Increased serological response against human herpesvirus 6A is associated with risk for multiple sclerosis

Elin Engdahl1,a, Rasmus Gustafsson1,a, Jesse Huang1,a, Martin Biström2, Izaura Lima Bomfim1, Pernilla Stridh1, Mohsen Khademi1, Nicole Brenner3, Julia Butt3, Angelika Michel1, Daniel Jons4, Maria Hortlund5, Lucia Alonso-Magdalena6, Anna Karin Hedström1,9, Louis Flamand7, Masaru Ihira8, Tetsushi Yoshikawa8, Oluf Andersen4, Jan A. Hillert1, Lars Alfredsson9,10, Tim Waterboer3, Peter Sundström2,b, Tomas Olsson1,b, Ingrid Kockum1,b, Anna Fogdell-Hahn1, b, *

1 Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden AND Center for Molecular Medicine (CMM), Stockholm, Sweden
2 Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden
3 Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
4 Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, the Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
5 Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden
6 Department of Neurology, Skåne University Hospital, Malmö, Sweden
7 Department of microbiology, infectious disease and immunology, Laval University, Québec city, Québec, Canada
8 Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Aichi, Japan
9 Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
10 Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden

Correspondence:
Anna Fogdell-Hahn, Associate Professor
Anna.Fogdell-Hahn@ki.se

These authors contributed equally to this work

Keywords: Human herpesvirus 6A, Human herpesvirus 6B, multiple sclerosis, association, risk, Epstein-Barr virus, human leukocyte antigen, serology
Human herpesvirus (HHV)-6A or HHV-6B involvement in multiple sclerosis (MS) etiology has remained controversial mainly due to the lack of serological methods that can distinguish the two viruses.

A novel multiplex serological assay measuring IgG reactivity against the immediate-early protein 1 from HHV-6A (IE1A) and HHV-6B (IE1B) was used in a MS cohort (8742 persons with MS and 7215 matched controls), and a pre-MS cohort (478 individuals and 476 matched controls) to investigate this further.

The IgG response against IE1A was positively associated with MS (OR = 1.55, p = 9x10^{-22}), and increased risk of future MS (OR = 2.22, p = 2x10^{-5}). An interaction was observed between IE1A and Epstein-Barr virus (EBV) antibody responses for MS risk (attributable proportion = 0.24, p = 6x10^{-6}). In contrast, the IgG response against IE1B was negatively associated with MS (OR = 0.74, p = 6x10^{-11}). The association did not differ between MS subtypes or vary with severity of disease. The genetic control of HHV-6A/B antibody responses were located to the Human Leukocyte Antigen (HLA) region and the strongest association for IE1A was the DRB1*13:01-DQA1*01:03-DQB1*06:03 haplotype while the main association for IE1B was DRB1*13:02-DQA1*01:02-DQB1*06:04.

In conclusion a role for HHV-6A in MS etiology is supported by an increased serological response against HHV-6A IE1 protein, an interaction with EBV, and an association to HLA genes.
INTRODUCTION

Human herpesvirus 6A (HHV-6A) and HHV-6B are closely related beta-herpesviruses with distinct biological and immunological properties as well as differences in epidemiology and disease associations (1). HHV-6B is acquired early in life (2, 3), with the vast majority of children infected before the age of two. Primary HHV-6B infection results in roseola, a disease characterized by high fever, rashes and occasional febrile seizures (3-5). As with all herpesviruses, HHV-6A and HHV-6B can establish latency and reactivate later in life, which can lead to severe diseases such as encephalitis (reviewed in (6)). Less is known about any clinical manifestations of the primary infection of HHV-6A, but this virus has repeatedly been reported to be associated with multiple sclerosis (MS) (7-12). As previous studies have been limited in size or unable to separate the HHV-6A from B serologically, a more definite view on their respective roles in MS would benefit from a comprehensive population based case-control study on the diverging serological response against these two viruses.

MS is characterized by central nervous system inflammation and demyelination, with several different disease courses: relapsing remitting MS (RRMS), secondary progressive MS (SPMS), and primary progressive MS (PPMS). The etiology of the disease includes a genetic predisposition (13, 14). Lifestyle/environmental factors, like virus infections and smoking also play a role, and they often interact with MS risk genes (15). Among virus infections, the gamma-herpesvirus Epstein-Barr virus (EBV) has remained the strongest suspect in the MS etiology (16-21). Another beta-herpesvirus, cytomegalovirus (CMV), has through serological analysis been negatively associated with MS risk (22). We here explore the potential associations of HHV-6A and B in MS, and interaction with serological response to EBV and CMV, using serology applied to both a very large incident and prevalent MS case-control material, and importantly, also a pre-MS case-control cohort.

Seroconversion against HHV-6 usually occurs in early childhood (23-25) but as the two viruses have similar proteomes, it has been difficult to distinguish anti-HHV-6A from anti-HHV-6B antibody responses. This inability is a major concern when investigating virus-specific disease associations and a possible explanation for the contradictory associations between HHV-6 IgG response and MS (26-33). However, even though HHV-6A and HHV-6B are 90% homologous, there are parts of their genome with more divergence (34, 35). The immediate-early 1 (IE1) proteins (termed IE1A for HHV-6A and IE1B for HHV-6B), encoded by the open reading frame (ORF) U90-U89, are among the most divergent with only 62% homology (36, 37) and with differences in biological properties. IE1A, but not IE1B, can transactivate several heterologous promoters (37, 38) while IE1B, but not IE1A can silence IFN-α/β signaling (39). The ORF U11 coding for p100 in HHV-6A and 101K in HHV-6B also exhibit relatively high divergence with only 81% amino acid identity (40). These structural proteins are essential for viral growth and propagation (41), and 101K has been identified as the dominant antigen recognized by anti-HHV-6B IgG (42). With the aim to discriminate between IgG responses against these two viruses, we developed a novel bead-based multiplex serology assay measuring IgG antibodies against IE1A, IE1B, p100 and 101K, selecting the most divergent parts of these protein sequences. This assay was used to screen serum or plasma samples from persons with MS, persons that later develop MS, and controls for HHV-6A and HHV-6B protein-specific antibodies.
RESULTS

High IE1A antibody response is positively associated with MS and is a risk factor for developing MS in youth

We used a novel multiplex serology assay to investigate the specific IgG responses against the HHV-6A protein IE1A and the HHV-6B protein IE1B. To investigate if persons with MS and controls differed in IgG responses, logistic regression analyses were used to compare strong and weak responders, defined as the highest or lowest quartile of each measured response. This revealed that a high IE1A antibody response was positively associated with MS (OR = 1.55, p = 9x10^-22) while a high IE1B antibody response was negatively associated with MS (OR = 0.74, p = 6x10^-11) (Table 1, Figures 1A and 1B).

To investigate if these differences also were present before MS onset, serum samples drawn from persons with RRMS at a median of 8.3 years before symptom onset and from matched controls were analyzed. Strong IE1A responders had a higher risk of developing MS later in life, compared to low responders (OR = 2.22, p = 2x10^-5) (Table 2, Figure 1D). No significant difference in IE1B IgG response was observed before MS onset (OR = 0.96, p = 0.8) (Table 2 and Figure 1E).

