



A Second Major Histocompatibility Complex Susceptibility Locus for Multiple Sclerosis

Citation

Yeo, Tai Wai, Philip L De Jager, Simon G Gregory, Lisa F Barcellos, Amie Walton, An Goris, Chiara Fenoglio, et al. 2007. A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Annals of Neurology* 61(3): 228-236.

Published Version

[doi://10.1002/ana.21063](https://doi.org/10.1002/ana.21063)

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4737553>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

A Second Major Histocompatibility Complex Susceptibility Locus for Multiple Sclerosis

The International Multiple Sclerosis Genetics Consortium

Objective: Variation in the major histocompatibility complex (MHC) on chromosome 6p21 is known to influence susceptibility to multiple sclerosis with the strongest effect originating from the *HLA-DRB1* gene in the class II region. The possibility that other genes in the MHC independently influence susceptibility to multiple sclerosis has been suggested but remains unconfirmed.

Methods: Using a combination of microsatellite, single nucleotide polymorphism, and human leukocyte antigen (HLA) typing, we screened the MHC in trio families looking for evidence of residual association above and beyond that attributable to the established *DRB1*1501* risk haplotype. We then refined this analysis by extending the genotyping of classical HLA loci into independent cases and control subjects.

Results: Screening confirmed the presence of residual association and suggested that this was maximal in the region of the *HLA-C* gene. Extending analysis of the classical loci confirmed that this residual association is partly due to allelic heterogeneity at the *HLA-DRB1* locus, but also reflects an independent effect from the *HLA-C* gene. Specifically, the *HLA-C*05* allele, or a variant in tight linkage disequilibrium with it, appears to exert a protective effect ($p = 3.3 \times 10^{-5}$).

Interpretation: Variation in the *HLA-C* gene influences susceptibility to multiple sclerosis independently of any effect attributable to the nearby *HLA-DRB1* gene.

Ann Neurol 2007;61:228–236

It is well established that the major histocompatibility complex (MHC) on chromosome 6p21 contains at least one gene that influences susceptibility to multiple sclerosis.^{1–6} Although this association was first identified more than 30 years ago¹ through the study of class I human leukocyte antigens (HLAs), it was quickly realized that this signal was predominantly, if not exclusively, the result of linkage disequilibrium (LD) with class II HLA genes, and that these exert the primary effect on susceptibility.^{7,8} The complex nature of the MHC, especially its high gene content, extreme polymorphism, and extensive LD,⁹ has con-

founded efforts to resolve the nature of the MHC association in multiple sclerosis, although progress and useful clarifications have been made, especially in recent years.

In virtually every population studied, multiple sclerosis is found to be associated with the *DRB1*1501* allele.¹⁰ The only exceptions are those populations where this allele has a low frequency, and analysis is therefore underpowered; but even in these situations, *DRB1*1501* is generally overrepresented in cases.¹¹ The *DRB1*1501* allele is carried on a particularly extensive haplotype,¹² the most common *DR15* haplo-

From the ¹Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom; ²Department of Neurology, Center for Neurologic Diseases, Brigham and Women's Hospital; ³Harvard Medical School, Boston; ⁴Program in Medical and Population Genetics, The Broad Institute at the Massachusetts Institute of Technology and Harvard University, Cambridge, MA; ⁵Duke University Medical Center, Center for Human Genetics, Durham, NC; ⁶Department of Neurology, School of Medicine, University of California San Francisco, San Francisco; ⁷Division of Epidemiology, School of Public Health, University of California at Berkeley, Berkeley, CA; ⁸Department of Neurological Sciences, Dino Ferrari Center, University of Milan, IRCCS Ospedale Maggiore Policlinico, Milan, Italy; ⁹Tissue Typing Laboratory, Addenbrooke's Hospital; ¹⁰Wellcome Trust Sanger Institute, Genome Campus, Hinxton; ¹¹Department of Pathology, Immunology Division, University of Cambridge, Cambridge, United Kingdom; ¹²Harvard Center for Neurodegeneration and Repair; ¹³The Center for Genome Research, Massachusetts General Hospital, Boston, MA; ¹⁴Center for Human Genetics Research,

Vanderbilt University Medical Center, Nashville, TN; ¹⁵Institute for Human Genetics, School of Medicine, University of California San Francisco, San Francisco, CA; and ¹⁶Montréal Heart Institute and Université de Montréal, Montréal, Québec, Canada.

Received Sep 14, 2006, and in revised form Nov 13. Accepted for publication Nov 20, 2006.

T.W.Y. and P.L.D. contributed equally to this work.

This article includes supplementary materials available via the Internet at <http://www.interscience.wiley.com/jpages/0364-5134/suppmat>

Published online Jan 24, 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ana.21063

Address correspondence to Dr Sawcer, University of Cambridge, Department of Clinical Neuroscience, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, United Kingdom.
E-mail: sjs1016@mole.bio.cam.ac.uk

type found in white Europeans. As a result, many variants from flanking genes, even some located quite a distance from *DRBI*, have sufficient LD with *DRBI*1501* that they invariably also show evidence for association with the disease in any population where association with *DRBI*1501* can be demonstrated.^{11,13–17} This extensive LD has made it difficult to establish which of the variants making up this haplotype is primarily responsible for the association. The distinction between *DRBI*1501* and *DQB1*0602* has been particularly taxing because the LD between these closely mapped genes is especially tight in those populations where the disease is frequent, that is, white Europeans and their migrant descendants. However, recent studies in the admixed African American population indicate the supremacy of the *DRBI*1501* allele.¹⁸

