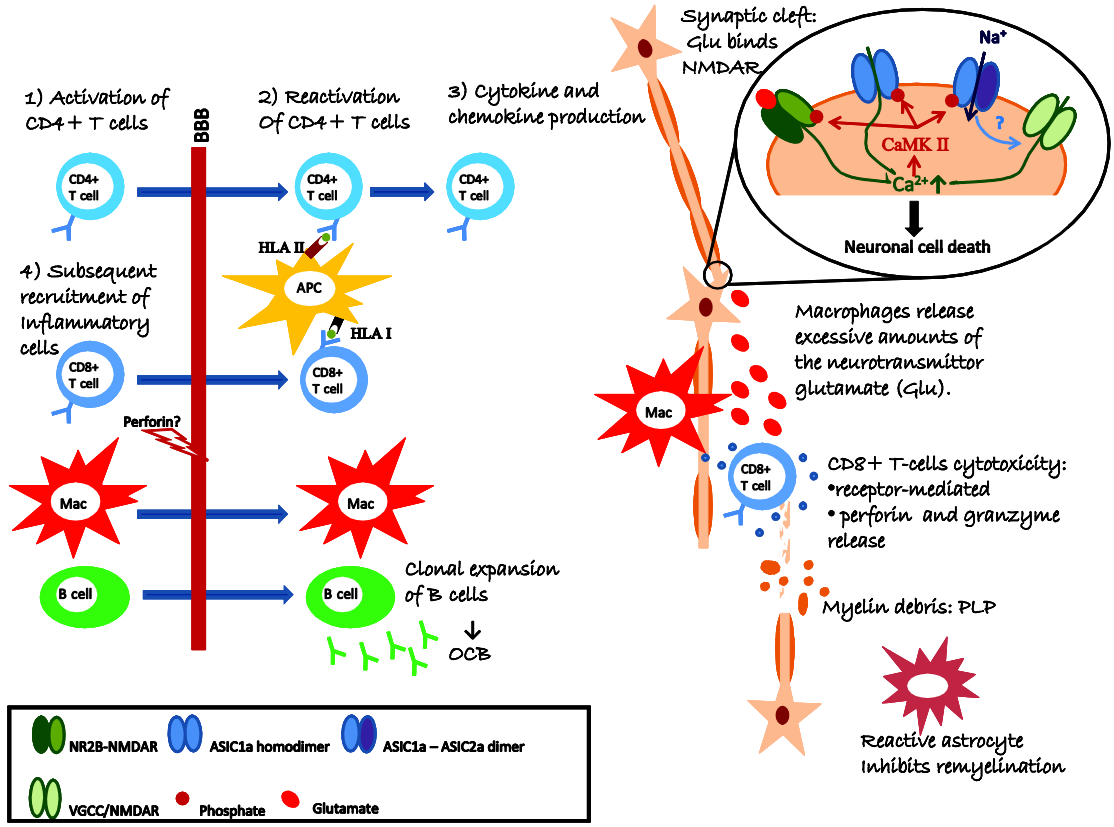


Figure 11. This picture shows MS pathogenesis and where the results evolving from the work in this thesis fit. The sections below take you through the pathogenic mechanism connected to the different studies in this thesis. Picture adapted[65; 66]*. BBB: blood-brain barrier, APC: antigen-presenting cell; Mac: macrophage;, HLA: human leukocyte antigen; OCB: oligoclonal bands; PLP: proteolipid protein; ASIC: acid-sensing ion channel; NMDARs: N-methyl-D-aspartate subtype of glutamate receptors; NR2B: regulatory subunit of NMDARs; CaMKII: Ca^{2+}/calmodulin-dependent protein kinase I;, VGCC: voltage-gated Ca^{2+} channels.*

The new, but still established view that axonal degeneration occurs already at an early stage of the disease fits with the results of study II: patients suffering from the primarily inflammatory (?) *HLA-DRB1*15*-positve, OCB-positive entity on the one hand and

patients suffering from the primarily degenerative (?) *HLA-DRB1*04*-positive, OCB-negative entity on the other reach one important disease milestone, the inability to walk 100 meter without unilateral aid, at the same age.

In active MS lesions, CD8+ T cells predominate (reviewed in [65]). In an inflammatory milieu, HLA class I molecules are up-regulated, and subsequently cytotoxic CD8+ T cells can recognize and destroy infected cells in the body. However if the antigen these cytotoxic cells react against is a self antigen such as a myelin peptide, their effect is destructive rather than beneficial. CD8+ T cells mediate their cytotoxicity in two ways, through the Fas receptor-Fas ligand interaction and by releasing granules containing perforin and granzymes. Perforin released by CD8+ T cells may also take part in the disruption of BBB tight junctions thus increasing vascular permeability in the CNS[67]. What underlies the genetic susceptibility conferred by the perforin gene[35] (*PRF1*; study III)? Could a possible mechanism b gain-of-function over-expression of perforin causing excessive damage to oligodendrocytes, axons and BBB tight junctions or a loss-of-function leading to abrogation of feedback control of CD4+ T cells?

Excessive release of glutamate, from macrophages or other cells, is emerging as an important determinant of neuronal injury in MS (reviwed in [66; 68]). Glutamate can cause excitoxicity and death of both oligodendrocytes and neurons and thus may play a part in the pathogenesis of MS, as seen in primarily neurodegenerative disorders. The neurotransmitter and its receptor are found up-regulated in MS lesions, and up-regulation is correlated with axonal damage; in addition, blockade of the receptor has beneficial effects in the animal model of MS.

The glutamate receptor (N-methyl-D-aspartate subtype of glutamate receptors) NMDAR is known to be involved in ischemic neuronal death[66] . Work by Gao et al.[66] has further connected NMDAR activation with ASIC receptors in the events leading to the loss of neurons. Activiation of NMDAR leads to a signal cascade activating Ca^{2+} /calmodulin-dependent protein kinase II (CaMKII), which phosphorylates the NMDAR and ASIC dimers. The phophorylation of NMADR and the ASIC1 homodimer facilitates influx of Ca^{2+}, which eventually reaches toxic levels where upon the cell dies. The ASIC1/ASIC2 heterodimer is also phophorylated; though not allowing the passage of Ca^{2+} it may affect voltage-gated Ca^{2+} channels. A nematode model system revealed that a gain-of-function increase in the influx of Na^{+} through corresponding

ion-channel, without permeability to Ca^{2+}, caused neurodegeneration[27]. Thus, does this connect the inflammatory events of MS to neurodegeneration?

Centonze et al.[68] uses the animal model experimental autoimmune encephalomyelitis (EAE) to actually connect the two features of MS: inflammation and neurodegeneration. AMPA receptors are also glutamate receptors and PICK-I has been shown to be able to regulate these (reviewed in [69]). Centonze et al.[68] describe a mechanism independent from demyelination in which phosphorylation and excessive expression of AMPA receptors follows upon activation of microglia. The subsequent signal cascade leads to synaptic degeneration and neuronal death.

PLP is the predominant protein compartment of myelin in the CNS. Genetic variations of the *PLP1* gene cause dysmyelinative disorders which are considered non-inflammatory. Two case reports of MS patients bearing mutations for the *PLP1* gene make us speculate in that this is the disease-causing gene in the family investigated in paper IV. An explanation for the inflammation seen in those MS patients bearing *PLP1* mutations is that it occurs secondary to dysmyelination. The *PLP1* gene has the potential to undergo *de novo* mutation. The resulting rare alleles could plausibly cause MS in a proportion of patients; however, such an effect would be impossible to detect by way of an association study.

2.7.3 Concluding remarks and future perspectives

GWAS identifies markers for contributory causes, or I could also put it this way: they identify separate slices of pies. For a complex trait, there are a large number of slices acting in different pies, i.e., having additive effects; and in the same pies, i.e., having more (or less) than an additive effect on a carrier's risk of disease. Several factors may act on a disease pathway to cause disease; however you seldom hear someone discuss whether the hypothesis presupposes that the same pathway is involved in all patients but patients carry different susceptibility factors acting in the pathway; or if the factors within a pathway act jointly to cause disease at an individual level. What are the contributory causes at an individual level? Is it most reasonable to believe they act in the same pathway or does a person need susceptibility factors in several pathways to develop a disease, e.g., in MS factors conferring susceptibility in inflammatory pathways in combination with factors conferring susceptibility in neurodegenerative pathways?

