Human Endogenous Retrovirus-K18 *env* as a risk factor in multiple sclerosis

AK Tai, PhD[1], **EJ O'Reilly, MSc**[2], **KA Alroy**[1], **KC Simon, DSc**[2], **KL Munger, MSc**[2], **BT Huber, PhD**[1], and **A Ascherio, MD, DrPH**[2,3,4]

[1] Department of Pathology, Tufts University School of Medicine, Boston, MA, USA

[2] Department of Nutrition, Harvard School of Public Health, Boston, MA, USA

[3] Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[4] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA

Abstract

To assess whether variation in human endogenous retrovirus (HERV)-K18 *env*, an Epstein-Barr virus (EBV) associated superantigen, is a risk factor for multiple sclerosis (MS), we developed a SNP-based genotyping method to determine the allelic and genotypic distribution of the three alleles of HERV-K18 *env*. We conducted a nested case-control study amongst 207 MS cases and 403 matched controls drawn from two large ongoing cohorts, the Nurses Health Study and Nurses Health Study 2 (NHS/NHS2). Analyses were replicated in an independent series of 909 MS cases and 339 controls. Overall, in the NHS/NHS2 there was a significant association between HERV K18 *env* genotype and risk of MS (χ^2 p-value=0.03). As compared with individuals homozygous for the K18.2 allele, risk of MS three fold higher among carriers of two K18.3 alleles (p=0.03). An increase in MS risk among carriers of the K18.3 allele was also observed in the replication study, but did not reach statistical significance. In pooled analyses, homozygous carriers of the K18.3 allele had a significantly increased risk of MS (RR comparing K18.3/K18.3 versus K18.2/K18.2 = 2.7; 95% CI: 1.1 to 6.4). These results suggest that variation in EBV-associated superantigen HERV-K18 *env* could influence the genetic susceptibility to MS.

Keywords

multiple sclerosis; endogenous retroviruses; HERV-K18; polymorphisms; HLA

Introduction

Multiple sclerosis (MS) is a chronic disabling neurodegenerative disease estimated to affect 350,000 individuals in the United States and approximately 2 million worldwide[1]. Common symptoms include, but are not limited to, visual disturbance, sensory problems, general weakness and poor coordination. These symptoms are the clinical manifestations of lesions accumulated in the CNS as a result of axonal demyelination and degeneration, thought to be mediated by an autoimmune process, possibly triggered by infection[2].

Although several infections have been implicated in MS, data supporting a role of specific microbes are inconclusive[2]. One of the strongest candidates is the Epstein-Barr Virus (EBV). A link between EBV and MS was suspected early, because of the similarity between the epidemiology of MS and infectious mononucleosis[3]. Epidemiologic evidence supports a role for EBV in the etiology of MS, including the observatiosn that risk of MS is extremely low

risk amongst individuals who are seronegative for antibodies to EBV[4,5] and increases following infectious mononucleosis, a common manifestation of EBV infection late in childhood or adulthood[6]. Further, IgG antibody titer against the Epstein Barr nuclear antigen 1 in healthy young adults is a strong, dose-dependent predictor of future MS risk[7]. In spite of a consistent association between EBV and MS, potential mechanisms to explain this association have so far been little investigated.

Recently, an EBV associated superantigen, the envelope protein (*env*) of a human endogenous retrovirus (HERV)-K18, has been identified and cloned[8]. Superantigens are a class of proteins capable of deregulating the immune response, and as such, an association may exist between HERV-K18 *env* and the development of MS. There are three alleles of the HERV-K18 *env,* and all of them demonstrate superantigen activity[9]. However, due to differences in the amino acid sequences, biochemical differences are predicted. The three alleles are not evenly distributed within the Caucasian population; in a recent study, the majority of individuals carried alleles K18.1 (46.6%) and K18.2 (42.5 %), while only 10.8% were carriers of the allele K18.3[9]. Given the association between HERV-K18 *env* and EBV and the plausibly relevant biological mechanism, we hypothesized that this EBV-associated superantigen may contribute to the association of EBV with MS, and that the three allelic forms of the superantigen may influence general susceptibility to MS. To address this hypothesis, we conducted a nested case-control study amongst participants in two longitudinal cohorts[10]. For validation, analyses were replicated in The Brigham and Women's Hospital (BWH) MS Genetics Collection, a sub-set of the sample used in a recent whole genome association scan of MS[11].