In the presence of one susceptibility allele it is difficult to identify effects attributable to a second allele,¹⁹ especially if the second allele exerts a more modest effect or has a low frequency, or both. However, by analyzing populations where *DR15* haplotypes are less common, and by using large cohorts, it has been possible to demonstrate that the *DRBI*0301* allele also confers susceptibility to multiple sclerosis, thereby confirming allelic heterogeneity at the *DRBI* locus.^{4,6,18,20} Furthermore available evidence suggests that the susceptibility effects of the *DRBI*1501* allele may be modulated by other *DRBI* alleles.^{6,20}

The relation between the MHC and multiple sclerosis is further complicated by the accumulating evidence suggesting that MHC loci mapping outside *DRBI* also influence susceptibility to the disease.^{11,13–16} Work in animal models suggests that clustering of susceptibility loci is a common phenomenon in complex disease,²¹ and it therefore appears reasonable to expect that other genes from the MHC region may influence susceptibility to multiple sclerosis. The observation of positive logarithm of odds scores in the MHC region in linkage studies stratified for the effects of *DRBI* supports the existence of secondary loci,^{22–24} although none of these data reaches a level providing statistical confidence. In considering these linkage data, it is important to remember that early linkage studies in multiple sclerosis^{25–27} were significantly underpowered,²⁸ to the point that they could not even convincingly demonstrate evidence for linkage resulting from the effects of *DRBI*. Confirmation of linkage in this region has been established only in more recent studies involving many hundreds of families.^{23,24,29} Given the inherently limited resolution of linkage-based studies,²⁸ the absence of statistically significant linkage in the MHC region after exclusion of primary effects attributable to *DRBI* does not exclude the presence of secondary loci. Several authors have attempted to identify secondary loci using more powerful association-based methods. Two groups

have typed dense microsatellite maps of the region and both found evidence for a secondary locus maximal in a region close to *HLA-A*: one group identifying the marker D6S1683 just telomeric of *HLA-A*,¹¹ and the second group implicating a region including *HLA-A* extending from MOGCA to D6S265 marker.¹⁴ Follow-up studies in Norway also found evidence implicating the D6S265 marker.¹⁶ In another smaller study, a microsatellite marker close to *HLA-C* (marker C1_3_2) also showed evidence for an independent effect.¹⁵ In contrast, a systematic effort to screen the MHC and flanking regions using single nucleotide polymorphisms (SNPs) found no evidence for association beyond that attributable to *DRBI*1501*, although this study was limited by a high genotyping failure rate (40%) and, more importantly, a distribution of markers leaving regions close to the classical loci essentially unexplored.¹⁷ In all of these studies, statistical power has inevitably been reduced by the processes required to filter out the primary effect attributable to *DRBI* and the large correction required for multiple testing. Unfortunately, none of the published studies has used sufficient samples to compensate for these statistical penalties, and thus none is able to provide unequivocal evidence supporting any particular secondary locus.

Subjects and Methods

UK Trio Families and Sporadic Cases

The 480 trio families (an affected individual and both parents) and 721 sporadic cases participating in our study were recruited from across the United Kingdom. All subjects involved in this study gave written informed consent and provided a venous blood sample from which DNA was extracted and normalized. Comparing data from the five classical loci in the trio family index cases ($n = 480$) with those from the sporadic cases ($n = 721$) showed that there was no statistically significant difference between these two cohorts. The clinical details for each set of cases are summarized in Table 1. All cases were diagnosed according to recognized criteria.^{30,31}

UK Extension Analysis Control Cohorts

The fully anonymous control data used in the extension analysis were derived from three sources: local organ donors, national organ donor records held by UK Transplant (UKT), and the 1958 birth cohort. Ethical permission for using these data was obtained from the appropriate respective research ethics committees. Data from all five classical loci (*HLA-A*, *-B*, *-C*, *-DRBI*, and *-DQB1*) were available for the donor and UKT cohorts, whereas these were available for only three loci in the 1958 birth cohort (*HLA-B*, *-DRBI*, and *-DQB1*). Only white individuals from these cohorts with complete data were included: 408 for the donor cohort, 2,201 for the UKT cohort, and 1,051 for the 1958 birth cohort (total 3660 individuals). There was no evidence for any statistically significant difference among these three cohorts in a pairwise comparison of the classical loci. We also compared each of the three control cohorts with the nontransmitted alleles

Table 1. Patient Demographics

Demographics	US Trio Index (n = 450)	UK Trio Index (n = 480)	UK Sporadic (n = 721)
Sex (M:F)	1:3.2	1:3.2	1:2.5
Mean age (yr)	39	38	48
Mean age at onset (yr)	29	25	33
Mean EDSS	4.0	4.4	4.6
Mean duration (yr)	11	13	15

The slightly younger age and greater proportion of female individuals seen in cases from the trio families in each population reflects the requirement for both parents to be alive and willing to take part. This necessarily means that these patients tend to be younger, and because the disease has a younger age at onset in female individuals, also results in an increased proportion of women.

EDSS = Extended Disability Status Scale.

from the 480 trio families. Again, there was no evidence for any statistically significant difference.

US Trio Families

All cases from the 450 US trio families were diagnosed according to the McDonald criteria.³¹ All individuals involved in this study gave written informed consent using documents approved by the institutional review board and provided a venous blood sample from which DNA was extracted and normalized. The clinical details for the 450 index cases from the trio families are summarized in Table 1.

