

From the Neuroimmunology Unit,
Department of Clinical Neuroscience
Karolinska Institutet, Stockholm, Sweden

COMMON VIRUSES AND HOST GENE INTERACTIONS IN MULTIPLE SCLEROSIS

Emilie Sundqvist



Stockholm 2013

Cover photo © Emilie Sundqvist, 2013.

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

Published by Karolinska Institutet. Printed by Larserics Digital Print AB, Stockholm Sweden.

© Emilie Sundqvist, 2013
ISBN 978-91-7549-067-0

ABSTRACT

Multiple sclerosis (MS) is a neurological disorder, characterised by demyelination and inflammation of the central nervous system, leading to sensory and motor symptoms. MS is thought to be complex disease, with both environmental and genetic risk factors underlying disease susceptibility. Epstein-Barr virus (EBV) and cytomegalovirus (CMV) infections are two environmental risk factors, one with a robust association to MS (EBV) and one where the results have been more inconclusive (CMV). The strongest genetic risk factors lies within the HLA genes, with *HLA-DRB1*15* as the strongest susceptibility factor, and *HLA-A*02*, as the most protective genetic factor.

In **paper I**, the role of EBV infection, and the interaction with *HLA-DRB1*15* and *HLA-A*02* was studied. Anti-EBNA1 IgG was measured, as was IgG antibodies towards 5 different epitopes of EBNA1. High levels of EBNA1 385-420 IgG antibodies were strongly associated with MS, independent of EBNA1 IgG antibody level. There was interaction on the additive scale between EBNA1 385-420 IgG and *HLA-DRB1*15* and absence of *HLA-A*02*. In **paper II**, we tried to replicate findings by Simon et al, where they found interaction on the multiplicative scale between EBNA1 IgG levels and smoking (never/ever), but our analysis showed no such interaction.

In **paper III**, the association between CMV and MS was studied, yielding a significant negative association between CMV and MS. To further validate our results, a meta-analysis of published retrospective studies was performed, which provided a similar negative association, supporting our results.

In **paper IV**, the focus shifted from the association of viruses to MS, to dissecting the host genetic influence on anti-JCV seropositivity and anti-JCV antibody levels. JC virus is the virus responsible for Progressive multifocal leukoencephalopathy (PML), a rare but potentially fatal side-effect seen in MS-patients treated with natalizumab. A meta-analysis of three genome wide association studies performed in two sets of MS cases, one Scandinavian and one German, and a set of Swedish controls, strongly indicated that the HLA class II region was involved in regulating anti-JCV antibody response, and anti-JCV antibody levels. Analysis of classically named HLA-alleles supported these findings. The alleles in the DRB1*15-DQB1*06:02-DQA1*01:02-haplotype were all strongly negatively associated with anti-JCV antibody status and low anti-JCV antibody levels. The alleles in the DRB1*13-DQB1*06:02-DQA1*01:03-haplotype were positively associated with anti-JCV antibody status. Several non-HLA loci were suggestively associated with anti-JCV antibody status and anti-JCV antibody levels ($p < 0.0001$). However, these findings will have to be replicated in an independent dataset.

This thesis highlights the interactions between environmental and genetic factors in modulating MS risk. It also shows that the HLA genes have a central role in the susceptibility to JCV infection.

LIST OF PUBLICATIONS

- I. **Epstein-Barr virus and multiple sclerosis: interaction with HLA**
SUNDQVIST E, Sundström P, Lindén M, Hedström AK, Aloisi F, Hillert J, Kockum I, Alfredsson L, Olsson T
Genes and Immunity. 2012 Jan;13(1):14-20
- II. **Lack of replication of interaction between EBNA1 IgG and smoking in risk for multiple sclerosis**
SUNDQVIST E, Sundström P, Lindén M, Hedström AK, Aloisi F, Hillert J, Kockum I, Alfredsson L, Olsson T
Neurology. 2012 Sep 25;79(13):1363-8
- III. **Cytomegalovirus seropositivity is negatively associated with multiple sclerosis**
SUNDQVIST E, Bergström T, Daialhosein H, Nyström M, Sundström P, Hillert J, Alfredsson L, Kockum I, Olsson T
Submitted to *Multiple Sclerosis Journal*
- IV. **The influence of host genetics on anti-JC virus serology status and titer in multiple sclerosis patients and controls**
SUNDQVIST EMILIE, Buck D, Warnke C, Albrecht E, Khademi M, Bomfim I, Fogdell-Hahn A, Alfredsson L, Bach Søndergaard H, Gieger C, Hillert J, International MS Genetics Consortium, Oturai A B, Carulli J P, Hemmer B, Kockum I, Olsson T
Manuscript

CONTENTS

1	Multiple Sclerosis	1
1.1	Genetic risk factors	2
1.1.1	HLA genes	2
1.1.2	Non-HLA genes	3
1.2	Environmental risk factors	4
1.2.1	Epstein - Barr virus	4
1.2.2	Cytomegalovirus	5
1.2.3	Other environmental risk factors	6
1.3	JC virus	7
1.4	the immune system	8
1.4.1	Cells and proteins of the immune system	8
1.4.2	Response to viral infections	9
1.4.3	The role of the immune system in MS	10
2	Epidemiological studies and statistical analysis	12
2.1	Cohort studies	12
2.2	Case-control studies	12
2.3	Odds Ratio	13
2.4	Regression analysis	14
2.5	Interaction	15
2.6	Statistical power	17
2.7	Meta – analysis	18
3	Genetics	20
3.1	Genetic variation and polymorphisms	20
3.1.1	The Human Leukocyte Antigen genes	20
3.2	Linkage disequilibrium	21
3.3	Hardy-Weinberg Equilibrium	23
3.4	How do we identify genetic risk factors?	23
3.4.1	Linkage studies	23
3.4.2	Candidate gene studies and GWAS	24
3.5	Population stratification	24
4	Materials and methods	26
4.1	Study populations	26
4.1.1	EIMS	26
4.1.2	IMSE I	27
4.1.3	Other datasets included in paper IV	27
4.2	Genotyping	27
4.2.1	Classical HLA-genotyping	27
4.2.2	Large scale SNP genotyping	28

4.3	Genotype imputation	28
4.4	Enzyme-linked immunosorbent assay (ELISA)	29
4.5	Statistical Analyses.....	30
4.5.1	Association and correlation tests	30
4.5.2	Interaction analysis.....	30
4.5.3	Conditional logistic regression.....	30
4.5.4	Principal component analysis	30
4.5.5	Genetic case-control analysis	30
4.5.6	Meta-analyses.....	30
4.5.7	Manhattan plots, marker annotation, etc	31
5	Aims of this thesis	32
6	Results and discussion.....	33
6.1	Common viruses in MS (Papers I – III)	33
6.1.1	Results from paper I.....	33
6.1.2	Results from paper II.....	34
6.1.3	Results from paper III.....	34
6.1.4	Discussion on papers I – III.....	35
6.2	Host genetics (Paper IV)	39
6.2.1	Results from paper IV	39
6.2.2	Discussion on paper IV.....	40
7	Concluding remarks.....	43
8	Future perspectives.....	45
9	Acknowledgements.....	47
10	References.....	50

LIST OF ABBREVIATIONS

APC	Antigen presenting cell
BMI	Body mass index
CD	Cluster of differentiation
CI	Confidence interval
CIS	Clinically isolated syndrome
CMV	Cytomegalovirus
CNS	Central nervous system
CSF	Cerebrospinal fluid
DAMP	Danger associated molecular pattern
DC	Dendritic cell
EA	Early antigen
EBV	Epstein-Barr virus
EBNA1	Epstein-Barr nuclear antigen 1
ELISA	Enzyme-linked immunosorbent assay
LD	Linkage disequilibrium
GWAS	Genome wide association study
HLA	Human leukocyte antigen
HWE	Hardy-Weinberg equilibrium
Ig	Immunoglobulin
IM	Infectious mononucleosis
IMSGC	International MS genetics consortium
JCV	JC virus
MAF	Minor allele frequency
MHC	Major histocompatibility complex
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
NLRs	Nucleotide oligomerisation domain-like receptors
NK cell	Natural killer cell
nOD values	Normalised optical density values
OCBs	Oligoclonal bands
OR	Odds ratio
PAMP	Pathogen associated molecular pattern
PML	Progressive multifocal leukoencephalopathy
PPMS	Primary progressive MS
PC	Principal component
PCA	Principal component analysis
RA	Rheumatoid arthritis
RLRs	Retinoic acid-inducible gene 1-like receptors

RRMS	Relapse-remitting MS
SNP	Single nucleotide polymorphism
SPMS	Secondary progressive MS
SSPs	Sequence specific primers
TLR	Toll-like receptor
VCA	Viral capsid antigen
VDR	Vitamin D receptor
VDRE	Vitamin D response element
WTCCC2	Wellcome Trust case-control consortium 2

1

MULTIPLE SCLEROSIS

Multiple Sclerosis, MS (OMIM #126200) is a neurological disorder, characterised by inflammation and demyelinating lesions in the central nervous system (CNS), axonal loss and progressive build-up of sclerotic plaques. MS affects twice as many women as men, and the age of onset is usually in the late twenties or early thirties. The average life time risk for women is around 1 in 400 [1], but in high-risk populations, such as in Northern Europe, it can be around 1 in 200 [2]. The prevalence in Sweden has recently been estimated to 188.9/100,000 individuals, and a female to male ratio of 2.35:1 [3]. Yearly incidence is around 600 new cases per year.

Around 80 % of patients present with a type of relapsing-remitting disease (RRMS), with bouts of disease activity between periods of remission, but later, after two to three decades, enter a secondary progressive stage (SPMS). A minority of patients present with a primary progressive type of disease (PPMS), with no remitting phases and a progressively worsening of the disease. There are also patients who only experience one event, and are referred to as belonging to the clinically isolated syndrome (CIS) group of patients; some of these later develop definite MS, and the factors contributing to the conversion from CIS to MS are often studied.

The risk of progressing to MS is increased if the patient also presents with brain lesions on MRI (magnetic resonance imaging) scans [4]. During relapses, patients commonly experience symptoms from the motor, sensory, visual, and autonomic systems [5].

The MS diagnosis is defined by a set of criteria on clinical history and MRI test results, usually the McDonald criteria [6], and earlier also the Poser criteria were used [7]. According to the McDonald criteria from 2011, if a patient has had more than two attacks, with either more than two lesions, or one lesion but with sufficient evidence of a prior attack, no further tests are needed. If the patient does not fulfil these criteria, supporting events, such as the presence of inflammatory lesions in the brain, disseminated in space and time, and additional clinical attacks, are needed to make a definite diagnosis [6].

Antibodies of the IgG and IgM isotypes can be found in the cerebrospinal fluid (CSF) of many MS patients. They are usually visualised by running electrophoresis on samples from CSF in a gel, where the immunoglobulins will be seen as bands, so called oligoclonal bands (OCBs). In fact, these bands are extremely common among MS patients, and it is estimated that around 90% of all patients have IgG-OCBs at the

time of their first examination [8]. Testing of CSF for OCBs is no longer part of the McDonald criteria for RRMS, but is used in the diagnosis of PPMS [6]. Since autoantibodies against CNS-specific epitopes [9, 10], are present in MS-patients, and most of the successful therapies for RRMS are immunomodulatory drugs, RRMS, at least, can be considered to have a strong inflammatory, probably autoimmune, component. Whether the same is true for PPMS, is debated (reviewed in [11]).

The β -interferons of different brands are today the first-line drugs available for RRMS. It has been estimated that they reduce the number of relapses with about 30% (reviewed in [5]). A negative side effect of these therapies is the generation of neutralising antibodies, which bind to and inhibit the effects of the drug, thereby reducing efficacy. Therefore, patients are monitored for the development of neutralising antibodies [12]. Glatirameracetate is also used as a first line treatment with a similar of efficacy.

For many years there were no alternatives to the interferons, until the release of monoclonal antibodies (mAbs) as a therapy for MS. mAbs bind to specific molecules on the cell surface, and the mAbs available or in late phase clinical trials today, all target immune system related molecules [13]. Natalizumab was the first mAb therapy released for MS (see section 1.3). Now, also low weight oral drugs with immunomodulatory effects are entering the field. Fingolimod was released in 2012, and many others are in phase II and III clinical trials, and will hopefully be released soon. However, most drugs act broadly in the immune system with potentials for long-term risks and even higher degrees of efficacy is desired. This motivates further research on the causes of MS, to achieve both prevention and more selective therapeutic strategies.

1.1 GENETIC RISK FACTORS

MS is considered to be a complex disease, where genetic and environmental/lifestyle factors both contribute to disease susceptibility. Parents, siblings and children of MS-patients have a higher age-adjusted risk than second or third-degree relatives. For monozygotic twins the concordance rate is around 30% and for dizygotic twins around 7% [1].

1.1.1 HLA genes

The Human leukocyte antigen (HLA) molecules are proteins expressed on cell surfaces, and their function is to present foreign and endogenous peptides on the cell surface. In animals, the HLA counterpart is usually called major histocompatibility complex (MHC).

There are two major classes of HLA-molecules, class I and II. HLA class I molecules are expressed on almost all cell types, while HLA class II molecules are predominantly expressed on antigen presenting cells (APCs). Class I present cytosolic antigens on the cell surface, whereas the class II molecules present extracellular proteins. The cell types that can be activated by class I and II molecules also differ, class I mainly activates CD8+ cytotoxic T-cells, and class II activates CD4+ T-cells.

The HLA class II *HLA-DRB1*15* haplotype, also known as HLA-DR2 (*DQB1*06:02-DQA1*01:02-DRB1*15:01-DRB5*01:01*) has been associated with an increased risk for MS in several populations [14], with an odds ratio of around 3.08, but the OR varies across populations [15]. The HLA class I alleles *HLA-A*02*, has been shown to have a *DRB1*1501*-independent association with MS, with an odds ratio of approximately 0.73 [15-17]. Other HLA-alleles associated with MS are *DRB1*13:03* (OR=2.43), *DRB1*03:01* (OR=1.26) and *DRB1*08:01* (OR=1.18), among others [15, 18]. Through dense SNP fine-mapping of the extended HLA-region, a polymorphism in the HLA-G gene has been associated to MS in white Americans [19].

The HLA-region has been associated to many other autoimmune diseases, such as Type 1 Diabetes [20-22], ankylosing spondylitis [23], rheumatoid arthritis [22, 24], Crohn's disease [22], and systemic lupus erythematosus [22], to name a few. Although one HLA-gene can be associated with many autoimmune diseases, the alleles could differ between diseases. One allele can also be positively associated in one disease, but negatively associated with another disease, such as in the case of *DRB1*15* which is a strong risk factor in MS, but acts as a protective allele in type 1 diabetes.

1.1.2 Non-HLA genes

The search for non-HLA genes associated with MS has been going on since the early 1970s initially with little success (reviewed in [25]). It was in 2007, with the publication of the association between the *interleukin-7 receptor (IL7R)* and MS from two different groups [26, 27], and the 2007 MS genome wide association study (GWAS), which found associated alleles in *IL2RA* and *IL7R* [28], that things started to change. The latest MS GWAS found 52 loci associated with MS, 29 were novel loci not previously published [15]. With the still unpublished data from the Immunochip, the number of associated loci is likely to increase. The authors of the 2011 GWAS concluded that many of the associated markers were close to genes involved in the immune system, with an overrepresentation of genes involved in T-cell maturation [15]. It is important to note that of the established 59 non-HLA loci, have a much smaller impact on MS risk with odds ratios in the region of 1.05 to 1.3, compared to *HLA-DRB1*15* with OR \approx 3.

1.2 ENVIRONMENTAL RISK FACTORS

Smoking, lack of sun exposure, low levels of vitamin D and different types of human herpes viruses have all been suggested as environmental risk factors associated with MS, with varying degrees of evidence to back those claims. My studies largely focus on selected viral factors in the etiology of MS.

Epstein-Barr virus and cytomegalovirus are both from the family of human herpes viruses (*herpesviridae*), a family of large enveloped DNA viruses. Other common viruses belonging to this family are varicella zoster virus, which causes chicken pox, and the herpes simplex viruses, which cause cold sores, and in more severe cases, can cause herpes simplex encephalitis. Some of them have been linked to MS, presence of antibodies against measles, rubella and varicella zoster in CSF has been implicated as important in the conversion from CIS to RRMS [29].