The strategies applied in this thesis focus on finding pieces acting in one or a small number of similar pies. If successful, such strategies may provide knowledge about how risk factors interplay at an individual level. The rationale behind reduced-heterogeneity strategies is to increase the signal-to-noise ratio by reducing the number of contributory causes at a population level. But how many contributory causes are needed for a sufficient cause of disease?

In paper III, we identified five chromosomal regions of interest; we further explored the genes in these regions in relation to what is already known about genetic susceptibility and pathogenesis in MS, which the obviously drawback of important genes possible being overlooked. Yet, I would like to believe that we perhaps have already succeeded in uncovering several genes in the same sufficient cause.

The results obtained in this thesis project are reasonable in that they are consistent with previous reports of genetic susceptibility in MS, discussed in detailed in each paper, I'll here just highlight a few: the OCB-negative entity is associated with *HLA-DRB1*04* both in our population and in Japan; the *ACCN1* gene is seen in our distantly related MS patients and in an isolated population in Sardinia; mutations in the *PLP1* gene are reported in two MS-case reports, and thus monogenic X-linked MS is plausible.

The results of GWAS greatly influence the focus of research in the field of complex genetics, yet I believe that the results derived from this thesis project also deserve a thorough follow-up:

- I would like to see that OCB-negative MS is addressed when analyzing data from genome-wide association studies in MS. This may identify other pieces in the pie specific for OCB-negative MS. Given the fact that *HLA-DRB1*04* is associated with sero-positive rheumatoid arthritis, the explanation for the lack of sign of intrathecal IgG production must be more complex than the *HLA-DRB1* association can explain. I would further propose a sub-stratification on *HLA-DRB1*04*04 positive OCB-negative patients, since I regard this as the prototypic OCB-negative entity. The homogeneity of such a group is likely to overcome the problem of insufficient statistical power in such a small stratum. I

would like also to further explore my hypothesis of OCB-negative MS being the same entity across populations.

- One question in the Lysvik cluster is how many loci that are involved. The conservativeness in our design gave us five. Do these contain common variants of small effects or is it reasonable to search for mutations? MS has already been shown to be associated to the *PRF1* gene and mutation in the *ACCN1* causes neurodegeneration in a model system. Thus, the first action I would like to take in following up this study is to sequence the *PRF1* and the *ACCN1* genes in the patients included in the study and in additional patients and controls from the same geographic area. If positive results are obtained then I'd also look for the mutation in larger cohorts. I would also like to more profoundly investigate the genes in the consensus regions and do pathway analysis.

- In the Huddinge kinship an important concern is the affection status of the undiagnosed sister. I will continue the sequencing of the *PLP1* gene and also investigate the existence of rearrangement on the X chromosome in this family.

- The shared-segment analysis is applicable in many studies and in the light of MS patients seeming to be more related to one another than controls, it would be interesting to go further looking for IBD sharing in so called sporadic cases and unaffected individuals. It is also possible through our Swedish registries to connect "sporadic" MS patients to common ancestors or to use genetic relatedness.

3 BACKGROUND

3.1 MULTIPLE SCLEROSIS

3.1.1 Demography

Two and a half million people worldwide are affected by MS. The disease most commonly presents in early adulthood and thus has great impact on both family and career planning. Typically, the following decade and a half features periods of symptoms, with varying degree of impact on life and working capacity, separated by symptom-free periods. This relapsing-remitting course (RRMS) eventually converts to irreversible progression (secondary progression; SPMS). One fifth of the patients display a progressive course from onset (PPMS), however with a disease onset, on average, a decade later. Age at the attainment of different levels of disability as measured by Kurtzke's Expanded Disability Status Scale (EDSS[70]), such as the inability to walk 100 meters without a cane (EDSS 6.0), is independent of the initial course of the disease[71]. Despite this, the disease at an individual level is extremely variable, ranging from very benign with little disability and impact on life to rapid progression. The inability to predict the specific course for an individual is of course stressfull for the newly diagnosed patient. One of the ultimate goals of MS research is to find a treatment that prevents patients from converting to the progression phase.

There are some demographic factors that increase the risk of being affected. Women are affected twice as often as men. The prevalence of MS is unevenly distributed in the world, with a high prevalence in the Nordic countries where about 1 in 1000 is affected. Country of residence before the age of 15 determines the individual risk (reviewed in [72]). Ethnicity is also a risk predictor, in fact some ethnicities, although living in high-prevalent areas, seem protected (reviewed in [72]).

Evidence that MS etiology include a genetic component is reviewed in[73] and includes studies of twins[74] and adoptees[75]. The comparison of monozygotic (MZ) and dizygotic twins (DZ) reveals that the concordance for the disease increases with genetic sharing; the concordance among MZ twins being 25% and DZ twins 5%. The adoptee study showed that adopted first-degree relatives to MS patients have the same risk as the background population; thus the familial aggregation of MS appears to depend on

shared genes and not shared environment. Although MZ twins share their whole genome, the relatively low concordance rates seen, do indicate the importance of a non-inheritable factor, i.e. environment.

The most commonly cited research providing evidence for the notion that this environmental factor is an infectious agent is the work of Kurtzke[76], describing an MS epidemic on the Faroe Islands. His thesis is that the British troops brought this agent to the islands and that it subsequently has caused MS cases in waves of epidemics, although this agent has never been identified. Last year I had the good fortune − in the name of MS research – to visit the Faroe Islands while helping a medicine student, Stefanie Binzer, with a project. We visited several people affected with MS and identified patients not fitting into the epidemic waves described by Kurtzke. Several of the patients told us about old people they had known who had displayed MS symptoms long before the arrival of British troops. These preliminary data, which will require a thorough follow-up, at least make me strongly question the Faroe Islands studies, although I hate I have to say that. (Two years ago Anne-Marie, Thomas and I took a lovely journey around Lysvik together with John Kurtzke and his wife. John Kurtzke has independently of the work of Anne-Marie described Lysvik as a high-prevalence area).

3.1.2 MS pathogenesis

MS is an immune-mediated demyelinating disease of the central nervous system. No causative agent has been identified, although several viruses have shown some evidences of playing a role. The mechanism for the proposed autoimmunity is unresolved. Debate concerning the primary event in MS is presently ongoing.

Last year, Trapp and Nave published a review arguing for the evolving concept that MS may be primary neurodegenerative disorder with secondary inflammatory demyelination[77]. I remember listening to Professor Trapp at the very first conference I attended: ISNI in Venice in 2004. I cannot say that I understood everything, but it was very clear that the data he presented about the existence of grey-matter lesions in MS brains raised a lot of questions from the audience. Traditionally MS has been considered a white-matter disease, which may be a consequence of that grey-matter lesions are not detected by T2-weighted MRI since the blood-brain barrier remains intact; they are also difficult to detect macroscopically, and with histological staining[77].

However, a recent study suggested that cortical lesions are a prominent feature in MS[78]; and the fact that cerebral cortex becomes demyelinated without a significant influx of immune cells is one of Trapp's and Nave's[77] arguments for the notion that inflammation is secondary to a degenerative process.

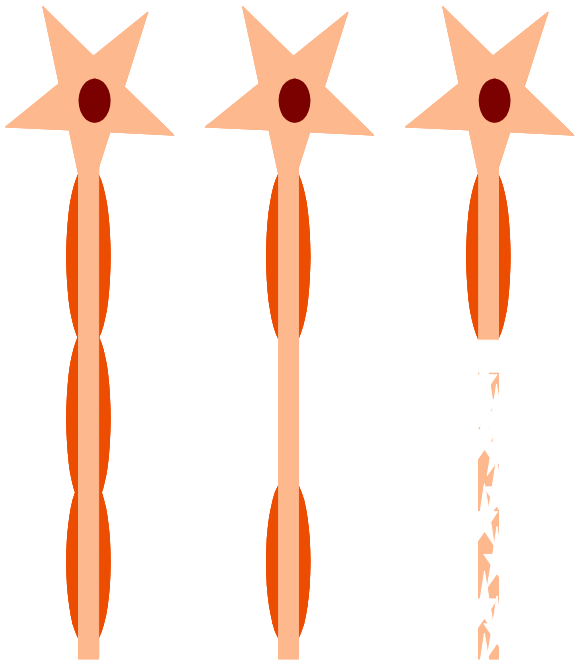


Figure 12. 1) Normal neuron. 2) Demyelinated neuron. 3) Degenerated neuron.
A neuron consists of the cell body with the nucleus, from which there are short outgrowths, dendrites, and a long outgrowth, the axon, through which the neuron communicates with other neurons. The axon is covered with myelin, from oligodendrocytes, for the purposes of keeping the chemo-electrical signal intact as it passes along the axon. In MS the myelin is broken down and thus the signal cannot be conducted properly. The plasticity of the brain compensates for impaired conduction by letting the message to be transduced in the affected neuron take another route through other neurons. Degeneration of axons was previously believed to be secondary to demyelination, but it has become increasingly evident that it occurs extensively at an early stage of the disease (reviwed in [65; 79]). The cell bodies are located in the cortex - the grey matter of the brain, while the white matter consists of the myelin-wrapped axons and supportive tissue.

