An extension to a statistical approach for family based association studies provides insights into genetic risk factors for multiple sclerosis in the HLA-DRB1 gene

Sreeram V Ramagopalan1,2, Roisin McMahon2, David A Dyment1,2, A Dessa Sadovnick3, George C Ebers1,2 and Knut M Wittkowski*4

Address: 1Department of Clinical Neurology, University of Oxford, The West Wing, The John Radcliffe Hospital, Oxford, OX3 9DU, UK, 2Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK, 3Department of Medical Genetics and Faculty of Medicine, Division of Neurology, University of British Columbia, G920, Detwiller Pavilion, VCHA – UBC Hospital, 2211 Wesbrook Mall, Vancouver, British Columbia, V6T 2B5, Canada and 4Center for Clinical and Translational Science, The Rockefeller University, 1230 York Ave Box 322, New York, NY 10021, USA

E-mail: Sreeram V Ramagopalan - sramagopalan@gmail.com; Roisin McMahon - roisin@strubi.ox.ac.uk; David A Dyment - ddyment@well.ox.ac.uk; A Dessa Sadovnick - dessa.sadovnick@gmail.com; George C Ebers - george.ebers@clneuro.ox.ac.uk; Knut M Wittkowski* - kmw@rockefeller.edu

*Corresponding author

Published: 04 February 2009 Received: 31 December 2008

BMC Medical Genetics 2009, 10:10  doi: 10.1186/1471-2350-10-10  Accepted: 4 February 2009

This article is available from: http://www.biomedcentral.com/1471-2350/10/10

© 2009 Ramagopalan et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Multiple sclerosis (MS) is a complex trait in which genes in the MHC class II region exert the single strongest effect on genetic susceptibility. The principal MHC class II haplotype that increases MS risk in individuals of Northern European descent are those that bear HLA-DRB1*15. However, several other HLA-DRB1 alleles have been positively and negatively associated with MS and each of the main allelotypes is composed of many sub-allelotypes with slightly different sequence composition. Given the role of this locus in antigen presentation it has been suggested that variations in the peptide binding site of the allele may underlie allelic variation in disease risk.

Methods: In an investigation of 7,333 individuals from 1,352 MS families, we assessed the nucleotide sequence of HLA-DRB1 for any effects on disease susceptibility extending a recently published method of statistical analysis for family-based association studies to the particular challenges of hyper-variable genetic regions.

Results: We found that amino acid 60 of the HLA-DRB1 peptide sequence, which had previously been postulated based on structural features, is unlikely to play a major role. Instead, empirical evidence based on sequence information suggests that MS susceptibility arises primarily from amino acid 13.

Conclusion: Identifying a single amino acid as a major risk factor provides major practical implications for risk and for the exploration of mechanisms, although the mechanism of amino acid 13 in the HLA-DRB1 sequence’s involvement in MS as well as the identity of additional variants on MHC haplotypes that influence risk need to be uncovered.
Background

Multiple sclerosis (MS) is a common inflammatory disease of the central nervous system characterized by myelin loss, axonal pathology and progressive neurological dysfunction [1]. The aetiology of MS is unknown, but it is clear that both genetic and environmental components are important [2].

The only consistent genetic association with MS in Northern Europeans had been with extended MHC haplotypes especially those containing HLA-DRB1*1501 [3]. Recently, the interleukin 7 receptor (IL7R) interleukin 2 receptor (IL2R), ecotropic viral integration site 5 (EVI5) [4] and kinesin family member 1B (KIF1B) [5] genes have been shown to be additional MS susceptibility loci [3, 6, 7]. However, the MHC is the largest contributor to MS risk [3].

The association between MS and MHC class II has been fine mapped to the extended haplotype HLA-DQA1*0102-DQB1*0602-DRB1*1501-DRB5*0101 [8]. Intense linkage disequilibrium within the MHC has prevented the exact susceptibility locus from being identified. Analysis of the MHC region with a large number of markers as well as classical typing show evidence for the involvement of the class II region only [9, 10]. The paradigm is more complex than one in which the HLA-DRB1*15 allele acts solely to increase MS risk. Our previous investigations have shown that HLA-DRB1*15 and HLA-DRB1*17 bearing haplotypes increase risk of MS, while HLA-DRB1*14 and HLA-DRB1*11 bearing haplotypes are protective [11, 12]. Additionally, HLA-DRB1*10, DRB1*01, and DRB1*08 interact with HLA-DRB1*15 to influence disease risk [11, 12].

MHC class II molecules present antigen to CD4+ T helper cells and are integral to successful maintenance of self tolerance by the immune system and the adaptive immune response to invading pathogens [13]. Each HLA-DRB1 allele forms, by the presence of defined amino acid anchors, a number of specific pockets comprising a peptide binding groove [14]. Different HLA-DRB1 alleles may thus have different binding affinities for disease-related peptides as determined by their protein sequence, subsequently influencing composition of T cell repertoires, ultimately resulting in HLA-DRB1 alleles having varying effects on disease risk. However, protein sequence analysis has failed to provide clarity. Class II alleles in MS patients are structurally no different to those in healthy controls [15]. While some studies have suggested that variable residues in the DR beta chain may determine MS susceptibility [16-18], others found no evidence that MS pathogenesis is mediated by allele overlapping antigen binding sites [19]. However, these studies were based on a relatively small number of individuals and, thus, may have been underpowered to detect any relevant effects [11, 12]. More recently, Barcellos et al. [20], by aligning the protein sequences of HLA-DRB1*1501, *1503, *1701, *0401, *0801, and *0803 with that of HLA-DRB1*140101, *140102, *140103, and *1404 have suggested that the amino acid at position 60 of the HLA-DRB1 protein sequence determines the effect of an HLA-DRB1 allele on MS susceptibility. This model however does not take into account other HLA-DRB1 resistance alleles, notably HLA-DRB1*11, *10 and *01 [12].