1.2.1 Epstein - Barr virus

Epstein-Barr virus (EBV) is a common herpes virus and it has been estimated that more than 90 % of all individuals worldwide will become infected during their lifetime [30]. Most people are infected during early childhood, and the infection is usually asymptomatic. However, in around 30-50 % of all individuals infected during adolescence or young adulthood, infectious mononucleosis (IM) develops after the primary infection [30, 31]. IM symptoms include fever, sore throat, and fatigue.

B-cells are the primary target of EBV, but it can also infect epithelial cells in the oropharynx. During the primary infection the virus enters the tonsils and infects naive B-cells, which become activated, migrate to the germinal centres where they proliferate and differentiate into latently infected memory B-cells. These memory B-cells exit the tonsils and enter the blood stream. Only during cell-division do they express EBV-proteins and only Epstein Barr Nuclear antigen 1 (EBNA1). The memory B-cells can re-enter the tonsils, and differentiate into plasma cells, which also starts a new lytic cycle. During the lytic phase of the infection, the cells produce EBV-particles that can infect nearby cells and thus spread the infection. During this phase, the immune system starts to respond to the infection; CD8+ T-cells kill infected B-cells and plasma cells, and soluble antibodies bind to free virus particles and inhibit further spread. The difference between asymptomatic EBV infection and IM is thought to be that in the latter case, there is a much higher number of infected B-cells, and the symptoms are caused by the mass-killing of B-cells by cytotoxic T-cells. Why the number of infected B-cells would increase with age is not known (reviewed in [32]).

Eventually, the virus establishes a latent infection in B-cells. During the latent infection, parts of the EBV genetic program is shut down, the lymphocytes proliferate and home to sites (e.g. bone marrow) where the infection persists for the

rest of life of the host. Periodically there will be a reactivation of the lytic EBV genetic program, to yield new virions that will infect new cells and maintain a constant infection.

Antibodies against EBV-related proteins are seen in all phases of the infection. During an acute lytic phase, IgM and later IgG antibodies against viral capsid antigen (VCA) and early antigen (EA) can be detected. During the latency phase, antibodies against EBV nuclear antigens 1-5 (EBNA) dominate [31]. It has also been shown that individuals can have antibody responses towards different amino acid domains of EBNA1 [33], and that these are also associated with increased MS risk [33, 34]. The connection between MS and EBV has been studied extensively. MS patients often have elevated levels of antibodies towards EBNA1 compared to controls [35, 36], which we also observed in our Swedish case-control material (paper I). Past history of IM has also been linked to a higher risk developing MS [37, 38], lending further support to a role of EBV in MS etiology.

1.2.2 Cytomegalovirus

Cytomegalovirus (CMV) is another human herpes virus that can establish a life-long latent infection in the host. The primary infection is usually asymptomatic, although congenital infection is a major cause of hearing loss and other neurological impairments in newborns worldwide [39]. In rare cases CMV can cause CMV mononucleosis, with symptoms similar to infectious mononucleosis. The seroprevalence among women varies from 45-100% worldwide, and not surprisingly, the seroprevalence increases with age. Women also seem to have a higher seroprevalence than men (reviewed in [40]). There is also a socioeconomic difference in CMV seroprevalence, with higher rate of seropositive individuals among low and middle socioeconomic groups compared to high [40, 41].

CMV usually established latency in epithelial cells [42] and hematopoietic cells [43, 44]. *In vitro*, CD33+ cells were able to support CMV latency, and CD33+ cell activation also lead to viral reactivation and production of viral particles [44]. CMV is thought to play a role in immunosenescence, and the CD8+ T-cell response EBV and CMV can induce in elderly people differ [44]. Approximately 10% of all CD4+ and CD8+ memory T-cells in seropositive individuals are CMV-specific [42].

Like EBV, CMV has been studied in regard to MS before, but many studies were small and unfortunately, yielded only inconclusive results [45-48]. However, a negative association between CMV seropositivity has been reported in adult [49] and pediatric MS cases [50]. Anti-CMV antibodies have also been shown to have a positive effect on MRI progression and clinical outcomes [47]. Seropositivity has also been associated with a decreased time to relapse and increased number of relapses [51]. In a murine mouse model called Theiler's murine encephalitis, co-infection with

murine CMV ameliorated the disease, and increased motor performance in mice. There was also a positive immunomodulatory effect seen in the proportion of CD3+ cells reaching the brain [52].

1.2.3 Other environmental risk factors

In 2001, Hernan *et al* published a report from the Nurses' Health study, associating smoking with an increased risk for MS. Before that, studies had shown inconsistent results [53]. In 2011, Hedström *et al* published data on smoking as a risk factor for MS using cases and controls from the EIMS study (see section 4.1.1) [54]. The increase in risk was evident up to five years after smoking cessation. MS patients who are also current smokers have a higher risk for all-cause mortality, compared to patients who are non-smokers [55]. Interestingly, the use of Swedish snuff for more than 15 years was associated with a decreased risk for MS [54]. Passive smoking has also been associated with an increased risk for MS [56]. Further analysis on the use of snuff showed that a combination of snuff and smoking was negatively associated with MS, suggesting a protective effect of nicotine in MS development [57].

A further study analysing possible interactions with *HLA-DRB1*15* and *A*02*, the two strongest genetic risk factors for MS was subsequently performed, also using data from EIMS. There was an interaction between smoking and *HLA-DRB1*15* on the additive scale with departure from additivity (see section 2.5). However, this interaction was confined to individuals not carrying *A*02*. There was also a significant interaction between absence of *HLA-A*02* and smoking among *HLA-DRB1*15* positive individuals [58]. This suggests that smoking and HLA genes jointly predisposes to MS in some individuals.

Lack of sun exposure, residing at a high latitude, and low vitamin D levels have also been suggested as risk factors in MS. Data from our own materials indicate that low levels of sun exposure is a stronger risk factor than low vitamin D levels, hinting at a more complex mechanism than just low vitamin D levels affecting MS risk [59].

Vitamin D is produced in an inactive form in the skin by keratinocytes. Vitamin D3 (colecalciferol) is created when 7-dehydrocholesterol is exposed to UV-B radiation. Vitamin D3 is also supplied via the diet. In the liver, vitamin D3 is hydroxylated and becomes 25-hydroxyvitamin D3. In the kidney, CYP27B1, a cytochrome P450 enzyme, catalyzes the hydroxylation of 25-hydroxyvitamin D3 into 1, 25-hydroxyvitamin D3, the bioactive form of vitamin D. 1, 25-hydroxyvitamin D3 can bind to the vitamin D receptor (VDR). VDR forms a heterodimer with the retinoid X receptor and the complex can bind to vitamin D response elements (VDRE) in the genome.

Vitamin D is thus involved in the regulation of the immune system through binding of the VDR to VDRE in the regulatory region of immune genes [60, 61]. It has been shown in lymphoblastoid cell lines that there seem to be an overlap between VDR binding sites and active regulatory genomic regions. This overlap was not as strong in non-immune cell types [61]. Further support for the involvement of vitamin D in MS comes from the association of markers in the *CYP27B1* and *CYP24A1* regions, which has been observed in several GWAS and candidate studies [15, 62, 63]. *CYP24A1* codes for an enzyme involved in vitamin D metabolism.

Another risk factors involved in MS is high body mass index (BMI) before the age of twenty [64]. High BMI was also associated with MS in female paediatric MS patients [65]. Shift work at age 20 has also been associated with an increased risk for MS [66].

1.3 JC VIRUS

JC virus is a nonenveloped icosahedral polyoma virus with a double-stranded DNA genome. JCV can infect several types of cells; B-cells in blood and tonsils, CD34+ hematopoietic progenitor cells and primary tonsillar stromal cells [67]. The primary infection is usually asymptomatic and most are infected during childhood [68]. The virus maintains a persistent infection in the kidneys and bone marrow, and in some instances, also in the brain [69]. The seroprevalence varies world-wide, from 3-75% in various populations in early studies [70], but later studies suggest a prevalence of 40-60% in Europe [71-73]. In 1971, the virus was isolated from a male patient with progressive multifocal leukoencephalopathy (PML) [74], the first evidence of a causal link between JCV and PML.

Natalizumab is a monoclonal antibody used in the treatment of MS, that can block VLA-4 (very late activation antigen-4), leading to inhibition of cell migration across the blood brain barrier (reviewed in [13]). Unfortunately, one of the most severe side-effects of natalizumab treatment is PML, a possibly fatal demyelinating disease of the central nervous system. Before the introduction of natalizumab on the market, PML was mostly seen in immunocompromised patients (AIDS, leukaemia or organ transplant recipients). The mechanisms by which JCV causes PML in a proportion of MS patients remain unclear [75, 76]. Studies have shown that the risk for PML among patients treated with natalizumab increases with treatment duration, and that seronegative individuals have a very low risk for PML [77].

There is no cure for PML, although antiviral drugs have been tested. The most important thing to do is to improve the patient's immune system so it can clear the viral infection, by removing natalizumab through dialysis. By doing so, there is always the risk of the so called immune reconstitution inflammatory syndrome, in

which the immune system responds extremely forcefully, and actually causes more damage to the infected tissue instead of clearing the infection. Thus, the opportunistic infection causes problems in conditions with immunosuppression and in cases of treatment with a series of immunomodulatory drugs for inflammatory disease. It is therefore important to learn more about factors leading to a persistent JCV infection, which in turn may pave the way for therapeutic and/or preventive strategies. Host genetics is hereby one area to explore.

1.4 THE IMMUNE SYSTEM

The immune system is there to protect us from foreign invaders, such as parasites, viruses and bacteria which we encounter daily. It is when the immune system starts attacking self-antigens that autoimmunity can develop.

1.4.1 Cells and proteins of the immune system

The immune system is the native defence we are all born with to combat potentially dangerous and damaging pathogens. Once activated, it produces a rapid and effective response.

The immune system recognizes patterns of molecular structures that are shared between microbial pathogens, so called pathogen associated molecular patterns, PAMPs. These PAMPs can be single or double-stranded RNA found in viruses, or bacterial proteins and molecules such as pilin and lipopolysaccharides. The innate immune system also recognises endogenous signals, DAMPs (danger-associated molecular patterns), which are released when the cells are damaged, such as the stress-induced protein heat shock proteins and cellular proteins released during necrosis.

Macrophages and neutrophils are cells that can phagocytise (devour) and kill microbes they encounter. They also release pro-inflammatory cytokines. The complement system consists of plasma proteins that bind to microbes and initiate a cascade of reactions with the ultimate goal to opsonise or kill microbes. Natural killer cells can recognise infected cells and kill them, as well as activate and stimulate macrophages to kill infected cells. Dendritic cells (DCs) are professional antigen presenting cells (APCs). After activation they migrate to the lymph nodes where they are potent activators of naive T-cells. There are two types of DCs, myeloid DCs and plasmacytoid DCs.

B and T lymphocytes are the main effector cells of the adaptive immune system, and they express receptors that have specificity for a certain antigen. The key to the vast repertoire of antigens that can be recognised by B and T-cells is the rearrangement of the B and T-cell receptor genes. This means that the immune system constantly

generates cells with new specificities. Both cell types start as lymphoid precursor cells in the bone marrow (B-cells), and thymus (T-cells), and undergo a complex chain of maturation stages before they become mature cells.

Mature B-cells leave the bone marrow and start to migrate between secondary lymphoid organs. When they encounter the antigen, they will be activated. Activation can either be independent of T-helper cells, through antigen binding directly to receptors on the B-cells, or T-helper cell dependent, where the T-cell can increase the B-cell response. B-cells have immunoglobulin receptors expressed on their cell-surface; upon binding of the antigen to these receptors, the B-cells become activated and undergo clonal expansion. These cells start to secrete antibodies, some switch from producing immunoglobulins of the M isotype to produce immunoglobulins of the G isotype. Through mutations in the immunoglobulin genes, the B-cells can also start to produce antibodies with higher affinity than the initial response antibodies, so called affinity maturation. A portion of the cells will become memory B-cells, available for a fast response upon a new encounter with the antigen. Most B-cells are follicular B-cells from the bone marrow, but another subset develops in the fetal liver and are called marginal zone B-cells.

Naive T-cells are formed in the thymus, and become activated in the peripheral lymphoid organs when they meet APCs, such as dendritic cells, that express their specific antigen on HLA molecules on the cell surface. After activation the T-cell expands and differentiates, and subsequently enter the circulation. CD8+ cytotoxic T-cells (CTLs) bind to and kill infected cells, and CD4+ T-helper cells produce cytokines that recruit more immune cells to the site, and promote phagocyte activation to clear the infection. T_H1 -cells activate macrophages through interferon- γ . INF- γ also promotes B-cell switching, and promote the differentiation of T_H1 -cells. T_H2 -cells are important in combating helminthic (parasites) infections, and through interleukin 4 and 13 they can also activate macrophages. T_H17 -cells produce IL-17 which recruits leukocytes to the site of infection. Regulatory T-cells is another subset of T-cells, and they are important in suppressing immune responses and control tolerance to self-antigens.

1.4.2 Response to viral infections

The Toll-like receptors (TLRs) are membrane bound PAMP recognition receptors and are expressed on dendritic cells, phagocytes and many other different cell types, and some of them can bind and recognise viral antigens. There are nine different TLRs, with different ligands and differences in the downstream activation of cellular pathways (reviewed in [78]. TLR3 uses the TRIF pathway, TLRs 1, 2, 5-9 activate MyD88, and TLR4 can use both pathways. Both pathways can lead to activation of interferon regulatory factors (IRFs) and production of type 1 interferons.

The cytoplasmic Retinoic acid-inducible gene 1-like receptors, RLRs, are important in detecting viral RNA. RIG-1 detects double stranded RNA and leads to activation of transcription factors NF- κ B and IRF-3, and subsequent production of type 1 interferons. RIG-1 can also have direct antiviral effects in the acute phase of infection [79].

Nucleotide oligomerisation domain-like receptors (NLRs) is another type of receptor capable of detecting viral components. Just as with TLRs stimulations, NLRs stimulation leads to type 1 interferon responses, an important tool in combating viral infections.

Type 1 interferons (α and β) lead to increased cellular resistance to viruses, increase the cytotoxicity of NK cells and upregulate the expression of MHC class I, increasing the probability of an infected cell to be detected and killed. Type 1 interferons also increase the number of lymphocytes contained in the lymph nodes. Cells have ways in which they can sense infection and undergo apoptosis, and infected cells can also be more sensitive to external apoptotic signals. Type 1 interferons can be produced by DCs, macrophages and other immune cells, as well as by fibroblasts [78].

Natural killer (NK) cells normally interact with MHC class I molecules on other cells, but there are other receptors that can activate NK-cells, one of them is NKp46. It has been shown that NKp46 can recognise hemagglutinins on cells infected with influenza A virus and lyse them [80]. Plasmacytoid DCs recognise viral nucleotides and produce type 1 interferons after activation. The expression of viral particles on antigen presenting cells will also activate T-and B-cells, which will lead cytokine production, CD8+ mediated killing of target cells, and to the recruitment of other cells to the area.

1.4.3 The role of the immune system in MS

T and B-cells can both recognise and bind self-antigens. Normally these cells are identified and killed, but when this mechanism fails, auto-reactive cells can enter the system and start attack the body's own cells. MS is thought to have an auto-immune component, where immune cells are attacking and destroying self-antigens.

In the thymus, T-cells are presented with self-antigens bound to MHC-complexes on the surface of epithelial cells. T-cells that can only weakly bind to these complexes receive stimulating signals and survive. T-cells that bind too strongly to these complexes of MHC and self-peptides either undergo apoptosis or differentiate into regulatory T-cells. It is when this complex process fails that self-reactive T-cells are free to enter the circulation. When they later encounter the self-antigen in the periphery, they will become activated and perform their immunological functions in a setting which is negative for the individual, i.e. autoimmunity. Immature B-cells

constantly need external signals for survival, and it is thought that only cells with functional membrane immunoglobulins can receive these survival signals. B-cells that have a high affinity for self-antigens may undergo receptor editing to change B-cell receptor affinity, or they may undergo apoptosis.