The pre-MS cohort was divided into age groups and further analyzed (Table 2). In individuals younger than 20 years old, strong IE1A responders had a 3.38 (p = 0.004) times higher risk of developing MS later in life. This MS risk decreased with age, reaching an OR of 1.51 (p = 0.3) in the oldest age group (30-39 years). Also in the established MS cohort the highest OR was seen in the youngest age group (Table 1).

Interaction between high IE1A and EBV antibody responses on MS risk

As antibody responses against other herpesviruses such as EBV and CMV have been associated with MS (19, 22), the interplay between HHV-6A/-6B and these viruses in MS was analyzed. The median antibody levels of EBV and CMV index among controls were used as cutoffs for strong or weak responders. Interaction analyses revealed a significant additive interaction between IE1A and EBV responses on MS risk (attributable proportion due to interaction (AP = 0.24, p = 6x10^-6, Figure 2A), meaning that 24% of the risk for developing MS in those with strong IE1A and strong EBV responses was due to interaction between these factors. No significant interaction between HHV-6A and CMV immune response was observed, neither was the HHV-6B IE1B or 101K responses interacting with EBV on MS risk (Figure 2B, SI appendix Figure S1A and B). An analysis of how the OR varied for the three HHV-6A, and B antigens depending on EBV index was investigated using a sliding window approach and shows that IE1A mediated risk for MS is limited to individuals with EBV response higher than the median among controls (SI appendix Figure S2).

Antibody levels against IE1A are higher in both established MS cases and pre-MS cases, compared to matched controls

Linear regression models were used to analyze the HHV-6A/6B IgG levels, and the results are in line with the association seen for high/low serological response. MS cases, both before
(pre-MS) and after MS onset (established MS), had higher IgG levels against IE1A than
controls (p = 6x10^{-10} and p = 9x10^{-10}, respectively; Figure 1, Table S1 and S2 in
Supplementary Material). In contrast, anti-IE1B IgG levels were lower in established MS
cases compared to controls (p = 9x10^{-14}), but this association was not observed before MS
onset. When dividing the study cohorts into age groups, IE1A reactivity was consistently
higher in MS cases compared to in controls (p = 6x10^{-10} and p = 9x10^{-30}, respectively; Figure
1, Table S1 and S2 in Supplementary Material). In contrast, anti-
IE1B IgG levels
were lower in established MS cases compared to controls (p = 9x10^{-14}), but this association was not observed before MS
onset. When dividing the study cohorts into age groups, IE1A reactivity was consistently
higher in MS cases compared to in controls (Figure 3A and B), while the pattern for the anti-
IE1B reactivity was more inconsistent (Figure S3A-D in Supplementary Material). The
associations of IE1A with MS were significant both with and without adjustment for EBV and
CMV responses, indicating that the increased IE1A response in MS was not confounded by
these two anti-viral responses (Table S1 and S2 in Supplementary Material). In addition to the
IE proteins, antibodies against the structural protein 101K (HHV-6B) and p100 (HHV-6B)
were measured. A high 101K serological response was not associated with MS or with later
development of MS in the pre-MS cohort (Table 1 and 2). Results from the p100 analysis
were excluded from further analyses due to the low reactivity against this antigen.

High IE1A response is associated with relapsing and progressive MS, but not disease
severity

The association of high IE1A responses was similar regardless of disease course (OR_{RRMS} =
1.62, p = 1x10^{-20}; OR_{SPMS} = 1.49, p = 7x10^{-7}; OR_{PPMS} = 1.53, p = 9x10^{-9}). The same was true
for high IE1B responders (OR_{RRMS} = 0.77, p = 2x10^{-6}; OR_{SPMS} = 0.67, p = 2x10^{-6}; OR_{PPMS} =
0.62, p = 7x10^{-4}). A high 101K response was associated with RRMS (OR = 1.18, p = 1x10^{-3}),
but negatively associated with PPMS (OR = 0.66, p = 2x10^{-3}).

HHV-6A and 6B serology was not associated with two MS severity scores, the Multiple
Sclerosis Severity Score (MSSS) and the Age Related Multiple Sclerosis Severity Score
(ARMSS) (43) (data not shown).

The IgG responses against HHV-6B proteins vary with age and sex

The level of IgG responses against IE1B and 101K decreased with age (p = 4x10^{-21} and p =
3x10^{-39}, respectively; Figure S3A-D in Supplementary Material). A sex difference could be
observed, with women eliciting a significantly stronger antibody response against IE1B (p =
4x10^{-4}) and 101K (p = 2x10^{-23}). Antibody levels against IE1A did not differ significantly
between the sexes nor with age.

Smoking associates with increased IE1A IgG response in persons with MS

As smoking has been reported to be a risk factor for MS disease (44) and has been associated
with higher HHV-6 IgG levels (32), the effect of smoking on HHV-6 protein-specific IgG
responses was investigated. Persons with MS and with a history of regular smoking showed
higher IE1A IgG levels compared to those who never smoked (p = 2x10^{-5}). This was not
observed in controls (p = 0.4). The responses against the other proteins were not affected by
smoking, neither in MS cases nor in controls (data not shown).
IgG responses against HHV-6 proteins are associated with different HLA haplotypes

To investigate the influence of genetic factors for the serological response against the HHV-6A and HHV-6B protein sequences, genome-wide association studies (GWAS) were performed for both IgG levels (Figure 4) and high/low response (Figure S4 in Supplementary Material). The primary genetic association was mapped to the Human Leukocyte Antigen (HLA) region (6p21). There was a clear difference in IE1A and IE1B response in regard to their association to SNPs located in the HLA region, where IE1A levels were associated (p < 5x10^-8) with 191 SNPs mapping to the HLA region while IE1B IgG levels were significantly associated with only two SNPs in this region.

Deciphering of the associations within the HLA region showed that IgG responses against the three different protein sequences were associated with different HLA haplotypes (Table 3, Tables S3 and S4 in Supplementary Material). There were some differences between cases and controls, but for most haplotypes the association was stronger when cases and controls were analyzed together (Table 3, Tables S3 and S4 in Supplementary Material). Several HLA alleles exhibited significant association with IE1A response at the cut-point level for GWAS (p < 5x10^-8). Higher IE1A IgG levels were associated with carrying the DRB1*13:01-DQA1*01:03-DQB1*06:03 haplotype in both MS cases and controls also after adjustments for HLA. The DPA1*02:01-DPB1*01:01 and the DRB1*04:01-DQA1*03-DQB1*03:02 haplotypes were associated with a lower IE1A response in both MS cases and all subjects. These associations where also still significant after adjustment for HLA (Table 3). The main HLA association for IE1B was DRB1*13:02-DQA1*01:02-DQB1*06:04, although none of the HLA associations reached the genome wide significance of p < 5x10^-8 (Table S3 in Supplementary Material). Anti-101K response showed associations with several SNPs in the HLA region, but for the individual HLA haplotypes detected, none of them reached genome-wide significance of p < 5x10^-8 (Table S4 in Supplementary Material).

Re-analyses were performed adjusting for the most associated HLA allele in each HLA haplotype reported in Table 3, and Tables S3 and S4 in Supplementary Material. The number of significant SNPs (p < 5x10^-8) decreased from 191 to 4 in the IE1A GWAS, and no SNP in the HLA locus remained significantly associated with the antibody responses against IE1B and 101K (Figure S5 and Tables S7 and S8 in Supplementary Material).