The HLA-DRB1 gene is unusual in that many loci in the coding sequence can have any one of the four nucleotides depending on the allelotype. Thus, empirical studies were difficult, first, because such variability requires large sample sizes, but – even more importantly – because most traditional statistical methods are limited to the more typical case of bi-allelic loci. We present here the largest systematic investigation to date and an extended more sensitive statistical approach, which, for the first time, will allow us to determine empirically whether or not the nucleotide or protein sequence of HLA-DRB1 can account for allelic susceptibility to MS.

Methods

Participants

All participants in the study were ascertained through the ongoing Canadian Collaborative Project on the Genetic Susceptibility to MS (CCPGSMS), for which the methodology has been previously described [21].

Genotyping

Total genomic DNA, extracted from whole blood as part of the CCPGSMs, was used to type HLA-DRB1 alleles by an allele-specific PCR amplification method [22]. All genotypes were generated blind to pedigree structure and disease status of the individual. Initially, HLA-DRB1 alleles were classified into 10 categories, HLA-DRB1*01 to HLA-DRB1*10. Four of them were then subdivided, *05 into *11/12, *06 into *13/14, *02 into *15/16, and *03 into *17/18. Since then, these "two-digit" genotypes have been refined by adding two or four more digits. The first two digits describe the type, which corresponds to the serological antigen carried by an allelotype. The third and fourth digits are used to list the allelotype subtypes, numbers being assigned in the order in which the DNA sequences were determined. Alleles whose numbers differ in the first four digits must differ in one or more nucleotide substitutions that change the amino acid sequence of the encoded protein. Alleles that differ only by synonymous nucleotide substitutions within the
coding sequence are distinguished by the use of the fifth and sixth digits.

In our population, 24 PCRs were carried out to amplify allelotypes corresponding to alleles HLA-DRB1*01, HLA-DRB1*04, HLA-DRB1*07, HLA-DRB1*08, HLA-DRB1*09, HLA-DRB1*10, HLA-DRB1*11, HLA-DRB1*12, HLA-DRB1*13, HLA-DRB1*14, HLA-DRB1*15, HLA-DRB1*16, HLA-DRB1*17, and HLA-DRB1*18. Each HLA-DRB1 genotype was scored twice by independent observers.

Sequence Information
HLA-DRB1 allele sequence information was obtained from the Immunogenetics database of the European Bioinformatics Institute [23]ftp://ftp.ebi.ac.uk/pub/databases/imgt/mhc/hla/Alignments_Rel_2.21.zip, whose nucleotide and amino acid numbering scheme will be used here.

Statistical Analysis
For terrestrial life forms, the number of alleles per single nucleotide polymorphism (SNP) is limited to five: A, C, G, T, and X (deletion), although only two alleles have been seen for most human SNPs. For HLA-DRB1 and HLA-DRB2 SNPs, however, up to 4 and 5 nucleotides, respectively, have been observed. As proven in [24], the informative data for each bi-allelic parental mating type (PMT) can be organized into three strata, with two informative filial constellations each:

| Mating Offspring Genotypes |
|---------------------------|
| A.A~A.C                   |
| A.A A.C                   |
| A.C~A.C                   |
| A.A C.C (A.C children are non-informative) |
| A.C~C.C                   |
| A.C C.C                   |
| ...                       |
| T.T~T.X                   |
| T.T T.X                   |
| T.X~T.X                   |
| T.T X.X (T.X children are non-informative) |
| T.X~X.X                   |
| T.X X.X                   |

With three alleles, only two constellations need to be considered:

| Mating Offspring Genotypes |
|---------------------------|
| A.A~A.G                   |
| A.C A.G                   |
| A.C~A.G                   |
| A.A A.C A.G C.G           |
| ...                       |
| X.X~T.G                   |
| G.X T.X                   |
| X.X~G.X                   |
| G.T G.X T.X X.X           |

With four or five alleles, each stratum contains four informative filial genotypes, e.g.:

| Mating Offspring Genotypes |
|---------------------------|
| A.C~G.T                   |
| A.G A.T C.G C.T           |

To determine, whether a particular allele confers risk, children with this allele need to be `paired` within the same PMT stratum with children having the same other allele. For bi-allelic PMTs, this leaves three informative mating types per combination of alleles, each with two informative filial genotypes [24]. For tri-allelic PMTs, two situations may arise. One homozygous parent, also yields only two filial genotypes, while two heterozygous parents, yield strata with four filial genotypes. For quad-allelic PMTs, both parents are necessarily heterozygous yielding, again, four possible filial genotypes.

After excluding non-informative children, all filial genotypes within a PMT stratum have the same expectation under the null hypothesis. Thus, for the special case of bi-allelic parents, it has been suggested [25] that the sign/McNemar test for exact ties be applied [26], i.e. to count how often either of the genes were transmitted, irrespective of the PMT. Stratification [24] yields a test statistic based on counts representing independently observed events (cases born, rather than alleles transmitted, which are subject to identical genetical and environmental confounders). Here, the need for stratification is even more apparent than for bi-allelic parents [27-29], as there are, for instance, twice as many A alleles to be expected among the children of A.C~A.G parents than either C or G alleles.