MS is considered by many to be a T-cell mediated autoimmune disease. In the CNS, auto-reactive T-cells encounter resident DCs and are reactivated. These T-cells secrete pro-inflammatory cytokines, leading to recruitment of other inflammatory cells. In the end, the inflammatory process will lead to the destruction of the myelin wrapping the axons and leading to neuronal death, causing sensory and motor symptoms [81]. EBV-positive B-cells have also been found in the brains of MS-patients [82, 83], a possible indication of a pathogenic role of EBV infection MS disease development and/or progression.

The search for MS auto-antigens has recently started to yield interesting results. Auto-antibodies against myelin oligodendrocyte glucoprotein have been found in MS patients (reviewed in [84]), and auto-antibodies against neurofascin can mediate axonal injury [85], while antibodies against contactin-2 have been suggested to be involved in gray matter pathology in MS [86]. Recently, the potassium channel KIR4.1, expressed on astrocytes and responsible for potassium and water balance [87], was found to be a target for auto-antibodies in a subset of MS patients [9]. Interestingly, KIR4.1 appears to co-localise with aquaporin 4, another membrane channel. Auto-antibodies towards aquaporin 4 are seen in neuromyelitis optica [88], which was previously thought to be a subtype of MS [87].

2

EPIDEMIOLOGICAL STUDIES AND STATISTICAL ANALYSIS

2.1 COHORT STUDIES

In a cohort study, a group of people are followed over a period of time, and during that time some will develop the phenotype of interest, while others will not. Exposures to environmental factors of interest are followed during that period, often through examinations and questionnaires. A closed cohort does not allow any new participants to be included, whereas a dynamic cohort allows for new people during the follow-up period. Examples of closed cohort studies are the British birth cohort of 1958 [89], the British 1970 birth cohort study [90], where they are following all individuals born in Britain during that particular year. Another example of a cohort study is the Framingham heart study, where individuals living in Framingham, Massachusetts, have been followed since 1948 [91].

For common diseases, the cohort study is an excellent and feasible choice of study design, where you can get a sufficient number of healthy and diseased individuals. For a very rare disease, such as MS, the cohort would need to be very large to yield a sufficient number of cases for any meaningful statistical analyses. For a disease with late onset, the cohort will also have to be followed for a very long time, before cases appear among the studied individuals.

2.2 CASE-CONTROL STUDIES

A more cost-efficient study design is the case-control study, which can, if planned and carried out properly, give the same type of information a more costly cohort study can. The basic principle of a case-control study is to compare cases and controls; this can be in regards to an environmental exposure or a certain genotype. Case-control studies can also be done within the framework of a cohort study, and are then usually referred to as nested case-control studies.

In a case-control study, cases are identified and a group of controls is sampled from the source population. Often the cases and controls are matched for certain variables, such as sex and age. To increase power in a cost-efficient way, more controls can be added to the study.

In our case, the EIMS cases and controls are matched for age at index (when the case is included), sex (male/female) and where they live. Matching will hopefully decrease the noise, as the groups are slightly more homogenous.

If matching is done properly, these matching variables will not show any association to the disease outcome. For example sex would be strongly associated with MS if the controls were drawn at random, with a high frequency of females among the cases (approx. 70%), and lower among randomly drawn controls (approx. 50%). Hence it is not possible to test for association to matching variables.

Recall bias is an important issue in case-control studies, when information regarding exposures is collected after disease onset. Cases can either have a tendency to over- or underestimate their exposures to certain factors, which has to be taken into account when analysing the results. There is also the risk of reverse causation, where cases differ in exposure to controls, but not because the exposure causes disease, but due to changed behaviour because they have the disease. An example would be that MS cases report lower sun exposures than controls. A quick conclusion would be that low sun exposure is associated with an increased risk for MS, but in reality it might be that cases have been in the sun less because they do not feel well in the heat (MS patients can be sensitive to heat) or have trouble moving (due to early symptoms) and stay indoors more.

Case-control studies are not prospective studies, something which can be important if one wants to examine disease causing mechanisms rather than just measure if an exposure is associated with a disease. Prospective data also removes the problem of reverse causation, since a chronological order of events can be defined, with the disease onset as the endpoint.

2.3 ODDS RATIO

Odds ratio (OR) can be calculated from case-control data where you have information on the number of exposed and unexposed cases and controls for a certain factor, for example smoking or a genetic factor, and gives the effect size for that particular risk factor. This is often presented in a 2-by-2 table, as shown below.

	Exposed	Non-exposed
Cases	<i>a</i>	<i>b</i>
Controls	<i>c</i>	<i>d</i>

The formula for odds ratio is simple:

$$\text{Odds among patients} / \text{odds among controls} = \frac{a/b}{c/d} = \frac{ad}{cb}$$

An odds ratio of 1 means that there is no difference in risk between cases and controls for that risk factor. An OR higher than 1 indicates an increased risk for disease if exposed to that variable. Conversely, an OR lower than 1 indicates that exposure to that variable is associated with a decreased risk for disease. Odds ratios are often presented with a confidence interval (CI), usually 95%, which gives an indication of the precision of the OR and the variance within the analysed dataset. To test for statistical significance of the observed OR, the 2-by-2 table can be utilised in a χ^2 -test.

2.4 REGRESSION ANALYSIS

To assess the relationship between an outcome variable and one or several explanatory variables (also known as dependent and independent variables, respectively), regression analysis can be used. If the outcome variable is continuous and follows a normal distribution, a linear regression is applied. A linear regression model with one outcome and one explanatory variable is called a simple linear regression, compared to a model in which you have more than one explanatory variable, called a multiple linear regression:

$$Y = \alpha + \beta_1 x_1 + \beta_k x_k + \varepsilon$$

Where α is the intercept, and the value of Y if all variables are 0, $\beta_1 \dots \beta_k$ are the regression coefficients. A linear regression model has some underlying assumptions, such as that the relationship between Y and x is linear, and that the dependent variable follows a normal distribution. In those instances when data does not follow a normal distribution, different methods of transforming the data should be considered. Alternatively a non-parametric method should be used. The most common is a use the natural logarithm scale when transforming the data, but there are other methods.

To analyse the effect of our predictor variables on a quantitative outcome, such as the levels of an antibody towards JC virus, a linear regression approach can be used, with GWAS genotypes as explanatory variables.

If the outcome variable is a binary trait, such as healthy and sick, a logistic regression may be used. An inherent characteristic of the logistic regression (sometimes called binary logistic regression), is that the $\exp(\beta)$ is the odds ratio, for an individual with $x_i=1$, if the explanatory variables are categorical or binary. It is then possible to draw up a table of all possible combinations of variables within the dataset. If the explanatory variable x is continuous, $\exp(\beta)$ is the increase in odds associated with a unit increase in x .

With a multivariate logistic regression, it is possible to adjust for variables affecting the outcome, or to adjust for matching variables in a case-control study.

Conditional logistic regression can be used in matched case-control datasets. A key (or strata) identifying each pair of matched case and control, is used instead of the matching variables. It is the ideal method for analysing a data set with matched cases and controls, but can mean loss of data points if not all cases have a matching control, which is the case in the EIMS study.

2.5 INTERACTION

Interaction in its most biological sense is the actual physical interaction between two proteins or molecules in a biological pathway or mechanism. These interactions are essential for the survival of any cell (and organism) and best studied *in vivo* and *in vitro*.

Epistasis is the term generally used to describe the interaction between two genetic loci, where the expression of one is dependent on the other. To confuse matters more, epistasis is also a term used to describe a statistical interaction between two genetic loci. In statistical epistasis there is no need for any actual physical and biochemical interaction between two genetic loci.

There are many ways in which you can analyse how two factors interact to modulate disease risk, also known as effect modification. Two very common ways to study interaction are the *multiplicative* and *additive models*. The basis of both these models is to compare an observed joint effect of two factors with an expected joint effect. In a multiplicative model, the expected joint effect would be $OR_A \times OR_B$, and in an additive model it would be expressed as $OR_A + OR_B$. If there is a difference between the observed and the expected, there is a *departure from* multiplicatively (or additivity). This departure can either be positive (the observed joint effect is larger than expected), called *synergism*, or negative (observed joint effect is smaller than expected), called *antagonism* [92].

The difference between the two models lies in the way the joint effect is calculated. In multiplicative interaction, an interaction term between factor A and B is added to the logistic regression model.

Multiplicative interaction

		A				A	
B	OR	-	+	B	OR	-	+
	-	1.0	3.0		-	1.0	3.0
	+	2.0	6.0		+	2.0	9.0
Expected joint OR = 2.0 x 3.0 = 6.0							
No interaction: Observed joint OR = 6.0				Positive interaction: Observed joint OR = 9.0			

Adapted from:

Szklo, M. and F.J. Nieto, Epidemiology: beyond the basics. 2nd ed. 2007

In departure from additivity, the sum, rather than the product, of the odds ratios of factors A and B, is measured and compared to the expected.

Additive interaction

		A				A	
B	OR	-	+	B	OR	-	+
	-	1.0	3.0		-	1.0	3.0
	+	2.0	4.0		+	2.0	6.0
Expected joint OR = 2.0 + 3.0 - 1.0 = 4.0							
No interaction: Observed joint OR = 4.0				Positive interaction: Observed joint OR = 6.0			

Adapted from:

Szklo, M. and F.J. Nieto, Epidemiology: beyond the basics. 2nd ed. 2007

The double negative (A-/B-) group is the reference group, with a baseline risk of OR=1.

Attributable proportion (AP) due to interaction is the proportion of cases caused by the interaction itself, and can be calculated as follows, using the numbers from the table above:

$$AP = \frac{(\text{observed OR} - \text{expected OR})}{(\text{observed OR} - \text{baseline risk})} = \frac{(6 - 4)}{(6 - 1)} = \frac{2}{5} = 0.4$$

AP=0.4 means that 40% of all cases who are exposed to both factor A and B, are caused by this interaction, or put another way they are cases because of the interaction between these factors.

The additive model is what many epidemiologists refer to biological interaction, and is founded on the concept of the sufficient cause model (also known as the “pie model”). This model dictates that a complex disease has several different causes that can lead to disease [93].

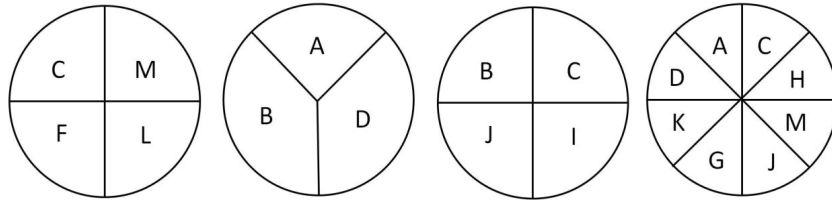


Figure 1. Four different sufficient causes where each piece of the pie is a genetic or an environmental risk factor. Each sufficient cause can be thought of as a separate individual. In all instances where the pie is complete, disease will develop. However, not all individuals will have the same components in their respective “pie”, and there will also be individuals with exactly the same “pie”. According to the additive model, when two components are present in the same pie, there is departure from additivity.

Through regression modelling, there is also the possibility to adjust for confounders, and can be applied to both the multiplicative and the additive models. The power to detect interaction depends on the OR for each factor, as well as the joint OR, and on the frequency of the factors studied [94].

There are other statistical methods to study interaction, such as data mining and machine learning methods, for example multifactor dimensionality reduction, Bayesian networks, and Random Forest. Just as the multiplicative model, these do not follow the sufficient cause model described above.

Even though departure from additivity has been called biological interaction, it is still a statistical method and interactions do not have to be physical, and the two interacting factors do not have to be involved in the disease-causing pathway *at the same time*. As such, statistical interactions do not imply causality in the biological sense, but can be a valuable method in pin-pointing disease causing mechanisms and pathways that warrant further investigations.

2.6 STATISTICAL POWER

Lack of statistical power can be the bane of any type of study, it doesn't matter how well planned and thought-out it initially was (although a good study design can help ameliorate the problem). So what is statistical power?

To answer that, we need to go back to the null hypothesis (H_0) and type I and II errors, all important concepts in statistics.

- The H_0 is the default position, while the alternative hypothesis usually postulates an alternative theory. H_0 often propose that there is *no difference* between groups,

or no effect, while the alternative hypothesis that there is a difference between groups.

- Type I error is that you reject the null hypothesis even though it is true (false positive).
- Type II error is the reverse, you fail to reject the null hypothesis even though it is actually false (false negative).

The probability of a type I error is commonly denoted α , and the probability of a type II error is called β . The power of a statistical test is the probability that the test will correctly reject the null hypothesis, when H_0 is false. The power of a study is calculated as $1 - \beta$, and β is usually set at 0.20, meaning that the minimum acceptable power of a study is 80%.

Power calculations can be used to estimate the power you have to detect a certain effect size (in case-control settings, the OR) of a given variable, given certain circumstances (disease prevalence in the population, minor allele frequency, number of cases and controls, rate of misclassification of disease, and the levels of α and β). Generally, the more cases and controls you have the more power you have for a given effect size. Traditionally, $p < 0.05$ have been used as a cut-off for significance.

When two groups, for example cases and controls, are tested repeatedly for differences exposures to several variables or genetic variants, there is an increased risk that a significant result is obtained just by random chance. A type I error is thus made. The more tests, or comparisons, that are made on the same dataset, the higher the number of false positive findings. Several methods to correct for multiple testing exist. In genome-wide association studies, where many comparisons are made at the same time, the cut-off is $p < 5 \times 10^{-8}$.

2.7 META - ANALYSIS

In a meta-analysis you combine and compare results from several studies into one analysis, for example to calculate a common OR in a set of case-control studies.

In the fixed model, the effect size is assumed to be equal across all studies, and all the factors influencing the effect size are the same. Each study is assigned a weight; the inverse of the variance within that study [95].

In the random effects model, it is assumed that the effect size can vary between studies. The distribution of these effect sizes is generally assumed to follow a normal distribution. The weight assigned to each study is the inverse of the variance within each study, plus the variance between studies [95].

Heterogeneity is the variability between the different studies included in a meta-analysis and is an important factor that needs to be addressed. If there are differences in estimates, or if the estimates are similar but the confidence intervals differ in magnitude, heterogeneity could be a problem. The meta-analysis performed in R 2.15 using the rmeta-package, assumes a fixed model and tests for heterogeneity. Other packages can be used for random model analyses. The meta-analysis command in plink 1.07 [96] gives the Cochrane's Q statistic value and I^2 -value to help assess heterogeneity for each marker. The I^2 -value gives the heterogeneity on a scale of 1-100. The power of both of these tests is dependent on the number of studies in the meta-analysis.

3

GENETICS

3.1 GENETIC VARIATION AND POLYMORPHISMS

Our genomes are not identical, and it is these differences that make us unique, and also separate us from other species. These variations come in many forms, some are whole genes that are present in one species but not in another. Others are smaller sequences of a few base pairs that are repeated several times, but the number of these repeats differs between individuals (e.g. microsatellites). The smallest variation is the single nucleotide polymorphisms (SNPs), where **one** base differs from one person to another:

Me:	AGTCATTCG T TACGACGA
You:	AGTCATTCG A TACGACGA

Through modern genotyping techniques, these can now be analysed in human DNA. Earlier techniques only allowed for discrimination of microsatellites. SNPs are often bi-allelic, that is there are only two possible variants, in the case above, A and T. The less common allele in a population is called the minor allele, and the more common is referred to as the major allele. SNPs with a minor allele frequency (MAF) <5% are called rare variants.

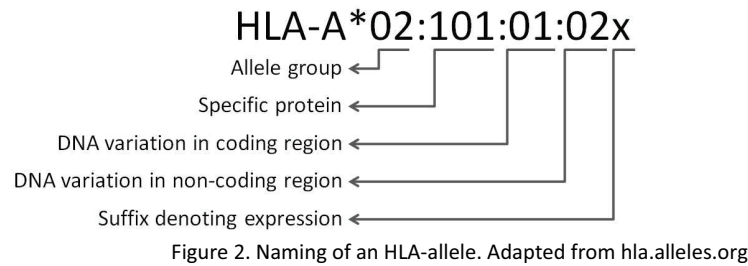
3.1.1 The Human Leukocyte Antigen genes

In the human genome, the extended HLA-region is located on the short arm of chromosome 6. It is a large region, around 7.6 Mbp, and the region also contains other immune system related genes, histone and tRNA genes, as well as genes coding for olfactory receptors (smell perception) [97]. The HLA-region is one of the more polymorphic gene regions in the human genome.