Although several HLA haplotypes were found to influence anti-HHV-6A/6B protein specific IgG responses, the association between serological responses and MS disease remained to a large extent unaltered when adjusting for carriage of associated HLA alleles (Tables S3 and S4 in Supplementary Material).

IgG response to HHV-6A and B proteins interacts with MS risk HLA alleles

We investigated the interaction between IE1A, IE1B and 101K IgG responses and the major MS risk HLA alleles DRB1*15:01 and A*02:01 in conferring risk to MS. High IgG response to IE1A interacted with both DRB1*15:01 and absence of A*02:01 (AP=0.31, p=2x10^-8 and AP=0.21, p=2x10^-8 respectively) while IE1B only interacted with DRB1*15:01 (AP=0.19, p=1x10^-3, Figure 5). The interaction between IE1A and DRB1*15:01/A*02:01 on MS risk was only observed in persons with high EBV levels, while the IE1B-DRB1*15:01 interaction was only significant in persons with low EBV levels (data not shown).
DISCUSSION

We show in a large national case-control cohort that persons with MS have higher IgG reactivity against the IE1 protein sequence from HHV-6A compared to controls, while an association in the opposite direction was observed for reactivity against the corresponding IE1 protein sequence from HHV-6B. Importantly, the positive association for IE1A was observed also in samples drawn before MS onset, indicating that it is not simply the state of chronic inflammatory disease that induces the higher level of anti-viral antibodies, but that differences in serological status precede clinical symptom onset. Since having a strong IE1A response during adolescence (< 20 years) conferred the highest risk of developing MS on average 10 years later in life (Table 2), our data argues against a reversed causation. We presume that the increased humoral immune reactivity to the selected viral proteins reflects a more intense primary infection and/or reactivation, resulting in higher viral load and therefore an increased anti-viral response. Assuming that HHV-6A infection does play a role in disease onset, this data suggests that acquisition of HHV-6A at a younger age might play an important role in triggering MS.

The positive association seen for HHV-6A (IE1A), but not for HHV-6B (IE1B), with MS disease is in line with some previous studies (8, 10-12) and is interesting considering the differences between the viruses. Both HHV-6A and 6B have the ability to remain latent in the brain (45), but only HHV-6A has been shown to infect and form latent infection in oligodendrocytes (46), the myelin-producing cell and the presumed target of the autoimmune reaction in MS. Speculatively, reactivation of the virus from oligodendrocytes could direct the immune system towards these target cells, suggesting a mechanism that would explain a selective association of HHV-6A with MS. Furthermore, human oligodendrocyte progenitor cells expressing the HHV-6A latency-associated viral protein U94A do not migrate accurately (47), which may yield insufficient myelin repair in the brain and hence could provide another potential link between HHV-6A infection and MS disease. An additional possibility for a potential causative role for HHV-6A in myelin tissue destruction is supported by in vitro data showing that supernatants from HHV-6A, but not from HHV-6B, infected cell cultures induce caspase-independent cell death (e.g. necroptosis) in oligodendrocytes (48). This virus-specific pattern is in line with the data in the present study where increased IE1A, but not IE1B, IgG levels are seen in MS plasma. Thus, an association to HHV-6A gives plausible explanations for both the myelin degradation and impaired re-myelination in MS.

Regarding interactions with already established risk factors for MS, we can report an additive interaction between strong IgG responses to IE1A and EBV (Figure 2A), not seen for IE1B (Figure S1A in Supplementary Material). This suggest that increased immune response to both viruses are involved in MS and would be consistent with studies reporting that HHV-6A, but not HHV-6B, infection of B cells resident in MS lesions can activate EBV replication (49-51). As HHV-6A infection can activate LMP-1 and EBV nuclear antigen (EBNA)-2 protein expression (50), two proteins important for EBV immortalization of B cells, one can speculate that the increased IE1A IgG levels seen in the present study may be a result of increased infection and transformation of EBV infected B cells. However, that would lead to a general increase of antibody of all specificities including the response against IE1B, which we did not find. An interaction between HHV-6A and EBV in MS has been suggested elsewhere (52), where the author hypothesizes that HHV-6A activates latent EBV in B-cells resident in MS lesions and that both viruses, together, are fundamental for the etiopathogenetic processes of MS. We could in addition observe interaction between the main MS HLA risk alleles and IE1A IgG response (Figure 5) indicating that these MS risk factors are acting jointly in

...
HHV-6A in MS

The specificity of the serological response we report here is interesting. As antigens, IE1A proteins are located in the cell nucleus and two relevant questions to consider are how the B-cell activates a response against these antigens and what functions these antibodies might have. Increased IgG responses against intra-nuclear herpesvirus proteins in MS have been reported previously for the HHV-6A and 6B protein p41, and the EBV protein EBNA-1 (53, 54). For B cells to be directed against nuclear or intracellular antigens the cell needs to be disrupted, for example through necroptosis, a form of immunogenic programmed cell death where cell swelling results in rupture of the cell membrane and release of intracellular components into the surrounding tissue (reviewed in (55, 56)). In line with this hypothesis, necroptosis markers have been observed in MS lesions (57) and the intrathecally produced antibodies characteristic of persons with MS often are directed against ubiquitous intracellular proteins (58). The role of these anti-nuclear antigen B-cell responses are less clear. Since the antibodies directed against nuclear antigens probably do not have any neutralizing effect on the viruses, they would not protect against infection, reactivation or dissemination. An anti-IEA response might be seen as a marker of both increased infection and increased tissue destruction and their function might be to clear cell debris. The relevance of these antibodies, protective or detrimental, in infections and regarding the association with autoimmune disease is yet to be determined.

Due to the high similarity and potential antibody cross-reactivity between HHV-6A and HHV-6B, methods discriminating between their serological responses have been difficult to develop. The IE1A and IE1B sequences used in our assay align to some extent (Figure S9 and Table S6 in Supplementary Material) and the possibility of cross-reactivity should not be neglected. However, the lack of correlation between the IE1A and IE1B serological measurements (Figure S6 in Supplementary Material) in combination with their associations with MS in opposite directions (Table 1 and 2), suggests that the method indeed has the potential to discriminate between HHV-6A and HHV-6B. Validating the method using serum from children with primary infection further supported this notion, where seroconversion after primary HHV-6B infection was seen for IE1B and 101K only (Table S5 in Supplementary Material).