When developing a test suitable for multi-allelic loci, it is useful to note that the sign test can be written in the two equivalent forms given at the ends of the following equation:

\[
\frac{(a-b)^2}{a+b} = \frac{a}{a+b} \left( \frac{a+b}{2} \right)^2 - \frac{2(a-b)^2}{a+b} = \frac{(a-b)^2}{a+b} \left( \frac{a+b}{2} \right)^2 - \frac{2(a-b)^2}{a+b} = \frac{(a-b)^2}{a+b} \left( \frac{a+b}{2} \right)^2 \]

As \( (a-b)/\sqrt{a+b} \approx \Delta \chi^2_1 \) and the "standardized residuals" \( (a-b)/\sqrt{a+b} \) and \( (b-a)/\sqrt{a+b} \) are symmetric, each follows asymptotically a standard Gaussian distribution, allowing the contributions from different PMT strata to be combined in an additive fashion.
To test the effect of allele A, for instance, the following strata are informative:

**Mating Informative Offspring Genotypes**

\[
\begin{align*}
A.A^+A.x: & \quad A.A \ vs \ A.x \\
A.x^+A.x: & \quad A.A \ vs \ x.x \\
A.x^+x.y: & \quad A.x \ vs \ x.y \\
A.x^+y.y: & \quad A.y \ vs \ x.y \\
A.x^+y.z: & \quad (A.y + A.z) \ vs \ (x.y + x.z) \\
\end{align*}
\]

The informative children AA in the A.x~A.x and A.x~A.y strata differ from the 'non-A' (x.x or x.y) counterparts by two alleles. As in [24], this will be accounted for by assigning twice the weight to these strata, rather than assuming that the effects of the two alleles transmitted to the same child are independently observed.

With more than two alleles, several comparisons could be made within the A.x~A.y stratum, e.g.,

**Mating Informative Offspring Genotypes**

\[
\begin{align*}
A.x^+A.y: & \quad (A.A + A.y) \ vs \ (A.x + x.y) \\
A.x^+x.y: & \quad (A.A + A.x) \ vs \ (A.y + x.y) \\
\end{align*}
\]

In both cases, the term to the left (A.A or A.x/A.y) would have more A alleles than the corresponding term on the right (A.x/A.y or x.y, respectively). Averaging across these comparisons yields the same term (A.x+A.y)/2 on both sides which, as the heterozygous children born to two heterozygous children in the biallelic case, can be ignored, resolving this seeming ambiguity. Again, stratification is essential for developing a sound statistical approach to deal with this complex situation.

From column card in Figure 1, the maximum number of informative strata for the influence of a given allele is \(3 \times (4 + 3) \times 12 + 6 = 54\) out of the 105. Of course, if a particular PMT is not observed, the corresponding strata need not be included. Column \(w\) gives the weight to be applied to this stratum’s contribution based on the number of alleles differing.

This stratification now allows for different degrees of 'dominance' to be considered. An allele is "dominant" or "recessive", if having a single copy confers the same risk as having two copies or no copy, respectively.

When both parents are heterozygous for the allele of interest, only those children are counted where both alleles are equal to or different from the allele of interest \((n^* = n_{A^{-A}} + n_{A^{-A}})\). Thus, among these strata, the effective sample size depends on the allele considered.

Using the results of [30], the strata’s contributions can be combined into an allele-specific test statistic by forming a single test statistic from the sum of the effect estimates and their variances, respectively, e.g.,

\[
T_n = \frac{\sum_{x,y,z} \sum_{PMT} n_{A^{-A}} PMT (A.A)^w \pm n_{A^{-A}}^{-A} \text{PMT}}{\sqrt{\sum_{x,y,z} \sum_{PMT} (A.A)^w \pm n_{A^{-A}}^{-A} \text{PMT}}} \times Z_{0.1}
\]

where \(x\) is the subset of the 54 PMTs informative (relevant and not empty) for nucleotide A, while \(n_{A^{-A}}^{-A}\text{PMT} \) and \(n_{A^{-A}}^{-A}\text{PMT} \) states the number of subjects with a preponderance of A or non-A alleles, respectively, within the stratum indicated (i.e., \(n_{A^{-A}} \), \(n_{A^{-A}} \), or \(n_{A^{-A}} \), see above).

From the above, nucleotide-specific test statistics can be obtained by performing the following steps:

1) Select the PMTs where at least one parent is heterozygous for the nucleotide considered (Figure 1).

2) Among the strata where both parents are heterozygous for the nucleotide considered, eliminate the counts of children that are also heterozygous for this nucleotide.

3) By number of parental copies of the nucleotide of interest (1, 2, or 3), count the number of children with more (1 or 2) vs the number of children with less (0 or 1, respectively) alleles of this nucleotide.

4) Perform the stratified McNemar test [24] based on these six numbers and the weights assigned to each of
the three categories of strata according to the degree of dominance to be considered (Figure 1).

Software
Statistical analyses were done using S-PLUS 8.0 http://www.insightful.com and the muStat library (available from http://csan.insightful and http://cran.r-project.org). Surface and cartoon representations were produced using PyMOL http://www.pymol.org. The program VOLUMES (R. Esnouf, unpublished data) was used with a 1.4 Å radius probe to map the extent of the P4 pocket cavity.

Results
Of 7333 genotyped subjects, 3178 children (1773 affected and 1405 unaffected) from 1352 families (1–15 children per family) had complete familial allele-type and disease status information available.

Sequence analysis identified 93 SNPs that suffice to characterize the differences between the 13 main allelic types of HLA-DRB1. As shown in Table 1, the relationship between nucleotides and allelic groups is complex. Thus, characterizing subjects by the 13 allelic groups (two-digit resolution) may not suffice to identify the genetic factors contributing to disease susceptibility (or resistance). For instance, for the second SNP (nucleotide position (N) 016), HLA-DRB1*04 (T) is uniquely different from all other groups (C), while the third SNP (N037) separates HLA-DRB1*01, *15, and *16 (A) from all other alleles (G). HLA-DRB1*09 is the only allelic group characterized by a single allele at each of the 93 SNPs – all other allelotypes have genetic variability, often with more than two alleles observed within each allelic group at a given SNP. The region between N256 and N308 is characterized by various allelotypes with three or four potential nucleotides per SNP.