The class I genes are HLA-A, B, C, E, F and G, and the class II genes are DRA, DRB (divided into DRB1-9), DQA1, DQB1, DPA1, DPB1, DMA, DMB, DOA, and DOB. There are currently 6,919 class I alleles, and 1,875 class II alleles listed in the IMGT/HLA database [98]. Of all of these, HLA-B is the most polymorphic gene, with over 2,800 alleles coding for more than 2,100 proteins.

Because of the many variations of alleles, the nomenclature of HLA-alleles has been reworked several times. An example of the current naming standard is given in figure 2 below. The gene is always given first, with the allele group given after an asterisk. Then follows digits denoting a specific protein variant, variation in coding

sequence and variation in non-coding sequence. At the very end is a position for a suffix denoting the protein expression. The suffixes used are: N = null, L = low, S = soluble, C = cytoplasm (not cell surface), A = aberrant (there is doubt whether the protein is expressed, and Q = questionable.



There is also extensive linkage disequilibrium (see section 3.2) between alleles in genes within the HLA, making the region extremely complex and fascinating. Many alleles from different genes within the HLA are often inherited together. One example are *A*01-B*08-DRB1*03*, which is a very common combination of alleles in Caucasians, often called the B*08-haplotype.

3.2 LINKAGE DISEQUILIBRIUM

Linkage disequilibrium (LD), a term first coined by Lewontin and Kojima in 1960 [99], is the non random association of two or more loci. Markers with high LD would be seen as inherited together more often than expected. The reason is that there has been no crossing over (no recombination) event during meiosis. Generally, markers that are close to each other on the chromosome, are more likely to be in LD, than markers that are far apart, and the strength of LD decreases as the distance between loci increases, but studies have shown that LD can also be preserved over large distances in certain regions of the genome (reviewed in [100]). There is also the assumption that LD will decay over time in a population, as an increase in the number of generations, and therefore, meioses, increases and hence more recombination have occurred between the loci.

There are several ways to measure LD, but D , D' , and r^2 are the most common. Deviation (D) is the LD-measurement proposed by Lewontin and Kojima. For two markers, A and B, D is calculated:

$$D = P_{AB} - p_1q_1$$

Where P_{AB} is the frequency of the AB-haplotype, p_1 and q_1 are the allele frequencies of A and B, respectively. D depends on allele frequencies and can vary between -0.25 to 0.25, and $D=0$ indicates no LD.

Another way to measure LD is D' , which is a normalised version of D and uses the theoretical maximum of D (D_{max}). For two markers A and B , with alternative alleles a and b , D' is calculated as follows:

Allele	Frequency	$D \geq 0$	$D' = D/D_{max}$
A	p_1	D_{max} : the smallest of p_1q_2 and p_2q_1	
a	p_2		
B	q_1	$D < 0$	$D' = D/D_{max}$
b	q_2		
		D_{max} : the smallest of p_1q_1 and p_2q_2	

If D' is +1 or -1, it implies that there is complete LD between markers.

r^2 , or Δ^2 , is the correlation coefficient between a pair of alleles. It is another very common measurement of LD, and also very popular among geneticists. It ranges from 0 – 1, where 1 indicates complete LD between loci.

$$r^2 = \frac{D^2}{p_1p_2q_1q_2}$$

A *haplotype* is a stretch of DNA where there have been no recombination events, and all variants at all loci along that stretch are inherited together.

Factors that affect LD on a population level are genetic drift, inbreeding, mutations and gene flow. Genetic drift is stochastic fluctuations in allele frequencies in a population due to random sampling [101]. In a small population, these fluctuations will cause non-random associations [99].

Inbreeding will impact the decay of LD, but it will be different in low inbreeding populations and in species where there are high levels of inbreeding, such as selfing species. Mutations are one way new variants are introduced into a population but probably have only small effects on the generation of LD [99].

Gene flow occurs when there is a mixture of two genetically different populations, and genetic material is shared. The likelihood of generating LD when two populations are mixed depends on the ratio of contribution by each population, and the differences in allele frequencies [99]. The level of LD is known to vary between regions of the genome. The HLA-region on chromosome 6 shows extreme levels of LD, due to the pressure of natural selection on this region [100], and also high level of polymorphisms. This high level of LD has made it very hard to distinguish the actual risk-bearing gene in the major genetic MS-risk haplotype, *DRB1*15:01-DQB1*06:02-DQA1*01:02*. LD can also be used to our advantage, as in imputation of

genetic data, and the possibility to replace one marker with another if the LD between them is high.

3.3 HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg principle is a central concept in population genetics, and was independently proposed in 1908 by G.H Hardy and W. Weinberg [102]. It states that the allele and genotype frequencies in a population will remain constant (in *equilibrium*) over generations. This holds for both bi-allelic and multi-allelic loci.

The principle relies on a few assumptions:

- Mating is random; no inbreeding
- The population is infinitely large
- No gene flow (no new genes are added through mating with other populations)
- No mutations, as they can alter allele frequencies
- No natural selection; all individuals have equal success of survival and reproduction

In any situation where these assumptions are not fulfilled, there will be a deviation from the Hardy-Weinberg equilibrium (HWE), when comparing the observed number of each genotype with the expected number of each genotype. In case-control analyses, any deviation from HWE in either cases or controls can influence the conclusions drawn from such a study. Deviations from HWE can be due to sampling bias, mistyping of genotypes, or spurious associations due to population stratification (see section 3.5) [103]. If a marker deviates from the HWE in cases but not in controls, it can lend further support to an association between the marker and the disease phenotype [103, 104].

3.4 HOW DO WE IDENTIFY GENETIC RISK FACTORS?

3.4.1 Linkage studies

Linkage studies analyses the pattern of inheritance between a marker in the genome and a certain phenotype, and thus utilises family materials. In the early days of linkage analysis, microsatellites were used as genetic markers, but nowadays the analysis of SNPs has become more common.

In MS, linkage analysis was able to identify the HLA-region as an important genetic risk factor, but so far, no non-HLA genes have been found through linkage analysis [25].

3.4.2 Candidate gene studies and GWAS

Before the genome wide association study (GWAS), there was the candidate gene study. A possible candidate gene, often chosen because of evidence from animal models, functional studies or through knowledge of the immune system and disease pathogenesis, was genotyped in a set of MS cases and controls. Often several markers in the gene were genotyped at once, and the association with disease was calculated.

Association tests are usually performed on the allele level, as in comparing the frequency of one allele in cases versus controls, using a χ^2 -test. In some cases, a genotypic association test can be warranted, where you compare the frequency of genotypes instead of alleles, between cases and controls. Many applications also allow for testing of recessive and dominant models.

So far, only common variants (MAF >5%) have been studied using GWAS. Rare variants can be analysed through targeted sequencing, and in some cases, with very low MAFs, even through the use of family materials and linkage analysis [105].

One drawback with genetic studies is that the causal SNP is sometimes not possible to genotype or not known, and instead one uses tagging-SNPs that are in LD with the causative marker(s).

3.5 POPULATION STRATIFICATION

Population stratification – inherent differences in allele frequency between cases and controls due to differences in ancestry, can cause false associations in case-control analyses. With advances in technology and statistics, it has become easy to handle population stratification in large-scale genotyping datasets.

There are different ways to deal with stratification:

Genomic control, where you, for each marker, adjust for population stratification with an inflation factor, denoted λ [106], which should be 1 or close to 1.

Structured associations, in which the data is divided into subclusters and associations are given for each cluster and then combined, for example in a meta-analysis [107], is another way.

Principal component analysis (PCA) is a method to reduce the dimensionality of a dataset by analysing the co-variance between variables. The analysis looks for directions in the data (in a 3D-space) and identifies the main direction of the data.

This first axis, that therefore explains most of the variability, is called the principal component. The next step in the analysis is to look for the next axis, orthogonal (90 degree angle) to the previous, which explains most of the remaining variation. This step is repeated until no more possible components remain. The significance of each component is tested, and significantly associated (explain part of population stratification) can be included as covariates in an association analysis. Something similar is performed in the Eigenstrat software where they apply a principal component analysis and then continuously correct for ancestry in the association analysis [107]. Some researchers do not look at the number of significant PCs, but rather use a fixed number of PCs to adjust for [108].

Genomic control and structured associations were previously very common ways to deal with population stratification. Even if not used to correct for population stratification, genomic control (λ) is often given in published papers, as way to easily indicate whether population stratification was a problem or not. Another way would be to state the number of significantly associated principal components in the PCA, as given by the Tracy-Widom statistics.

Principal component analysis, though becoming more popular within the genetic research community, is not without faults. If the principal components are equally distributed between cases and controls, or if self-reported ancestry overlaps with the information carried by the PCs, some data will be superfluous and will reduce power in the association analysis. Other ways to select which PCs to include have been proposed [108].

4

MATERIALS AND METHODS

4.1 STUDY POPULATIONS

In my papers I have used data from two ongoing epidemiological studies, EIMS and IMSE, as well as a Danish set of MS cases, and I have had access to results from a German dataset of MS cases.

All cases included in study I-III, and all Swedish cases in study IV, fulfil the McDonald criteria for MS (a defined set of criteria for MS diagnosis) [109]. All cases and controls in study I-III are of Swedish, Danish, Norwegian or Finnish ancestry. Both IMSE and EIMS have been approved by the Ethical Review Board at Karolinska Institutet and all participants in these studies have provided written informed consent.

4.1.1 EIMS

The Epidemiological investigation in MS is an ongoing Swedish study led by researchers at Karolinska Institutet. Started in 2005, the study aims to collect all newly diagnosed patients in Sweden, and more than 30 clinics from all over the country, including all university hospitals, are now submitting new cases. It is the neurologist at the recruiting unit that makes all examinations and final diagnosis. All cases fulfil the McDonald criteria for MS [110].

As a new case is recruited, two controls are randomly selected from the national population registry, matched for sex, age at index, and area of residence. The study population includes ages 16-70. Both cases and controls are asked to answer an extensive questionnaire regarding lifestyle (diet, exercise, occupation, etc), and to donate blood samples.

The response rate for the questionnaire is around 98% for cases, and 73% for controls, and of these responders, 98% of cases leave blood sample, and 58% of controls [54]. From these blood samples, DNA, serum and plasma are extracted at the Karolinska Institutet Biobank. To increase the number of DNA samples received from controls, they are now given the option to leave a saliva sample instead of blood.

A small proportion of the study (around 1000 individuals) have been characterised for serological responses to CMV, EBV (EBNA1, EBNA1 fragments), varicella zoster virus, Human herpes virus 6 A/B, and herpes simplex 1 and 2. Samples from the

EIMS study have been genotyped as part of the IMSGC/WTCCC2 GWAS [15], and on the Immunochip [111]. Many have also been genotyped with classical HLA-DRB1 and A genotyping.

4.1.2 IMSE I

The Immunomodulatory MS study I is a post-marketing (phase IV) study of patients currently on natalizumab treatment, with over 40 clinics in Sweden recruiting patients.

Cases are included as they start treatment, and examinations are performed at baseline, and then every six months. At every examination a blood sample is taken, and cognitive and functional tests are performed to evaluate treatment effect. Continuous follow-up on adverse events and treatment side-effects is performed and data entered into the Swedish MS-registry [112].

Samples from the IMSE I study have been included on the Immunochip, and also, genotyped in the IMSGC/WTCCC2 GWAS [15] and with classical HLA-genotyping in our group.

4.1.3 Other datasets included in paper IV

The Danish dataset consists of 158 cases on natalizumab treatment, with serum taken at baseline (before treatment). 84 of these cases had imputed HLA-genotypes and GWAS genotypes, as they were part of the IMSGC/WTCCC2 GWAS [15].

As part of the meta-analysis, I also had access to the results of the analysis on 751 German MS cases, collected from collaborators in Munich. These cases had also been analysed in the GWAS [15] and were part of a study similar to IMSE.

4.2 GENOTYPING

4.2.1 Classical HLA-genotyping

Low resolution genotyping for the *HLA-DRB1*, *A* and *C* genes was performed with kits from Olerup SSP AB [113]. The method uses sequence specific primers (SSP). A polymerase chain reaction is run for around 30 cycles, and the products are loaded to an agarose electrophoresis gel stained with GelRed (Biotium, Hayward, CA, USA), run at 200 volts for 15 minutes and visualised and photographed on a UV-table. The wells with a positive signal (band) are then interpreted according to a chart, and a person's *HLA-DRB1* and *A* and *C* genotypes were established at the two-digit level.

For *HLA-B*, a Luminex-based genotyping method was used [18], which uses SSPs attached to micro-beads [114].

4.2.2 Large scale SNP genotyping

Instead of genotyping each sample for 500,000 markers, one at a time, technology now allows us to genotype each sample for hundreds of thousands of markers at once, saving DNA, reagents, and time. Many life science technology companies now offer chips with genome wide coverage, where the markers are evenly spaced over the whole genome, chosen to tag common polymorphisms. There are also custom solutions that can answer more specific research questions; the so called Immunochip is one of these custom chips.

The Immunochip is a collaborative effort between consortia, like the IMSGC, working on different autoimmune diseases (rheumatoid arthritis, psoriasis, celiac disease to name a few). The goal was to fine map disease associated genes, i.e. find the primary disease association in each region by typing many markers in each gene/region. All genes/regions that had reached genome-wide significance in at least one autoimmune disease were included. Each consortium had the opportunity to select a list of genes to be included on the chip, which was then produced by Illumina [115]. Around 200,000 markers in around 250 different risk genes are included on the chip. Each consortium has been responsible for the analysis for their own samples, and the results on RA [111] and celiac disease [116] have already been published, and, hopefully, the report on MS is forthcoming.

Genotypes and imputed HLA-genotypes from the IMSGC/WTCCC2 GWAS from 2011 [15] were used in paper IV, as well as imputed HLA-genotypes from the Immunochip. The IMSGC/WTCCC2 GWAS was performed using Illumina HumanQuad 660 genotyping chips. Samples were finger printed with Sequenom markers, and concentration measured with Picogreen (in duplicates). Samples were also analysed for call rate, heterozygosity and relatedness. Principal component analysis was performed to analyse population stratification. Samples that did not meet any of the requirements were excluded. SNPs were automatically called and the clusters checked [15].

4.3 GENOTYPE IMPUTATION

Imputation is a method with which you infer genotypes of one marker, based on the genotypes of other polymorphisms in high LD with that marker. By imputing genotypes, you don't even need to genotype them in your material; but what you do need is a reference panel of individuals genotyped for all markers (even those that your own material is missing). Many groups now use data from the HapMap project [117] or 1000 genomes [118] as reference panels. These data are available for everyone to use.

HLA-imputation is slightly different to normal SNP imputation, in that you do not predict genotypes of a certain SNP, but rather the haplotype of SNPs determining on which specific HLA-allele it is situated. The reference panel will also have to be classically genotyped for the HLA-alleles you want to impute, in addition to SNP genotyped. Several programs have been designed to be able to do this, and for the HLA-imputation used in this thesis, HLA*IMP [119] has been used on IMSGC/WTCCC2 genotypes, and HLA*IMP:02 [120] was used to impute genotypes from the Immunochip.

4.4 ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)

Enzyme-linked immunosorbent assay (ELISA) was used to measure the IgG antibody response. For anti-EBNA1 IgG a commercial ELISA-kit was used on serum (Biotest, Dreieich, Germany), and the assays performed according to the protocol [36].

ThioredoxinA-EBNA-1 fusion proteins of the EBNA1 fragments were generated and then expressed in *E. coli* bacterial culture. The six fusion-proteins covered the following amino acids in EBNA1 protein sequence, 1-90, 61-90, 385-420, 410-440 and 421-449. The fusion-proteins were purified and then coated onto 96-well plates. An in-house ELISA was used to assess the antibody response to each of the six fusion-proteins [33].

Anti-CMV IgG was measured with an in-house ELISA. The CMV antigen was prepared from embryonic fibroblasts infected with strain AD169 [121]. Human serum samples were analysed in duplicates, at a dilution of 1/200, using ELISA [122]. Alkaline phosphatase-conjugated, purified F(ab')₂ fragment goat-antihuman IgG was used as conjugate (Jackson ImmunoResearch, West Grove, PA). The plates were read at 405nm and 650nm in a multiscan FC reader (Thermo Scientific, Vantaa, Finland). Cut-off was calculated by adding 0.2 OD units to the mean absorbance value of 160 runs of negative control samples. Serum samples with OD values at cut-off were run in 2-fold dilutions (1/100-1/12800) and then judged as positive or negative.