The treatment of MS involves anti-inflammatory and immunosuppressive drugs, classical immuno-modulating treatments being recombinant interferon beta and glatiramer acetate (GA). The reason for treating MS with interferon beta was because it was believed that MS was caused by the reactivation of a viral infection. The beneficial effect of interferon beta in MS is unclear but the drug potentially decreases antigen presentation, modulates apoptosis and the entry of immune cells into the CNS. GA on the other hand relied on the idea of mimicry, being a polymer that resembles myelin basic protein. Both of these drugs have effects of the number of annual relapses in patients and number of lesions detected with MRI; however none of them have large impacts on progression. Thus, progression may occur independently from inflammation; however, Lassman in a review from 2007[80] instead explains this by that there are different patterns of inflammation in different stages of the disease. In relapsing MS, inflammatory events occur at the site of acute lesions; however, such lesion are rare in primary or secondary progressive MS. In the progressive phase of MS there is instead atrophy of both white and grey matter as well as diffuse axonal loss in normal-appearing white matter. Lassmann interprets the lack of infiltrating T and B cells in cortical lesions differently from Trapp and Nave, and argues that soluble factors produced by these cells indeed cause myelin damage directly or indirectly through microglia activation. Lassmann also refers to a study by Serafini et al.[81], which showed the formation of lymphoid follicles in meninges and perivascular spaces in progressive MS; thus inflammation processes can go on in MS even with a normal or repaired blood-brain barrier (whose duty is to keep the CNS immune-privileged). The existence of lymphoid follicles within the CNS correlates well with another hallmark of MS: the intrathecal synthesis of immunoglobulins.

3.1.3 Diagnosing MS

The diagnosis of MS requires two important considerations. The patients must have presented with at least two relapses disseminated in time and space. In the McDonald criteria[7], the diagnosis can be based on one clinical relapse if a second subclinical relapse is detectable with MRI. It is also important that the diagnostic work up excludes all other possible diagnoses. Two paraclinical tests aid the clinician in diagnosing MS: the detection of oligoclonal bands (OCB) and magnetic resonance imaging (MRI).

3.1.4 Oligoclonal bands

In the Nordic countries, it is routine to test for intrathecal immunoglobin G synthesis during the diagnostic work up of MS. Figure 13 shows two outcomes of this test.

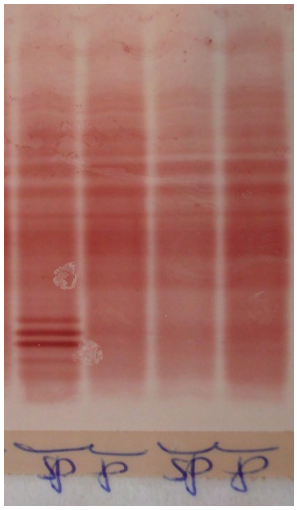


Figure 13. The samples from one OCB-positve patient and one OCB-negative patient on the same gel. From the patient a CSF sample and a blood sample are taken at the same occasion. The definition of testing positive for OCB is the present of two or more IgG bands in CSF that are absent in plasma. The left are first the CSF sample and then the paired plasma sample from an OCB-positive patient. To the right are the CSF sample and the paired plasma sample from an OCB-negative patient. Most commonly OCB-negative patients are doubly negative, but a patient can also be defined as OCB-negative upon the presence of the same OCB both in CSF and in plasma.

3.1.5 MRI

The different stages of MS can be visualized using magnetic resonance imaging (MRI), with different weightings: T1, T2 and T1 with gadolinium contrast. T1 provides a good contrast between white and grey matter and visualizes atrophy (hypointense lesions). T2 visualizes edema, inflammation, remyelination and unspecific changes (hyperintense lesions). The gadolinium (Gd) molecule passes the blood-brain barrier; thus gadolinium-enhanced T1 scans visualize ongoing inflammation. T2 and Gd-enhanced lesions are both included of the diagnostic criteria of McDonald et al.[7-9] (at least 2 T2 lesions on brain MRI).

In the revised McDonald criteria[82] the diagnosis CIS (clinical isolated syndrome) can be based on one clinical relapse if a second subclinical relapse is detectable with MRI

and other MRI criteria are fulfilled (at least three of the following: 1 infratentorial lesion, 1 juxtacortical lesion, 3 periventricular lesions, 1 Gd-enhancing lesion).

A typical MRI scan of an MS brain shows multiple white spots constituting the lesions. If serial MRI scans are taken in a patient at different time points, one can see how the lesions increase and decrease in size and number.

3.2 EPIDEMIOLOGICAL / BIOSTATISTICAL CONCEPTS – PAPERS I & II

The information presented in this section can probably also be found in the book by Rothman[1] on epidemiology and in the paper by Ahlbom and Alfredsson[83]; these references may be good to read - even for me. I've learnt about these topics during a series of very well conducted courses on epidemiology and biostatistics held at the Department of Medical Epidemtiology and Biostatistics (MEB) of Karolinska Institutet, which several of my colleagues and I have attended. More important for my understanding of these concepts is the discussion always ongoing in our group to which we denote a lot of time, but which in the end rewards us with deeper understanding of the problems we are facing. That is true problem-based learning!

3.2.1 Outcomes and exposures

The outcome is very often the same thing as the disease or another phenotype such as obesity. The outcome can be dichotomous (affected/unaffected) or quantitative (for example blood pressure). In longitudinal studies the outcome can be recovery from disease (in non-chronic diseases) or death. In my thesis project, I have used the following outcomes: in paper I, OCB-positive MS and OCB-negative MS; in paper II, the longitudinal outcomes attainment of EDSS 6.0 and onset.

The exposure is the thing whose role in causing the outcome we want to investigate. In reality we do not look at causal relationships when studying epidemiology; we only detect associations. The exposure can be environmental or genetic and can be measured in many different ways. The exposure can be the answers from a questionnaire about smoking habits; it can be the measurement in blood of levels of vitamin D; or it can be carriage of a certain genetic variant detected by genotyping. The exposures I have investigated in my thesis are the following: In paper I, carriage of certain *HLA-DRB1* alleles; in paper II, carriage of *HLA-DRB1* risk alleles and OCB status.

3.2.2 Odds ratio

Odds ratios (OR) are the cross-sectional approximation of the relative risk (RR). The RR measurement first measures the exposure and then examine whether the disease-phenotype is more common among exposed than unexposed. This requires a large cohort that is followed longitudinally. In a disease like MS, where the incidence (number of new cases in the population per time unit) is very low, the study design required for measuring a RR is very inconvenient. An OR can be calculated using a

case-control design, whereby instead of looking at the whole population, one collects approximately equal numbers of patients and more or less matched control individuals from the same population. The OR measures if the exposure is more common among cases than controls or vice versa.

Once I had a teacher during one of my courses who said: “You Swedes have such difficulties with the concept of odds. Go to Britain; because they bet on everything, they know everything about odds.” Perhaps odds are a bit unintuitive for a Swede, but they merely represents the chance for one thing to occur compared to another thing, for example the odds of being exposed compared to not being exposed among patients. The odds ratio is the ratio between two sets of odds, for example the odds of being exposed compared to not being exposed among cases divided in the odds of being exposed compared to not being exposed among controls. Note that the result of the calculation will be exactly the same even if you flip the outcome and exposure: the odds of being a case in exposed compared to unexposed divided by the odds of being unaffected in exposed compared to unexposed.

Have you ever been considered changing your lifestyle after reading a tabloid headline stating something like: “Eating tomatoes doubles your risk for colon cancer”? Think again, or at least be aware of what an odds ratio really tells you (because it is odds ratios or related effect measures these tabloids are reporting). The OR is always related to the background risk. If one in a hundred people will be affected by colon cancer during his or her lifetime, your risk while eating tomatoes is two in a hundred. Would you stop eating tomatoes because of that?