Materials and Methods

Study population

The investigation was conducted among participants in the Nurses' Health Study (NHS) and Nurses' Health Study II (NHS II), two ongoing cohort studies for the investigation into risk factors of chronic diseases. The NHS began in 1976 when 121,700 women aged 30–55 returned mailed questionnaires regarding lifestyle factors and disease history, and the NHS II began in 1989 when 116,671 women aged 25–42 returned similar questionnaires. Biennial questionnaires are mailed to update information on risk factors and disease occurrence, and follow-up rates above 90% have been consistently maintained. All participants in the cohorts were invited to provide blood samples for investigations of biomarkers and disease outcomes. Blood was collected from women between 1989 and 1990 in NHS (32, 826 women) and from 1996 to 1999 in NHS II (29, 613 women). In the NHS, women who did not provide a blood sample were invited to give a cheek-cell sample for genetic analyses; over 32,000 cheek-cells samples were received and stored for genetic analyses. The BWH MS Genetic Collection is a clinic based sample, the primary purpose of which is to explore the role of genetic factors in relation to MS. MS patients (227 men and 682 women, mean age=45.9 years) were recruited by their treating neurologist and controls are unrelated, healthy individuals recruited at the same site (172 men and 166 women, mean age=46.4 years). Details of the larger case-control sample, of which the BWH cohort is a sub-set, have been described previously[11].

Case ascertainment

The ascertainment of MS cases in the NHS and NHS II cohort has been previously described [10,12]. A total of 148 incident cases of definite and probable MS were documented among women with available blood samples, and 66 cases among women with DNA available from cheek-cell samples. For each case, we randomly selected as controls two women without MS, matched by year of birth and study cohort. Of the total 214 cases and 428 controls, seven cases and twelve controls did not have genotype information and were not included in the current

analysis. An additional 13 controls were excluded because they were matched to a case without genotype information. A total of 207 women with MS (145 with blood and 62 with cheek-cells) and 403 controls (281 with blood, and 122 with cheek-cells -- 4 controls with cheek cells could not be genotyped) were included in the analysis. Over 90% of the women included in the study reported having a white ancestry. All the BWH cases were examined by an MS specialist at the study site. The sample population for analysis includes 909 cases of MS and 339 controls with HERV-K18 genotype information.

TaqMan-based SNP genotyping

To characterize the potential association between the HERV-K18 *env* and MS, we have developed a highly specific, sensitive and efficient screening method, using TaqMan MGB probe-based SNP genotyping for the determination of the allelic and genomic distribution of the three alleles of HERV-K18 *env*. The sequences of the three alleles of HERV-K18 *env* were retrieved from the Genebank (accession number for K18.1, K18.2 and K18.3 are AF134984, Y18890 and AF333069, respectively). The sequences were then used to blast against the Genebank to retrieve sequences from other HERVs that are highly homologous to the HERV-K18 *env*. A primer pair was selected for sequences that are unique to the HERV-K18, but common between the three alleles (SNP forward: 5′-CTAAATTCCATTCTAACGGTTCCTTT-3′; SNP reverse: 5′-GTTTGGGAGGCTGGTTTAATAACTAT-3′). One TaqMan MGB probe was designed for each allele around a single nucleotide polymorphism (SNP) that is unique to that particular allele (K18.1: 5′-AAAGTTGCCTAAAGC-3′; K18.2: 5′-AAAGTTGCGTAAAGC-3′; K18.3: 5′-AAAGTTGCATAAAGC-3′). SNP PCR was carried out on an ABI 7300 Sequence Detection System. The reaction was performed in a volume of 10 μl containing ABI TaqMan PCR MasterMix, DNA template, 450 nM of each primer and 125 nM of each TaqMan MGB probe. The PCR was carried out with 10 min initial denaturation at 95°C, followed by 60 cycles of 95°C for 15 s and 62°C for 1 min. The data were then analyzed with the system software and subjected to statistical analysis.

Covariate assessment

A single SNP (rs3135005) that is strongly correlated with the haplotype *DRB1*1501, DQA1*0102, DQB1*0602* [13], was used for *HLA-II* genotyping. Smoking history, latitude of birthplace, and ethnicity, were obtained as part of the general follow-up in the NHS and NHS2 [12].