For anti-JCV IgG a two-step method was applied. The method consisted of an ELISA followed by a second confirmation step in which the samples were tested for cross-reactivity with the similar BK virus, and/or low-affinity binding, to distinguish true and false positives [123].

4.5 STATISTICAL ANALYSES

4.5.1 Association and correlation tests

SPSS 19 (IBM, Armonk, NY, USA) was used for logistic regression analysis to assess association of environmental and genetic factors with MS (papers I-IV), and for linear regression analysis for papers III and IV. SPSS 19 was also used for Mann-Whitney tests (paper III), χ^2 -tests (paper III), and to test for correlations between transformed anti-EBNA1 IgG and anti-CMV IgG (paper III).

4.5.2 Interaction analysis

R 2.4.0 and 2.15 was used to study interactions on the additive scale using a slightly modified version of a previously published R-script [124]. The modification of the script allowed the calculation of p-value of AP as well as permutation in the calculation of this p-value. SPSS 19 was used to calculate interaction on the multiplicative scale (papers I-III).

4.5.3 Conditional logistic regression

R was also used for conditional analysis using the clogit command (paper II and III), in R 2.15.0.

4.5.4 Principal component analysis

Before the PCA, SNP-pruning to remove dependent markers in LD with each other, with $r^2 > 0.2$ as cut-off, was performed in plink 1.07 [96]. Eigensoft software package [107] was used to perform principal component analysis of GWAS data in paper IV, and to calculate genomic control (λ). Cut-off in the Tracy-Widom statistics for principal components were set at $p < 0.05$, PCs with lower p-values were included in subsequent analyses.

4.5.5 Genetic case-control analysis

Genetic case-control association and linear regression in paper IV was performed using plink 1.07 [96]. Hardy-Weinberg equilibrium $p < 0.001$; and individual missingness less than 10% were used as cut-offs in the case-control association analysis.

4.5.6 Meta-analyses

Meta-analysis in paper III was performed in R 2.15.0 using the rmeta-package, assuming a fixed model and testing for heterogeneity ($p < 0.05$).

The meta-analysis in paper IV was performed using plink 1.07, which gives estimates and p-values for both the fixed and the random model, as well as the p-value for Cochran's Q-test and the I^2 -value (both are measures of heterogeneity).

4.5.7 Manhattan plots, marker annotation, etc

In paper III and IV, the `rntransform` algorithm from the GenABEL package [125] in R 2.15.0 was used to transform anti-viral antibody levels to achieve normal distribution.

R 2.15.0 was used to generate manhattan plots for paper IV [126].

Gene annotation of markers in paper IV was performed with Vario watch [127].

Pathway analysis was performed using the tools at webgestalt.com (Vanderbilt university) [128].

Haploview [129] was used to assess LD between nearby markers in paper IV, and $r^2 < 0.2$ was used as a cut-off for independence between markers.

GeneCards [130] was used to research gene expression of certain genes.

5

AIMS OF THIS THESIS

The overall aims of this thesis was to study EBV and CMV as MS risk factors in a Swedish case-control material, and to study host genetic factors influencing JC virus serostatus and response.

The more specific aims for each paper were:

Paper I:

- Study the association of antibodies directed towards EBNA1 IgG and different epitopes of EBNA1 IgG to MS.
- Study interactions between EBNA1; 385-420 IgG and *HLA-DRB1*15* and *A*02*, the strongest genetic risk factors associated with MS

Paper II:

- Study the interaction between smoking and EBNA1 IgG on the additive and multiplicative scale. A previous report described such an interaction on the multiplicative scale.

Paper III:

- Study the association between CMV serostatus and MS in a Swedish case-control material.
- Study possible interactions between CMV serostatus and *HLA-DRB1*15* and *A*02*.
- Confirm our findings by performing meta-analyses on the association of CMV with MS risk, using data from previously published retrospective and prospective studies.

Paper IV:

- Perform a genome wide association study on the host genetic influence on anti-JCV antibody status and JCV normalised optical density (nOD)-values in Scandinavian MS cases and Swedish controls
- Perform a meta-analysis on the results from the GWAS with data from a German set of MS cases
- Study associations of classical HLA-alleles on anti-JCV antibody status and nOD values in Scandinavian and German MS cases and Swedish controls

6

RESULTS AND DISCUSSION

6.1 COMMON VIRUSES IN MS (PAPERS I - III)

Papers I-III concerns the associations of the two common viruses EBV (paper I and II) and CMV (paper III).

6.1.1 Results from paper I

It has previously been reported by Sundström *et al* in a dataset of approximately 100 MS cases and 200 controls that there was reactivity against different domains of the EBNA1 protein [33]. We collaborated with Sundström *et al*, and plasma samples from the EIMS cohort were sent for analysis of the IgG antibody response towards EBNA1 and the fragments of EBNA1 in their laboratory. As these individuals were also asked to answer the EIMS-questionnaire, we had information (self-reported) on past history of infectious mononucleosis (IM). For 678 MS cases and 1490 controls, we had all demographic data we needed (age, gender, area of residence (urban/countryside). For 570 MS cases and 673 controls we had anti-EBNA1 antibody levels, and HLA-DRB1 and HLA-A genotypes.

Of the MS cases, 17 % reported past history of IM, and only 11 % of the controls, $OR=1.89$ (1.45-2.48), $p=3 \times 10^{-6}$. Next we analysed the possible interaction between IM and *HLA-DRB1*15* and absence of *A*02*. There was a weak interaction between IM and *HLA-DRB1*15*, AP due to interaction=0.34 (0.01-0.68 95% CI), $p=0.05$, seen in 512 MS cases and 672 controls, while adjusting for smoking status (current/never), age, sex, area of residence.

Since the majority of individuals in Sweden have had an EBV-infection, seropositivity (absorbance unit of 10 and up) was not used, instead we used the median among controls as the cut-off for high versus low antibody response. The median for all IgG types were lower among controls than MS cases (paper I, table 1).

High levels of EBNA1 IgG, and all the EBNA1 fragment IgGs were associated with MS risk, with EBNA1 385-420 IgG being the most strongly associated antibody type, $OR=3.79$ (2.93-4.87 95% CI), $p=9 \times 10^{-12}$, independent of EBNA1 IgG level. Since EBNA1 385-420 IgG was more strongly associated with MS than EBNA1 IgG, we decided to continue our analyses with this variable.

We observed an interaction between anti-EBNA1 IgG levels and *HLA-DRB1*15* on the additive scale, AP due to interaction=0.60 (0.45-0.76 95% CI), $p=5 \times 10^{-6}$ after permutation. This interaction was observed when we stratified on *A*02* carriership. There was also an interaction between EBNA1 IgG and absence of *A*02*, in all cases and controls and in *DRB1*15*-positive cases and controls that survived 10,000 permutations, AP=0.39 (0.18-0.61 95% CI), $p=9 \times 10^{-4}$, and AP=0.49 (0.21-0.76 95% CI), $p=2 \times 10^{-3}$, respectively.

A risk group analysis combining *HLA-DRB1*15*, *A*02* and EBNA1 385-420 IgG indicated that individuals carrying *DRB1*15*, EBNA1 385-420 IgG high, and lacking *A*02*, had around 16 times higher risk for MS compared to individuals without these risk factors, OR=16.03 (9.42-27.30 95% CI), $p=2 \times 10^{-24}$.

6.1.2 Results from paper II

A previous report had indicated that there was an interaction on the multiplicative scale between EBNA1 IgG and smoking, modulating the risk for MS [131]. We wanted to analyse this in our EIMS material, as well as analyse interactions on the additive scale.

We found no interaction, on either scale, between smoking and EBNA1 IgG in the EIMS material. We did see interactions on the additive scale between EBNA1 IgG and *DRB1*15*, AP=0.34 (0.11-0.57 95% CI), $p=5 \times 10^{-3}$, and between EBNA1 IgG and absence of *A*02*, AP=0.36 (0.13-0.59 95% CI), $p=2 \times 10^{-3}$. We did not see any interactions on the multiplicative scale in the entire dataset. We did, however, see an interaction between smoking and absence of *A*02* in a conditional analysis.

A risk group analysis indicated that subjects with all four risk factors, *HLA-DRB1*15*, absence of *A*02*, high EBNA1 IgG, and smoking, had approximately 26 times higher risk, OR=26.32 (10.20-67.94 95% CI), $p=2 \times 10^{-11}$, compared to individuals with none of these risk factors. However, the number of individuals in the high-risk group was very small, only 69 cases and 18 controls, apparent from the large confidence interval.

6.1.3 Results from paper III

CMV seropositivity has been linked to MS previously [50], and has, as mentioned above, a protective effect on clinical and MRI outcomes [47]. We wanted to investigate the association between CMV and MS in our EIMS material.

CMV seropositivity was associated with a decreased MS risk, OR=0.73 (0.58-0.92 95% CI), $p=0.008$, adjusted for index age, gender, smoking, sun exposure, EBNA1 IgG levels, *HLA-A*02* and *DRB1*15*. Since the study by Waubant *et al* included paediatric MS cases, we also did the same analysis in all individuals with index age >18, with the negative association still apparent.

To validate our findings of a negative association, we also performed a meta-analysis on previously published data, which confirmed our results, OR=0.79 (0.67-0.93 95% CI). We also performed a meta-analysis on prospective studies, to investigate a causal relationship. This meta-analysis only included three very small studies [35, 132, 133], and showed no significant association between CMV and MS.

6.1.4 Discussion on papers I - III

Our results of increased antibody levels against EBNA1 IgG among MS patients compared to controls are consistent with previous findings, both from previous Swedish reports [36], and two American studies in which they had the opportunity to study serum samples taken both pre and post-diagnosis, and the elevated titers were apparent in both types of samples [35, 133]. Interaction between *HLA-DRB1*15* and EBNA1 385-420 IgG had been seen previously, by stratifying on *HLA-DRB1*15* [33]. However, they did not study interaction by testing any deviations from neither additivity, nor multiplicativity. In paper II we also observed an interaction between EBNA1 IgG and *HLA-DRB1*15* positivity, weaker than the interaction between EBNA1: 385-420 IgG and *DRB1*15*, which we would expect, since we already established EBNA1: 385-420 IgG as the stronger risk factor of the two.

The increased risk for MS in those patients with past history of IM is also in line with previous data (reviewed in [134]). IM was not as strongly associated with MS as high EBNA1 IgG or EBNA1 385-420 IgG levels. This could be due to the fact that CMV mononucleosis has similar symptoms. As an interesting side-note, our data has also been used in a sort of meta-analysis, where they confirm the interaction on the additive scale, between IM and *HLA-DRB1*15* positivity [135].

The lack of replication in paper II could, as discussed in the manuscript, be due to the differences in the study populations. Our EIMS-participants are younger than those of the Simon *et al* study, which could influence the risk for recall bias, and also implies longer disease duration, which could lead to higher disability.

The presence of elevated titer levels against EBNA1 in MS patients before disease onset could indicate that EBV has a role in MS before disease onset; there could also be aberrations in the immune system well before first of symptoms, partly expressed as increased response towards the latent EBV-infection.

Previous studies on the possible association between CMV and MS have been small, and our study is the largest to date. The negative association was quite strong, OR=0.73, p=0.008. The fact that the meta-analysis *without* our material also indicates a negative association, OR=0.79, further supports our results. The meta-

analysis on the prospective data, unsurprisingly, gave no clear results which could be due to lack of power.

The data from Zivadinov *et al* indicated a favorable role of CMV positivity in regards to clinical and MRI outcomes [47]. A positive effect of murine CMV has been seen in an animal model called Theiler's murine encephalitis, in which murine CMV was used as a co-infection together with the disease-inducing murine RNA-virus. These animals showed increased motor performance and decreased brain atrophy [52]. On the other hand, the results from Horakova suggests that the role of CMV is not entirely positive, as they observed an increased relapse rate and decreased time to first relapse after the initial CIS-event [51]. Based on these results, the role of CMV in MS is not straightforward, and further investigations are needed to establish a) a true causal relationship, if such exists, b) define the role CMV has in MS progression and development, which can differ based on genetic background (both of the host and the virus), exposure to other environmental factors, e.g. smoking, vitamin D, and sun exposure, and other co-infections, such as EBV and varicella zoster. Thus, we do not know why a CMV infection would protect against MS. A potential mechanism would be that CMV induces a very intense immune response. In infected individuals, a very high proportion of circulating T-cells would be directed against CMV, leaving little room for a potential autoimmune response against the CNS, as seen in MS.

So far we have established that EBV is positively associated with MS and CMV negatively associated with MS. One theory behind this could be so called immunodominance, where the response to one virus, most likely EBV, uses up the resources of the immune system, so there are none left to combat the other virus, CMV. Individuals co-infected with CMV and EBV would then have high levels of anti-EBV antibodies, but low levels of anti-CMV antibodies, possibly as low as being classified as negative in a laboratory test. If this immunodominance is more pronounced in MS patients, many would falsely be classified as anti-CMV antibody negative, and CMV thus showing a negative association with MS. As mentioned in paper III, we do not see any difference between CMV positive and negative MS cases in terms of EBNA1 IgG levels, and only a trend towards a difference among the CMV positive and negative controls, $p=0.06$. If there is immunodominance, the correlation between anti-CMV IgG and anti-EBNA1 IgG should be negative; high levels of anti-EBNA1 IgG would entail low levels of anti-CMV IgG. In our test of normalised antiviral IgGs, there was a positive correlation between anti-CMV IgG and anti-EBNA IgG in both cases and controls, arguing against a competition of immune resources. There has also been debate over the flexibility of the immune system, and that infection with both EBV and CMV will increase the total number of CD8+ T-cells, with an expansion of CMV-specific CD8+ T-cells but no effect on the number of EBV-specific CD8+ T-cells, seen in transplant patients [136]. Others have seen no such alteration in T-cells

numbers after infection, and argue that this might be a specific characteristic of transplant patients and not present in healthy individuals [137].

Molecular mimicry has been suggested as one mechanism through which both EBV and CMV can influence MS development. There are protein sequence similarities between EBNA1 and myelin basic protein [138], which could lead to cells attacking the myelin around axons. As for CMV, a CMV-associated peptide has been shown to cross-react with MOG in a marmoset experimental autoimmune encephalomyelitis model [139].

Another protein with similarities to EBNA1 is $\alpha\beta$ -crystallin, which is similar to the EBNA1 384-420 fragment [140]. In the brain $\alpha\beta$ -crystallin is a strong inhibitor of inflammation, leading to decreased activation of NF κ B in T-cells, macrophages, DCs and glial cells, and MS patients have a strong T-cell response to $\alpha\beta$ -crystallin [141]. One hypothesis could be that in EBV-positive individuals, the T-cell response to $\alpha\beta$ -crystallin is increased further, leading to increased inflammation in the brain, and tissue damage. Another hypothesis regarding $\alpha\beta$ -crystallin is the “mistaken self” hypothesis. $\alpha\beta$ -crystallin is absent from healthy lymphoid tissues, where the T-cells undergo selection. The immune tolerance towards $\alpha\beta$ -crystallin is not developed, and cells that are reactive towards $\alpha\beta$ -crystallin are free to enter the system. Any infection that causes upregulation of $\alpha\beta$ -crystallin in oligodendrocytes, will lead to inflammation of the CNS. EBV is a candidate, because it can upregulate the expression of $\alpha\beta$ -crystallin in B-cells, which in turn can present the protein to CD4+ T-cells and activate them (reviewed in [142]). MS patients also have a decreased CD8+ T-cell reactivity towards EBV-infected B-cells, and decreased T-cell control of Ig-secreting B-cells [32].

As my three studies are retrospective studies, and lack samples before disease onset we can't really answer any questions regarding causality, only report findings of associations between MS and EBV and CMV. This is an inherent limit of all retrospective studies. The only way to get past that would be to either perform a vast, expensive cohort study, or to try and locate samples taken before disease onset. This has been done as a part of the Nurses' Health study in the US [35] and at Kaiser Permanente Hospital in California [133]. Researchers at Umeå university also have a biobank of prospectively collected samples from MS cases and controls [36]. The lack of significant findings in the prospective CMV meta-analysis in paper III could be due to lack of power, with only three small studies, it is a possibility. It is however also possible that there really is no difference in CMV seropositivity prior to onset of the MS disease. The changes in the immune system leading to MS usually start years before the disease symptoms manifest themselves. An aberrant immune system can lead to decreased ability to respond to a CMV infection, leading to lower

levels of anti-CMV antibody response, and lower frequency of anti-CMV antibody positivity among MS-patients.