Serological responses against all three antigens investigated in this study were mainly influenced by genetic factors in the HLA region. HLA associations with serological responses have been seen before (59) and is expected, since long lasting and IgG isotype switched B-cell response is T-cell dependent and facilitated through interaction between HLA and the T-cell receptor. The associated HLA haplotypes were relatively similar in MS and controls, suggesting that the influence of HLA had more to do with control of the viral infection than MS disease. Overall, the known MS-associated HLA haplotype DRB1*15:01 (60) was not associated with serological levels and the associations to MS were still significant after correction for the major MS associated HLA alleles, which would indicate that the serological response could not be explained solely by previous known genetic risk factors for MS. An interesting exception was the association of IE1A and presence of HLA-A*02:01 that was seen in controls, but not for persons with MS. Moreover, we found an interaction between both the DRB1*15:01 and absence of HLA-A*02 (the extended HLA haplotype confirming the highest risk for MS) with IE1A only in persons with high EBV levels, and an interaction of DRB1*15:01 with IE1B only in persons with low EBV levels. Thus, the interaction between HHV-6A and MS associated HLA alleles seems to have effect on the risk for MS only in persons with high anti-EBV response.
The IE1A antigen had the strongest HLA association in comparison with the other antigens investigated, but this might only reflect the ability of the HLA systems to respond differently to different infections on a population level. Furthermore, individuals respond differently to antigens from the same virus (Figure S6B in Supplementary Material). Difference in protein structure, location, phase of expression, and function for the IE1 or 101K proteins (37) possibly makes the immune system encounter them under divergent conditions.

In the present study we could confirm some of our previously data using the HHV-6B lysate based commercial ELISA, namely lower serological response against HHV-6B lysate (32) and the 101K protein in males and in HLA-A*02 carriers. Female sex has been associated with increased acquisition of HHV-6B in children (5), and it is possible that the HHV-6B lysate IgG (32) and 101K IgG responses reflect this difference in primary HHV-6B infection. The association with HLA-A*02, found in our previous study (32) and suggestively confirmed for anti-101K IgG response in the present study, indicates a role for CD8+ T cells in the immune response against HHV-6B. In line with this notion, 101K peptides have been shown to be presented by HLA-A*02:01 on HHV-6B infected cells and these cells are recognized and killed by CD8+ T cells (59). One can hypothesize that if infected cells are removed, the systemic viral burden may decrease, thus possibly explaining the lower levels of anti-HHV-6A/6B (32) and anti-101K IgG levels in individuals with the HLA-A*02 allele.

In conclusion, we provide strong serological data supporting a role for HHV-6A in MS etiology, though causality, as with all forms of association studies remain to be proven.

MATERIALS AND METHODS

Study subjects

Two different patient cohorts were used, one with samples taken during MS disease (Established MS, Table 4) and one where samples had been collected before MS onset (pre-MS, Table 5). Some individuals (348 of the RRMS patients and 1 control) included in the Established MS cohort were also included in the pre-MS cohort. These individuals were excluded from the Established MS cohort when association between serological response and MS disease was analyzed in both cohorts, but were included in other analyses.

Established MS cohort

The established MS cohort included persons with MS and matched controls from the Epidemiological Investigation of Multiple Sclerosis (EIMS (44), n = 5674), Genes and Environment in Multiple Sclerosis (GEMS (61), n = 8903), Immunomodulation and Multiple Sclerosis Epidemiology study (IMSE (62), n = 1079) and Stockholm Prospective Assessment of Multiple Sclerosis (SPASM / STOPMS (63), n = 301). EIMS and STOPMS have an incidence design, i.e. newly diagnosed persons with MS were invited to join these studies, whereas GEMS and IMSE have a prevalence design, i.e. already diagnosed persons with MS were invited to join these studies. To each person with MS included in GEMS and EIMS, several non-MS individuals, randomly selected from the National population register, matched for age at diagnosis, sex and residency, were invited to participate as control subject.

In total, 8742 persons with MS and 7215 controls were included (Table 4). Blood samples...
HHV-6A in MS

were either drawn and shipped over night in room temperature prior to plasma isolation or frozen within a few hours after sampling. Plasma samples were stored at −80 °C until analysis. Information regarding smoking habits were obtained for 6534 persons with MS and 5996 controls through self-reported questionnaires (64). Data regarding disease characteristics was obtained mainly through the national Swedish MS registry (65) and in some instances from medical records. All MS cases fulfilled the McDonald Criteria (66, 67) and all study participants provided written, informed consent. Some (n=497) individuals had given more than one sample and for these only the earliest collected sample was included. This study was conducted in line with the aims of the EIMS, IMSE, GEMS and SPASM/STOPMS studies, all which were approved by the Regional Ethical Review Board in Stockholm and performed according to the ethical standards of the Declaration of Helsinki.

Disease severity, Multiple Sclerosis Severity Scores (MSSS) (68) and Age Related Multiple Sclerosis Severity scores (ARMSS) (69), was calculated from expanded disability status scale (EDSS) measurements reported to the Swedish MS registry by the treating neurologist.

Pre-MS cohort

The pre-MS cohort is a prospective case-control study on biobank samples drawn before symptom onset (Table 5, Figure S8 in Supplementary Material). As of February 2012, the Swedish MS registry (65) containing 11,196 MS cases was cross-linked with three Swedish microbiological biobanks which contain the remainders of sera after clinical microbiological analyses performed at the University Hospitals of Skåne and Gothenburg, and the Public Health Agency of Sweden. The serum samples had been stored at −20 °C in these biobanks until analysis.

From individuals who later developed RRMS, only samples drawn before the age of 40 were included (n=478). Individuals who did not develop MS served as controls (n=476). Controls were matched for biobank, sex, date of blood sampling and date of birth, in order of decreasing priority. The un-even number of cases and controls was due to dropout of study participants after the matching process, leaving 474 matched sets of cases and controls. Between cases and controls the mean difference in age at sampling and date of serum collection was 65 days and 6 days, respectively. For six individuals there were no matching case or control but these individuals were still included in the analyses. In cases where registry data were incomplete, notably time of onset, the registered data were checked and corrected at one of the contributing center. The study was approved by the Regional Ethical Review Board in Umeå and performed according to the ethical standards of the Declaration of Helsinki.

Measurement of IgG antibodies

For detection of IgG antibodies against the different viral proteins, a multiplex serological assay using beads coated with recombinant glutathione s-transferase (GST) fusions proteins was used. The assay procedure has been described in detail elsewhere (70). In short, antigens were expressed as GST fusion proteins using modified pGEX vectors in E. coli. Four different HHV-6 protein sequences were expressed: HHV-6A and -6B specific regions of the IE1 protein, IE1A and IE1B respectively, and a divergent region of the structural protein 101K
(HHV-6B) and p100 (HHV-6A) (Figure S9 and Table S6 in Supplementary Material). The antigen expressing bacteria were lysed, the lysate cleared of insoluble components and thereafter in situ purified on specific polystyrene beads set (SeroMap, Luminex Corporation) coupled to glutathione-casein (GC). Plasma/serum from the study subjects were diluted 1:1000 and pre-incubated with GST lysate (71) to remove antibodies specific for GST or wildtype bacterial proteins present in the lysate. Beads coated with different antigens were mixed and incubated with pre-incubated plasma/serum in filter bottom 96-well plates. After washing, a biotinylated goat-anti human IgG secondary antibody (Dianova) was added and, after additional washing steps, detected by streptavidin-R-phycoerythrin (Moss). Median fluorescence intensity (MFI) was measured with a Luminex 200 analyzer. On every plate, four plate controls were tested to assess assay variation, as described in detail elsewhere (72, 73), yielding antigen-specific coefficients of variation of 13.2% – 18.7% for the EBV and CMV antigens. The plate controls did not react with any of the HHV-6A or B antigens.

