Evidence for *CRHR1* in multiple sclerosis using supervised machine learning and meta-analysis in 12,566 individuals

Farren B.S. Briggs¹, Selena E. Bartlett², Benjamin A. Goldstein¹,3, Joanne Wang⁴,5, Jacob L. McCauley⁶, Rebecca L. Zuvich⁷, Philip L. De Jager⁸,9, John D. Rioux¹⁰, Adrian J. Ivinson¹¹, Alastair Compston¹², David A. Hafler⁸,¹³, Stephen L. Hauser⁴,5, Jorge R. Oksenberg⁴,5, Stephen J. Sawcer¹², Margaret A. Pericak-Vance⁶, Jonathan L. Haines⁷, International Multiple Sclerosis Genetics Consortium and Lisa F. Barcellos¹,*

¹Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, CA 94720-7356, USA, ²Ernest Gallo Clinic and Research Center, ³Department of Neurology, School of Medicine and ⁴Institute for Human Genetics, School of Medicine, University of California, San Francisco, CA 94143-0435, USA, ⁵Division of Biostatistics, School of Public Health, University of California, Berkeley, CA 94720-7356, USA, ⁶John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA, ⁷Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37232-0700, USA, ⁸Program in NeuroPsychiatric Genomics, Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital, Boston, MA 02115, USA, ⁹Program in Medical and Population Genetics, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA 02139, USA, ¹⁰Laboratory in Genetics and Genomic Medicine of Inflammation, Montreal Heart Institute, Université de Montréal, Montreal, Quebec, Canada H1T1C8, ¹¹Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA 02155, USA, ¹²Department of Clinical Neurosciences, Addenbrooke's Hospital, University of Cambridge, PO Box 165, Hills Road, Cambridge CB2 2QQ, UK and ¹³Department of Neurology, Yale School of Medicine, New Haven, CT 06520-8018, USA

Received May 3, 2010; Revised July 12, 2010; Accepted July 29, 2010

The primary genetic risk factor in multiple sclerosis (MS) is the *HLA-DRB1* 1501 allele; however, much of the remaining genetic contribution to MS has yet to be elucidated. Several lines of evidence support a role for neuroendocrine system involvement in autoimmunity which may, in part, be genetically determined. Here, we comprehensively investigated variation within eight candidate hypothalamic–pituitary–adrenal (HPA) axis genes and susceptibility to MS. A total of 326 SNPs were investigated in a discovery dataset of 1343 MS cases and 1379 healthy controls of European ancestry using a multi-analytical strategy. Random Forests, a supervised machine-learning algorithm, identified eight intronic SNPs within the corticotrophin-releasing hormone receptor 1 or *CRHR1* locus on 17q21.31 as important predictors of MS. On the basis of univariate analyses, six *CRHR1* variants were associated with decreased risk for disease following a conservative correction for multiple tests. Independent replication was observed for *CRHR1* in a large meta-analysis comprising 2624 MS cases and 7220 healthy controls of European ancestry. Results from a combined meta-analysis of all 3967 MS cases and 8599 controls provide strong evidence for the involvement of *CRHR1* in MS. The strongest association was observed for rs242936 (OR = 0.82, 95% CI = 0.74–0.90, *P* = 9.7 × 10⁻⁵). Replicated *CRHR1* variants appear to exist on a single associated haplotype. Further investigation of mechanisms involved in HPA axis regulation and response to stress in MS pathogenesis is warranted.

*To whom correspondence should be addressed at: Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, 209 Hildebrand Hall, University of California, Berkeley, CA 94720, USA. Tel: +1 5106427814; Fax: +1 5106430580; Email: barcello@genepi.berkeley.edu

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
INTRODUCTION

Multiple sclerosis (MS; MIM no. 126200) is a clinically heterogeneous, autoimmune disease of the central nervous system including two distinct, though intersecting, neuropathological phases: inflammation and neurodegeneration (1). The pathogenesis of MS involves a substantial genetic component, with HLA class II genes within the major histocompatibility complex (MHC) on chromosome 6p21 conferring ~50% of the genetic risk (2). Admixture analyses have shown the primary susceptibility locus for MS within the MHC is HLA-DRB1, and more specifically, the HLA-DRB1*15 allele (3,4). The complexities surrounding the HLA-DRB1 contribution are the focus of multiple ongoing research efforts within the field (5–7). The identification of non-MHC susceptibility loci, while progressing, is far from complete. Recent genome-wide association (GWA) and replication studies have begun to unravel the polygenic etiology of MS. Modest associations for over a dozen variants, including SNPs within IL2RA, IL2RA, CLEC16A, CD58, TNFRSF1A, IRF8, KIF21B and TMEM39A, have been described (8–14). To date, GWA studies using currently available technology have had limited success refining the genetic etiology for most complex diseases, including MS (15,16). Candidate gene approaches based on strong hypotheses and well-powered datasets still remain an important strategy for detection of disease-associated variants (17,18).