This would render current statistical analysis strategies, comparing one allelic group against all others, highly inefficient. To resolve this conundrum, we extended a –93 SNPs allelic group characterized by a single allele at each of the radius probe to map the extent of the P4 pocket cavity.

Software
Statistical analyses were done using S-PLUS 8.0 http://www.insightful.com and the muStat library (available from http://csan.insightful and http://cran.r-project.org). Surface and cartoon representations were produced using PyMOL http://www.pymol.org. The program VOLUMES (R. Esnouf, unpublished data) was used with a 1.4 Å radius probe to map the extent of the P4 pocket cavity.

Results
Of 7333 genotyped subjects, 3178 children (1773 affected and 1405 unaffected) from 1352 families (1–15 children per family) had complete familial allele-type and disease status information available.

Sequence analysis identified 93 SNPs that suffice to characterize the differences between the 13 main allelic types of HLA-DRB1. As shown in Table 1, the relationship between nucleotides and allelic groups is complex. Thus, characterizing subjects by the 13 allelic groups (two-digit resolution) may not suffice to identify the genetic factors contributing to disease susceptibility (or resistance). For instance, for the second SNP (nucleotide position (N) 016), HLA-DRB1*04 (T) is uniquely different from all other groups (C), while the third SNP (N037) separates HLA-DRB1*01, *15, and *16 (A) from all other alleles (G). HLA-DRB1*09 is the only allelic group characterized by a single allele at each of the 93 SNPs – all other allelotypes have genetic variability, often with more than two alleles observed within each allelic group at a given SNP. The region between N256 and N308 is characterized by various allelotypes with three or four potential nucleotides per SNP.

This would render current statistical analysis strategies, comparing one allelic group against all others, highly inefficient. To resolve this conundrum, we extended a –93 SNPs allelic group characterized by a single allele at each of the radius probe to map the extent of the P4 pocket cavity.

## Results

### Of 7333 genotyped subjects, 3178 children (1773 affected and 1405 unaffected) from 1352 families (1–15 children per family) had complete familial allele-type and disease status information available.

Sequence analysis identified 93 SNPs that suffice to characterize the differences between the 13 main allelic types of HLA-DRB1. As shown in Table 1, the relationship between nucleotides and allelic groups is complex. Thus, characterizing subjects by the 13 allelic groups (two-digit resolution) may not suffice to identify the genetic factors contributing to disease susceptibility (or resistance). For instance, for the second SNP (nucleotide position (N) 016), HLA-DRB1*04 (T) is uniquely different from all other groups (C), while the third SNP (N037) separates HLA-DRB1*01, *15, and *16 (A) from all other alleles (G). HLA-DRB1*09 is the only allelic group characterized by a single allele at each of the 93 SNPs – all other allelotypes have genetic variability, often with more than two alleles observed within each allelic group at a given SNP. The region between N256 and N308 is characterized by various allelotypes with three or four potential nucleotides per SNP.

This would render current statistical analysis strategies, comparing one allelic group against all others, highly inefficient. To resolve this conundrum, we extended a –93 SNPs allelic group characterized by a single allele at each of the radius probe to map the extent of the P4 pocket cavity.

## Results

**Of 7333 genotyped subjects, 3178 children (1773 affected and 1405 unaffected) from 1352 families (1–15 children per family) had complete familial allele-type and disease status information available.**

Sequence analysis identified 93 SNPs that suffice to characterize the differences between the 13 main allelic types of HLA-DRB1. As shown in Table 1, the relationship between nucleotides and allelic groups is complex. Thus, characterizing subjects by the 13 allelic groups (two-digit resolution) may not suffice to identify the genetic factors contributing to disease susceptibility (or resistance). For instance, for the second SNP (nucleotide position (N) 016), HLA-DRB1*04 (T) is uniquely different from all other groups (C), while the third SNP (N037) separates HLA-DRB1*01, *15, and *16 (A) from all other alleles (G). HLA-DRB1*09 is the only allelic group characterized by a single allele at each of the 93 SNPs – all other allelotypes have genetic variability, often with more than two alleles observed within each allelic group at a given SNP. The region between N256 and N308 is characterized by various allelotypes with three or four potential nucleotides per SNP.

This would render current statistical analysis strategies, comparing one allelic group against all others, highly inefficient. To resolve this conundrum, we extended a –93 SNPs allelic group characterized by a single allele at each of the radius probe to map the extent of the P4 pocket cavity.

## Results

**Of 7333 genotyped subjects, 3178 children (1773 affected and 1405 unaffected) from 1352 families (1–15 children per family) had complete familial allele-type and disease status information available.**

Sequence analysis identified 93 SNPs that suffice to characterize the differences between the 13 main allelic types of HLA-DRB1. As shown in Table 1, the relationship between nucleotides and allelic groups is complex. Thus, characterizing subjects by the 13 allelic groups (two-digit resolution) may not suffice to identify the genetic factors contributing to disease susceptibility (or resistance). For instance, for the second SNP (nucleotide position (N) 016), HLA-DRB1*04 (T) is uniquely different from all other groups (C), while the third SNP (N037) separates HLA-DRB1*01, *15, and *16 (A) from all other alleles (G). HLA-DRB1*09 is the only allelic group characterized by a single allele at each of the 93 SNPs – all other allelotypes have genetic variability, often with more than two alleles observed within each allelic group at a given SNP. The region between N256 and N308 is characterized by various allelotypes with three or four potential nucleotides per SNP.