In addition to the HHV-6 antigens, antibody responses against other viral proteins were measured with this multiplex assay (as described above) (74). The IgG responses against four CMV antigens (pp150, pp52, pp28, pp65) and two EBV antigens (one EBNA1 peptide sequence, aa 385-420 (75), and VCA p18) were used to calculate CMV and EBV indexes. These indexes reflect an overall response against several epitopes of the viruses on a continuous scale. The indexes for one individual were calculated as the sum of the fractions of MFI determined by the measured virus protein over the median MFI among controls of that specific protein. In addition, three HHV-6A proteins (major capsid protein (MCP), U94A and p100) were first included as antigens but were excluded due to low reactivity.

### Batch control

The samples were run in different batches over different days. To compensate for the batch-to-batch variation, inter-batch controls (two plates with 180 samples) were analyzed within each run and used to correct for the variations. Standard linear model, or non-linear models (e.g. logarithmic, exponential) where appropriate, were used to adjust the batch variation and only batch-corrected MFI values are presented in this study.

### Validation of assay

To investigate the specificity of the Luminex assay, samples from 10 children with exanthema subitum (ES; HHV-6B primary infection) were investigated. For all individuals, one sample was available from the acute ES phase and one from the convalescence phase. The median time between these samples was 7.5 days. If considering 7 days as the minimum time to mount an antibody response, only 5 individuals were retained in the analysis as individuals that should be HHV-6B positive. Comparing the samples collected during the acute ES phase with the samples collected during the convalescence phase of these five HHV-6B infected children, a ≥10 times increase in MFI value was observed for 4 individuals (80%) for the 101K antigen, and for 2 individuals (40%) for IE1B. The antibody responses against the two HHV-6A protein sequences (IE1A and p100) were not increased in the samples from the convalescence phase. Although based on few individuals, these results suggest that the assay can detect seroconversions upon primary HHV-6B infection and that negligible cross-reactivity occurs between 101K and p100, or between IE1A and IE1B.
Measured reactivities against IE1A, IE1B, 101K and p100 for the ES samples are presented in Table S3 in Supplementary Material. The ES control samples were analyzed separately from the MS study samples using a secondary antibody directed against IgA+IgM+IgG (Dianova) instead of the anti-IgG antibody described above, and the results from 1:100 dilutions were used instead of 1:1000 as for the MS samples. Thus, a direct comparison of the MFI values between the MS and ES samples is not possible, instead the relative MFI shifts between the paired samples are more informative.

To validate if the differential seroresponse between acute and convalescent ES (n=5) constitute a true seroconversion, the difference in measures were compared to the general variability of antibody response from 39 paired (established MS) samples collected <2 months apart, but with a minimum time of 7 days between the two samples. The response against 101K was significantly elevated in the convalescent sample from the ES cohort compared to the second sample of the adult MS cohort (p=0.0003, Mann Whitney U test). The relative changes of the other antigens were not significantly different.

Together, these results indicate that the assay detect HHV-6B IgG responses correctly, but whether an HHV-6A infection elicit only IE1A and/or p100 IgG responses could not be determined.

Statistics

Statistical analyses were performed in the software R, version 3.4, on data from the established MS cohort, and with SPSS, version 23, on data from the pre-MS cohort. All graphs were constructed in R version 3.4.

To control for potential confounders, the antibody responses were analyzed using regression models adjusted for age and sex. The established MS cohort was also adjusted for study design (incidence or prevalence design) in all regression analyses. As the p100 IgG responses were mainly in the technical noise area of the assay, the anti-p100 response was therefore excluded from the analyses.

Associations between MS and antibody status were investigated using logistic regression models comparing the MS frequency in strong and weak responders. Individuals with MFI values in the 4th quartile among controls were regarded as strong responders and individuals with MFI values in 1st quartile among controls were regarded as weak responders. To get a clear separation between strong and weak responders, the in-between responders (individuals with MFI values in between 25th and 75th percentile) were excluded in this analysis. Strong and weak responders may reflect seropositive and seronegative individuals, but true serostatus cannot be determined due to lack of validated control samples. The quartiles were determined separately for the two cohorts and are indicated in Figure 1.

To assess if additive interaction occurred between antibody responses to different viruses and between MS HLA risk alleles and HHV-6 immune response in and MS risk, and between MS HLA risk alleles and antibody response, interaction analyses were performed in the established MS cohort. This was done by calculating the proportion attributable to interaction (AP) using logistic regression analysis, where the odds ratios were calculated in relation to the group with the lowest MS risk (76).
Antibody levels were analyzed using linear regression models. The IgG responses were heavily left-skewed. To obtain a more normal distribution, the antibody levels (MFI values) were transformed using a log base 10 transformation prior to statistical analysis. When investigating how MS disease influenced the antibody levels, the models were conducted with unadjusted or adjusted data for potential confounders in order to see if the association between MS and antibody levels remained (Tables S1 and S2 in Supplementary Material). In this analysis, both study cohorts were adjusted for EBV and CMV index and the established MS cohort was also adjusted for carriage of HLA alleles (MS risk alleles HLA-A*02:01, HLA-DRB1*15:01 and the alleles associated with each antibody response indicated in Table 3 and Table S3 and S4 in Supplementary Material).

Correlations between antibody measurements were investigated with Spearman correlation tests (Figures S5 and S6 in Supplementary Material).

The main hypothesis tested in this study is that HHV-6A/6B IgG responses are associated with MS disease. To test this hypothesis, three antibody specificities were measured in two different study cohorts, corresponding to 6 independent tests thus using the Bonferroni correction on an alpha level of 0.05 yields a threshold of 0.008 for significance. When other hypotheses were answered, the same alpha value of 0.008 was regarded as significant.

GWAS

Genotypes for ~720,000 single nucleotide polymorphisms (SNPs) were determined using an Illumina OmniExpress BeadChip for 6,396 MS cases and 5,530 controls from the established MS cohort. SNPs with a minor allele frequency of < 2%, with a call-rate of less than 98% or those which were not in Hardy-Weinberg equilibrium among controls (p < 0.0001) were removed from analysis. Individuals with > 2% failed genotype calls, with increased heterozygosity (> mean + 2 SD), related individuals (increased identity by descent, IBD), or individuals where the recorded sex differed from the genotype result were removed from analysis. Population outliers identified using the SmartPCA program were removed. A principal component analysis was conducted using Eigensoft (77) and five PCA components were used to control for population stratification. Linear regression models investigating the association between SNPs and Log10-transformed IE1A, IE1B and 101K antibody levels were analyzed both separately for MS cases and controls, and for all subjects together, using PLINK v1.9 (78) adjusted for 5 PCA vectors, age, sex, cohort type (and MS disease when all subjects were analyzed together). Logistic regression models, adjusted as the linear regression models, were performed to investigate the association between SNPs and strong or weak response (comparing the 3rd and 1st quartile respectively) against IE1A, IE1B, and 101K. Associations for SNPs with p < 5x10^-8 were regarded as significant, but all SNPs with suggestive association (p < 10^-5) are reported in Tables S4 and S5 in Supplementary Material. Genome build GRCh37 was used.