3.2.3 Hazard ratio

A hazard ratio (HR) describes the odds for one group compared to another *per time unit* and is the effect measure in survival analysis. Such analysis is commonly used for measuring the effect of a treatment; patients with and without treatment are followed for a certain amount of time and (for a serious condition) the number of deaths are counted. In the example, the *exposure* is treatment, the *outcome* is death and the *time variable* may be months. The question the survival analysis addresses is; at a monthly level, how much higher are the odds of surviving in a treated patient compared to an untreated? (You may immediately think that it is not likely that the effect of treatment

on survival will be the same the first month and the 12th month; that is something you can handle in the analysis, although I'm not going into it.)

In MS, we commonly use survival analysis with the attainment of a disability level as the outcome and disease duration measured in years as the time variable. The question we want to answer may be: does this genetic variation affect the progression? A HR of 2 indicates that at each year, exposed patients have, compared to non-exposed, a doubled risk in attaining the specified EDSS score. Some EDSS scores are more stable than others and therefore better to use; for example EDSS 6.0, which is when a patient no longer can walk 100 meters without unilateral aid (a cane). The problem with duration as a time variable in MS is that clinical onset is preceded by the detrimental biological events underlying the disease and that the time from the biological to the clinical onset may vary. The clinical onset al.so has some other uncertainties due to the features of the disease and the diagnostic criteria. Therefore, in study II, we used *age* as the time variable; *onset* and *EDSS 6.0.0* as two different outcomes; and carriage of *HLA-DBR1*15*, *HLA-DBR1*04* and *the presence of OCB* as three different exposures.

3.2.4 Confounding

Confounding is easy to define – but may be extremely complicated to sort out in reality (which Eva, Thomas and I experienced while working on paper II). A confounder is something that is both connected to your exposure of interest *and* affects the outcome. A confounder skews the association between the exposure and the outcome. Perhaps you are interested in the effect of coffee drinking on heart disease and you find an association. But wait; there may be a connection between drinking a lot of coffee and smoking (imaging the time before smoking was forbidden in cafés)? We already know that smoking has an adverse effect with regard to heart problems; thus smoking is potentially a confounder. If it turns out to be so, then the effect of coffee drinking on heart disease may disappear when smoking is controlled for. But a confounder can also mask a true effect.

Currently in the field of complex genetic, I suspect that we have problems with population stratification as a confounding factor in our association studies.

3.2.5 Interaction

It may also be so that coffee drinking has a small (or no) effect on heart-disease, smoking has a fairly large effect on heart-disease, but if you both drink coffee and smoke then you are almost certain to have a heart attack (this is just a fictitious example). Coffee drinking and smoking in this case would have an interacting effect on heart-disease. Interaction could also entail that exposure B reduces the risk of exposure A.

In part 2.1.2, I introduced the pie concept. Let's discuss interaction using that model. As discussed in part 2, the effect of genetic susceptibility variants are small, and most individuals carrying these variants remain healthy. Accordingly patients may also carry risk alleles without that these are contributing to the disease (this is connected to the concept of the attributable risk or etiological fraction of a risk factor). Using the pie model, we can picture this as an individual carrying pie pieces belonging to different pies. If we consider the outcome "dying young" (to add some drama), a young person who always exceeds the speed limit when driving, and in addition, is devoted to parachuting may have a largely increased risk of dying young compared to law-abiding drivers who do not parachute from airplanes. The sufficient cause of dying young may, however, consist of the contributory causes: being young, speed-driving, slippery road and sharp curve. Parachuting is not in that pie of sufficient causes. Still a person who both speeds and parachutes has an increased risk of dying compared to persons who only do one of these things. The effect of different pies adds up to one's total risk. Speeding and parachuting statistically interact. Biological interaction, however, occurs when things interact within the same pie, thus deviating from an additive effect. Let's say that parachuters tend to speed much more frequently than average youngsters, consequently, there is a biological interaction; in fact they have a genetic predisposition for taking risks. Such a biological interaction can be revealed by studying whether risk taking runs in families, which the anecdotal example of the legendary Kennedy family would seem to suggest.

Interaction is commonly investigated using interaction terms in logistic regression models. This is problematic if you are interested in biological interaction (which I consider to be the more important kind to study), since the logistic regression model uses a multiplicative scale. Consider exposure A and exposure B. Let's say they have ORs of 3 and 4, respectively. You add to the model the interaction between these: A*B.

Each variable in a logistic regression model is to be read as the effect of that variable adjusted for the effect of all other variables; consequently if the effect of A*B alone is 3*4=12, then the model containing all three variables will give an interaction term of the value of 1, which means no interaction. However, with regard to an additive affect; 12 heavily deviates from 3+4=7, and thus biologically interacts (I have oversimplified the calculations here to make my point).

A statistical interaction is simply a significant interaction term in which ever the selected model is. Often the interaction is both statistical and biological. But in the case of interaction terms in a logistic model, where interaction is deviation from the multiplicative effect, there may be situations where the interaction deviates from either the additive or multiplicative effect both not both. This is very important to keep in mind! We discussed ad nauseum the concept of interaction during our work on the analysis in paper II, in which the Cox regression model was used to derive the hazard ratios.

3.3 GENETIC ANALYSIS – PAPERS III & IV

3.3.1 SNPs

The DNA code consists of the four *nucleotides* or *bases*: *adenine (A), thymine (T), guanine (G)* and *cytosine (C)*. These are attached to the two strings in the DNA-molecule and can form two different base-pairs; adenine binds to thymine and guanine to cytosine. The human genome contains $3*10^9$ base-pairs [=3 million kilobases (kb) = 3000 mega bases (Mb)]. The sequence of these bases constitutes the life-code. The strings have different directions called five prime to three prime and three prime to five prime (5'-3' and 3'-5'), these are commonly referred to as *leading* and *lagging* strands and are of importance as some genes are transcribed in one direction and some in the other. At about one thousandth position in the sequence, the base is variable, meaning that individual A will carry, for example, an adenine on the leading strand while individual B will carry a thymine on the same strand. Individual A and individual B thus carry different *alleles* or *variants* at that position (and that explains why some of us have blue eyes and some of us brown eyes as well as most of all the other differences seen between individuals). A variant can also be referred to as a *polymorphism*. The name of these polymorphisms occurring at a single nucleotide in the sequence is simply: *single-nucleotide-polymorphism (SNP)*. SNPs are the most common variation in the genome.

SNPs can be located inside of genes or outside of genes (most of our genome is a gene-desert). Inside genes they can be located in transcribed segments (exons) or un-transcribed segments (introns). A gene is made up of several exons between which the introns lies. Even if a SNP is located within an intron it can still possibly impact the transcription of the gene. A SNP within an exon may affect the sequence of the protein for which the gene codes. Each set of three consecutive nucleotides in the exon sequence correspond to an amino acid. The amino acids are the pearls out of which proteins are built. A specific amino acid is coded by a small number of specific nucleotide triplets; thus a SNP in an exon may change a pearl in the protein-chain but it may also contribute to a triplet encoding the same amino-acid as the alternative variant.

3.3.2 Evolutionary impact on genetic variants

SNPs originate from the occurrence once upon a time, in a single germ cell of a shift in the nucleotide. Unless this shift turned out to be detrimental to the next generation, it

was transmitted to the following generations. If the new variant turned out to be beneficial for the individual in terms of survival or reproduction it then may rapidly increase in frequency in the population, i.e. underwent what is called positive selection. Other factors having a impact on the SNP-allele frequency are demographic events such as non-random mating. In humans, for example, our preferences of partner are not random; we tend to marry people within our closest surroundings (which may lead to inbreeding in remote areas such as the Orkney Islands), it may also be cultural preferences such as language barriers in a bilingual area, or a preference to marry someone sharing your religion or belonging to your church-community (giving rise to isolate populations such as the Amish and the Hutterites but also less extreme examples) or a wish to keep wealth in the family making marriages between cousins preferable (leading to inbreeding). More random events may also shift the allele frequencies, such as rapid expansion or rapid decrease in population size (bottlenecks). The latter may also have a non-random effect; for example during the plague (which had a tremendous demographic impact since it in some areas caused the death of up to one third of the population) some alleles may have been more beneficial in sustaining the disease. Yet, other alleles not connected to the ones under selection may also have undergone a random shift in frequencies. A shift in allele-frequencies – random or non-random –is called genetic drift. The age of a SNP also influences its distribution among populations, both with regard to frequencies and geographical distribution.