This would render current statistical analysis strategies, comparing one allelic group against all others, highly inefficient. To resolve this conundrum, we extended a –93 SNPs allelic group characterized by a single allele at each of the radius probe to map the extent of the P4 pocket cavity.
| Allelotype: | N# | 02 | 01 | 04 | 03 | 05 | 06 | 07 |
|------------|----|----|----|----|----|----|----|----|
| -25        | 14 | A  | A  | A  | G  | G  | G  | G  | A  | A  |
| -24        | 16 | C  | C  | C  | T  | C  | C  | C  | C  | C  |
| -17        | 37 | A  | A  | A  | G  | G  | G  | G  | G  | G  |
| -16        | 41 | C  | C  | C  | T  | T  | T  | T  | T  | T  |
| -16        | 42 | G  | G  | G  | T  | T  | T  | T  | T  | T  |
| -1         | 85 | T  | G  | G  | G  | G  | G  | G  | G  | G  |
| -1         | 86 | C  | C  | C  | C  | C  | C  | C  | C  | C  |
| -1         | 87 | T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 4          | 97 | C  | C  | C  | C  | C  | C  | C  | A  | A  |
| 4          | 98 | G  | G  | G  | G  | G  | G  | G  | A  | A  |
| 8          | 109| C  | C  | C  | T  | T  | T  | T  | T  | T  |
| 9          | 112| T  | T  | T  | G  | G  | G  | G  | G  | G  |
| 10         | 115| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 11         | 127| C  | G  | G  | G  | G  | G  | G  | G  | G  |
| 11         | 129| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 16         | 132| C  | C  | C  | C  | T  | T  | T  | T  | T  |
| 19         | 144| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 26         | 154| T  | T  | T  | A  | T  | T  | T  | T  | T  |
| 26         | 165| C  | G  | C  | C  | C  | C  | C  | C  | C  |
| 26         | 169| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 28         | 171| C  | C  | A  | C  | C  | C  | C  | A  | A  |
| 30         | 175| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 30         | 176| A  | A  | G  | C  | C  | C  | C  | C  | C  |
| 31         | 178| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 33         | 184| A  | A  | A  | C  | C  | C  | C  | A  | A  |
| 34         | 189| G  | G  | A  | A  | A  | A  | A  | A  | A  |
| 37         | 196| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 38         | 199| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 38         | 200| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 47         | 227| T  | T  | A  | A  | A  | A  | A  | A  | A  |
| 53         | 246| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 57         | 256| G  | G  | A  | G  | G  | G  | G  | G  | G  |
| 57         | 257| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 57         | 258| C  | T  | C  | C  | C  | C  | C  | C  | C  |
| 58         | 260| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 58         | 261| C  | T  | C  | C  | C  | C  | C  | C  | C  |
| 60         | 265| T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 60         | 266| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 60         | 267| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 67         | 286| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 69         | 294| A  | G  | G  | G  | G  | G  | G  | G  | G  |
| 70         | 295| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 70         | 296| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 70         | 297| G  | C  | C  | C  | C  | C  | C  | C  | C  |
| 71         | 298| A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 72         | 303| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 73         | 305| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 74         | 307| G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 74         | 308| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 77         | 317| C  | C  | A  | C  | C  | C  | C  | C  | C  |
| 77         | 318| C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 78         | 320| A  | A  | A  | A  | A  | A  | A  | A  | A  |
In this study, we performed the first empirical investigation of the nucleotide sequence of HLA-DRB1 with respect to determining disease risk. The statistical analysis employed in this study differs fundamentally from previous analyses conducted in that it is based on the actual nucleotides at a SNP, rather than HLA allelotypes. In other words, it allows at each locus the cases (or controls) to be categorized differently according to the nucleotides observed. In principle, we might have done the same analysis based on individual sequence data, irrespective of the two- (four- or six-) digit allele categories. In fact, using individual sequence data would avoid the ambiguities which reduced the sample size to fewer than 300 for the SNPs between N196 and N372 (Table 2) due to some allelotypes having many different nucleotides at these loci.

For amino acid 13, for instance, we count cases with unambiguous parental mating type for the following HLA-DRB1 allelotypes (see Table 3):

As the categories are refined or individual sequence data becomes available, ambiguous cases will become fewer or avoided altogether, respectively.

As in the case of case-control studies [31], counting alleles transmitted to cases, as suggested by [25] in the

---

### Table 1: Nucleotides (N#: nucleotide number, #A: amino acid number) found in 94 HLA-DRB1 SNPs discriminating the 13 main two-digit allelic groups (rows)

| N# | Nucleotide | Amino Acid |
|----|------------|------------|
| 78 | 321        | C          |
| 85 | 341        | T          |
| 86 | 344        | T          |
| 90 | 357        | A          |
| 93 | 365        | C          |
| 95 | 372        | C          |
| 96 | 373        | G          |
| 96 | 374        | A          |
| 98 | 379        | G          |
| 101| 388        | G          |
| 101| 389        | T          |
| 104| 397        | T          |
| 106| 405        | C          |
| 114| 429        | C          |
| 117| 438        | C          |
| 120| 446        | G          |
| 133| 484        | C          |
| 133| 485        | T          |
| 133| 486        | G          |
| 134| 489        | C          |
| 140| 505        | G          |
| 143| 511        | A          |
| 142| 512        | T          |
| 142| 513        | G          |
| 145| 522        | A          |
| 149| 534        | G          |
| 152| 543        | C          |
| 166| 583        | C          |
| 166| 584        | G          |
| 166| 585        | A          |
| 169| 594        | G          |
| 179| 624        | C          |
| 180| 625        | G          |
| 181| 629        | C          |
| 181| 630        | A          |
| 190| 656        | A          |
| 207| 707        | T          |
| 218| 739        | C          |
| 232| 783        | C          |
| 234| 789        | C          |