Statistical analysis

The assumption of Hardy-Weinberg Equilibrium (HWE) was tested using a χ^2 test comparing observed to expected genotype frequencies. The distribution of HERV-K18 *env* genotypes in the NHS/NHS2 violated the HWE assumption (p=0.02) with an excess of K18.3 homozygotes. To assess the possibility of a genotyping error, the 17 samples homozygous for the K18.3 allele were re-genotyped to confirm the result. Agreement was 100% between the two runs, suggesting that the deviation from HWE was not due to genotyping error. The assessment of HWE in the BWH showed no significant deviation at the p=0.05 level.

Conditional logistic regression models were used in the NHS/NHS2 cohort to calculate relative risks (RRs) and 95% confidence intervals (CIs) assessing the relationship between HERV-K18 *env* genotype and risk of MS, with or without adjustment for DRB1*1501 and other risk factors for MS. Adjustment for smoking, latitude, and ancestry did not materially change the results, and therefore only unadjusted results have been reported. Unconditional logistic regression models were used to assess the relationship between HERV-K18 *env* genotype and risk of MS in the BWH study consistent with the case-control ascertainment. To assess the overall association between HERV-K18 genotype and risk of MS, we used a likelihood ratio test,

comparing a model including all genotypes to the same model without genotypes. Analyses were conducted using SAS v9.1 and SAS Genetics (SAS Institute, Cary, NC). For the purposes of determining pooled estimates, unconditional logistic regression models were used to conduct the same analyses described above in the NHS/NHS2 cohort, adjusting for the matching factors. Results using this method were virtually identical to results obtained using conditional logistic regression models. Log RRs of results obtained from the NHS and NHS2 cohorts, and from the BWH case-control study were pooled using STATA version 9 (StataCorp LP, College Station, TX) using inverse variance weights. Fixed effects estimates are presented as tests of heterogeneity were all non-significant at the $\alpha=0.05$ level.

Results

Characteristics of participants in the NHS and NHS2 nested case-control study are presented in Table 1. As expected, cases were more likely to live at a northerly latitude and to be HLA-DRB1*1501 positive than controls. Overall, there was a significant association between HERV-K18 *env* genotype (χ^2 p-value=0.03) and risk of MS (Table 2). Risk of MS was higher among individuals homozygous for the K18.3 allele, relative to K18.2/K18.2 individuals (Table 2). As expected, the HLA-DRB1*1501 allele was strongly associated with MS risk (RR for dominant model 2.8; 95% CI: 1.9 to 4.1; p<0.0001), but adjustment for HLA-DRB1*1501 did not materially change the association between HERV-K18 *env* genotype and MS risk (Table 2). In an additive model, increasing number of K18.3 alleles relative to the number of K18.2 alleles was associated with an 80% increase in risk of MS (RR=1.8; 95% CI:1.3 to 2.6; p=0.001), whereas a weaker association was found for increasing number of K18.1 alleles relative to number of K18.2 alleles (RR=1.3; 95% CI:1.0 to 1.6; p =0.05). Restricting the analyses to individuals reporting white ancestry resulted in similar estimates.

In the replication study, based on the BWH case and controls, results for the association between HERV-K18 genotype and risk of MS were non-significant, but also suggested an increased risk of MS among individuals homozygous for the K18.3 allele (Table 3). In an additive model, there was a 40% increased risk of MS associated with the K18.3 allele (RR=1.3, 95% CI: 0.9, 1.8; p=0.14). Overall, significant increases in risk of MS were seen in the pooled analysis for individuals carrying one or two K18.3 alleles (Table 3).

Discussion

In this nested case-control study, we found evidence for an association between variation in HERV-K18 *env* and risk of MS. This association was independent from known genetic and non-genetic risk factors for MS, including the HLA-DRB1*1501 allele, cigarette smoking, latitude at birth, and ancestry. Limitations of this investigation include the relatively small sample size, which prevents a thorough exploration of possible interactions between variations at the HERV-K18 *env* and the HLA-DRB1 loci, and the observed deviation from HWE. The latter was, however, modest and is likely to represent a chance finding. Further, our primary results are based on genotype analyses, which, unlike analyses based on allele frequencies, are more robust to deviations from HWE[14].