An old discrimination between a SNP and a point-mutation was the cutoff of 1%: the minor allele of a SNP should be present in at least 1 of 100 individuals, otherwise it was called a mutation. Now a days this is more variable. One important thing to keep in mind about whole-genome genotyping using microarrays, as used in papers III and IV and also used in genome-wide-association-studies (GWAS) is that the markers on these chips are common SNPs. Thus GWAS are conducted under the hypothesis of common-variant−common-disease. In my studies, by contrast, I've used the same type of arrays to look for rare variants, i.e., by considering haplotypes. The way I do this is very different in paper III and paper IV, but the basic concept in both studies is the probability of identical-by-descent (IBD) sharing.

3.3.3 IBS and IBD

For a biallelic SNP an individual can carry three different genotypes: AA, AB, BB. If two individuals have exactly the same genotype, then they share both alleles *identical-*

by-state (IBS). Two individual can also share one allele IBS. Two individuals homozygous for different alleles do not share anything IBS. For whole-genome data, overall IBS sharing of 0, 1 and 2 alleles is the basis of genetic relatedness among individuals.

Sharing identical-by-descent (IBD) is the sharing of alleles IBS on account of shared ancestry (having a common ancestor). That is what we are interested in when studying affected individuals in families, since the idea is that the disease locus has been transmitted from the common ancestor.

If two siblings both share the genotype AB, and their mother has AB and their father has BB, then you know that they share A IBD, but you don't know if they inherited the same allele from their father or if that one is just IBS. You gain more information with multi-allelic markers (observe that the maximum number alleles transmitted from two parents to their offspring is four), and by genotyping several markers in several members in the family. Yet, in reality, it is rare to have a fully informative data, i.e. when you can tell for sure that an allele shared in two individuals is IBD. *We need to rely on good estimations of IBD sharing based on IBS status.*

3.3.4 Linkage

Non-parametric linkage is basically a test of excess IBD sharing in affected related individuals. *The maximum logarithm of the odds (LOD)-score probability*, is based on IBS-sharing among affected individuals, the genotypes of all family members and the population frequency of the allele. Under no linkage we assume that the mean sharing in two affected siblings is one allele.

In *parametric linkage* we add a few parameters. The parametric model includes assumptions about the disease model and disease-allele frequency. The disease model is specified in three parameters of disease-penetrance: 1) the risk of being affected in individuals not carrying the risk allele. 2) The risk of disease if carrying one copy of the risk-allele. 3) The risk of disease if carrying two copies of the risk-allele. Traditionally, a recessive model assumes only a risk for disease if carrying two alleles and a dominant model assumes that the risk is equal if carrying one or two alleles. An additive model is something in-between. Still you may want to specify that even if carrying a dominant

risk allele you have a chance to remain unaffected which is modulated by changing the penetrance of the disease allele.

However, we do not assume that we've managed to genotype the exact spot where the disease locus is located. With parametric linkage we try to estimate where that spot is. Two loci close to each other are likely to be inherited (linked!) together, while two distant loci may end up on different chromatides during meiosis and thus only one will be transmitted to the offspring. The cross overs taking place in the germ cells during meiosis may result in recombination. Two affected individuals from the same family may carry allele A at a locus, while a third affected individual in that family carries allele B; we would then call the first two non-recombinants and the third recombinant. In the same family, we have also genotyped unaffected individuals, but we assume that they have inherited the opposite allele at the disease loci, therefore B carriers in the unaffected are considered non-recombinant and A-carriers recombinant. Note that the example is simplified by only discussing one (and not both) chromosome per individual. Using the *recombinant fraction* guides us to the true disease locus position; if everybody is non-recombinant we are probably at the right spot. The greater the fraction of recombinants, the further away we are. The deviation from the null distribution at a locus then becomes a function of the distance to the true disease locus, which is also included in the LOD-score calculation.

One important thing that distinguish linkage from association studies is that in association studies you always regard one allele as being the risk allele in all affected, while in linkage you assume that the disease locus may be inherited together with allele A in family one and allele B in family two (allele A and B being alleles of the same locus).

The analysis in paper IV is based on both non-parametric and parametric linkage with different models. The analysis was based on bi-allelic SNPs covering the whole genome. On top of pruning this SNPset for errors we also excluded non-informative SNPs. Yet, the analysis on single SNPs would be meaningless since we randomly would see excess sharing here and there considering the large number of SNPs tested. Thus, the calculations of IBD sharing also involved the estimation of haplotypes (the sequence on a chromosome of consecutive SNP-alleles).

3.3.5 Segmental sharing

Also in paper III the basic idea is that the disease locus is shared IBD in affected individuals. However, even though we used the same platform for genotyping, the design was by necessity completely different. So can we make assumptions about IBD sharing, based on IBS sharing in distantly related individuals?

Regarding SNPs, the IBS sharing of an allele in two individuals is very likely. However, the sharing of a long segment of consecutive alleles is highly unlikely, at least if the individuals are distantly related, each allele has a modest frequency in the population and the loci are inherited independently of each other. How all these requirements were addressed in our work is discussed in detail in paper III; in this background section I'll just go through the basics of segmental sharing as constructed in Plink 1.05[22; 23] .

The advantage of the Plink way of looking at sharing is that the program independently considers the two dimensions of the probability of sharing; Thomas and I conceptualized this in terms of the "length" and the "width" (see figure 14 and 15). The "length" refers to how long a shared segment needs to be to make it in highly unlikely that it merely is just IBS. Plink considers both the physical length measured in kilobases (kb) and the number of (independent) consecutive SNPs. "Width" refers to the fact that the probability of a segment being shared *by chance* decreases with the number of chromosomes sharing it.

In Plink, you can by using permutation obtain a probability value for the excess sharing in cases compared to controls. However, in an affected-only design no probability value can be calculated in Plink. We instead used a formula[25] to calculate the probability of sharing among n number of chromosomes (individuals) given the pedigree structure and correcting for the fact that it was a whole genome scan. The formula is discussed in detail in the paper.

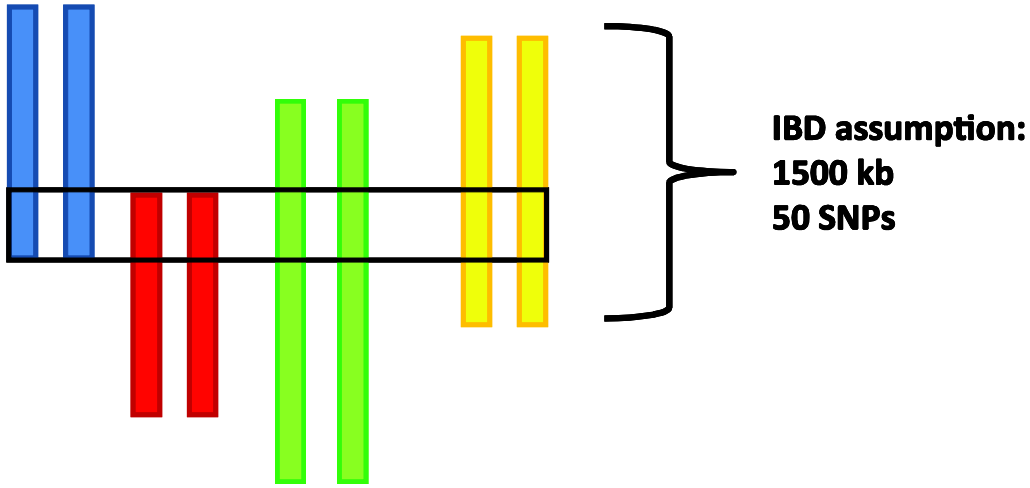


Figure 14. Homozygous sharing. Each color symbolizes chromosomes from the same individual. First, long runs of homozygosity within individuals are retrieved. These intra-individual homozygous segments are then compared inter-individually to identify segments overlapping between individuals. The overlaps can be identical between individuals but different individuals can also be homozygote for different alleles in the segment. Information about that is also available in the output from the Plink analysis and in paper III we consider only allelically matching overlaps.