Screening Single Nucleotide Polymorphisms

The recently completed resequencing of the MHC region from consanguineous homozygous cell lines carrying specific disease-associated haplotypes provided a comprehensive and detailed description of these haplotypes.^{32,33} By comparing the sequence from the PGF line, which carries the multiple sclerosis-associated DR15 haplotype (HLA-A3-B7-Cw7-DR15), with the other completed haplotypes, COX (HLA-A1-B8-Cw7-DR3) and QBL (A26-B18-Cw5-DR3), we were able to identify 241 coding variants (outside the hypervariable regions). These variants distinguish the multiple sclerosis-associated haplotype from alternatives and are therefore especially promising candidate susceptibility variants. To increase the coverage provided by this set of markers, we also developed assays for SNPs already identified as tagging common haplotypes in the MHC region³⁴ and supplemented this list with variants from the class III region. Working assays were established for a total of 110 SNPs, including 5 from the extended class I region, 1 from the extended class II region, and 104 from the classical MHC. Seventy SNPs were genotyped using a Sequenom MassArray MALDI-TOF platform,³⁵ whereas the remaining 40 were genotyped using TaqMan allelic discrimination assays on an ABI7900HT genotyping platform (Applied Biosystems, Foster City, CA).³⁶ The primer sequences used and basic performance characteristics for each marker are included in Supplementary Table S1.

Screening Microsatellites

To generate a screening set of microsatellites, we first identified an exhaustive list of such markers (n = 248) lying within the extended MHC using published³⁷⁻⁴¹ and publicly available resources: National Center for Biotechnology Infor-

mation (NCBI) UniSTS (Build 34.3; <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unists>), GDB Human Genome Database (<http://www.gdb.org>), and Ensembl (Build 22.34d.1; <http://www.ensembl.org/index.html>). From Ensembl BLAST (<http://www.ensembl.org/Multi/blastview>) and electronic polymerase chain reaction (ePCR) analysis⁴² (NCBI; <http://www.ncbi.nlm.nih.gov/sutils/e-pcr>), with emphasis on markers previously suggested to be of relevance in multiple sclerosis,^{11,14,15} we then selected markers from this list to produce an informative map with a density of approximately 1 marker per 50 to 100kb across the classical MHC region (29.8–33.2Mb⁹). The 69 markers selected were then typed in a trial set of 122 trio families. Seventeen markers were found to be monomorphic and two assays failed; we used the remaining 50 markers in our study. These 50 included 8 from the extended class I region, 4 from the extended class II region, and 38 from the classical MHC itself. Each microsatellite was amplified by PCR using TrueAllele PCR Premix and the manufacturer's standard conditions (Applied Biosystems). The PCR products were genotyped on a 3700 Genetic Analyzer (Applied Biosystems) using GENESCAN version 3.5 (Applied Biosystems) and GENOTYPER version 3.7 (Applied Biosystems) software. The primer sequences used and basic performance characteristics for each marker are included in Supplementary Table S1.

Human Leukocyte Antigen Typing

In the screening of trio families, four-digit (medium-resolution) typing of *HLA-DRB1* and *HLA-DQB1* was performed in all the UK trios and 60% of the US trios.^{20,43} In the remaining US trios, lower resolution typing was performed as described later. For the analysis of all 930 trio families together, the resolution was down grouped to the low-resolution level. The class I loci in the UK trio families and all the HLA typing performed in the sporadic cases was low-resolution typing based on PCR with sequence-specific primers, as developed by Olerup and Zetterquist⁴⁴ and previously used by us.⁴⁵ Alleles were assigned using World Health Organization nomenclature. The *HLA-DRB1*15* alleles were subtyped into *HLA-DRB1*1501*, *1502*, or *1503*. The primer sequences and combinations for each PCR reaction are given in Supplementary Table S2.

LD between *DRB1* and the other classical loci was confirmed in both the trio family and extension data sets; estimates for Global D' and Cramer's V (measures of LD be-

Table 2. Multiallelic Measures of Linkage Disequilibrium between *DRB1* and the Other Classical Loci

Locus	Cases	Control Subjects	Transmitted	Nontransmitted
Global D'				
<i>HLA-A</i>	0.24	0.24	0.29	0.32
<i>HLA-C</i>	0.42	0.44	0.51	0.53
<i>HLA-B</i>	0.49	0.47	0.56	0.56
<i>HLA-DQB1</i>	0.91	0.91	0.97	0.95
Cramer's V				
<i>HLA-A</i>	0.18	0.19	0.27	0.28
<i>HLA-C</i>	0.26	0.28	0.31	0.36
<i>HLA-B</i>	0.34	0.36	0.31	0.33
<i>HLA-DQB1</i>	0.78	0.79	0.82	0.82

The higher resolution of *HLA-DRB1* and *HLA-DQB1* typing together with the greater degree of phase information explains why the estimates for linkage disequilibrium are generally higher in this cohort than in the case-control analysis.

tween multiallelic loci) are summarized in Table 2. As expected, the extent of LD is inversely correlated with the distance from *DRB1*. It should be noted that although the extent of LD between class I loci and *DRB1* is modest, this level of LD is sufficient to produce an association signal at class I loci secondary to *DRB1* effects. Indeed, this is precisely how the association between the MHC and multiple sclerosis was first recognized. Allele counts and frequencies for the five classic loci in the screening and extension data sets are provided in Supplementary Table S3.