Note that many SNPs have more than two nucleotides and some SNPs have several nucleotides for the same allelic group. For instance, Nucleotide 257 can be either A or T within allelotype *12 and A, T, or G within allelotype *13. Given the high prevalence of the HLA-DRB1*15 (15) allele, the nucleotide found in this group of alleles was used as the reference, unless several nucleotides were found in the alleles constituting the HLA-DRB1*15 group.
| #N | #A | A   | C   | G   | T   | #N | #A | A   | C   | G   | T   |
|----|----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|.....
widely used Transmission-Disequilibrium-Test (TDT) has recently been proven to result in a suboptimal analysis strategy even in the case of bi-allelic parents [24], because the TDT’s variance estimate is based, in part, on the counts of non-informative heterozygous children born to two heterozygous parents. As with Student’s t- vs. the Gauss’ z-test (which are also asymptotically equivalent), this additional variance would need to be accounted for. The stratified McNemar test [24], in contrast, avoids this problem by replacing $2 \times (n_{P. P} + n_{P. Q} + n_{Q. Q})$ in the denominator by $4 \times (n_{P. P} + n_{Q. Q})$. For tri- or tetra-allelic parents, stratification is even more important, because it also leads to novel analysis strategies.

Barcellos et al. [20] have suggested that amino acid 60 accounts for the protective effect of the HLA-DRB1*14 allele, because this locus is close to the P4 binding pocket and HLA-DRB1*1401 encodes histidine (CAC), while HLA-DRB1*1701, *0801, and *1501 all encode tyrosine (TAC). Upon close inspection of the SNP profiles, however, the observed relationship between HLA-DRB1*15/*15, HLA-DRB1*15/*08 and HLA-DRB1*14 seems to be more complex.

Among the 4-digit HLA-DRB1 allelotypes, the majority of alleles code TAC for tyrosine, as does HLA-DRB1*15 (all 4-digit allelotypes).

### Table 2: Test statistics for specific alleles at any of the 93 SNPs increasing (positive) or decreasing (negative) the risk of MS (Continued)

| SNP | n | Allele 1 | Allele 2 | Risk | p-value |
|-----|---|---------|---------|------|---------|
| 73  | 438| 117     | 6.723   | -6.723| 1231    | 0.506   | -0.506  |
| 74  | 446| 120     | -4.595  | 4.595  | 769     | -0.422  | 0.422   |
| 75  | 484| 133     | -15.117 | 15.117 | 1171    | -2.781  | 2.781   |
| 76  | 489| 134     | 6.565   | -6.565 | 1236    | 0.562   | -0.562  |
| 77  | 505| 140     | -9.607  | 9.607  | 1325    | -0.829  | 0.829   |
| 78  | 511| 142     | 15.117  | -15.117| 1171    | 2.781   | -2.781  |
| 79  | 522| 145     | 2.573   | -2.573 | 121     | -1.342  | 1.342   |
| 80  | 534| 149     | -6.565  | 6.565  | 1236    | -0.562  | 0.562   |
| 81  | 543| 152     | 3.750   | -3.750 | 544     | 1.403   | -1.403  |
| 82  | 584| 166     | -1.000  | 1.000  | 36      | 0.316   | -0.316  |
| 83  | 585| 166     | 15.117  | -15.117| 1171    | 2.781   | -2.781  |
| 84  | 594| 166     | -3.206  | 3.206  | 696     | -1.308  | 1.308   |
| 85  | 624| 179     | 5.203   | -5.203 | 1003    | 1.941   | -1.941  |
| 86  | 625| 180     | -4.485  | 4.485  | 747     | -0.512  | 0.512   |
| 87  | 629| 181     | 3.348   | -3.348 | 718     | 1.199   | -1.199  |
| 88  | 630| 181     | 7.597   | -7.597 | 1248    | 2.093   | -2.093  |
| 89  | 656| 190     | -0.450  | 0.450  | 180     | -0.616  | 0.616   |
| 90  | 707| 207     | -6.168  | 6.168  | 1032    | -1.428  | 1.428   |
| 91  | 739| 218     | 2.961   | -2.961 | 657     | 1.418   | -1.418  |
| 92  | 783| 232     | 1.000   | -1.000 | 36      | 0.316   | 0.316   |
| 93  | 789| 234     | 6.813   | -6.813 | 1207    | 0.771   | -0.771  |

Left side: analysis of cases, right side: analysis of unaffected siblings.

**Figure 2**

**Position of candidate risk polymorphic residues in HLA-DRB1*1501.**

a. Cartoon representation of HLA-DRB1 (PDB ID 1BX2) in which the Cα atoms of candidate risk polymorphic residues (Table 2) are represented as small spheres (green). The β1 and β2 domains are labeled.

b. Surface and cartoon representation of HLA-DRB1 viewed from above the binding groove. The three-dimensional arrangement of the P4 pocket is shown as a grey mesh. The Cα atoms of the polymorphic residues 13 β (green) at the base of the P4 pocket and residue 60 β (orange) are represented as small spheres.
In the population of [32], 30% of the population were genotyped at the 4-digit level. Among them, 82% were *1401 and 9% were *1404. If one assumes that these 30% were representative and that the same proportion holds for our population, H at P60 could exert its "protective" effect in about 90% of all 14xx subjects. However, the purported disease risk increasing TAC codon is also seen in *0101, *0103, *0104, *0110, and *0111. Thus, if this amino acid should cause *08 to have its special role in *15/*08 heterozygous subjects, this effect should be seen with other alleles, yet this is not the case [11, 12].