HLA-imputation and Associations with HLA haplotypes

HLA allele variants for MHC class I and II were imputed by the software HLA*IMP:02 (79) for 7,063 MS cases and 6,098 controls using genotypes from the MS Replication Chip (80).
This chip densely covered in the MHC region. Associations between HLA alleles and IgG responses against each HHV-6 antigen were determined by linear regression models using R version 3.3.1. Associated alleles were combined into haplotypes using previously reported common haplotypes in the Caucasian population (81). Analyses were stratified by MS affection status, but were also conducted on persons with MS patients together with controls. Models were adjusted for age, sex, cohort type (incidence or prevalence), 6 PCA vectors, and for MS disease status when all subjects were analyzed together. To determine the allele with the strongest effect in each associated haplotype, stepwise conditional analyses were performed. Secondary analyses were made where associated HLA alleles were added as co-variables to test for independent associations.

ACKNOWLEDGEMENTS

The authors thank Dr. Steven Jacobson, Dr. Elisabetta Caselli and Dr. Dario Di Luca for providing serum for method validation. The authors also thank Ingileif Jonsdottir and Kári Stefánsson for genotyping done by deCODE, and Alexander T Dithey at Wellcome Trust Centre for Human Genetics, University of Oxford, for contribution in HLA imputation. Swedish Research Council (grant no 2015-02419), Swedish Brain Foundation, KAW Foundation, Margareta af Ugglas Foundation. J Huang, P Stridh and I Kockum were supported partly by funding from MultipleMS Horizon 2020 grant number 733161. J Huang was also partly supported by EndMS doctoral studentship EGID: 3045 from the Multiple Sclerosis society of Canada.

AUTHOR CONTRIBUTIONS STATEMENT

EE contributed to design and interpretation of the work, acquisition, statistical analyses, design of tables and figures for the establish MS cohort. RG identified the antigens used. JH contributed to genetic analyses and choice of statistical methods. MB was responsible for the pre-MS cohort. ILB, PS, MK were responsible for the MS case control cohort database and biobank. ILB did HLA imputation and selection of viral antigens. PS managed the GWAs genotyping. AKH was responsible for the EIMS study and HLA interactions. NB, JB and AM set up and validated the multiplex serological assay. DJ, MH and LAM contributed to the pre-MS cohort. LF helpful in designed the protein antigens. MI and TY were contributing to the validation. OA contributed to the pre-MS and MS cohort. JAH and LA initiated of register and biobank. TW was responsible for the multiplex serological assay. PS initiated the pre-MS cohort and supervised interpretation. TO initiated the serological analyses, initiated and governed the MS case control register, genetic analyses and biobank. IK initiated the serological analyses and governed the MS case control register, genetic analyses and biobank. AFH supervised the initiation, execution and finalization of the HHV-6A/B serology study.
CONFLICT OF INTEREST STATEMENT

Jan Hillert has received honoraria for serving on advisory boards for Biogen, Sanofi-Genzyme and Novartis and speaker's fees from Biogen, Novartis, Merck-Serono, Bayer-Schering, Teva and Sanofi-Genzyme. He has served as P.I. for projects, or received unrestricted research support from, BiogenIdec, Merck, Novartis and Sanofi-Genzyme. Lucia Alonso Magdalena has received speaking honoraria from Merck-Serono and served at the advisory board for Merck-Serono and Biogen. Lars Alfredsson received speaker honoraria from Biogen Idec and TEVA. Tomas Olsson has received lecture and/or Advisory board honoraria, and unrestricted MS research grants from Biogen, Novartis, Sanofi, Merck-Serono and Roche. Anna Fogdell-Hahn has received speaker's fees from Pfizer, Biogen, Merck-Serono, and Sanofi-Genzyme. She has served as P.I. for projects, or received unrestricted research support from, Biogen Idec and Pfizer. All the other co-authors declare no conflict of interest.
REFERENCES


HHV-6A in MS


PubMed Central PMCID: PMCPMC4667104.


PubMed PMID: 4128558.


PubMed Central PMCID: PMCPMC3898140.


Consortium IMSG. The Multiple Sclerosis Genomic Map: Role of peripheral immune cells and resident microglia in susceptibility(2017).

FIGURE LEGENDS

FIGURE 1

Antibody responses against HHV-6A and 6B proteins in MS cases and controls. Log10-transformed antibody levels measured as median fluorescence intensity (MFI) are visualized with bean plots for established MS cohort (n=8742 persons with MS (blue) and n=7215 controls (pink)) (A, B, C) and pre-MS cohort (n=478 persons with MS (blue) and n=476 controls (pink)) (D, E, F) for HHV-6A IE1A IgG (A and D); HHV-6B IE1B IgG (B and E); HHV-6B anti-101K IgG (C and F). The 1st and 3rd quartiles are indicated with dotted lines and solid lines indicate median.

FIGURE 2

Interaction of antibody response against different herpesviruses in association to MS.
Odds ratios (OR) and confidence intervals (CI) for IE1A and EBV (A) and IE1A and CMV (B), were obtained through logistic regression models adjusted for age, sex and cohort type analyzing the Established MS cohort (n=8742 persons with MS and n=7215 controls). OR were calculated in relation to the group with the lowest MS risk. Plus (+) indicates being a strong responder while minus (-) indicates being a weak responder. Strong IE1A response is defined as having an MFI value being in the upper quartile of measured response, while a low response is having an antibody measurement being in the lower quartile of measured response. Strong EBV / CMV response is defined as having a higher EBV / CMV index than the median among controls, while a weak response is having a lower index compared to the median among controls.

FIGURE 3

Median MFI response against HHV-6A IE1A protein in different age groups. Median of median fluorescence intensity (MFI) in different age groups for A) pre-MS cohort (n=478 persons who later developed MS (green) and n=476 controls (pink)) and B) established MS cohort (n=8394 persons with MS (green) and n=7214 controls (pink)). Statistics were calculated with linear regression. Significant (p<0.008) differences in IgG levels between MS cases and controls within each age group are indicated with *.

FIGURE 4

Manhattan plots visualizing associations between SNPs and anti-HHV-6A/6B protein IgG response levels. GWAS data (n=6,396 MS cases and n=5,530 controls from the established MS cohort) obtained through linear regression models showing associations between SNPs and IgG response (Log10 levels) against (A) IE1A, (B) IE1B and (C) 101K. Red lines indicate GWAS significance level of 5x10^-8 (-log10 = 7.3 on the y-axis) and blue lines indicate suggestive association (p=10^-5). Analysis was carried out jointly in MS cases and controls and adjusted with age, sex, cohort type and case status.