Statistical Analysis

Transmission disequilibrium testing (TDT) of data from trio families was performed using the TDTPHASE program, part of the UNPHASED suite.⁴⁶ Before analysis, Mendelian errors were zeroed out using the PEDCHECK program,⁴⁷ and basic performance characteristics were established using the PEDSTATS program.⁴⁸ In the extension study, case-control association testing was performed using the COCAPHASE program, also part of the UNPHASED suite.⁴⁶ In all tests, we used the EM option and grouped rare alleles (haplotypes) with expected counts of less than 10 in both case (transmitted) and control subjects (nontransmitted). This cutoff was chosen to ensure that we did not include alleles where the available data were insufficient to provide any power. In conditional analysis, we used the main-effects test. Measures of LD were also calculated using the relevant UNPHASED program. For multiallelic markers, UNPHASED calculates Global D' and Cramer's V as measures of the overall extent of LD. Nontransmitted alleles in the trio families were identified using the MERLIN program,⁴⁹ ignoring alleles where phase was uncertain.

Results

To refine the relation between the MHC and multiple sclerosis, we first screened the region by genotyping 110 SNPs and two class II HLA loci (*HLA-DRB1* and *HLA-DQB1*) in a cohort of 930 trio families (an affected individual and their parents): 480 from the United Kingdom and 450 from the United States. In addition, we also typed 50 microsatellite markers and 3

class I HLA loci (*HLA-A*, *-B*, and *-C*) in the same 480 UK trios. As expected, multiple markers showed highly significant evidence for association (see Supplementary Table S4 for individual results). To eliminate the confounding effects of LD with the *DRB1*1501* allele, we excluded all families where either parent carried a *DRB1*1501* allele and reanalyzed the data from the remaining 318 families (146 from the United Kingdom and 172 from the United States). The results generated in this subgroup analysis are also listed in Supplementary Table S4. After stringent Bonferroni correction for the number of markers tested ($n = 165$), we found that two markers continued to show statistically significant evidence for association: *HLA-C* ($p_{\text{corrected}} = 0.04$) and rs3132552 ($p_{\text{corrected}} = 0.006$), a synonymous coding polymorphism in the corneodesmosin gene. Because rs3132552 lies just 151kb telomeric of *HLA-C*, it was not surprising to find substantial LD between these loci in the UK trios (Global D' = 0.61). Testing rs3132552 in the *DRB1*1501*-negative trios from the United States and United Kingdom independently demonstrated that the marker shows association in both populations ($p_{\text{uncorrected}} = 0.001$ in the US cohort; $p_{\text{uncorrected}} = 0.009$ in the UK cohort). These data demonstrate that there is significant residual association within the MHC region above and beyond that attributable to the well-established association with *DRB1*1501*. However, the power available in this modest *DRB1*1501*-negative subgroup, together with the extensive LD between the various MHC loci, makes it impossible to establish with any confidence which locus is primarily responsible for the observed effect.

Consequently, to further refine the nature of the secondary MHC association identified in our screening experiment, we typed the five classical HLA loci (*HLA-A*, *-B*, *-C*, *-DRB1*, and *-DQB1*) in an additional 721 sporadic UK multiple sclerosis patients and estab-

Table 3. Association of Classical Human Leukocyte Antigen Loci in the Case–Control Data Sets

Cohort Subgroups	HLA-A	HLA-C	HLA-B	HLA-DRB1	HLA-DQB1	Cases, N	Control Subjects, N	Power, % ^a
Extension Analysis								
Full	3.3 × 10⁻⁷	1.4 × 10⁻¹⁶	4.3 × 10⁻²⁸	9.1 × 10⁻⁹⁴	1.2 × 10⁻²¹	1,201	3,660	99.9
DRB1*15 excluded	3.4 × 10⁻⁴	3.2 × 10⁻⁵	5.3 × 10⁻⁶	5.2 × 10⁻⁷	3.1 × 10⁻⁴	481	2,635	95.2
DRB1*15 and *03 excluded	0.042	2.3 × 10⁻⁵	0.017	2.9 × 10⁻⁴	0.016	297	1,802	75.8
DRB1*15, *03, and *0103 excluded	0.053	5.9 × 10⁻⁵	0.014	0.18	0.0075	264	1,724	69.5
Replication analysis								
Full	1.1 × 10⁻⁴	1.2 × 10⁻¹⁰	7.9 × 10⁻¹³	6.7 × 10⁻⁵³	4.2 × 10⁻¹⁴	1,201	3,660	99.6
DRB1*15 excluded	5.4 × 10 ⁻³	1.5 × 10⁻³	6.3 × 10⁻⁴	9.7 × 10⁻⁴	9.6 × 10⁻⁴	481	2,635	79.8
DRB1*15 and *03 excluded	0.21	2.6 × 10⁻⁴	0.010	0.16	0.012	297	1,802	50.2
DRB1*15, *03, and *0103 excluded	0.36	1.2 × 10⁻³	0.017	0.61	8.2 × 10 ⁻³	264	1,724	44.9

As each of the 5 markers has been tested 4 times in each approach, a Bonferroni correction factor of no more than 20 is required. Applying this to the nominal *p* values included in the table indicates that only those in bold are significant after this conservative correction for multiple testing.