MS pathogenesis is thought to involve multiple biological pathways that contribute to inflammatory and neurodegenerative components of the disease. Genetic variation within these pathways is therefore likely to be associated with disease predisposition. One biological pathway involved in MS is the hypothalamic–pituitary–adrenal (HPA) axis, a principal component of the neuroendocrine system that regulates individual response to physical and emotional stress, and maintains homeostasis with strong neuroimmune-modulating properties (19,20). Various measures of HPA axis activity have been investigated in small studies of clinically heterogeneous MS cases, where impaired HPA axis activity in MS cases appears to be dependent on clinical phenotypes and disease duration (21). A recent comprehensive analysis demonstrated strong evidence for HPA hyperactivity in 173 well-defined MS patients and 60 healthy controls (22).

The HPA axis is further implicated in pathology of MS by evidence supporting psychological stress as a risk factor for MS onset and exacerbations in several studies, despite methodological differences (23). MS cases are more likely to have experienced stressful life events prior to symptom onset than matched healthy controls for the same prodromal time period (24,25) or controls with other neurologic or rheumatic diseases (26). Several studies suggest that stressful life events often precede disease relapse in MS cases (27–33); and interestingly, MS cases that experienced family or work conflicts have an increased risk of developing a new gadolinium-enhancing (Gd+) MRI lesion (27). Psychological stress has also been associated with both susceptibility and disease progression for other autoimmune diseases (34–39). In addition, a recent retrospective cohort study examined hospital discharge records and reported significant associations between individuals who experience traumatic childhood stress and increased rates of first hospitalizations for any of 21 autoimmune diseases, including MS (40).

Although evidence supports the involvement of the HPA axis in the pathology of MS, the exact nature of the relationship and the underlying biological mechanism(s) remain unclear. In this study, we investigated variation within eight candidate genes involved in mediating and/or regulating the neuroendocrine activity of the HPA axis using a supervised machine-learning algorithm (Random Forests). A total of 326 informative SNPs within BDNF, CRHBP, corticotrophin-releasing hormone receptor 1 (CRHR1), CRHR2, HCRTR1, HCRTR2, OPRD1 and OPRK1 were genotyped in 1343 MS cases and 1379 healthy controls of European ancestry. Random Forests was first used to rank important SNP predictors of case status. Univariate association tests, haplotype and logistic regression analyses were then used to further characterize relationships between important variants and risk for MS. We confirmed our significant associations in a large meta-analysis of 2624 MS and 7220 control subjects of European ancestry (13). We report for the first time CRHR1 as a candidate locus for MS susceptibility based on results from a multi-stage analysis (Fig. 1).

RESULTS

Stage I: discovery analysis

A discovery dataset was composed of 1343 MS cases and 1379 healthy controls of European ancestry (14). The rs3135388 (A/G) SNP was used to determine the presence of the HLA-DRB1*1501 allele in MS cases and controls as described previously (8). As expected, the A variant was significantly associated with MS risk [odds ratio (OR) = 2.7, 95% confidence interval (CI): 2.4–3.1, \( P = 7.8 \times 10^{-49} \)] and demonstrates that our case–control sample is representative of those described in other studies. This SNP had a risk allele frequency of 29.8% in cases and 13.6% in controls. Using this dataset, Random Forests identified eight intronic SNPs within CRHR1 on chromosome 17q21.31 as important predictors of MS (Table 1 and Fig. 2). Six of these eight CRHR1 SNPs were significantly associated with decreased risk of MS in logistic regression models (\( P_{\text{corrected}} < 0.05 \))(Table 1), adjusted for gender and HLA-DRB1*1501. A false discovery rate (Benjamini–Hochberg) (FDR-BH) correction for multiple testing (eight tests) was applied. All eight SNPs demonstrated similar magnitude and direction of effect.