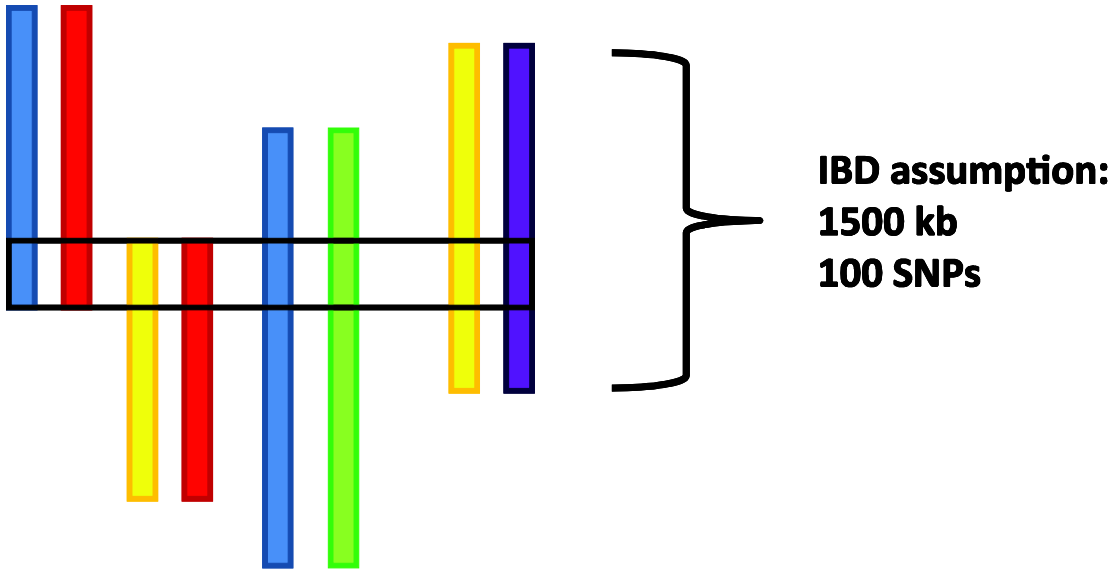


Figure 15. Heterozygous sharing. Each color symbolizes chromosomes from the same individual. First long runs of heterozygous sharing is retrieved between pairs of individuals. These pair-wise heterozygous segments are then compared between pairs to identify overlapping segments. In the example, in the figure, four pairs share the same segment, but it is shared among five individuals. In paper III, we considered only allelically matching overlaps.

4 ACKNOWLEDGEMENTS

Stort tack till min huvudhandledare Thomas Masterman, för att du alltid lyssnat intresserat på mina idéer och åsikter. Du har en förmåga att lyssna öppensinnet som få besitter. Du har varit min guide. Våra diskussioner har varit oändligt många och rört de flesta aspekter av avhandlingsarbetet men även mer filosofiska sidor av vetenskapen, språk och personliga ämnen.

Stort tack till min handledare Professor Jan Hillert för att jag fick en position i din grupp. Din sunda skepsis till forskningsidéer har tvingat mig att vässa mina argument och tänka några steg längre. Vi delar sällan åsikt även om vi med åren kommit att förstå varandra ganska väl. Jag är oändligt tacksam för att jag fått friheten att utveckla mina idéer under ditt beskydd.

Stort tack till min handledare Anne-Marie Landtblom, för din enorma entusiasm. Ditt brinnande engagemang för forskning och för att studera, vad du definierar som, atypisk MS fick mig på kroken för ett litet projekt rörande OCB-negative MS. Och det var så det hela började…. Du har också varit ett fantastiskt stöd under resans gång i Lysviks-projektet. Du har ett enormt engagemang för patienterna i Lysvik med en massa pågående forskningsprojekt. Det är sant inspirerande!

Stort tack till min handledare Ingrid Kockum, för att du kom in och med goda råd styrde upp och såg till att jag blev klar. Utan dina praktiska handfasta råd och all den praktisk hjälp jag fått av dig hade jag inte kommit i hamn redan nu. Din insats i sista artikeln i avhandlingen är ovärderlig! Jag har verkligen uppskattat både våra djupa genetik-diskussioner och de kloka råd du delar med dig av rörande alla aspekter av tillvaron.

Ett jättestort tack till alla människor som genom blodprov bidragit till forskningen!

Utan alla fantastiska kollegor hade inte denna avhandling blivit av:

Tack Izaura “Iza” Lima Bomfim, du har fått mig att växa som forskare och som människa.

Tack Boel Brynedal för alla våra temperamentsfulla diskussioner som förgyllt forskartillvaron. Vi började nästan samtidigt och då hade vi väldigt olika syn på saker och ting, men nu så har vi möts och håller rätt ofta med varandra. Det kanske är ett tecken på att tiden är inne för att finna nya kollegiala sammanhang. Jag önskar dig all lycka till i Boston!

Tack Frida Lundmark för både vetenskapliga och förtroliga samtal. Det blev tomt utan dig sedan du disputerat. Jag hoppas vi kommer ses framöver och gå barnvagnspromenader.

Tack Eva Greiner för god vänskap och gott samarbete som inneburit många diskussioner som berikats av dina kliniska kunskaper.

Tack Jenny Link. En dag var du bara en av oss! Glad och alltid otroligt hjälpsam, nyfiken och kunskapstörstande.

Wangko Lundström, vet inte om det är Göteborgaren i dig men din humor och skärpa satte guldkant på tillvaron i gruppen, tack! All lycka till med dina doktorandstudier på andra sidan atlanten.

Tack Virginija Karrenbauer för vårt lilla team som Thomas doktorander. Våra olika kunskaper kompletterar varandra bra. Du är så klok att prata med!

Homayoun Roshanisefat, tack för ditt stora engagemang för Huddinge familjen.

Tack MS immunologi/virologi gruppen för gränsöverskridande och berikande diskussioner: Anna Fogdell-Hahn, Malin Lundkvist, Rasmus Gustafsson, Ajith Sominanda, Mathula Thangaraj och Jenny Ahlqvist.

Tack Eva Lindström, Kristina Duvefelt och Helena Modin, det gamla MS genetik-gänget för det arv vi fått förvalta.

Tack Elin Karlberg och Helga Westerlind, för att ni tar vid och engagerar er i MS forskningen.

Tack Eva Johansson, Marjan Jahanpanah och Annelie Porsborn, för trevliga samtal i vårt lilla lunchrum och på pick-nick filten i gröngräset utanför sjukhuset.

Tack Merja Kanerva, Cecilia Svaren-Quiding, Ingegerd Löfving Arvholm, Anna Mattsson, Faezeh Vejdani för trevlig kollegial samvaro och hjälp med praktiska saker i labbet.

Tack Gunnel Larsson och Yvonne Sjölind, för hjälp med administration.

Stort tack till alla nuvarande och före detta kollegor som bidragit till de vetenskapliga diskussionerna: Kristina Gottberg, Sverker Johansson, Christina Sjöstrand, Konstantinos ”Kosta” Kostulas, Leszek Stawiarz, Raymond Press, Sten Fredriksson, Lotta Ytterberg, Lotta Widén Holmqvist, Lena von Koch, Mabel Cruz, Åke Seigert, Sebastian Yakisich, Orus Rot, Andreia Gomes och Yassir Hussein.

Tack Anna Nilzén, för din stora hjälpsamhet när en hel klan av patienter strömmande in för att lämna forskningsprov.

Tack Benjamin Lamb för din hjälp med sekvenseringen av PLP1 genen. Det var väldigt trevligt att ha dig som student och med din kliniska bakgrund hade du även mycket att lära mig.

Tack Stefanie Binzer, för att jag fick handleda dig under ditt Färöarna projekt. Det var en jättespännande och lärorik resa. Jag hoppas på fortsatt samarbete.

Också en hel skara människor utanför Neurologen möjliggjorde detta avhandlingsarbete:

Tack till Margarita Callander, Inger Boström och alla som på olika sätt varit engagerade i Lysviks projektet.

Stort tack till våra släktforskare i Lysviks projektet som möjliggjorde denna studie. Särskilt tack till Arne och Moira Linnarud, det är alltid lika roligt att åka och hälsa på er.

Marco Zucchelli, thank you for statistical help.

Jerome Wojcik, thank you for helping me with the sporadic case-control material in the Lysvik project.

Tack Magnus Lekman, för genetik-diskussioner i allmänhet och kopplings-diskussioner i synnerhet och för att jag fick låna din superdator.

Tack Gudrun J. Bergman, för alla diskussioner kring statistik och genetik.

Tack alla på plan 00 för att ha välkomnat våran grupp.

Tack Lina Rosvall, för intressanta epi och biostatdiskussioner, promenader och vänskap!