^aThis column indicates the power of each analysis to identify a common allele (frequency 10%) conferring a risk with an odds ratio of 1.6 under a multiplicative model at a level of significance sufficient to survive Bonferroni correction (nominal *p* = 0.0025).⁵⁷

lished UK control data from a cohort of 3,660 individuals (see Subjects and Methods). Given that the US trio families provided independent evidence implicating the HLA-C region, we analyzed these new data from additional UK samples together with those from the 480 UK index cases used in the first screening experiment in the form of an extension analysis as opposed to a replication study.⁵⁰ Results from the first-pass unstratified analysis of all 1,201 cases and 3,660 control subjects are shown in Table 3. Association with *DRB1* is overwhelmingly the most significant, with the majority of this effect attributable to the *DRB1*1501* allele ($p_{\text{uncorrected}} = 4.5 \times 10^{-88}$). After excluding all individuals carrying *DRB1*1501*, analysis of the remaining data continues to show highly significant evidence for association with the most significant effect still appearing to come from the *DRB1* locus (see Table 3). As expected, the majority of the residual *DRB1* effect is attributable to the *DRB1*03* allele ($p_{\text{uncorrected}} = 6.3 \times 10^{-5}$). Therefore, we next excluded all individuals carrying *DRB1*03* alleles. Analysis of the remaining data shows that significant evidence for association is still evident but only at *HLA-C* and *DRB1* (see Table 3). At *DRB1*, the only allele showing significant evidence for association is the *DRB1*0103* allele ($p_{\text{uncorrected}} = 1.8 \times 10^{-5}$). Even though this allele is relatively uncommon, we also excluded all individuals carrying the *DRB1*0103* allele. In this final stratified analysis, statistically significant association is only apparent at the *HLA-C* locus (see Table 3). Analysis of *HLA-C* after conditioning on *DRB1* has no important effect on the evidence for association at this locus, whereas conditioning on *HLA-C* confirms that none of the four other loci exerts any residual main effects.^{19,51} Inspection of the individual *HLA-C* alleles in the extension analysis shows that the *HLA-C*05* allele is the most significantly associated, being underrepresented in

cases (Table 4). Following a replication approach, in which the original 480 index cases from the trio families are excluded, has little effect on the interpretation of the results. The results from this replication approach are shown in the lower half of Table 3 for comparison. Our analysis indicates that three *DRB1* alleles (**1501*, **03*, and **0103*) and one *HLA-C* allele (**05*) exert independent effects on susceptibility to multiple sclerosis. To determine the risk associated with each *DRB1* susceptibility allele in haplotypes with and without *HLA-C*05*, we reanalyzed the full data set three times, first excluding all individuals carrying *DRB1*03* or **0103*, next excluding all individuals carrying *DRB1*1501* or **0103*, and then finally after excluding all individuals carrying *DRB1*1501* or **03*. Table 5 shows the relative risk associated with the various haplotypic combinations of *DRB1* and *HLA-C* alleles, as

Table 4. Individual HLA-Cw Allele Associations in Final Subgroup of Data (in Which All Individuals Carrying *DRB1*1501*, **03*, and **0103* Alleles have been excluded; see Table 3)

Allele	Cases, N (%)	Control Subjects, N (%)	<i>p</i>
01	12 (2)	116 (5)	0.0061
02	34 (7)	133 (6)	0.38
03	76 (15)	403 (17)	0.22
04	67 (13)	252 (10)	0.12
05	37 (7)	320 (13)	3.3×10^{-5}
06	71 (14)	278 (12)	0.19
07	110 (21)	470 (20)	0.41
08	32 (6)	113 (5)	0.18
12	18 (3)	75 (3)	0.69
14	6 (1)	21 (1)	0.56
15	24 (5)	49 (2)	0.0017
16	29 (6)	153 (6)	0.49
17	3 (1)	18 (1)	0.66

compared with haplotypes carrying no associated allele at either locus. The relative risk associated with haplotypes carrying *HLA-C*05* without *DRB1* risk alleles is consistent across the three analyses and is significantly less than 1 in each case, confirming the protective nature of this allele. The data suggest that the risk associated with the *DRB1*1501* allele is reduced but not abolished by the presence of an *HLA-C*05* allele on the same haplotype. Unfortunately, the frequency of the other combined haplotypes (*DRB1*03* with *HLA-C*05* and *DRB1*0103* with *HLA-C*05*) was too low to provide sufficient statistical power to make a judgment about whether the *HLA-C*05* allele alters the risk associated with the secondary *DRB1* risk alleles.

Discussion

Using a combination of microsatellite, SNP, and HLA typing in family-based and case-control cohorts from two different populations, we have shown that *HLA-C* exerts an independent effect on susceptibility to multiple sclerosis above and beyond any effects attributable to the nearby *DRB1* gene. We found no support for effects attributable to *HLA-A* or any of the microsatellite loci previously suggested by other researchers, although such effects cannot be excluded at this time.^{11,13,14,16} It remains possible that the observed association with *HLA-C* is secondary to LD with a nearby but as yet untyped variant. The absence of any residual main effects at *DQB1*, *HLA-B*, and *HLA-A* make it unlikely that the observed association results from these loci despite their strong LD with *HLA-C*. As with all genetic analyses of complex diseases, our

study has a number of limitations that need to be considered when evaluating its conclusions.

Adequate correction for multiple testing is particularly important in studies that involve systematic screening and stratification because the number of tests performed is generally large.⁵² However, calculating an appropriate correction factor can be difficult when there is LD between markers or when subsets (strata) of data are analyzed in addition to total data sets. In these situations, tests are partially correlated rather than fully independent, and simply counting the number of tests performed provides an excessively conservative correction factor. Application of such a crude Bonferroni correction⁵³ runs the risk for inflating the type II error rate unless the sample size used is sufficient to compensate for this conservatism. However, inaccurate attempts to assess the degree of interdependence between tests might underestimate the correction required, thereby resulting in a type I error. Because these plague the genetic analysis of multiple sclerosis, we elected to apply the conservative Bonferroni corrections at each stage. It might be argued that even this approach is insufficient and that we should account for all of the tests performed across the whole study and not just those used at each stage. The evidence for association we identified concerning *HLA-C* would remain significant even if such a project-wide correction factor (approximately 300) were to be applied.