HERV-K18 *env* was chosen as a candidate gene in MS given that it is transactivated by EBV. Although EBV is a strong risk factor for MS, the mechanisms that relate EBV infection to MS remain unknown. In a reversal of previous smaller studies, large numbers of EBV infected B-cells have been recently found in the brain of a large majority of patients with MS[15]. These cells were more numerous in areas with active inflammatory infiltrates, where cytotoxic CD8 + T-cells displaying an activated phenotype were seen contiguous to the EBV infected cells, suggesting that an EBV-specific cellular immune response could be driving MS pathology. Molecular mimicry between a viral epitope and a self-antigen in the CNS has also been

proposed[16–18]. On the other hand, the possibility that EBV activates a superantigen that increases the risk of MS has been little explored. Analyses of the Vβ regions of autoreactive T-cells in MS do not seem to support a polyclonal expansion of the Vβ7 and Vβ13 subsets that are activated by HERV-K18 *env*[8,19]. Nevertheless, activation by HERV-K18 *env* of a large number of T-cells could have pro-inflammatory effects and cause an increase in endothelial permeability or otherwise facilitate an inflammatory process in the CNS or provide activation and/or survival signals to autoimmune T cells in a bystander manner. Alternatively, variations in HERV-K18 *env* could be in linkage disequilibrium with mutations in the neighboring genomic regions. The HERV-K18 gene is embedded in the first intron of the CD48 gene (chromosome 1), which is itself regulated by EBV infection[20]. In a previous study, the K18.3 allele was found to be part of a haplotype that extends for 30 Kb and comprises part of the exonic sequences of CD48, and to be inversely associated with risk of type I diabetes[21]. Further investigation of the CD48 gene region would thus be important.

In summary, we found that sequence variations in the HERV-K18 *env*, which encodes a superantigen transactivated by EBV, are significantly associated with risk of MS. This finding, combined with the strong evidence linking EBV infection to MS, supports the possibility that HERV-K18 *env* is a susceptibility gene for MS.

Acknowledgments

References

1. Rosati G. The prevalence of multiple sclerosis in the world: an update. Neurol Sci 2001;22(2):117–39. [PubMed: 11603614]

2. Hunter SF, Hafler DA. Ubiquitous pathogens: links between infection and autoimmunity in MS? Neurology 2000;55(2):164–5. [PubMed: 10908883]

3. Warner HB, Carp RI. Multiple sclerosis and Epstein-Barr virus (letter). Lancet 1981;2:1290. [PubMed: 6118702]

4. Ascherio A, Munch M. Epstein-Barr Virus and Multiple Sclerosis. Epidemiology 2000;11(2):220–24. [PubMed: 11021623]

5. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: The role of infection. Ann Neurol 2007;61(4):288–99. [PubMed: 17444504]

6. Thacker EL, Mirzaei F, Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: A meta-analysis. Ann Neurol 2006;59(3):499–503. [PubMed: 16502434]

7. Levin LI, Munger KL, Rubertone MV, Peck CA, Lennette ET, Spiegelman D, 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):2496–500. [PubMed: 15914750]

8. Conrad B, Weissmahr RN, Böni J, Arcari R, Schüpbach J, Mach B. A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. Cell 1997;90:303–13. [PubMed: 9244304]

9. Stauffer Y, Marguerat S, Meylan F, Ucla C, Sutkowski N, Huber B, et al. Interferon-alpha-induced endogenous superantigen. a model linking environment and autoimmunity. Immunity 2001;15(4):591–601. [PubMed: 11672541]

10. Hernán MA, Olek MJ, Ascherio A. Geographic variation of MS incidence in two prospective studies of US women. Neurology 1999;53(8):1711–18. [PubMed: 10563617]

11. The International Multiple Sclerosis Genetics Consortium. Risk Alleles for Multiple Sclerosis Identified by a Genomewide Study. N Engl J Med 2007;357(9):851–62. [PubMed: 17660530]

12. Ascherio A, Zhang SM, Hernan MA, Olek MJ, Coplan PM, Brodovicz K, et al. Hepatitis B vaccination and the risk of multiple sclerosis. N Engl J Med 2001;344(5):327–32. [PubMed: 11172163]

13. de Bakker PIW, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 2006;38(10):1166–72. [PubMed: 16998491]

14. Schaid DJ, Jacobsen SJ. Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. Am J Epidemiol 1999;149(8):706–11. [PubMed: 10206619]

15. Serafini B, Rosicarelli B, Franciotta D, Magliozzi R, Reynolds R, Cinque P, et al. Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. J Exp Med 2007;204(12):2899–912. [PubMed: 17984305]