Tack Lotta Fredholm, för att jag fick göra praktik hos dig på Forskning&Framsteg och för vårt fortsatta samarbete som jag ser mycket fram emot.

Tack Maria Kvarnström, för att du smittade mig med ditt intresse för MS-forskning. Du var min första handledare för ett forskningsprojekt.

Tack MedBi-utbildning i Linköping för den grund jag fick att utgå ifrån som forskare.

Tack Linis-gänget: Ullis, Pernilla, Lovisa, Madeleine, Nettan, Maria, Nina och Bodil. Det är skönt med vänner utanför KI som ändå förstår exakt vad det är för skumt man sysslar

med när man doktorerar. Nu sprids vi över världen på olika post-doc men jag är övertygad om att vänskapen ändå består!

Csilla och Johan, jag saknar er så nu när ni är på post-doc i Japan. Tack för alla våra äventyr.

Stort tack till barndomsvänner, övriga vänner och familj som stöttat mig dessa år! Särskilt tack till min älskade mamma och mina älskade systrar Katarina och Sofia. Tack också till min framlidna pappa och framlidna morföräldrar som betytt oerhört mycket för mig.

Min tacksamhet mot dig Johan är större än vad jag kan uttrycka i ord. Du har varit fantastisk dessa månader då jag har varit helt uppslukat av avhandlingen och du har stöttat mig i alla år även om du ibland suckat över att du tvunget skulle fastna för en forskare.

Tack Joel (det är ditt arbetsnamn), du har gett mig energi och fokus. Längtar efter dig!

5 REFERENCES

1. Rothman K (2002) Epidemiology : an introduction. Oxford University Press, New York
2. Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallstrom 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 39:1108-1113
3. Andersson M, Alvarez-Cermeno J, Bernardi G, Cogato I, Fredman P, Frederiksen J, Fredrikson S, Gallo P, Grimaldi LM, Gronning M, et al. (1994) Cerebrospinal fluid in the diagnosis of multiple sclerosis: a consensus report. J Neurol Neurosurg Psychiatry 57:897-902
4. Fukazawa T, Kikuchi S, Sasaki H, Hamada K, Hamada T, Miyasaka K, Tashiro K (1998) The significance of oligoclonal bands in multiple sclerosis in Japan: relevance of immunogenetic backgrounds. J Neurol Sci 158:209-214
5. Kikuchi S, Fukazawa T, Niino M, Yabe I, Miyagishi R, Hamada T, Hashimoto SA, Tashiro K (2003) HLA-related subpopulations of MS in Japanese with and without oligoclonal IgG bands. Human leukocyte antigen. Neurology 60:647-651
6. Modin H, Masterman T, Thorlacius T, Stefansson M, Jonasdottir A, Stefansson K, Hillert J, Gulcher J (2003) Genome-wide linkage screen of a consanguineous multiple sclerosis kinship. Mult Scler 9:128-134
7. 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 50:121-127
8. Tintore M, Rovira A, Martinez MJ, Rio J, Diaz-Villoslada P, Brieva L, Borras C, Grive E, Capellades J, Montalban X (2000) Isolated demyelinating syndromes: comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis. AJNR Am J Neuroradiol 21:702-706
9. Barkhof F, Filippi M, Miller DH, Scheltens P, Campi A, Polman CH, Comi G, Ader HJ, Losseff N, Valk J (1997) Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. Brain 120 (Pt 11):2059-2069
10. Roxburgh RH, Seaman SR, Masterman T, Hensiek AE, Sawcer SJ, Vukusic S, Achiti I, et al. (2005) Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology 64:1144-1151
11. Marrosu MG, Muntoni F, Murru MR, Spinicci G, Pischedda MP, Goddi F, Cossu P, Pirastu M (1988) Sardinian multiple sclerosis is associated with HLA-DR4: a serologic and molecular analysis. Neurology 38:1749-1753
12. Coraddu F, Reyes-Yanez MP, Parra A, Gray J, Smith SI, Taylor CJ, Compston DA (1998) HLA associations with multiple sclerosis in the Canary Islands. J Neuroimmunol 87:130-135
13. Saruhan-Direskeneli G, Esin S, Baykan-Kurt B, Ornek I, Vaughan R, Eraksoy M (1997) HLA-DR and -DQ associations with multiple sclerosis in Turkey. Hum Immunol 55:59-65
14. Orth-Gomer K, Lewandrowski E, Westman LP, Wang HX, Leineweber C (2005) [Who can manage until the age of 65? Not the women--more than a half of all women stop working prematurely]. Lakartidningen 102:2296-2299, 2301, 2302-2293
15. Landtblom AM, Riise T, Boiko A, Soderfeldt B (2002) Distribution of multiple sclerosis in Sweden based on mortality and disability compensation statistics. Neuroepidemiology 21:167-179
16. Giedraitis V, Modin H, Callander M, Landtblom AM, Fossdal R, Stefansson K, Hillert J, Gulcher J (2003) Genome-wide TDT analysis in a localized

population with a high prevalence of multiple sclerosis indicates the importance of a region on chromosome 14q. Genes Immun 4:559-563
17. Callander M, Landtblom AM (2004) A cluster of multiple sclerosis cases in Lysvik in the Swedish county of Varmland. Acta Neurol Scand 110:14-22
18. Virolainen E, Wessman M, Hovatta I, Niemi KM, Ignatius J, Kere J, Peltonen L, Palotie A (2000) Assignment of a novel locus for autosomal recessive congenital ichthyosis to chromosome 19p13.1-p13.2. Am J Hum Genet 66:1132-1137
19. Houwen RH, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380-386
20. Puffenberger EG, Kauffman ER, Bolk S, Matise TC, Washington SS, Angrist M, Weissenbach J, Garver KL, Mascari M, Ladda R, et al. (1994) Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet 3:1217-1225
21. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci U S A 104:19942-19947
22. Purcell S PLINK 1.05. http://pngumghharvardedu/purcell/plink
23. 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 81:559-575
24. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF (2008) Runs of homozygosity in European populations. Am J Hum Genet 83:359-372
25. Durham LK, Feingold E (1997) Genome scanning for segments shared identical by descent among distant relatives in isolated populations. Am J Hum Genet 61:830-842
26. Bernardinelli L, Murgia SB, Bitti PP, Foco L, Ferrai R, Musu L, Prokopenko I, Pastorino R, Saddi V, Ticca A, Piras ML, Cox DR, Berzuini C (2007) Association between the ACCN1 gene and multiple sclerosis in Central East Sardinia. PLoS ONE 2:e480
27. Waldmann R, Champigny G, Voilley N, Lauritzen I, Lazdunski M (1996) The mammalian degenerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in Caenorhabditis elegans. J Biol Chem 271:10433-10436
28. Waldmann R, Voilley N, Mattei MG, Lazdunski M (1996) The human degenerin MDEG, an amiloride-sensitive neuronal cation channel, is localized on chromosome 17q11.2-17q12 close to the microsatellite D17S798. Genomics 37:269-270
29. Baron A, Deval E, Salinas M, Lingueglia E, Voilley N, Lazdunski M (2002) Protein kinase C stimulates the acid-sensing ion channel ASIC2a via the PDZ domain-containing protein PICK1. J Biol Chem 277:50463-50468
30. Deval E, Salinas M, Baron A, Lingueglia E, Lazdunski M (2004) ASIC2b-dependent regulation of ASIC3, an essential acid-sensing ion channel subunit in sensory neurons via the partner protein PICK-1. J Biol Chem 279:19531-19539
31. 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:e42
32. D'Alfonso S, Nistico L, Zavattari P, Marrosu MG, Murru R, Lai M, Massacesi L, Ballerini C, Gestri D, Salvetti M, Ristori G, Bomprezzi R, Trojano M, Liguori M, Gambi D, Quattrone A, Fruci D, Cucca F, Richiardi PM, Tosi R (1999) Linkage analysis of multiple sclerosis with candidate region markers in Sardinian and Continental Italian families. Eur J Hum Genet 7:377-385