Association testing can be confounded by population stratification and other phenomena that lead to inadequate matching of cases and control subjects. Our use of trio families and transmission disequilibrium testing

Table 5. Haplotype Analysis of *HLA-C* and *HLA-DRB1* in Subsets of the Case-Control (Extension) Cohort Showing the Relative Risk with 95% Confidence Intervals for Each Haplotype

<i>HLA-C</i>	<i>HLA-DRB1</i>	Cases, N (%)	Control Subjects, N (%)	RR	CI
Analysis 1 (excluding all individuals carrying <i>DRB1*03</i> or <i>*0103</i>)					
*R ^a	*X ^a	900.3 (54.7)	2,589.0 (70.3)	1.00	—
*05	*X	75.7 (4.6)	395.7 (10.7)	0.55	0.40–0.76
*R	*1501	634.7 (38.6)	652.7 (17.7)	2.80	2.42–3.23
*05	*1501	35.3 (2.1)	46.3 (1.3)	2.19	1.18–4.10
Analysis 2 (excluding all individuals carrying <i>DRB1*1501</i> or <i>*0103</i>)					
*R ^a	*Y ^a	611.1 (70.1)	2,521.0 (70.8)	1.00	—
*05	*Y	50.9 (5.8)	396.5 (11.1)	0.53	0.37–0.75
*R	*03	182.9 (21.0)	579.5 (16.3)	1.30	1.08–1.57
*05	*03	27.2 (3.1)	64.5 (1.8)	1.74	0.96–3.15
Analysis 3 (excluding all individuals carrying <i>DRB1*1501</i> or <i>*03</i>)					
*R ^a	*Z ^a	514.0 (87.7)	2,121.0 (84.7)	1.00	—
*05	*Z	39.0 (6.7)	332.0 (13.3)	0.49	0.34–0.69
*R	*0103	33.0 (5.6)	51.0 (2.0)	2.96	1.63–4.37
*05	*0103	≈0 (≈0.0)	≈0 (≈0.0)	—	—

*R indicates any *HLA-C* allele except *05, *X any *DRB1* allele except *1501, *Y any *DRB1* allele except *03, and *Z any *DRB1* allele except *0103.

RR = relative risk; CI = confidence interval.

protects against these confounders in the screening phase. However, the extension phase experiments used case-control analysis; therefore, results emerging from these efforts could have been confounded by such errors. The classical loci are highly polymorphic and provide a powerful means to test for stratification. The absence of any difference between the three UK control cohorts or the two sets of UK cases is therefore extremely reassuring in this respect. Perhaps even more reassuring is the absence of any difference between the unrelated UK control cohorts and nontransmitted alleles from the UK trio families, suggesting that our cases are drawn from the same genetic background as the control subjects and making it unlikely that hidden population stratification accounts for our observations. However, replication of the findings in independent cohorts will be necessary to fully exclude this possibility.

Efforts to further replicate our findings will require considerable resources because exclusion of individuals carrying *DRB1* susceptibility alleles means that only 47% of control subjects and 22% of cases will ultimately be informative for the study of *HLA-C*. Large initial cohorts would need to be selected to provide realistic power. The absence of any statistically significant difference in an underpowered study should not be misinterpreted as evidence against this effect. Just as it proved difficult to establish that *DRB1*1501* is responsible for the primary effect in this region, it may prove even more difficult to refine this secondary effect in detail. It remains possible that the observed association is secondary to LD with a nearby but as yet untyped variant. The absence of any residual main effects at *DQB1*, *HLA-B*, and *HLA-A* make it unlikely that the observed association results from these loci despite their strong LD with *HLA-C*.

If *HLA-C* is, in fact, the locus primarily responsible for our observations, this would implicate novel pathways in disease pathogenesis, in particular, the innate immune system. HLA-C molecules, loaded with nonamer peptides, act as ligands for the killer cell immunoglobulin-like receptors (KIRs). KIR receptors contain two or three immunoglobulin-like domains and a short (*KIR2DS*, *KIR3DS*) or long (*KIR2DL*, *KIR3DL*) cytoplasmic tail, corresponding to an activating or inhibitory action on the natural killer cells and T-cell subsets on which they are expressed. Whereas the ligands for activating KIRs remain elusive, and may have relatively weak binding affinities, those for the inhibitory KIRs are well defined. The subset of *HLA-C* alleles with Ser at position 77 and Asn at position 80 (C1 group) bind to the *KIR2DL2* and *KIR2DL3* receptors, whereas those with Asn at position 77 and Lys at position 80 (C2 group) bind to the *KIR2DL1* receptor. *HLA-C* peptide binding specificity may further influence interaction with KIRs.⁵⁴ No evidence for associa-

tion was seen in our data after grouping *HLA-C* alleles according to this functional categorization, suggesting that the protective effect seen for *HLA-C*05* is specific to this allele, and not a consequence of its group function. Alternately, as the effect appears to be allele rather than functional group specific, it may indicate that susceptibility to multiple sclerosis is conferred by an aspect of *HLA-C* function that is independent of its interaction with KIR.