Self-reported environmental factors, such as infectious mononucleosis, BMI, and smoking, to name a few, are all subject to recall bias. It is known that the MS-patients spend a longer time to fill in the questionnaire. The reasons could be that they are more interested in getting correct answers, calling parents etc do discuss previous infections and vaccinations, or that they, although incident cases, have slight cognitive impairments and need more time. The tendency to over-or under-report certain “bad” habits, such as drinking, smoking etc, can also differ between individuals, and between MS cases and controls, in which case they can influence the estimates. For each case, two controls are chosen at random from the population registry (matched as described) but in some instances no control, or only one of the controls, participate in the study. The EIMS study recruits around 300-350 MS patients per year, out of the approximately 600 new cases diagnosed in Sweden each year. If these cases and controls that actually answer differ somewhat from non-respondents, it could introduce some kind of selection bias. These factors are important as they all influence the estimates, and should be taken into consideration when assessing the results.

6.2 HOST GENETICS (PAPER IV)

In paper IV we analysed the possible influence of host genetics on anti-JCV antibody positivity and anti-JCV normalised OD-values, using meta-analysis of the results from three different populations; a set of Scandinavian (Sweden and Denmark) MS cases, a set of German MS cases, and a set of Swedish controls.

6.2.1 Results from paper IV

The demographic data on the three datasets is presented in table 1, paper IV. They have similar age at sampling and percentage of women. Differences in JCV nOD median between the Scandinavian and German cases could be due to different viral strains and type of material (plasma, serum) tested, as well as other differences in the two groups (disease duration, previous treatments etc). The controls have higher percentage of anti-JCV antibody positive individuals than any of the case groups. Indeed, a χ^2 -test indicated a weak, but significant, negative association, $p=0.02$.

Non-HLA adjusted meta-analysis

In the first meta-analysis we did not adjust for the HLA-region, and it soon became apparent that the strongest signal in both analyses, the anti-JCV antibody status and the anti-JCV nOD levels, came from the MHC/HLA region on chromosome 6 (figure 1A and B, paper IV). Zooming in on the HLA-region indicated that the strongest signal came from markers in or near the class II genes *DRB1*, *DQA1* and *DQB1* (figure 2A and B, paper IV). In both tests, markers in the HLA-region reached genome-wide significance (p-value less than 5×10^{-8}), with $p=4 \times 10^{-14}$ for anti-JCV antibody status, and $p=4 \times 10^{-12}$ for anti-JCV nOD-levels, respectively (fig 2, paper IV).

Using a cut-off of $p < 1 \times 10^{-4}$, 24 non-HLA loci were associated with anti-JCV antibody positivity (table 2, fig 2A, paper IV) and 43 were associated with anti-JCV nOD values (table 3, fig 2B, paper IV).

Since this strong signal from chromosome 6 warranted further analyses, and we had access to classically typed or imputed HLA-genotypes from three different sources in a much larger material (table 1, paper IV), a logical step was to analyse association of these to the two outcomes, anti-JCV antibody status and anti-JCV nOD values.

HLA-associated with anti-JCV antibody status

Our data indicates that the strongest associated HLA-haplotype was the *DRB1*15-DQA1*01:02-DQB1*06:02*-haplotype; for *HLA-DRB1*15*, OR=0.42 in Scandinavian MS cases, OR=0.54 in German MS cases and OR=0.54 among the Swedish controls. This strong negative association is interesting, as it is the strongest MS risk susceptibility haplotype. The two other alleles on this haplotype, *DQB1*06:02* and *DQA1*01:02*, also showed strong protective associations. The *DRB1*15*-

*DQB1*06:02-DQA1*01:02* haplotype showed the strongest association to transformed anti-JCV nOD levels in the linear regression analysis, $p=0.001$ for all alleles, β -0.109 for *DRB1*15*, and -0.117 for *DQA1*0102*. The protective effect seemed stronger in controls (table 5, paper IV).

A haplotype which showed positive association with anti-JCV antibody positivity was the *DRB1*13-DQB1*06:03-DQA1*01:03-haplotype*; with an OR= 1.62 in Scandinavian MS cases, OR=1.58 in German MS cases, and OR=1.55 in Swedish controls, for *DRB1*13* (table 4, paper IV).

HLA-adjusted meta-analysis

The HLA-alleles that showed independent association in multivariate analyses were included as covariates in a second, HLA-adjusted, meta-analysis performed. There was no remaining association ($p<0.0001$) in the HLA-region in the HLA-adjusted analysis. Thirty-eight non-HLA loci were associated ($p<0.0001$) with anti-JCV antibody status (table 6, figure 3A, paper IV), and 40 loci associated with transformed anti-JCV nOD levels (table 7, figure 3B, paper IV).

6.2.2 Discussion on paper IV

To my knowledge, this is the first time a host genetic influence on JCV infection has been proven. The strong signals from the class II genes in the HLA-region are interesting, and this region has also been shown to influence the anti-EBNA1 response [143], indicating that this region is important in the response to certain viral infections. As around 40-60% of all individuals are carriers of the virus, it is possible that many will encounter this virus repeatedly, but never develop antibody responses. For individuals with a certain genetic background, the combination of viral antigen and HLA class II molecule, can lead to a more efficient clearing of the virus and any traces of an infection will not be measureable. Conversely, a different genetic background will lead to latency being established, and presence of high levels of anti-viral antibodies in blood.

The top (lowest p -value and low I^2 -score) non-HLA loci presented in tables 2, 3, 6 and 7 in paper IV were used in an extensive literature search, to try and identify some of the pieces of the puzzle. The most interesting genes are outlined in the discussion in paper IV, but I will comment on two interesting candidates here.

SAP18 was one of the top hits in the analysis of anti-JCV antibody status. *SAP18* is part of a protein-complex called histone Deacetylase (HDAC). This complex is involved in cellular transcription and can be incorporated into HIV-virions. *SAP18* can interact with a protein called *INI1* (another part of the HDAC complex). *INI1* seems to be necessary for successful HIV-1 replication [144]. If this is true for JCV remains

to be seen, but unsuccessful replication of JCV could lead to abrogation of the primary infection, and the individual could remain anti-JCV antibody negative.

Of the loci associated with anti-JCV nOD values, Rho GTPase activating protein 44, encoded by *ARHGAP44*, also known as *RICH2*, was one of the most interesting loci. It has been associated with HIV progression to AIDS [145] and prevention of HIV virion budding and release from cells. A faulty budding of virions will lead to inhibited spread of the infection. A GeneCards [130] analysis of *ARHGAP44* shows that it is expressed in many cell types known to be susceptible to JCV infection.

The negative association between anti-JCV antibody positivity and MS needs to be studied further, to adjust for other genetic and environmental MS risk factors. The protective effect of *HLA-DRB1*15* on anti-JCV antibody positivity, which is a very common variant among MS cases, supports our results that the MS cases are, to a lesser extent, anti-JCV antibody positive than controls. Either they are more resistant to the infection, or they are able to efficiently clear the infection, leaving no traces of an infection (such as antibodies in serum or plasma). It has also been shown that natalizumab treatment leads to a slight decrease in anti-JCV antibody levels (Warnke et al, accepted in Journal of Neurology, Neurosurgery, and Psychiatry).

The analysis with classically named HLA-alleles included HLA-genotypes from three different methods: a) Sequence-specific primer PCR, a classic way of genotyping HLA, b) imputed HLA using HLA*IMP and MS GWAS genotypes, and c) imputed HLA using Immunochip genotypes. All of these methods have their own margins of error and uncertainties, but where there was more than one type of HLA-genotype available for an individual, and they mismatched, I consistently chose the classic SSP-PCR-based genotypes. The HLA-imputation used data from HapMap, 1958 British Birth Cohort, and CEPH-CEU+ (IMSGC/WTCCC2 GWAS) or HapMap, CEPH-CEU+, and 1000 genomes (Immunochip) as the reference panels. These are not entirely representative of the Scandinavian population in terms of haplotype structure, and some alleles present in Scandinavian populations are extremely rare in the reference panels. The consequence would be that some of our individuals will end up with the wrong genotype, especially on the four-digit level (i.e. DRB1*15:01-level and not two-digit, DRB1*15). This could be avoided by using a reference panel based on the Scandinavian populations, but no such dataset was available at the time of the imputation. However, this is an important point, and a Swedish or Scandinavian dataset with really dense mapping of the HLA-region could be beneficial, not only to MS-research, but to all areas where HLA and HLA-imputation are analysed and performed.

The possible heterogeneity between the three datasets could have influenced the results from the meta-analysis. The type of viral strain has not been analysed in this

study, but could be one possible variable affecting the host response. One other draw-back from this study is that one of the conditions that each marker had to meet was that it was associated in all three datasets when we did the meta-analysis. It is entirely possible, that we have missed markers that are only important in MS cases, or conversely, only important in controls. Although it would have been interesting to see these as well, since we would have smaller groups and lack power, they could have been false positive findings. The lack of power is a problem in our study, especially in the analysis of non-HLA markers, as we have very few cases and controls for a proper genome-wide association study.

7

CONCLUDING REMARKS

As mentioned earlier, MS is a complex disease with many genetic and environmental factors playing important roles in susceptibility. In figure 3 I have made an attempt to visualise the role of mainly EBV in MS pathogenesis, as I feel that there is a strong body of evidence supporting a mechanistic role of EBV in MS. The same level of evidence is not, at present, available for CMV.

In persons with MS, a decreased killing of EBV-infected B-cells may lead to a more uncontrolled infection and the generation of many EBV-positive memory B-cells remaining in the system. CD4+ and CD8+ T-cells that are reacting against EBV-epitopes may start to act on cells expressing similar molecules (MBP, $\alpha\beta$ -crystallin). Perhaps a genetic predisposition in the form of *HLA-DRB1*15* also helps in this process, with *HLA-DRB1*15* being better at presenting certain EBV and host peptides showing molecular mimicry than other HLA-alleles. Hence tolerance will be broken easier in the presence of *HLA-DRB1*15*. Vitamin D deficiency could also cause the immune system to have a lower threshold for stimulation, and act more aggressively. Vitamin D receptors are found in many cells of the immune system and many genes in the immune system have VDRE and presumably therefore their expression is affected by vitamin D levels. Smoking can induce post-translational modifications on proteins, as seen in RA [146], rendering them more immunoreactive.

EBV-infected B-cells have been found in the brain of MS-patients. It is a possibility that a reservoir of EBV-infected B-cells can be a cause of repeated influx of other immune cells, such as macrophages and T-cells, trying to remove the EBV-infected B-cells, but at the same time causing damage to surrounding tissues and nerves. Putting CMV in the context of MS pathogenesis is slightly less straightforward. Our results indicate that the negative association between CMV and MS was independent of EBNA1 levels. CMV and EBNA1 antibody levels correlated positively. In individuals with both infections, there may be some type of communication between CMV-specific T-cells and EBV-specific T-cells, making the latter less likely to also react to similar self-antigens. In those who are only CMV-positive, the immune system may acquire a more self-tolerant phenotype.

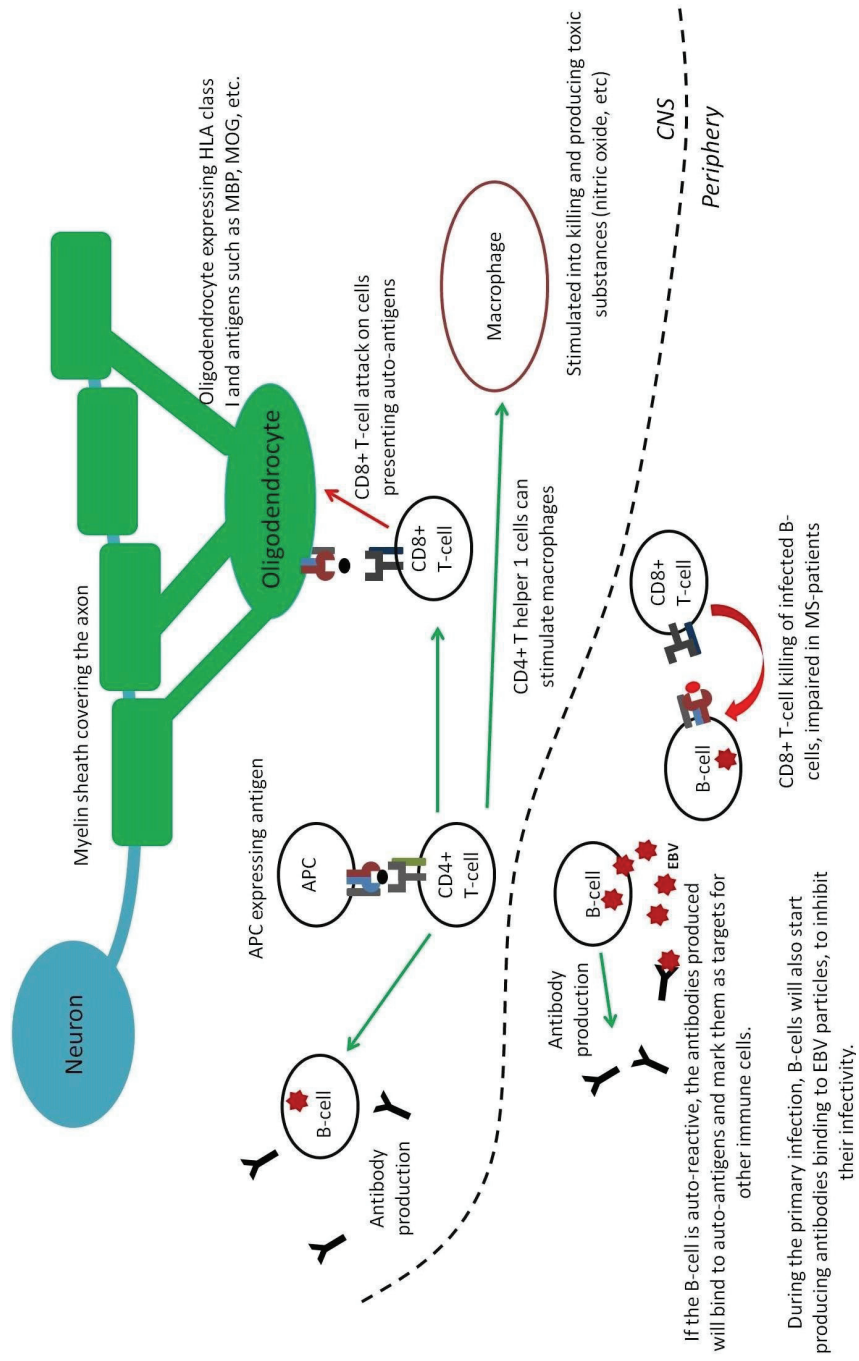


Figure 3: The interplay between EBV and the immune system and the pathogenesis of MS.

8

FUTURE PERSPECTIVES

There can be very little doubt that viruses, in particular EBV, are associated with MS. However, if viruses are a cause for MS, or a product of an altered immune response caused by the disease itself, remains to be seen. Unfortunately, the EIMS study was not designed to be able to answer those questions.

Assuming the Rothman pie model is applicable for MS, this thesis only deals with four slices of pie, CMV, EBV, and *HLA-DRB1*15* and *A*02*. It is likely that there will be pies in which all four are present, and also pies where none of these are included.

Many potential environmental risk factors have been studied in regards to their association to MS, and more are likely to be found, as more environmental epidemiological studies are being performed. With the plethora of MS-associated genes, there is great potential for performing gene-environment interactions, and gene-gene interactions. It may also become easier to identify both environmental and genetic risk factors for MS when both are considered at the same time, as not doing so may mask the association of some true risk factors.

In order to shed light on the question if the antibody response to any virus is more common in those who would later develop disease, a large prospective cohort study can be performed, locating and using samples collected before disease onset. But even then, if an environmental factor is affecting MS pathogenesis during a very short period of time, samples collected at the “wrong” time could yield no associations.

To investigate whether MS patients have different immune responses to viral epitopes than controls, cells from patients can be collected and stimulated *in vitro*. It would also be interesting to look at the viral strains and if there are any correlations between type of strain and disease status.