FIGURE 5

Interaction analysis between IE1A and IE1B IgG response and main MS risk HLA alleles. Odds ratios (OR) and confidence intervals (CI) for A) IE1A and DRB1*15:01, B) IE1A and A*02:01, C) IE1B and DRB1*15:01, and D) IE1B and A*02 were obtained through logistic regression models adjusted for age, sex and cohort type analyzing the Established MS cohort (n=7,063 MS cases and n=6,098 controls). OR were calculated in relation to the group with the lowest MS risk. Plus (+) indicates being a strong responder while minus (-)
indicates being a weak responder defined as having an MFI value in the upper quartile of measured response. AP=attributable proportion due to interaction, p is p-value for interaction. No adjustment for EBV was done in these figures.
### Table 1. Association of IE1A, IE1B and 101K antibody response to MS in established MS cohort

<table>
<thead>
<tr>
<th>Median disease duration</th>
<th>IE1A OR (95% CI)</th>
<th>p</th>
<th>IE1B OR (95% CI)</th>
<th>p</th>
<th>101K OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>11.2</td>
<td>1.55 (1.42-1.69)</td>
<td>9E-22</td>
<td>0.74 (0.67-0.81)</td>
<td>6E-11</td>
<td>1.06 (0.97-1.16)</td>
</tr>
<tr>
<td>Age &lt;30</td>
<td>2.0</td>
<td>1.80 (1.35-2.40)</td>
<td>7E-05</td>
<td>0.66 (0.50-0.87)</td>
<td>0.03</td>
<td>1.22 (0.92-1.62)</td>
</tr>
<tr>
<td>Age 30-39</td>
<td>5.4</td>
<td>1.56 (1.28-1.90)</td>
<td>1E-05</td>
<td>0.86 (0.70-1.05)</td>
<td>0.13</td>
<td>1.34 (1.10-1.64)</td>
</tr>
<tr>
<td>Age 40-49</td>
<td>10.7</td>
<td>1.48 (1.24-1.77)</td>
<td>1E-05</td>
<td>0.85 (0.70-1.02)</td>
<td>0.078</td>
<td>1.19 (0.99-1.43)</td>
</tr>
<tr>
<td>Age 50-59</td>
<td>17.4</td>
<td>1.46 (1.21-1.76)</td>
<td>6E-05</td>
<td>0.73 (0.60-0.89)</td>
<td>0.0015</td>
<td>0.92 (0.77-1.12)</td>
</tr>
<tr>
<td>Age ≥60</td>
<td>26.2</td>
<td>1.64 (1.35-2.00)</td>
<td>8E-07</td>
<td>0.57 (0.46-0.70)</td>
<td>1E-07</td>
<td>0.80 (0.65-0.98)</td>
</tr>
</tbody>
</table>

Odds ratios (OR) and p-values for MS risk were calculated using logistic regression models, comparing high and low IgG responders, both with all individuals included in the analyses and stratified into groups based on age at sampling. High responders were defined to fall in the upper quartile while low responders were found in the lower quartile of measured responses. The models using the established MS cohort were adjusted for sex and cohort type, as well as and age at sampling in the un-stratified group. Significant p-values are highlighted in bold. Data from 8394 MS cases and 7214 controls included in the analysis. The number of people with MS in each age group was for the established MS cohort in ages 20-29 n=1550; 30-39 n=3184; 40-49 n=3821; 50-59 n=3568; 60+ n=3321.
Table 2. Association of IE1A, IE1B and 101K antibody response to MS in pre-MS cohort

<table>
<thead>
<tr>
<th></th>
<th>median disease duration</th>
<th>IE1A</th>
<th>p</th>
<th>IE1B</th>
<th>p</th>
<th>101K</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>-8.3</td>
<td>2.22 (1.54-3.19)</td>
<td>2E-05</td>
<td>0.96 (0.66-1.38)</td>
<td>0.81</td>
<td>1.51 (1.06-2.17)</td>
<td>0.024</td>
</tr>
<tr>
<td>Age &lt;20</td>
<td>-9.8</td>
<td>3.38 (1.46-7.81)</td>
<td><strong>0.004</strong></td>
<td>1.09 (0.46-2.57)</td>
<td>0.84</td>
<td>1.36 (0.58-3.15)</td>
<td>0.48</td>
</tr>
<tr>
<td>Age 20-29</td>
<td>-8.4</td>
<td>2.29 (1.43-3.67)</td>
<td><strong>0.001</strong></td>
<td>0.88 (0.54-1.44)</td>
<td>0.62</td>
<td>1.60 (1.00-2.54)</td>
<td>0.048</td>
</tr>
<tr>
<td>Age 30-39</td>
<td>-5.7</td>
<td>1.51 (0.65-3.49)</td>
<td>0.33</td>
<td>1.09 (0.51-2.35)</td>
<td>0.83</td>
<td>1.42 (0.64-3.14)</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Odds ratios (OR) and p-values for MS risk were calculated using logistic regression models, comparing high and low IgG responders, both with all individuals included in the analyses and stratified into groups based on age at sampling. High responders were defined to fall in the upper quartile while low responders were found in the lower quartile of measured responses. The pre-MS cohort analyses are adjusted for sex, and age. Significant p-values are highlighted in bold. Data from 478 individuals who later developed MS and 476 matched controls included in the analysis.
Table 3. Association between IgG levels and HLA haplotypes and IE1A

|                | MS cases | | | | Controls | | | | All subjects | | | |
|----------------|----------|----------|----------|----------|----------|----------|----------|----------|------------|----------|----------|
|                | Adjustment^A | Adjustment #1 | Adjustment #2 | | Adjustment #1 | Adjustment #2 | | Adjustment #1 | Adjustment #2 | | | |
|                | β | p | β | p | | β | p | β | p | | β | p | |
| DRB1*13:01-DQA1*01:03-DQB1*06:03 | | | | | | | | | | | | |
| DRB1*13:01 | 0.13 | 8E-09 | 0.12 | 9E-08 | 0.11 | 5E-09 | 0.10 | 5E-07 | 0.11 | 4E-14 | 0.09 | 1E-10 |
| DQA1*01:03 | 0.12 | 6E-09 | 0.12 | 7E-08 | 0.10 | 5E-08 | 0.09 | 3E-06 | 0.11 | 2E-13 | 0.09 | 3E-10 |
| DQB1*06:03 | 0.11 | 2E-07 | 0.10 | 2E-06 | 0.12 | 6E-10 | 0.11 | 5E-08 | 0.11 | 4E-14 | 0.09 | 1E-10 |
| DPA1*02:01-DPB1*01:01 | | | | | | | | | | | | |
| DPA1*02:01 | -0.10 | 2E-08 | -0.09 | 3E-07 | -0.04 | 0.04 | -0.03 | 0.15 | -0.08 | 2E-09 | -0.07 | 2E-07 |
| DPB1*01:01 | -0.15 | 1E-08 | -0.12 | 5E-06 | -0.08 | 2E-04 | -0.06 | 0.01 | -0.12 | 1E-12 | -0.10 | 2E-08 |
| DRB1*04:01-DQA1*03-DQB1*03:02 | | | | | | | | | | | | |
| DRB1*04:01 | -0.05 | 0.01 | -0.04 | 0.02 | -0.07 | 4E-05 | -0.05 | 1E-03 | -0.07 | 4E-08 | -0.06 | 3E-07 |
| DQA1*03 | -0.03 | 0.06 | -0.03 | 0.08 | -0.03 | 0.02 | -0.02 | 0.15 | -0.04 | 2E-04 | -0.04 | 4E-04 |
| DQB1*03:02 | -0.03 | 0.04 | -0.03 | 0.05 | -0.05 | 3E-03 | -0.03 | 0.03 | -0.04 | 2E-04 | -0.04 | 5E-04 |
| A*01:01-B*08:01-C*07:01-DRB1*03:01-DQA1*05:01-DQB1*02:01 | | | | | | | | | | | | |
| A*01:01 | -0.02 | 0.26 | 0.01 | 0.68 | -0.06 | 1E-04 | -0.05 | 5E-03 | -0.04 | 2E-03 | -0.02 | 0.16 |
| B*08:01 | -0.07 | 9E-05 | -0.03 | 0.08 | -0.05 | 1E-03 | -0.03 | 0.09 | -0.06 | 8E-08 | -0.04 | 4E-03 |
| C*07:01 | -0.05 | 2E-03 | -0.02 | 0.19 | -0.03 | 0.02 | -0.01 | 0.43 | -0.04 | 7E-05 | -0.02 | 0.07 |
The allele variants in a given HLA haplotype are presented. Underlined alleles were conferring the strongest effect, determined by stepwise conditional analyses. HLA genotype data was available for 7,063 MS cases and 6,098 controls. Adjustment #1 = Age, sex, 5 PCA vectors, cohort type (incidence or prevalence). When adjusted for HLA, one associated HLA allele from each associated haplotype (underlined in the table) were added to the model in order to test for independent association of the investigated HLA allele. Adjustment #2 = Adjustment #1 + HLA. When all subjects were analyzed together, MS affection status were added as a covariable. Haplotypes associated to IE1A indicated in bold if any allele in it has a p-value <0.001.