16. Rand KH, Houck H, Denslow ND, Heilman KM. Epstein-Barr virus nuclear antigen-1 (EBNA-1) associated oligoclonal bands in patients with multiple sclerosis. J Neurol Sci 2000;173(1):32–9. [PubMed: 10675577]

17. Lang HL, Jacobsen H, Ikemizu S, Andersson C, Harlos K, Madsen L, et al. A functional and structural basis for TCR cross-reactivity in multiple sclerosis. Nat Immunol 2002;3(10):940–3. [PubMed: 12244309]

18. Lunemann JD, Edwards N, Muraro PA, Hayashi S, Cohen JI, Munz C, et al. Increased frequency and broadened specificity of latent EBV nuclear antigen-1-specific T cells in multiple sclerosis. Brain 2006;129(Pt 6):1493–506. [PubMed: 16569670]

19. Sutkowski N, Palkama T, Ciurli C, Sekaly R-P, Thorley-Lawson DA, Huber BT. An Epstein-Barr virus-associated superantigen. J Exp Med 1996;184:971–80. [PubMed: 9064357]

20. Klaman LD, Thorley-Lawson DA. Characterization of the CD48 gene demonstrates a positive element that is specific to Epstein-Barr virus-immortalized B-cell lines and contains an essential NF-kappa B site. J Virol 1995;69(2):871–81. [PubMed: 7815555]

21. Marguerat S, Wang WY, Todd JA, Conrad B. Association of human endogenous retrovirus K-18 polymorphisms with type 1 diabetes. Diabetes 2004;53(3):852–4. [PubMed: 14988274]
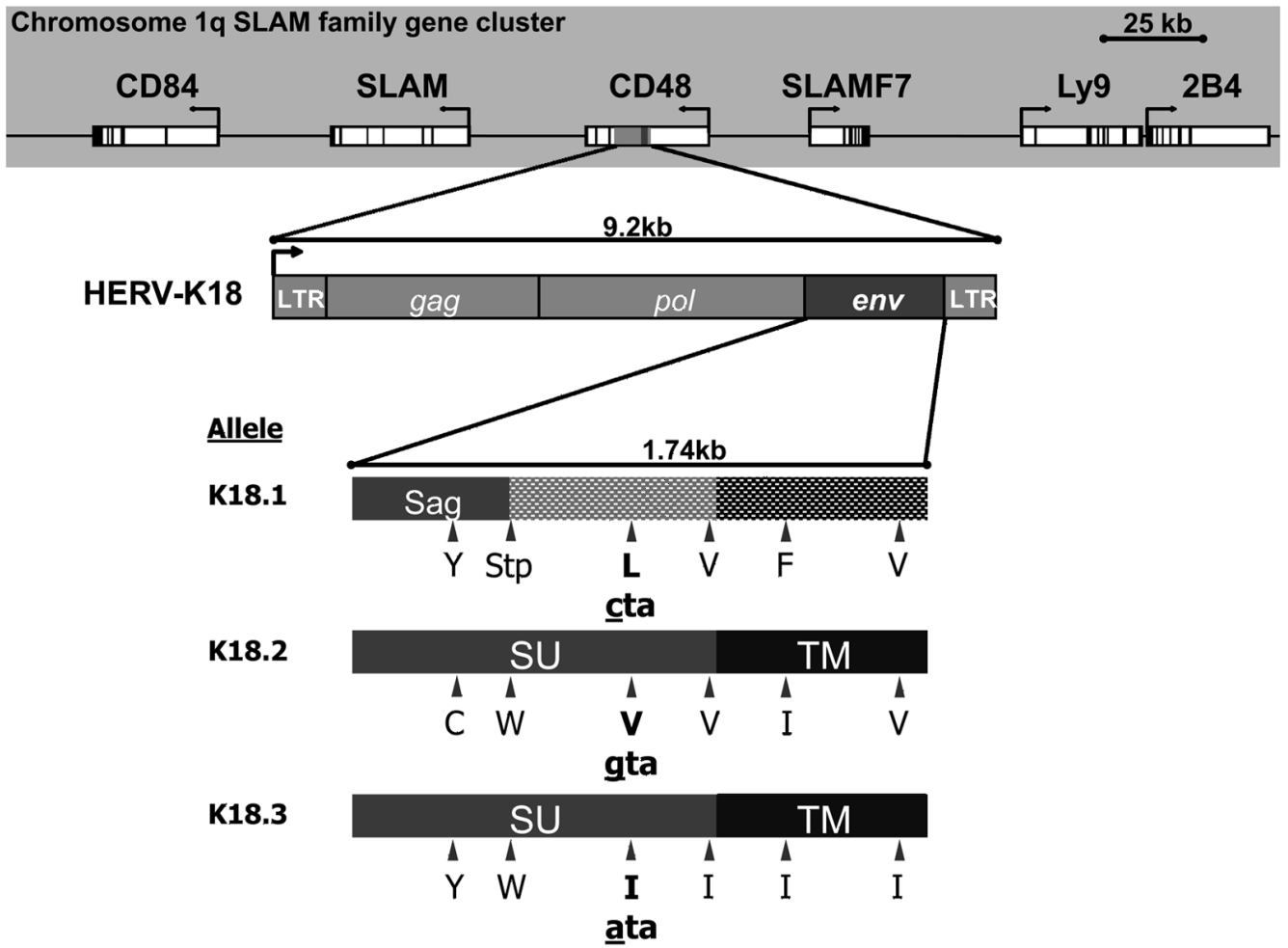
Figure 1.