As for the host genetic analysis in paper IV, replication is essential, especially of the non-HLA loci with suggestive associations ($p < 0.0001$) in our meta-analyses. However, financial limitations apply, as it is very costly to both genotype and run anti-JCV antibody analyses on large number of individuals. However, suggestive non-HLA hits from this study have been included on an upcoming MS-chip, and data from that will be used to replicate our present findings.

JCV response is usually measured in individuals who are on natalizumab treatment, and samples for host genetic studies should use samples collected at baseline, as the treatment can influence the antibody levels. Further analyses on the association between MS and JCV need to be performed, adjusting for factors that are potential confounders, such as anti-EBV and anti-CMV antibody status, smoking, and sun exposure, to name a few.

As genes are established as associated with anti-JCV antibody status and anti-JCV nOD levels, it would be interesting to study genotype-expression correlations in MS cases and healthy controls, stratifying for JCV status, and also analyse differences between *HLA-DRB1*15* carriers and non-carriers.

9

ACKNOWLEDGEMENTS

I would like to thank all patients and controls, who have taken their time to answer the EIMS and IMSE questionnaires, and given their blood. I would also like to thank all physicians and nurses at all the neurology clinics responsible for the recruitment of the MS patients.

My main supervisor **Ingrid Kockum**; I can honestly say this thesis would not have been if it weren't for you. Thank you for all your support, guidance, and most of all, patience, during all these years. I have learned a lot from you.

My co-supervisor **Tomas Olsson**, thank you for all the support and interest in my projects. Your passion for science and deep knowledge in the MS/autoimmune field is remarkable. You have a unique group with lots of possibilities and a wonderful atmosphere.

My co-authors at IMM, **Anna Karin Hedström, Lars Alfredsson, and Maria Bäärnhielm**. Lars, thank you for reading and commenting my manuscripts, I've learned a lot about epidemiology from you. Maria, thank you for all the good tips about articles and our vitamin D-collaborations, good luck with your future projects.

Collaborators and co-authors **Peter Sundström, Tomas Bergström, Francesca Aloisi, Helle Bach Søndergaard, Annette B Oturai and John Carulli**. My JCV-collaborators in Germany, **Dorothea Buck, Eva Albrecht and Bernhard Hemmer**. **Maria Nyström and Hadi Daialhosein** for performing the anti-EBV and anti-CMV antibody measurements, respectively.

Nina Nordin, Karin Kai-Larsen, Karin Lycke, Linda Jonsson and Carolina Holmén for nice meetings and the wonderful job you do in coordinating and organising the EIMS, IMSE I, IMSE II and GEMS studies.

All my fellow colleagues on floor 05:

Magdalena Lindén, you came to the group just a month or two after me. Thank you for always being so friendly and helpful, for all the nice discussions on life, science and everything in between. Good luck with all your projects.

Magnus Lekman you are always willing to take the time to help out even though you have more than enough to do. You know so much about a variety of disciplines.

Izaura Lima Bomfim for all the nice talks about science and academia, epidemiology, and genetics and being an exemplary sounding board. I think you have found your place in the world. If I feel half as much at home as you do in academia, when I find my place in the world, I will be very lucky.

Jenny Link, our resident HLA-expert and fellow computer game-aficionado. Thank you for reading and commenting on an early draft of this thesis. Good luck with your photography!

Samina Asad and **Alexandra Gyllenberg** for nice discussions about life and science. Samina, thank you for all the helpful tips in writing my thesis. Best of luck in your new group and new project. Sandra, good luck with the cell cultures and finding that CIITA-expression.

Tojo James, with your background in computer science, you bring a new perspective to the group. I wish you all the best in your projects.

Cecilia Dominguez, with your HLA-project you're almost one of us now! Thank you for all the nice talks (over the electrophoresis trays on floor 04 and later over the lunch table on 05) about life in general, and a career beyond academia in particular; good luck with all your future endeavors.

Kerstin Imrell, thank you for all the nice talks and excellent company during lunch breaks and coffee breaks. You have a passion for science that I sometimes envy.

Pernilla Appelquist, you have a great energy, thank you for all the nice chats, and for trying to make us all dress a little bit better on Fridays.

Present and past **members of the Neuroimmunology group** on floor 04, especially **Mohsen Khademi** and **Venus Azhary**, for keeping the Neuro-lab running and keeping us in check. It is not an easy task! **Nada Abdelmagid**, for a great collaboration and being so nice and friendly. **Pernilla Stridh Forsgren**, for always taking the time to answer my questions on statistics. **Biborka Bereczky-Veress**, for nice company in the writing room on floor 05. **Rita Nohra**, for taking care of me during my first week in the lab. **Bob Harris** for encouraging talks about the future beyond our world at CMM, your passion for science and research education is inspiring. I would also like to thank **Alan, Amennai, André, Andreas, Ann-Marie, Brinda, Clas, Cynthia, Faiez, Fredrik, Hannah, Hannes, Harald, Karl, Kelly, Lisa, Lou, Maja, Maria B, Margarita, Marie, Melanie, Mikael, Patrick, Petra, Rasmus, Rickard, Roham, Rux, Sabrina, Shahin, Sereeni, Sohel, Xingmei, Yu Li**, and anyone else I forgot to mention for creating a wonderful atmosphere. **Lara**, continue to explore Swedish cuisine – there's more to it than Kalles kaviar! **Elin Stridh**, thank you for helping me with the Open Array genotyping project.

Everyone at the CNS-administration, especially **Gullan** for all the help with travel and expense reports, and **Elizabeth All, Rebecca Wadell**, and **Elin Johansson**. Thank you **Marianne Youssefi** for administrating the department's PhD-education.

Everyone at KI Gene; **Selim** for introducing me to the "Taqman-machine", **Annika** and **Anna-Lee** for valuable help with the Open Array project.

Olle, Daniel, Dagmar, Henrik, and **Rudolf** at the CMM IT-department for all the help with IT-related questions. **Maria, Elnaz** and **Elizabeth** at the CMM-administration.

I would also like to thank the nice people of the MS group on floor 00, especially **Elin Engdahl, Rasmus Gustafsson, Ryan Ramanujam** and **Anna Fogdell-Hahn**, for a nice collaboration. **Clemens Warnke**, for taking the time to read the JCV manuscript and

to answer all my questions regarding JCV. **Jan Hillert**, for reading and commenting on all my manuscripts. **Sahl Bedri**, good luck with your PhD-projects! **Cecilia Svarén Quiding**, for introducing me to Olerup HLA-typing. **Helga Westerlind**, good luck with raising agility's next superstar. **Malin Ryner**, **Eva Greiner**, **Wangko Lundström** and **Boel Brynedal**, thank you for making me feel welcome when I came to Huddinge. **Christina Hermanrud**, thank you for all the tips on things to see in Boston. **Anna Glaser**, for nice chats. **Virginija Danylaite Karrenbauer**, if I got a box of chocolate every time I read someone's thesis and wrote down some questions I would be very happy.

Present and past **members of the Rheumatology group** on floor 04, especially **Vijole, Alex, Shankar, Hanna, Aurelie, Emeli, and Omri** for creating a busy, but nice, atmosphere in the writing room on floor 04. **Mei**, for being so friendly and helping me when I fell and dislocated my shoulder. **Ingela**, thank you for all the nice chats. **Jayesh** good luck with your thesis and I wish you and **Tina** all the best in raising your wonderful children. **Heidi**, for your positive attitude and abundance of energy. **Maria**, for all your help with the robot. **Leonid**, for your knowledge in the field of genetics and making sure the 384-well PCR-machines run smoothly.

Friends and family outside of work; even if you only read the acknowledgments, rest assured you have in some ways contributed to this thesis.

Anna, Cecilia, Helén, Sofie and **Sophia**, we go way back. Thank you for all the nice times through the years.

Joakim and Louise, Andreas and Karolina, Katarina and Biagio, Bjarne and Kerstin thank you for welcoming me into your family.

My family, for making me who I am, my brother **Viktor and pappa Christer** thank you for all your support. **Berit**, for all your advice on everything in life, big and small. **Farmor** and all my **aunts, uncles and cousins**. **Mamma Christina**, a few years ago I realised the career path I have chosen in life is *almost* the same as yours, so maybe it is all thanks to you?

Alex, my love, thank you for everything. You are my best friend and your support has been essential during these years.

REFERENCES

1. Compston, A. and A. Coles, *Multiple sclerosis*. Lancet, 2002. **359**(9313): p. 1221-31.
2. Hernan, M.A., M.J. Olek, and A. Ascherio, *Geographic variation of MS incidence in two prospective studies of US women*. Neurology, 1999. **53**(8): p. 1711-8.
3. Ahlgren, C., A. Oden, and J. Lycke, *High nationwide prevalence of multiple sclerosis in Sweden*. Mult Scler, 2011. **17**(8): p. 901-8.
4. Fisniku, L.K., et al., *Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis*. Brain, 2008. **131**(Pt 3): p. 808-17.
5. Compston, A. and A. Coles, *Multiple sclerosis*. Lancet, 2008. **372**(9648): p. 1502-17.
6. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria*. Ann Neurol, 2011. **69**(2): p. 292-302.
7. Poser, C.M., et al., *New diagnostic criteria for multiple sclerosis: guidelines for research protocols*. Ann Neurol, 1983. **13**(3): p. 227-31.
8. Krumbholz, M., et al., *B cells and antibodies in multiple sclerosis pathogenesis and therapy*. Nat Rev Neurol, 2012. **8**(11): p. 613-23.
9. Srivastava, R., et al., *Potassium channel KIR4.1 as an immune target in multiple sclerosis*. N Engl J Med, 2012. **367**(2): p. 115-23.
10. Elliott, C., et al., *Functional identification of pathogenic autoantibody responses in patients with multiple sclerosis*. Brain, 2012. **135**(Pt 6): p. 1819-33.
11. Stys, P.K., et al., *Will the real multiple sclerosis please stand up?* Nat Rev Neurosci, 2012. **13**(7): p. 507-14.
12. Sorensen, P.S., et al., *Clinical importance of neutralising antibodies against interferon beta in patients with relapsing-remitting multiple sclerosis*. Lancet, 2003. **362**(9391): p. 1184-91.
13. Bielekova, B. and B.L. Becker, *Monoclonal antibodies in MS: mechanisms of action*. Neurology, 2010. **74 Suppl 1**: p. S31-40.
14. Sawcer, S., *The complex genetics of multiple sclerosis: pitfalls and prospects*. Brain, 2008. **131**(Pt 12): p. 3118-31.
15. Sawcer, S., et al., *Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis*. Nature, 2011. **476**(7359): p. 214-9.
16. Fogdell-Hahn, A., et al., *Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease*. Tissue Antigens, 2000. **55**(2): p. 140-8.
17. Brynedal, B., et al., *HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis*. PLoS One, 2007. **2**(7): p. e664.
18. Link, J., et al., *Importance of human leukocyte antigen (HLA) class I and II alleles on the risk of multiple sclerosis*. PLoS One, 2012. **7**(5): p. e36779.
19. Cree, B.A., et al., *A major histocompatibility Class I locus contributes to multiple sclerosis susceptibility independently from HLA-DRB1*15:01*. PLoS One, 2010. **5**(6): p. e11296.
20. Barrett, J.C., et al., *Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes*. Nat Genet, 2009. **41**(6): p. 703-7.
21. Nejentsev, S., et al., *Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A*. Nature, 2007. **450**(7171): p. 887-92.
22. Fernando, M.M., et al., *Defining the role of the MHC in autoimmunity: a review and pooled analysis*. PLoS Genet, 2008. **4**(4): p. e1000024.

23. Yang, T., et al., *Association of HLA-B27 genetic polymorphisms with ankylosing spondylitis susceptibility worldwide: a meta-analysis*. *Mod Rheumatol*, 2013.
24. Kurko, J., et al., *Genetics of Rheumatoid Arthritis - A Comprehensive Review*. *Clin Rev Allergy Immunol*, 2013.
25. Gourraud, P.A., et al., *The genetics of multiple sclerosis: an up-to-date review*. *Immunol Rev*, 2012. **248**(1): p. 87-103.
26. Lundmark, F., et al., *Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis*. *Nat Genet*, 2007. **39**(9): p. 1108-13.
27. Gregory, S.G., et al., *Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis*. *Nat Genet*, 2007. **39**(9): p. 1083-91.
28. Hafler, D.A., et al., *Risk alleles for multiple sclerosis identified by a genomewide study*. *N Engl J Med*, 2007. **357**(9): p. 851-62.
29. Brettschneider, J., et al., *IgG antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome*. *PLoS One*, 2009. **4**(11): p. e7638.
30. Farrell, P.J., *Role for HLA in susceptibility to infectious mononucleosis*. *J Clin Invest*, 2007. **117**(10): p. 2756-8.
31. Goodin, D.S., *The causal cascade to multiple sclerosis: a model for MS pathogenesis*. *PLoS One*, 2009. **4**(2): p. e4565.
32. Pender, M.P., *The essential role of Epstein-Barr virus in the pathogenesis of multiple sclerosis*. *Neuroscientist*, 2011. **17**(4): p. 351-67.
33. Sundstrom, P., et al., *Antibodies to specific EBNA-1 domains and HLA DRB1*1501 interact as risk factors for multiple sclerosis*. *J Neuroimmunol*, 2009. **215**(1-2): p. 102-7.
34. Sundqvist, E., et al., *Epstein-Barr virus and multiple sclerosis: interaction with HLA*. *Genes Immun*, 2012. **13**(1): p. 14-20.
35. Ascherio, A., et al., *Epstein-Barr virus antibodies and risk of multiple sclerosis: a prospective study*. *JAMA*, 2001. **286**(24): p. 3083-8.
36. Sundstrom, P., et al., *An altered immune response to Epstein-Barr virus in multiple sclerosis: a prospective study*. *Neurology*, 2004. **62**(12): p. 2277-82.
37. Thacker, E.L., F. Mirzaei, and A. Ascherio, *Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis*. *Ann Neurol*, 2006. **59**(3): p. 499-503.
38. Ascherio, A. and K.L. Munger, *Environmental risk factors for multiple sclerosis. Part I: the role of infection*. *Ann Neurol*, 2007. **61**(4): p. 288-99.
39. Dollard, S.C., S.D. Grosse, and D.S. Ross, *New estimates of the prevalence of neurological and sensory sequelae and mortality associated with congenital cytomegalovirus infection*. *Rev Med Virol*, 2007. **17**(5): p. 355-63.
40. Cannon, M.J., D.S. Schmid, and T.B. Hyde, *Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection*. *Rev Med Virol*, 2010. **20**(4): p. 202-13.
41. Cannon, M.J., *Congenital cytomegalovirus (CMV) epidemiology and awareness*. *J Clin Virol*, 2009. **46 Suppl 4**: p. S6-10.
42. Vescovini, R., et al., *Different contribution of EBV and CMV infections in very long-term carriers to age-related alterations of CD8+ T cells*. *Exp Gerontol*, 2004. **39**(8): p. 1233-43.
43. Stern, J.L. and B. Slobedman, *Human cytomegalovirus latent infection of myeloid cells directs monocyte migration by up-regulating monocyte chemotactic protein-1*. *J Immunol*, 2008. **180**(10): p. 6577-85.
44. Hahn, G., R. Jores, and E.S. Mocarski, *Cytomegalovirus remains latent in a common precursor of dendritic and myeloid cells*. *Proc Natl Acad Sci U S A*, 1998. **95**(7): p. 3937-42.
45. Mowry, E.M., et al., *Vitamin D status and antibody levels to common viruses in pediatric-onset multiple sclerosis*. *Mult Scler*, 2011. **17**(6): p. 666-71.