For amino acid 13, instead, *15 (all K) is disjoint from both *08 (all G/glycine) and *14 (mostly S/serine)

**Codon\_HLA-DRB1 allele type carrying codon**

| Codon | HLA-DRB1 allele type |
|-------|---------------------|
| AGG/K (15xx): (all) | |
| GGT/G (08xx): (all) | |
| TCI/S (14xx): 01,02,03,05,06,07,08,09,12,13,14,16,17,18,19,20,21,22,23,24,25,26,27,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,51,53,54,55,56,58,59,60,62,63,64,65,66,67,69,70,72,74,75,77 | |
| GGT/G (14xx): 04,11,15,28,29,30,50,52,61,68,71,73,76 | |
| CAT/H (14xx): 10,31,57 | |

From this data, it seems unlikely that histidine at amino acid 60 contributes substantially to the protective effect of this haplotype. Instead, amino acid 13 emerges as a more likely explanation for a disease association gradient [32]. It is the only amino acid that shows sufficient variation to explain a hierarchy of disease associated alleles, raising the possibility that TCI/S is, in fact, associated with protection, while GTI/G is associated with increased risk. Being located right in the center of the P4 binding pocket, amino acid 13 is a residue that is potentially in contact with presented peptides [29, 32]. Another interesting finding of this approach is the implication that the promoter region, may play an important role. This finding lends notion to the idea of regulatory variants contributing to MS risk. As we have previously shown [33], HLA-DRB1 and therefore amino acid 13 cannot fully explain the MHC class II associated MS risk and these nearby variants remain to be uncovered.

It has been argued that "low resolution allele grouping [...] maximize [s] statistical power" [20', p. 2821] when comparing one group against all others. When increasing the sample size comes at the expense of increasing within group variance, however, statistical power often suffers. Let us assume that group *02 had not been separated into *15 (3994 cases) and *16 (208) cases. Then, according to Table 4, 208 'C' cases would have been added to the 3,994 'T' cases to a total of 4202, but the size of the 'G' group would have been reduced from 10,674 to 10,466. Even though the number of informative trios would slightly increase (from 1139 to 1171, see Table 2), the test statistic would drop from 11.022 to 10.689. The proposed statistical test, in contrast, utilized the information obtained through high resolution allelotyping (or sequencing) to increase group sizes by combining allelotypes with the same nucleotide at a given locus.

As we have demonstrated above, introducing the concept of stratification by parental mating type has not only quantitative advantages [24] over the original more simplistic approach, but also qualitatively different results for SNPs with more than two alleles in the population. As more data are collected, the proposed method could even be extended to address the difference between association (in trans) and protein function (in cis). To do so, one could categorize filial genotypes by pairs of amino acids, rather than nucleotides, so that differences in nucleotides coding for the same protein would be ignored. Of course, to ensure that genetic confounders are eliminated, parental mating types should still be defined based on the genetic code, rather than the amino acid for which it is coding. While this strategy is based on a large number of possible combinations between three-nucleotide parental mating types and combinations of filial amino acids, the number of combinations observed at each locus is likely to be substantially smaller.

**Conclusion**

In conclusion, an extended statistical approach allowed us to identify A013 at the center of the P4 pocket of HLA-DRB1 as a potentially important (although unlikely exclusive) risk factor for MS.
Table 4: Variation of genotypes for the amino acids marked in Table 2

| Allele: | N# | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 |
|---------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| -1      | 85 | T  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  |
| -1      | 86 | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  |
| -1      | 87 | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 13      | 124| A  | A  | T  | C  | T  | T  | T  | G  | T  | G  | T  | T  | T  |
| 13      | 125| G  | G  | T  | A  | C  | C  | G  | C  | G  | T  | T  | A  | T  |
| 13      | 126| G  | G  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 60      | 265| T  | T  | T  | T  | T  | T  | CT | T  | T  | T  | T  | T  | T  |
| 60      | 266| A  | A  | A  | A  | A  | A  | AC | AC | A  | A  | A  | A  | A  |
| 96      | 374| A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 96      | 375| A  | G  | T  | T  | T  | T  | T  | T  | T  | A  | T  | T  | T  |
| 101     | 388| G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 101     | 389| T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 133     | 484| G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  |
| 133     | 485| A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 142     | 511| A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  | A  |
| 142     | 512| T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  | T  |
| 166     | 583| C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  | C  |
| 166     | 584| G  | G  | G  | G  | G  | G  | G  | A  | G  | G  | G  | G  | G  |
| 166     | 585| A  | A  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  | G  |

#Cases: 3994 208 1396 1982 1873 1068 146 1519 243 408 80 98 1653

(Rare observations, such as TCC in only two out of 77 14xx subcategories are omitted).

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
KMW and GCE conceived and designed the experiments. SVR, RM, DAD, ADS and KMW performed the experiments. SVR, GCE and KMW analyzed the data and wrote the paper.

Acknowledgements
The authors would like to thank all patients who generously participated in this study and physicians participating in the CCPGMS. Experiments performed for this investigation comply with current guidelines and ethics. This work was funded by the Multiple Sclerosis Society of Canada Scientific Research Foundation and the Multiple Sclerosis Society of the United Kingdom. SVR is funded by the Medical Research Council of the United Kingdom. The project was also supported in part by Grant Number UL1RR024143 from the U.S. National Center for Research Resources (NCRR). The funding bodies had no role in the design of this study, the collection, analysis, and interpretation of data, the writing of the manuscript and in the decision to submit the manuscript for publication.