Table 1

Selected Characteristics of NHS/NHS2 MS Cases and Matched Controls

	Cases N=207	Controls N=403
Mean age at sample collection, (years)	53.5	53.6
Mean age at MS onset (years)	39.2	-
Residence in north tier		
At birth (%)	88 (47.1)	137 (37.5)
At age 15 (%)	94 (50.0)	134 (36.4)
At age 30 (%)	77 (40.5)	114 (30.4)
Scandinavian ancestry (%)	10 (4.9)	20 (5.1)
Ever-smoker at sample collection (%)	124 (59.9)	210 (52.4)
HLA-*DRB1*1501* positive (%)	110 (55.3)	119 (30.9)

Numbers may not sum to total number of cases and controls due to missing values.

Percentages calculated amongst non-missings.

Table 2

HERV-K18 Genotype Frequency and Genotypic Relative Risks for Multiple Sclerosis in the NHS/NHS2 study

Genotype	Cases n (%)	Controls n (%)	RR (95% CI)	p-value	RR[*] (95% CI)	p-value
1/1	58 (28.0)	95 (23.6)	1.8 (1.0, 3.1)	0.03	1.8 (1.0, 3.1)	0.05
1/2	59 (28.5)	142 (35.2)	1.2 (0.7, 2.1)	0.5	1.2 (0.7, 2.1)	0.6
1/3	25 (12.1)	36 (8.9)	1.9 (1.0, 3.8)	0.05	1.8 (0.9, 3.7)	0.08
2/2	31 (15.0)	89 (22.1)	Ref		Ref	
2/3	25 (12.1)	33 (8.2)	2.2 (1.1, 4.3)	0.02	2.1 (1.0, 4.3)	0.04
3/3	9 (4.4)	8 (2.0)	3.3 (1.1, 9.6)	0.03	2.9 (1.0, 8.8)	0.06
LRT				0.03		

[*] Adjusted for number of HLA-DRB1*1501 alleles; LRT=Likelihood Ratio test

Table 3

Genotype Frequencies and Relative Risks for Multiple Sclerosis in the Replication Study (BWH) and Pooled Risk Estimates

Genotype	BWH			Pooled	
	Cases n (%)	Controls n (%)	RR (95% CI)[*]	RR (95% CI)	p-value
1/1	160 (17.6)	67 (19.8)	1.0 (0.7, 1.5)	1.2 (0.9, 1.7)	0.22
1/2	376 (41.4)	134 (39.5)	1.2 (0.9, 1.7)	1.2 (0.9, 1.6)	0.21
1/3	88 (9.7)	25 (7.4)	1.5 (0.9, 2.6)	1.7 (1.1, 2.6)	0.01
2/2	184 (20.2)	82 (24.2)	Ref	Ref	
2/3	93 (10.2)	29 (8.6)	1.3 (0.8, 2.2)	1.6 (1.0, 2.3)	0.03
3/3	8 (0.9%)	2 (0.6%)	1.7 (0.3, 8.6)	2.7 (1.1, 6.4)	0.03

[*] Adjusted for gender