46. Myhr, K.M., et al., *Altered antibody pattern to Epstein-Barr virus but not to other herpesviruses in multiple sclerosis: a population based case-control study from western Norway*. J Neurol Neurosurg Psychiatry, 1998. **64**(4): p. 539-42.
47. Zivadinov, R., et al., *Positivity of cytomegalovirus antibodies predicts a better clinical and radiological outcome in multiple sclerosis patients*. Neurol Res, 2006. **28**(3): p. 262-9.
48. Banwell, B., et al., *Clinical features and viral serologies in children with multiple sclerosis: a multinational observational study*. Lancet Neurol, 2007. **6**(9): p. 773-81.
49. Bray, P.F., et al., *Epstein-Barr virus infection and antibody synthesis in patients with multiple sclerosis*. Arch Neurol, 1983. **40**(7): p. 406-8.
50. Waubant, E., et al., *Common viruses associated with lower pediatric multiple sclerosis risk*. Neurology, 2011. **76**(23): p. 1989-95.
51. Horakova, D., et al., *Environmental Factors Associated with Disease Progression after the First Demyelinating Event: Results from the Multi-Center SET Study*. PLoS One, 2013. **8**(1): p. e53996.
52. Pirko, I., et al., *CMV infection attenuates the disease course in a murine model of multiple sclerosis*. PLoS One, 2012. **7**(2): p. e32767.
53. Hernan, M.A., M.J. Olek, and A. Ascherio, *Cigarette smoking and incidence of multiple sclerosis*. Am J Epidemiol, 2001. **154**(1): p. 69-74.
54. Hedstrom, A.K., et al., *Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis*. Neurology, 2009. **73**(9): p. 696-701.
55. Lalmohamed, A., et al., *Causes of death in patients with multiple sclerosis and matched referent subjects: a population-based cohort study*. Eur J Neurol, 2012. **19**(7): p. 1007-14.
56. Hedstrom, A., et al., *Exposure to environmental tobacco smoke is associated with increased risk for multiple sclerosis*. Mult Scler, 2011.
57. Hedstrom, A., et al., *Nicotine might have a protective effect in the etiology of multiple sclerosis*. Mult Scler, 2013.
58. Hedstrom, A.K., et al., *Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis*. Brain, 2011. **134**(Pt 3): p. 653-64.
59. Baarnhielm, M., et al., *Sunlight is associated with decreased multiple sclerosis risk: no interaction with human leukocyte antigen-DRB1*15*. Eur J Neurol, 2012. **19**(7): p. 955-62.
60. Ramagopalan, S.V., et al., *A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution*. Genome Res, 2010. **20**(10): p. 1352-60.
61. Disanto, G., et al., *Vitamin D receptor binding, chromatin states and association with multiple sclerosis*. Hum Mol Genet, 2012. **21**(16): p. 3575-86.
62. *Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20*. Nat Genet, 2009. **41**(7): p. 824-8.
63. Sundqvist, E., et al., *Confirmation of association between multiple sclerosis and CYP27B1*. Eur J Hum Genet, 2010. **18**(12): p. 1349-52.
64. Hedstrom, A.K., T. Olsson, and L. Alfredsson, *High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women*. Mult Scler, 2012. **18**(9): p. 1334-6.
65. Langer-Gould, A., et al., *Childhood obesity and risk of pediatric multiple sclerosis and clinically isolated syndrome*. Neurology, 2013.
66. Hedstrom, A.K., et al., *Shift work at young age is associated with increased risk for multiple sclerosis*. Ann Neurol, 2011. **70**(5): p. 733-41.
67. Monaco, M.C., et al., *JC virus infection of hematopoietic progenitor cells, primary B lymphocytes, and tonsillar stromal cells: implications for viral latency*. J Virol, 1996. **70**(10): p. 7004-12.

68. Knowles, W.A., et al., *Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40*. J Med Virol, 2003. **71**(1): p. 115-23.
69. Ferenczy, M.W., et al., *Molecular biology, epidemiology, and pathogenesis of progressive multifocal leukoencephalopathy, the JC virus-induced demyelinating disease of the human brain*. Clin Microbiol Rev, 2012. **25**(3): p. 471-506.
70. Brown, P., T. Tsai, and D.C. Gajdusek, *Seroepidemiology of human papovaviruses. Discovery of virgin populations and some unusual patterns of antibody prevalence among remote peoples of the world*. Am J Epidemiol, 1975. **102**(4): p. 331-40.
71. Warnke, C., et al., *Anti-JC-virus antibody prevalence in a German MS cohort*. Mult Scler, 2012. **18**(7): p. 1054-5.
72. Trampe, A.K., et al., *Anti-JC virus antibodies in a large German natalizumab-treated multiple sclerosis cohort*. Neurology, 2012. **78**(22): p. 1736-42.
73. Outteryck, O., et al., *Anti-JCV antibody prevalence in a French cohort of MS patients under natalizumab therapy*. J Neurol, 2012. **259**(11): p. 2293-8.
74. Padgett, B.L., et al., *Cultivation of papova-like virus from human brain with progressive multifocal leukoencephalopathy*. Lancet, 1971. **1**(7712): p. 1257-60.
75. Warnke, C., et al., *Natalizumab and progressive multifocal leukoencephalopathy: what are the causal factors and can it be avoided?* Arch Neurol, 2010. **67**(8): p. 923-30.
76. Monaco, M.C. and E.O. Major, *The link between VLA-4 and JC virus reactivation*. Expert Rev Clin Immunol, 2012. **8**(1): p. 63-72.
77. Bloomgren, G., et al., *Risk of natalizumab-associated progressive multifocal leukoencephalopathy*. N Engl J Med, 2012. **366**(20): p. 1870-80.
78. Takeuchi, O. and S. Akira, *Innate immunity to virus infection*. Immunol Rev, 2009. **227**(1): p. 75-86.
79. Yoneyama, M., et al., *The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses*. Nat Immunol, 2004. **5**(7): p. 730-7.
80. Mandelboim, O., et al., *Recognition of haemagglutinins on virus-infected cells by NKp46 activates lysis by human NK cells*. Nature, 2001. **409**(6823): p. 1055-60.
81. Nuyts, A., et al., *Dendritic cells in multiple sclerosis: key players in the immunopathogenesis, key players for new cellular immunotherapies?* Mult Scler, 2013.
82. Serafini, B., et al., *Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis*. Brain Pathol, 2004. **14**(2): p. 164-74.
83. Magliozzi, R., et al., *B-cell enrichment and Epstein-Barr virus infection in inflammatory cortical lesions in secondary progressive multiple sclerosis*. J Neuropathol Exp Neurol, 2013. **72**(1): p. 29-41.
84. Mayer, M.C. and E. Meinl, *Glycoproteins as targets of autoantibodies in CNS inflammation: MOG and more*. Ther Adv Neurol Disord, 2012. **5**(3): p. 147-59.
85. Mathey, E.K., et al., *Neurofascin as a novel target for autoantibody-mediated axonal injury*. J Exp Med, 2007. **204**(10): p. 2363-72.
86. Derfuss, T., et al., *Contactin-2/TAG-1-directed autoimmunity is identified in multiple sclerosis patients and mediates gray matter pathology in animals*. Proc Natl Acad Sci U S A, 2009. **106**(20): p. 8302-7.
87. Racke, M.K., *Multiple sclerosis: The potassium channel KIR4.1-a potential autoantigen in MS*. Nat Rev Neurol, 2012.
88. Lennon, V.A., et al., *IgG marker of optic-spinal multiple sclerosis binds to the aquaporin-4 water channel*. J Exp Med, 2005. **202**(4): p. 473-7.
89. Power, C. and J. Elliott, *Cohort profile: 1958 British birth cohort (National Child Development Study)*. Int J Epidemiol, 2006. **35**(1): p. 34-41.

90. Elliott, J. and P. Shepherd, *Cohort profile: 1970 British Birth Cohort (BCS70)*. Int J Epidemiol, 2006. **35**(4): p. 836-43.
91. Kannel, W.B., et al., *Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study*. Ann Intern Med, 1961. **55**: p. 33-50.
92. Szklo, M. and F.J. Nieto, *Epidemiology : beyond the basics*. 2nd ed. 2007, Sudbury, Mass.: Jones and Bartlett Publishers. xiii, 489 p.
93. Rothman, K.J., *Epidemiology : an introduction*. 2002, New York, N.Y.: Oxford University Press. viii, 223 p.
94. Marras, C. and S.M. Goldman, *Genetics meets environment: evaluating gene-environment interactions in neurologic diseases*. Semin Neurol, 2011. **31**(5): p. 553-61.
95. Borenstein, M., *Introduction to meta-analysis*. 2009, Chichester, U.K.: John Wiley & Sons. xxviii, 421 p.
96. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
97. Horton, R., et al., *Gene map of the extended human MHC*. Nat Rev Genet, 2004. **5**(12): p. 889-99.
98. Robinson, J., et al., *The IMGT/HLA database*. Nucleic Acids Res, 2013. **41**(Database issue): p. D1222-7.
99. Hedrick, P.W., *Genetics of populations*. 4th ed. 2011, Sudbury, Mass.: Jones and Bartlett Publishers. xiv, 675 p.
100. Pritchard, J.K. and M. Przeworski, *Linkage disequilibrium in humans: models and data*. Am J Hum Genet, 2001. **69**(1): p. 1-14.
101. Slatkin, M., *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future*. Nat Rev Genet, 2008. **9**(6): p. 477-85.
102. Futuyma, D.J., *Evolution*. 2nd ed. 2009, Sunderland, Mass.: Sinauer Associates. xvi, 633, [62] p.
103. Esser, C. and J. Tomluk, *Reporting Hardy-Weinberg tests in case-control studies: reasons for caution but not for panic reactions*. J Invest Dermatol, 2005. **124**(5): p. 1082-3.
104. Gyorffy, B., I. Kocsis, and B. Vasarhelyi, *Missed calculations and new conclusions: re-calculation of genotype distribution data published in Journal of Investigative Dermatology, 1998-2003*. J Invest Dermatol, 2004. **122**(3): p. 644-6.
105. Gloyn, A.L. and M.I. McCarthy, *Variation across the allele frequency spectrum*. Nat Genet, 2010. **42**(8): p. 648-50.
106. Devlin, B. and K. Roeder, *Genomic control for association studies*. Biometrics, 1999. **55**(4): p. 997-1004.
107. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies*. Nat Genet, 2006. **38**(8): p. 904-9.
108. Li, Q., et al., *Genetic background comparison using distance-based regression, with applications in population stratification evaluation and adjustment*. Genet Epidemiol, 2009. **33**(5): p. 432-41.
109. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria"*. Ann Neurol, 2005. **58**(6): p. 840-6.
110. Polman, C.H., J.S. Wolinsky, and S.C. Reingold, *Multiple sclerosis diagnostic criteria: three years later*. Mult Scler, 2005. **11**(1): p. 5-12.
111. Eyre, S., et al., *High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis*. Nat Genet, 2012. **44**(12): p. 1336-40.
112. Holmen, C., et al., *A Swedish national post-marketing surveillance study of natalizumab treatment in multiple sclerosis*. Mult Scler, 2011. **17**(6): p. 708-19.
113. Olerup, O. and H. Zetterquist, *HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in*

- clinical practice including donor-recipient matching in cadaveric transplantation.* Tissue Antigens, 1992. **39**(5): p. 225-35.
114. Heinemann, F.M., *HLA Genotyping and Antibody Characterization Using the Luminex Multiplex Technology.* Transfus Med Hemother, 2009. **36**(4): p. 273-278.
 115. Cortes, A. and M.A. Brown, *Promise and pitfalls of the ImmunoChip.* Arthritis Res Ther, 2011. **13**(1): p. 101.
 116. Trynka, G., et al., *Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.* Nat Genet, 2011. **43**(12): p. 1193-201.
 117. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
 118. Abecasis, G.R., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
 119. Dilthey, A.T., et al., *HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes.* Bioinformatics, 2011. **27**(7): p. 968-72.
 120. Dilthey, A., et al., *Multi-Population Classical HLA Type Imputation.* PLoS Comput Biol, 2013. **9**(2): p. e1002877.
 121. Krishna, R.V., et al., *Solid-phase enzyme immunoassay for determination of antibodies to cytomegalovirus.* J Clin Microbiol, 1980. **12**(1): p. 46-51.
 122. Svennerholm, B., et al., *Herpes simplex virus type-selective enzyme-linked immunosorbent assay with Helix pomatia lectin-purified antigens.* J Clin Microbiol, 1984. **19**(2): p. 235-9.
 123. Gorelik, L., et al., *Anti-JC virus antibodies: implications for PML risk stratification.* Ann Neurol, 2010. **68**(3): p. 295-303.
 124. Kallberg, H., A. Ahlbom, and L. Alfredsson, *Calculating measures of biological interaction using R.* Eur J Epidemiol, 2006. **21**(8): p. 571-3.
 125. Aulchenko, Y.S., et al., *GenABEL: an R library for genome-wide association analysis.* Bioinformatics, 2007. **23**(10): p. 1294-6.
 126. Turner, S. *Annotated Manhattan plots and QQ plots for GWAS using R, Revisited.* Getting Genetics Done 2011 MONDAY, APRIL 25, 2011; Available from: <http://gettinggeneticsdone.blogspot.se/2011/04/annotated-manhattan-plots-and-qq-plots.html>.
 127. Cheng, Y.C., et al., *VarioWatch: providing large-scale and comprehensive annotations on human genomic variants in the next generation sequencing era.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W76-81.
 128. Zhang, B., S. Kirov, and J. Snoddy, *WebGestalt: an integrated system for exploring gene sets in various biological contexts.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W741-8.
 129. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
 130. Rebhan, M., et al., *GeneCards: integrating information about genes, proteins and diseases.* Trends Genet, 1997. **13**(4): p. 163.
 131. Simon, K.C., et al., *Combined effects of smoking, anti-EBNA antibodies, and HLA-DRB1*1501 on multiple sclerosis risk.* Neurology, 2010. **74**(17): p. 1365-71.
 132. Levin, L.I., et al., *Temporal relationship between elevation of epstein-barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis.* JAMA, 2005. **293**(20): p. 2496-500.
 133. DeLorenze, G.N., et al., *Epstein-Barr virus and multiple sclerosis: evidence of association from a prospective study with long-term follow-up.* Arch Neurol, 2006. **63**(6): p. 839-44.
 134. Ascherio, A. and K.L. Munger, *Epstein-barr virus infection and multiple sclerosis: a review.* J Neuroimmune Pharmacol, 2010. **5**(3): p. 271-7.
 135. Disanto, G., et al., *Assessing interactions between HLA-DRB1*15 and infectious mononucleosis on the risk of multiple sclerosis.* Mult Scler, 2013.

136. van Leeuwen, E.M., et al., *Differential usage of cellular niches by cytomegalovirus versus EBV- and influenza virus-specific CD8+ T cells*. J Immunol, 2006. **177**(8): p. 4998-5005.
137. Schwanninger, A., B. Weinberger, and B. Grubeck-Loebenstein, *Comment on "Differential usage of cellular niches by cytomegalovirus versus EBV- and influenza virus-specific CD8+ T cells"*. J Immunol, 2007. **178**(5): p. 2611-2; author reply 2612.
138. Niller, H.H., H. Wolf, and J. Minarovits, *Regulation and dysregulation of Epstein-Barr virus latency: implications for the development of autoimmune diseases*. Autoimmunity, 2008. **41**(4): p. 298-328.
139. Khan, N., et al., *Herpesvirus-specific CD8 T cell immunity in old age: cytomegalovirus impairs the response to a coresident EBV infection*. J Immunol, 2004. **173**(12): p. 7481-9.
140. Rand, K.H., et al., *Molecular approach to find target(s) for oligoclonal bands in multiple sclerosis*. J Neurol Neurosurg Psychiatry, 1998. **65**(1): p. 48-55.
141. Steinman, L., *A molecular trio in relapse and remission in multiple sclerosis*. Nat Rev Immunol, 2009. **9**(6): p. 440-7.
142. Lucas, R.M., et al., *Epstein-Barr virus and multiple sclerosis*. J Neurol Neurosurg Psychiatry, 2011. **82**(10): p. 1142-8.
143. Rubicz, R., et al., *A Genome-Wide Integrative Genomic Study Localizes Genetic Factors Influencing Antibodies against Epstein-Barr Virus Nuclear Antigen 1 (EBNA-1)*. PLoS Genet, 2013. **9**(1): p. e1003147.
144. Sorin, M., et al., *Recruitment of a SAP18-HDAC1 complex into HIV-1 virions and its requirement for viral replication*. PLoS Pathog, 2009. **5**(6): p. e1000463.
145. Le Clerc, S., et al., *Screening low-frequency SNPs from genome-wide association study reveals a new risk allele for progression to AIDS*. J Acquir Immune Defic Syndr, 2011. **56**(3): p. 279-84.
146. Klareskog, L., et al., *A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination*. Arthritis Rheum, 2006. **54**(1): p. 38-46.