References
1. Noseworthy JH, Lucchinetti C, Rodriguez M and Weinsenker BG: Multiple sclerosis. N Engl J Med 2000, 343(13):938–952.
2. Ebers GC: Environmental factors and multiple sclerosis. Lancet Neurol 2008, 7(3):268–277.
3. Ramagopalan SV and Ebers GC: Genes for multiple sclerosis. Lancet 2008, 371:283–285.
4. Hoppenbrouwers IA, Aulchenko YS, Ebers GC, Ramagopalan SV, Oostra BA, van Duijn CM and Hintzen RJ: EVI5 is a risk gene for multiple sclerosis. Genes Immun 2008, 9(4):334–337.
5. Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, Broer L, Jafari N, Hillert J, Link J, Lundstrom W, Greiner E and Desa Sadovnick A, et al: Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. Nat Genet 2008, 40(12):1402–1403.
6. Häller DA, Compton A, Sawyer S, Lander ES, Daly MJ, De Jager PL, de Bakker PIW, Gabriel SB, Mirel DB and Lwinon AJ, et al: Risk alleles for multiple sclerosis identified by a genomewide study. New England Journal of Medicine 2007, 357(9):851–862.
7. Lundmark F, Dufvefelt K, Isacbaeus E, Kockum I, Wallstrom E, Khademi M, Oturai A, Ryder LP, Saarela J and Harbo HF, et al: Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet 2007.
8. Fogdell A, Hillert J, Sachs C and Olerup O: The multiple sclerosis-and narcolepsy-associated HLA class II haplotype includes the DRB5*0101 allele. Tissue Antigens 1995, 46(4):333–336.
9. Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiitalar M, Ferretti V, Tienari P, Sadovnick AD and Petonen L, et al: A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. Nat Genet 2005, 37(10):1108–1112.
10. Chao MJ, Barnardo MC, Bu GZ, Lincoln MR, Ramagopalan SV, Herrera BM, Dyment DA, Sadovnick AD and Ebers GC: Transmission of class I/II multi-locus MHC haplotypes and multiple sclerosis susceptibility: accounting for linkage disequilibrium. Hum Mol Genet 2007.
11. Dyment DA, Herrera BM, Cader MZ, Willer CJ, Lincoln MR, Sadovnick AD, Risch N and Ebers GC: Complex interactions among MHC haplotypes in multiple sclerosis susceptibility and resistance. Hum Mol Genet 2005, 14(14):2019–2026.
12. Ramagopalan SV, Morris AP, Dyment DA, Herrera BM, Deluca GC, Lincoln MR, Orton SM, Chao MJ, Sadovnick AD and Ebers GC: The Inheritance of Resistance Alleles in Multiple Sclerosis. PLoS Genet 2007, 3(9):e150.
13. Watts C: The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. Nat Immunol 2004, 5(7):685–692.
14. Jones EY, Fugger L, Strominger JL and Siebold C: MHC class II proteins and disease: a structural perspective. Nat Rev Immunol 2006, 6(4):271–282.

15. Cowan EP, Pierce ML, McFarland HF and McFarlin DE: HLA-DR and -DQ allelic sequences in multiple sclerosis patients are identical to those found in the general population. Hum Immunol 1991, 32(3):203–210.

16. Ghabanbasani MZ, Gu XX, Saepeen M, Vandeveyver C, Raus J, Marynen P, Carton H and Cassiman JJ: Importance of HLA-DRB1 and DQA1 genes and of the amino acid polymorphisms in the functional domain of DR beta 1 chain in multiple sclerosis. J Neuroimmunol 1995, 59(1–2):77–82.

17. Allen M, Sandberg-Wollheim M, Sigoren K, Erlich HA, Pettersson U and Gylensten U: Association of susceptibility to multiple sclerosis in Sweden with HLA class II DRB1 and DQB1 alleles. Hum Immunol 1994, 39(1):41–48.

18. Teutsch SM, Bennetts BH, Buhler MM, Heard RN and Stewart GJ: The DRB1 Val86/Val86 genotype associates with multiple sclerosis in Australian patients. Hum Immunol 1999, 60(8):715–722.

19. Zipp F, Windemuth C, Pankow H, Dichgans J, Wienker T, Martin R and Muller C: Multiple sclerosis associated amino acids of polymorphic regions relevant for the HLA antigen binding are confined to HLA-DR2. Hum Immunol 2000, 61(10):1021–1030.

20. Barcellos LF, Sawcer S, Ramsay PP, Baranzini SE, Thomson G, Briggs F, Cree BC, Begovich AB, Villalada L and Montalban X, et al: Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet 2006, 15(18):2813–2824.

21. Sadovnick AD, Risch NJ and Ebers GC: Canadian collaborative project on genetic susceptibility to MS, phase 2: rationale and method. Canadian Collaborative Study Group. Can J Neurol Sci 1998, 25(3):216–221.

22. Olerup O and Zetterquist H: 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(3):225–235.

23. Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P and Marsh SG: IMGT/HLA and IMGT/ MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 2003, 31(1):311–314.

24. Wittkowski KM and Liu X: A statistically valid alternative to the TDT. Human Heredity 2002, 54(3):157–164.

25. Spielman RS, McGinnis RE and Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993, 52(3):506–516.

26. Wittkowski KM: Versions of the sign test in the presence of ties. Biometrics 1998, 54:789–791.

27. Ewens WJ and Spielman RS: The TDT is a Statistically Valid Test: Reply to Wittkowski and Liu. Human Heredity 2004, 58(1):59–60.

28. Ewens WJ and Spielman RS: Reply to Wittkowski and Liu (Beyond the TDT: rejoinder to Ewens and Spielman). Human Heredity 2004, 58(1):61–62.

29. Wittkowski KM and Liu X: Beyond the TDT: Rejoinder to Ewens and Spielman. Human Heredity 2004, 58(1):60–61.

30. Wittkowski KM: Friedman-type statistics and consistent multiple comparisons for unbalanced designs. Journal of the American Statistical Association 1988, 83:1163–1170.

31. Sassi PD: From genotypes to genes: Doubling the sample size. Biometrics 1997, 53(4):1253–1261.

32. Oksenberg JR, Baranzini SE, Sawyer S and Hauser SL: The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. Nat Rev Genet 2008, advanced online publication.

33. Zhao MJ, Barnardo M, Lincoln MR, Ramagopalan SV, Herrera BM, Dyment DA, Montpetit A, Sadovnick AD, Knight JC and Ebers GC: HLA class I alleles tag HLA-DRB1*1501 haplotypes for differential risk in multiple sclerosis susceptibility. Proc Natl Acad Sci USA 2008, 105(35):13069–13074.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2350/10/10/prepub