The established association of *HLA-C*06* with susceptibility to psoriasis provides a clear precedent for the involvement of *HLA-C* in complex inflammatory disease. In a Sardinian study of psoriasis, *HLA-C*05* was found to be significantly underrepresented (protective) in patients, although it was not established whether this effect is independent of or secondary to the overrepresentation of *HLA-C*06* and/or haplotypes containing risk alleles at the nearby corneodesmosin (CDSN) gene,⁵⁵ in which our most highly associated SNP, rs3132552, is found. Recently, the presence of the activating *KIR2DS1* and *KIR2DS2* genes was reported to be a novel risk factor for psoriasis and an interaction between *HLA-C* and KIR observed with the overall combination of activating and inhibitory genotypes influencing susceptibility.⁵⁶ Association of combinations between HLA class I and *KIR* genes has also been reported for a range of other infectious and autoimmune diseases.⁵⁴ This provides a strong rationale for investigating the *KIR* gene cluster as a candidate susceptibility locus in multiple sclerosis subsequent to our finding of association with *HLA-C*. The fact that the *KIR* gene cluster lies on chromosome 19 where modest evidence of linkage was observed in the recent high-density screen for linkage in multiple sclerosis²⁴ lends further support. In concordance with these results, this chromosome 19 linkage signal only declares itself in an ordered subset analysis based on those families not showing linkage at MHC, essentially those where the effects of *DRB1*1501* have been excluded.²⁴

In conclusion, we show that the class I gene *HLA-C*, or a locus in tight LD with it, confers additional effects on susceptibility to multiple sclerosis, substantially adding to our understanding of the MHC region in this disease and offering a clear roadmap to further experiments that will refine these observations in larger data sets.

Appendix

The International Multiple Sclerosis Genetics Consortium members are Tai Wai Yeo, MSc,¹ Philip L. De Jager, PhD,²⁻⁴ Simon G. Gregory, PhD,⁵ Lisa F. Barcellos, PhD,^{6,7} Amie Walton, BSc,¹ An Goris, PhD¹ Chiara Fenoglio, PhD,^{1,8} Maria Ban, PhD,¹ Craig J. Taylor, PhD,⁹ Reyna S. Goodman, BSc,⁹ Emily Walsh, PhD,⁴ Cara S. Wolfish, BSc,^{2,4} Roger Horton, MSc,¹⁰ James Traherne, PhD,¹¹ Stephan Beck, PhD,¹⁰ John Trowsdale, PhD,¹¹ Stacy

J. Caillier, BSc,⁶ Adrian J. Ivinson, PhD,^{3,12} Todd Green, BSc,^{4,13} Susan Pobywajlo, MPH,^{2,4} Eric S. Lander, PhD,⁴ Margaret A. Pericak-Vance, PhD,⁵ Jonathan L. Haines, PhD,¹⁴ Mark J. Daly, PhD,^{4,13} Jorge R. Oksenberg, PhD,^{6,15} Stephen L. Hauser, MD,^{6,15} Alastair Compston, PhD,¹ David A. Hafler, MD,²⁻⁴ John D. Rioux, PhD,^{2-4,16} and Stephen Sawcer, PhD¹

This study was supported by the following awards: Wellcome Trust Prize Studentship (T.W.Y.), St. Edmund's College (T.W.Y.), Cambridge Commonwealth Trust and Cambridge Philosophical Society (T.W.Y.), the William C. Fowler scholarship in Multiple Sclerosis (P.L.D.), the National Institutes of Health (K08 NS46341, P.L.D.; NS049477, S.L.H.; NS026799, S.L.H.; NS032830, J.L.H., M.A.P-V.), a GlaxoSmithKline Clinical Fellowship (S.S.), a Postdoctoral Fellowship of the Research Foundation–Flanders (FWO–Vlaanderen, A.G.), a European Neurological Society fellowship (C.F.), Cancer Research Institute fellowship (E.W.), the Medical Research Council (United Kingdom) (G0000648, A.C.), the Wellcome Trust (048880 and 057097, A.C.), a National Multiple Sclerosis Society Center Grant (AP 3758-A-16, D.A.H., A.J.I.), and The Penates Foundation (D.A.H., A.J.I.).

We are grateful to the individuals with multiple sclerosis and their families for making this study possible. We also thank UK Transplant and the 1958 birth cohort for access to their control data, J. Todd for his contribution to the MHC sequencing effort and for allowing us early access to the HLA typing data from the 1958 birth cohort, F. Dudbridge for his help calculating the confidence intervals for the relative risks associated with each haplotype, R. Lincoln for management of the US trios, D. Campbell for providing details of SNPs from the class III region, as well as Medical Research Council Geneservices, which genotyped these in the UK trios.