<table>
<thead>
<tr>
<th>HLA Alleles</th>
<th>Frequency</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB1*03:01</td>
<td>-0.06</td>
<td>5E-04</td>
</tr>
<tr>
<td>DQA1*05:01</td>
<td>-0.05</td>
<td>1E-03</td>
</tr>
<tr>
<td>DQB1*02:01</td>
<td>-0.06</td>
<td>5E-04</td>
</tr>
<tr>
<td>B*35:01</td>
<td>0.09</td>
<td>2E-04</td>
</tr>
<tr>
<td>C*04:01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>DRB1*01:01</td>
<td>0.08</td>
<td>2E-04</td>
</tr>
<tr>
<td>DQA1*01:01</td>
<td>0.05</td>
<td>3E-03</td>
</tr>
<tr>
<td>DQB1*05:01</td>
<td>0.06</td>
<td>1E-03</td>
</tr>
<tr>
<td>A*02:01</td>
<td>0.03</td>
<td>0.07</td>
</tr>
<tr>
<td>B*51:01</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>B*39</td>
<td>-0.17</td>
<td>1E-05</td>
</tr>
<tr>
<td>DRB1*12:01</td>
<td>-0.11</td>
<td>0.02</td>
</tr>
</tbody>
</table>

933 The allele variants in a given HLA haplotype are presented. Underlined alleles were conferring the strongest effect, determined by stepwise conditional analyses. HLA genotype data was available for 7,063 MS cases and 6,098 controls. Adjustment #1 = Age, sex, 5 PCA vectors, cohort type (incidence or prevalence). When adjusted for HLA, one associated HLA allele from each associated haplotype (underlined in the table) were added to the model in order to test for independent association of the investigated HLA allele. Adjustment #2 = Adjustment #1 + HLA. When all subjects were analyzed together, MS affection status were added as a covariable. Haplotypes associated to IE1A indicated in bold if any allele in it has a p-value <0.001.
<table>
<thead>
<tr>
<th></th>
<th>MS cases</th>
<th>Controls</th>
<th>All subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>8742</td>
<td>7215</td>
<td>15957</td>
</tr>
<tr>
<td>% Females</td>
<td>72%</td>
<td>75%</td>
<td>74%</td>
</tr>
<tr>
<td>Age at sampling [mean ± SD]</td>
<td>47.1 ± 14.0</td>
<td>48.2 ± 13.4</td>
<td>47.6 ± 13.8</td>
</tr>
<tr>
<td>Age category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20 years</td>
<td>112 [1%]</td>
<td>53 [1%]</td>
<td>165 [1%]</td>
</tr>
<tr>
<td>20-29 years</td>
<td>972 [11%]</td>
<td>611 [8%]</td>
<td>1583 [10%]</td>
</tr>
<tr>
<td>30-39 years</td>
<td>1831 [21%]</td>
<td>1471 [20%]</td>
<td>3302 [21%]</td>
</tr>
<tr>
<td>40-49 years</td>
<td>2115 [24%]</td>
<td>1832 [25%]</td>
<td>3948 [25%]</td>
</tr>
<tr>
<td>50-59 years</td>
<td>1918 [22%]</td>
<td>1707 [24%]</td>
<td>3625 [23%]</td>
</tr>
<tr>
<td>&gt;60 years</td>
<td>1794 [21%]</td>
<td>1541 [21%]</td>
<td>3335 [21%]</td>
</tr>
<tr>
<td>Median (IQR) age at MS onset(^{A})</td>
<td>32 (15)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Median (IQR) years between symptom onset and serum collection(^{B})</td>
<td>10.9 (17.3)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Disease course at sampling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRMS</td>
<td>5586 [64%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SPMS</td>
<td>1804 [21%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PPMS</td>
<td>549 [6%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Missing data / Other</td>
<td>803 [9%]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ever smokers(^{C})</td>
<td>3336 [51%]</td>
<td>2461 [41%]</td>
<td>5797 [46%]</td>
</tr>
</tbody>
</table>

\(^{A}\)Median age when the first MS symptom was reported to have occurred, i.e. not the same as age at MS diagnosis, data available for 8,505 persons with MS.

\(^{B}\)Median disease duration, in years, calculated as the time from age at onset to age at sampling.

\(^{C}\)Past and/or present regular smoking habits, smoking data was obtained for 12,530 individuals.
Table 5. Demographic data of the Pre-MS cohort

<table>
<thead>
<tr>
<th></th>
<th>MS Cases</th>
<th>Controls</th>
<th>All Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>478</td>
<td>476</td>
<td>954</td>
</tr>
<tr>
<td>% Female</td>
<td>83%</td>
<td>83%</td>
<td>83%</td>
</tr>
<tr>
<td>Age at sampling [mean ± SD]</td>
<td>24.7 ± 6.4</td>
<td>24.6 ± 6.4</td>
<td>24.7 ± 6.4</td>
</tr>
<tr>
<td>Age category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20 years</td>
<td>93 [20%]</td>
<td>91 [19%]</td>
<td>184 [19%]</td>
</tr>
<tr>
<td>20-29 years</td>
<td>274 [57%]</td>
<td>279 [59%]</td>
<td>553 [58%]</td>
</tr>
<tr>
<td>30-39 years</td>
<td>111 [23%]</td>
<td>106 [22%]</td>
<td>217 [23%]</td>
</tr>
<tr>
<td>Median (IQR) age at MS onset</td>
<td>34 (12)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Median (IQR) years between symptom onset and serum collection</td>
<td>- 8.3 (10)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Disease course at MS onset</td>
<td>RRMS</td>
<td>478 [100%]</td>
<td>-</td>
</tr>
</tbody>
</table>
A) IE1A and EBV interaction

- IE1A+ EBV+
- IE1A- EBV+
- IE1A+ EBV-
- IE1A- EBV-

OR (95% CI)

AP = 0.24
p = 6e-6

B) IE1A and CMV interaction

- IE1A+ CMV-
- IE1A+ CMV+
- IE1A- CMV-
- IE1A- CMV+

OR (95% CI)

AP = 0.05
p = 0.5