Stage II: replication analysis

Four of the eight CRHR1 SNPs identified in Stage I were examined in an independent group of 2624 MS and 7220 control subjects of European ancestry in an attempt to replicate our Stage I results (Table 1). Data from MS cases and controls were composed of three GWA scans generated on different SNP arrays, and subsequently each scan was imputed using a single panel of 2.56 million SNPs for meta-analysis, as described previously (13). Meta-analysis for CRHR1 SNPs was conducted using a fixed-effects logistic model based on the observed and expected allele dosage, and adjusted for cohort of origin, gender and HLA-DRB1*1501 (13). Notably, all four SNPs within CRHR1 were also
significantly associated with decreased risk of MS in this independent dataset (Table 1; rs242939: OR = 0.82, 95% CI = 0.71–0.95, \( P = 0.0078 \); rs242936: OR = 0.82, 95% CI = 0.73–0.93, \( P = 0.0024 \); rs173365: OR = 0.89, 95% CI = 0.83–0.96, \( P = 0.0029 \); rs17689966: OR = 0.89, 95% CI = 0.83–0.96, \( P = 0.0030 \)).

**Stage III:** combined analysis

The data from Stages I and II were combined to generate final effect estimates for four of the eight CRHR1 SNPs. The meta-analysis from Stage II was rerun, including the Stage I data as an additional cohort. In the combined meta-analysis of 3967 MS cases and 8599 controls subjects of European ancestry, the four CRHR1 variants were significantly associated with decreased risk of MS, and the association was stronger (\( P < 5 \times 10^{-4} \)) when compared with results from Stage I or Stage II (Table 2; rs242939: OR = 0.81, 95% CI = 0.72–0.91, \( P = 4.1 \times 10^{-4} \); rs242936: OR = 0.82, 95% CI = 0.74–0.90, \( P = 9.7 \times 10^{-5} \); rs173365: OR = 0.89, 95% CI = 0.84–0.95, \( P = 1.9 \times 10^{-4} \); rs17689966: OR = 0.89, 95% CI = 0.84–0.95, \( P = 3 \times 10^{-4} \)).

Table 1. CRHR1 SNP variants identified as important predictors of MS by Random Forests in Stage I analysis

<table>
<thead>
<tr>
<th>VI rank</th>
<th>SNP</th>
<th>Chr Base-pair location</th>
<th>Function</th>
<th>Minor allele</th>
<th>Stage I MAF cases</th>
<th>MAF controls</th>
<th>( P )-value</th>
<th>Adj. ( P )-value</th>
<th>OR (95% CI)</th>
<th>Stage II ( P )-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>rs12940065</td>
<td>17 41221636</td>
<td>Intron 1</td>
<td>A</td>
<td>0.16</td>
<td>0.17</td>
<td>0.45</td>
<td>0.45</td>
<td>0.94 (0.81–1.09)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>rs34174655</td>
<td>17 41231188</td>
<td>Intron 1</td>
<td>A</td>
<td>0.08</td>
<td>0.10</td>
<td>0.0031</td>
<td>0.024</td>
<td>0.74 (0.61–0.90)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>rs171442</td>
<td>17 41247686</td>
<td>Intron 2</td>
<td>A</td>
<td>0.06</td>
<td>0.08</td>
<td>0.0091</td>
<td>0.024</td>
<td>0.74 (0.59–0.93)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>rs242939</td>
<td>17 41251360</td>
<td>Intron 3</td>
<td>C</td>
<td>0.06</td>
<td>0.08</td>
<td>0.016</td>
<td>0.030</td>
<td>0.76 (0.61–0.95)</td>
<td>0.0078</td>
<td>0.82 (0.71–0.95)</td>
</tr>
<tr>
<td>1</td>
<td>rs242938</td>
<td>17 41251717</td>
<td>Intron 3</td>
<td>A</td>
<td>0.06</td>
<td>0.08</td>
<td>0.0074</td>
<td>0.024</td>
<td>0.74 (0.60–0.92)</td>
<td>0.0024</td>
<td>0.82 (0.73–0.93)</td>
</tr>
<tr>
<td>7</td>
<td>rs242936</td>
<td>17 41254990</td>
<td>Intron 4</td>
<td>A</td>
<td>0.10</td>
<td>0.12</td>
<td>0.019</td>
<td>0.030</td>
<td>0.81 (0.68–0.97)</td>
<td>0.0024</td>
<td>0.82 (0.73–0.93)</td>
</tr>
<tr>
<td>4</td>
<td>rs173365</td>
<td>17 41256855</td>
<td>Intron 4</td>
<td>A</td>
<td>0.43</td>
<td>0.46</td>
<td>0.030</td>
<td>0.040</td>
<td>0.89 (0.79–0.99)</td>
<td>0.0029</td>
<td>0.89 (0.83–0.96)</td>
</tr>
<tr>
<td>5</td>
<td>rs17689966</td>
<td>17 41266236</td>
<td>Intron 8</td>
<td>G</td>
<td>0.43</td>
<td>0.46</td>
<td>0.051</td>
<td>0.059</td>
<td>0.90 (0.80–1.00)</td>
<td>0.0030</td>
<td>0.89 (0.83–0.96)</td>
</tr>
</tbody>
</table>

VI, variable importance; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

*Important (top-ranking) predictors from the Random Forests analysis were determined based on the distribution of the VI scores (Fig. 2).