References

- Jersild C, Svejgaard A, Fog T. HL-A antigens and multiple sclerosis. *Lancet* 1972;1:1240–1241.
- Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* 1991;38:1–15.
- Stewart GJ, Teutsch SM, Castle M, et al. HLA-DR, -DQA1 and -DQB1 associations in Australian multiple sclerosis patients. *Eur J Immunogenet* 1997;24:81–92.
- Marrosu MG, Murru MR, Costa G, et al. DRB1-DQA1-DQB1 loci and multiple sclerosis predisposition in the Sardinian population. *Hum Mol Genet* 1998;7:1235–1237.
- Barcellos LF, Oksenberg JR, Begovich AB, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am J Hum Genet* 2003;72:710–716.
- Dyment DA, Herrera BM, Cader MZ, et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum Mol Genet* 2005;14:2019–2026.
- Olerup O, Carlsson B, Wallin J, et al. Genomic HLA-typing by RFLP-analysis using DR beta and DQ beta cDNA probes reveals normal DR-DQ linkages in patients with multiple sclerosis. *Tissue Antigens* 1987;30:135–138.
- Vartdal F, Sollid LM, Vandvik B, et al. Patients with multiple sclerosis carry DQB1 genes which encode shared polymorphic amino acid sequences. *Hum Immunol* 1989;25:103–110.
- Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet* 2004;5:889–899.
- Compston A, Confavreux C, Lassmann H, et al. *McAlpine's multiple sclerosis*. 4th ed. London: Churchill Livingstone, 2006.
- Marrosu MG, Murru R, Murru MR, et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. *Hum Mol Genet* 2001;10:2907–2916.
- Miretti MM, Walsh EC, Ke X, et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am J Hum Genet* 2005;76:634–646.
- Fogdell-Hahn A, Ligiers A, Gronning M, et al. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens* 2000;55:140–148.
- Rubio JP, Bahlo M, Butzkueven H, et al. Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. *Am J Hum Genet* 2002;70:1125–1137.
- de Jong BA, Huizinga TW, Zanelli E, et al. Evidence for additional genetic risk indicators of relapse-onset MS within the HLA region. *Neurology* 2002;59:549–555.
- Harbo HF, Lie BA, Sawcer S, et al. Genes in the HLA class I region may contribute to the HLA class II-associated genetic susceptibility to multiple sclerosis. *Tissue Antigens* 2004;63:237–247.
- Lincoln MR, Montpetit A, Cader MZ, et al. A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nat Genet* 2005;37:1108–1112.
- Oksenberg JR, Barcellos LF, Cree BA, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet* 2004;74:160–167.
- Koeleman BP, Dudbridge F, Cordell HJ, Todd JA. Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the Conditional Extended Transmission/Disequilibrium Test. *Ann Hum Genet* 2000;64:207–213.
- Barcellos LF, Sawcer S, Ramsay PP, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Hum Mol Genet* 2006;15:2813–2824.
- Rogner UC, Avner P. Congenic mice: cutting tools for complex immune disorders. *Nat Rev Immunol* 2003;3:243–252.
- Haines JL, Terwedow HA, Burgess K, et al. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group. *Hum Mol Genet* 1998;7:1229–1234.
- Ligers A, Dyment DA, Willer CJ, et al. Evidence of linkage with HLA-DR in DRB1*15-negative families with multiple sclerosis. *Am J Hum Genet* 2001;69:900–903.
- International Multiple Sclerosis Genetics Consortium. A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet* 2005;77:454–467.
- Sawcer S, Jones HB, Feakes R, et al. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet* 1996;13:464–468.
- Haines JL, Ter-Minassian M, Bazyk A, et al. A complete genomic screen for multiple sclerosis underscores a role for the major histocompatibility complex. The Multiple Sclerosis Genetics Group. *Nat Genet* 1996;13:469–471.
- Ebers GC, Kukay K, Bulman DE, et al. A full genome search in multiple sclerosis. *Nat Genet* 1996;13:472–476.
- Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516–1517.
- GAMES and the Transatlantic Multiple Sclerosis Genetics Consortium. A meta-analysis of whole genome linkage screens in multiple sclerosis. *J Neuroimmunol* 2003;143:39–46.
- Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;13:227–231.

31. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50:121–127.
32. Stewart CA, Horton R, Allcock RJ, et al. Complete MHC haplotype sequencing for common disease gene mapping. *Genome Res* 2004;14:1176–1187.
33. Traherne JA, Horton R, Roberts AN, et al. Genetic analysis of completely sequenced disease-associated MHC haplotypes identifies shuffling of segments in recent human history. *PLoS Genet* 2006;2:e9.
34. Walsh EC, Mather KA, Schaffner SF, et al. An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 2003;73:580–590.
35. Florez JC, Burtz N, de Bakker PI, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* 2004;53:1360–1368.
36. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc Natl Acad Sci U S A* 1991;88:7276–7280.
37. Tamiya G, Ota M, Katsuyama Y, et al. Twenty-six new polymorphic microsatellite markers around the HLA-B, -C and -E loci in the human MHC class I region. *Tissue Antigens* 1998;51:337–346.
38. Tamiya G, Shiina T, Oka A, et al. New polymorphic microsatellite markers in the human MHC class I region. *Tissue Antigens* 1999;54:221–228.
39. Foissac A, Salhi M, Cambon-Thomsen A. Microsatellites in the HLA region: 1999 update. *Tissue Antigens* 2000;55:477–509.
40. Matsuzaka Y, Makino S, Nakajima K, et al. New polymorphic microsatellite markers in the human MHC class II region. *Tissue Antigens* 2000;56:492–500.
41. Matsuzaka Y, Makino S, Nakajima K, et al. New polymorphic microsatellite markers in the human MHC class III region. *Tissue Antigens* 2001;57:397–404.
42. Schuler GD. Sequence mapping by electronic PCR. *Genome Res* 1997;7:541–550.
43. Rajalingam R, Krausa P, Shilling HG, et al. Distinctive KIR and HLA diversity in a panel of north Indian Hindus. *Immunogenetics* 2002;53:1009–1019.
44. Olerup O, 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;39:225–235.
45. Coraddu F, Reyes-Yanez MP, Parra A, et al. HLA associations with multiple sclerosis in the Canary Islands. *J Neuroimmunol* 1998;87:130–135.
46. Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003;25:115–121.
47. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;63:259–266.
48. Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005;21:3445–3447.
49. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101.
50. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38:209–213.
51. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002;70:124–141.
52. Freimer N, Sabatti C. The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nat Genet* 2004;36:1045–1051.
53. Bonferroni CE. *Teoria statistica delle classi e calcolo delle probabilita*. 8 vol: Pubblicazioni del R. Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936:3–62.
54. Rajagopalan S, Long EO. Understanding how combinations of HLA and KIR genes influence disease. *J Exp Med* 2005;201:1025–1029.
55. Orru S, Giuressi E, Carcassi C, et al. Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). *Am J Hum Genet* 2005;76:164–171.
56. Nelson GW, Martin MP, Gladman D, et al. Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. *J Immunol* 2004;173:4273–4276.
57. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149–150.