33. Sawcer S, Ban M, Maranian M, Yeo TW, Compston A, Kirby A, Daly MJ, et al. (2005) A high-density screen for linkage in multiple sclerosis. Am J Hum Genet 77:454-467
34. 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 13:464-468
35. 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, Figa-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 9:438-444
36. Goertsches R, Comabella M, Navarro A, Perkal H, Montalban X (2005) Genetic association between polymorphisms in the ADAMTS14 gene and multiple sclerosis. J Neuroimmunol 164:140-147
37. Goertsches R, Baranzini SE, Morcillo C, Nos C, Camina M, Oksenberg JR, Montalban X, Comabella M (2008) Evidence for association of chromosome 10 open reading frame (C10orf27) gene polymorphisms and multiple sclerosis. Mult Scler 14:412-414
38. Howe CL, Adelson JD, Rodriguez M (2007) Absence of perforin expression confers axonal protection despite demyelination. Neurobiol Dis 25:354-359
39. Malmestrom C, Lycke J, Haghighi S, Andersen O, Carlsson L, Wadenvik H, Olsson B (2008) Relapses in multiple sclerosis are associated with increased CD8+ T-cell mediated cytotoxicity in CSF. J Neuroimmunol 196:159-165
40. Hiraki K, Park IK, Kohyama K, Matsumoto Y (2009) Characterization of CD8-positive macrophages infiltrating the central nervous system of rats with chronic autoimmune encephalomyelitis. J Neurosci Res 87:1175-1184
41. Baron A, Schaefer L, Lingueglia E, Champigny G, Lazdunski M (2001) Zn2+ and H+ are coactivators of acid-sensing ion channels. J Biol Chem 276:35361-35367
42. Baron A, Waldmann R, Lazdunski M (2002) ASIC-like, proton-activated currents in rat hippocampal neurons. J Physiol 539:485-494
43. Chu XP, Wemmie JA, Wang WZ, Zhu XM, Saugstad JA, Price MP, Simon RP, Xiong ZG (2004) Subunit-dependent high-affinity zinc inhibition of acid-sensing ion channels. J Neurosci 24:8678-8689
44. Gao J, Wu LJ, Xu L, Xu TL (2004) Properties of the proton-evoked currents and their modulation by Ca2+ and Zn2+ in the acutely dissociated hippocampus CA1 neurons. Brain Res 1017:197-207
45. Wang W, Duan B, Xu H, Xu L, Xu TL (2006) Calcium-permeable acid-sensing ion channel is a molecular target of the neurotoxic metal ion lead. J Biol Chem 281:2497-2505
46. Staruschenko A, Dorofeeva NA, Bolshakov KV, Stockand JD (2006) Subunit-dependent cadmium and nickel inhibition of acid-sensing ion channels. J Neurobiol 67:97-107
47. Banna NR, Jabbur SJ (1970) Increased transmitter release induced by convulsant phenols. Brain Res 20:471-473
48. Abecasis GR MERLIN-1.1.2. http://wwwsphumichedu/csg/abecasis/merlin/indexhtml
49. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101
50. Gorman MP, Golomb MR, Walsh LE, Hobson GM, Garbern JY, Kinkel RP, Darras BT, Urion DK, Eksioglu YZ (2007) Steroid-responsive neurologic relapses in a child with a proteolipid protein-1 mutation. Neurology 68:1305-1307
51. Warshawsky I, Rudick RA, Staugaitis SM, Natowicz MR (2005) Primary progressive multiple sclerosis as a phenotype of a PLP1 gene mutation. Ann Neurol 58:470-473
52. Osaka H, Kawanishi C, Inoue K, Onishi H, Kobayashi T, Sugiyama N, Kosaka K, Nezu A, Fujii K, Sugita K, Kodama K, Murayama K, Murayama S, Kanazawa I, Kimura S (1999) Pelizaeus-Merzbacher disease: three novel mutations and implication for locus heterogeneity. Ann Neurol 45:59-64

53. Lee JA, Carvalho CM, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131:1235-1247
54. Jersild C, Fog T (1972) Histocompatibility (HL-A) antigens associated with multiple sclerosis. Acta Neurol Scand Suppl 51:377
55. Goldstein DB (2009) Common genetic variation and human traits. N Engl J Med 360:1696-1698
56. Kraft P, Hunter DJ (2009) Genetic risk prediction--are we there yet? N Engl J Med 360:1701-1703
57. Hirschhorn JN (2009) Genomewide association studies--illuminating biologic pathways. N Engl J Med 360:1699-1701
58. Hardy J, Singleton A (2009) Genomewide association studies and human disease. N Engl J Med 360:1759-1768
59. Lennon VA, Kryzer TJ, Pittock SJ, Verkman AS, Hinson SR (2005) IgG marker of optic-spinal multiple sclerosis binds to the aquaporin-4 water channel. J Exp Med 202:473-477
60. Lennon VA, Wingerchuk DM, Kryzer TJ, Pittock SJ, Lucchinetti CF, Fujihara K, Nakashima I, Weinshenker BG (2004) A serum autoantibody marker of neuromyelitis optica: distinction from multiple sclerosis. Lancet 364:2106-2112
61. Saikali P, Cayrol R, Vincent T (2009) Anti-aquaporin-4 auto-antibodies orchestrate the pathogenesis in neuromyelitis optica. Autoimmun Rev
62. Kallio SP, Jakkula E, Purcell S, Suvela M, Koivisto K, Tienari PJ, Elovaara I, Pirttila T, Reunanen M, Bronnikov D, Viander M, Meri S, Hillert J, Lundmark F, Harbo HF, Lorentzen AR, De Jager PL, Daly MJ, Hafler DA, Palotie A, Peltonen L, Saarela J (2009) Use of a genetic isolate to identify rare disease variants: C7 on 5p associated with MS. Hum Mol Genet 18:1670-1683
63. 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 (2007) Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. Nat Genet 39:1083-1091
64. Lalive PH (2008) Autoantibodies in inflammatory demyelinating diseases of the central nervous system. Swiss Med Wkly 138:692-707
65. McQualter JL, Bernard CC (2007) Multiple sclerosis: a battle between destruction and repair. J Neurochem 100:295-306
66. Gao J, Duan B, Wang DG, Deng XH, Zhang GY, Xu L, Xu TL (2005) Coupling between NMDA receptor and acid-sensing ion channel contributes to ischemic neuronal death. Neuron 48:635-646
67. Suidan GL, McDole JR, Chen Y, Pirko I, Johnson AJ (2008) Induction of blood brain barrier tight junction protein alterations by CD8 T cells. PLoS ONE 3:e3037
68. Centonze D, Muzio L, Rossi S, Cavasinni F, De Chiara V, Bergami A, Musella A, D'Amelio M, Cavallucci V, Martorana A, Bergamaschi A, Cencioni MT, Diamantini A, Butti E, Comi G, Bernardi G, Cecconi F, Battistini L, Furlan R, Martino G (2009) Inflammation triggers synaptic alteration and degeneration in experimental autoimmune encephalomyelitis. J Neurosci 29:3442-3452
69. Xu J, Xia J (2006) Structure and function of PICK1. Neurosignals 15:190-201
70. Kurtzke JF (1983) Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33:1444-1452
71. Confavreux C, Vukusic S (2006) Age at disability milestones in multiple sclerosis. Brain 129:595-605
72. Compston A (1997) Genetic epidemiology of multiple sclerosis. J Neurol Neurosurg Psychiatry 62:553-561
73. Sawcer S (2006) A new era in the genetic analysis of multiple sclerosis. Curr Opin Neurol 19:237-241
74. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC (2003) Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci U S A 100:12877-12882

75. Sadovnick AD, Ebers GC, Dyment DA, Risch NJ (1996) Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. Lancet 347:1728-1730
76. Kurtzke JF, Heltberg A (2001) Multiple sclerosis in the Faroe Islands: an epitome. J Clin Epidemiol 54:1-22
77. Trapp BD, Nave KA (2008) Multiple sclerosis: an immune or neurodegenerative disorder? Annu Rev Neurosci 31:247-269
78. Bo L, Vedeler CA, Nyland HI, Trapp BD, Mork SJ (2003) Subpial demyelination in the cerebral cortex of multiple sclerosis patients. J Neuropathol Exp Neurol 62:723-732
79. Trapp BD, Peterson J, Ransohoff RM, Rudick R, Mork S, Bo L (1998) Axonal transection in the lesions of multiple sclerosis. N Engl J Med 338:278-285
80. Lassmann H (2007) Multiple sclerosis: is there neurodegeneration independent from inflammation? J Neurol Sci 259:3-6
81. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F (2004) Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. Brain Pathol 14:164-174
82. 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, Weinshenker BG, Wolinsky JS (2005) Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Ann Neurol 58:840-846
83. Ahlbom A, Alfredsson L (2005) Interaction: A word with two meanings creates confusion. Eur J Epidemiol 20:563-564