*All alleles and base-pair location are specified with respect to the forward (+) strand of the National Center for Biotechnology Information’s Build 36 (University of California at Santa Cruz Genome Browser).

*FDR-BH corrected \( P \)-value (for eight tests) for each SNP variant from a logistic regression model where the variant is coded to reflect genotypes (0, 1 and 2, where 0 represents homozygous dominant genotype) and HLA-DRB1*1501 and gender were included in the model (PLINK v1.06).

*Adjusted proportional ORs from logistic regression models with HLA-DRB1*1501 and gender in each model (PLINK v1.06).

*Fixed-effects meta-analysis of three GWA scans for MS susceptibility based on the observed (imputed) and expected alleles dosage of each SNP, taking into account the empirically observed variance of the allele dosage, was conducted using logistic regression models, adjusted for location, HLA-DRB1*1501 and gender (STATA v9.0).
haplotype block spanning nearly the entire CRHR1 gene (46 kb; Supplementary Material, Fig. S1) and including 81 SNPs was observed. H2 was identified by the presence of the rs1396862T variant; with a prevalence of 22.9%, comparable to frequencies observed in European Americans and most western, central and southeastern Europeans (42). A total of 13 sub-haplotypes with a frequency ≥1% were observed, including 11 H1 and 2 H2 sub-haplotypes (Fig. 5). The H1.5 sub-haplotype frequency differed in MS cases and healthy controls (Fig. 5 and Table 3); it was significantly associated with decreased risk of MS when compared with all other H1 sub-haplotypes (OR = 0.64, 95% CI: 0.50–0.83, P = 4.1 × 10⁻⁴). Results did not differ when compared with all other H1 and H2 haplotypes (data not shown). Interestingly, H1.5 was the only sub-haplotype with all eight SNPs identified in Stage 1, and thus, concurred with results obtained from Random Forests analyses (Fig. 5).

**DISCUSSION**

Results from the current study of variants within eight candidate genes involved in mediating and/or regulating the neuroendocrine activity of the HPA axis demonstrate strong evidence for association between CRHR1 (GeneID 1394) and MS susceptibility. A large, well-powered study design, including both discovery and replication stages, was utilized. European ancestry estimates based on genetic markers were used to remove population outliers for case–control datasets in both stages; therefore, the potential impact of population stratification was minimized. The 17q21 region has been a key locus of interest in MS: it was identified as a possible MS locus in a meta-analysis of several genome-wide linkage screens (45). However, subsequent published GWA scans in MS, which to date are smaller than the current study, have not successfully identified CRHR1 or any other susceptibility locus in this region (8,46,47).

The neuroendocrine, autonomic and behavioral stress response is centralized to a common corticotropin-releasing-hormone (CRH)-mediated mechanism (48). CRH is released by the hypothalamus, and delivered to the anterior pituitary where it binds to CRHR1 receptors. This interaction of CRH and the G protein-coupled CRHR1 receptors at the pituitary is exclusively responsible for activating the biosynthesis and release of the adrenocorticotropic hormone, which culminates in steriodogenesis and the release of glucocorticoids (GCs; i.e. cortisol) from the adrenal gland (49). GC receptors (expressed by all nucleated cells) transduce the stress response signal in end organs, directly affecting transcription of GC-sensitive genes, with activational and/or inhibitory actions in various systems, including negative feedback at the pituitary to stop the response to stressful stimuli. CRH is also found peripherally, and regulates inflammation through direct activation of CRHR1 receptors on mast cells and monocytes/macrophages and via in vivo secretion of TNF, IL-1 and IL-6 (50–53). In addition, CRHR1 is expressed on resting human B and T lymphocytes and several other immune-related cells and tissues (54–60). Furthermore, urocortin, a CRHR1 ligand structurally similar to CRH, promotes microvessel permeability, and CRHR1 is significantly overexpressed in early stages of MS pathology (61). The functional assessment of the identified H1.5 alleles in cell lines and in vivo model systems would provide valuable information regarding the mechanisms of these variants in MS susceptibility.
development at the blood–brain barrier in animal models (61,62). A total of eight CRHR1 isoforms have been described and have varied distribution and potentially function in various biological mechanisms, as well as complex responsiveness to CRHR1 ligands (52,63,64). For example, signaling of the primary CRHR1 isoform (CRHR1a) is attenuated or amplified by other soluble isoforms (65).

Close to 100% of the common genetic variation in CRHR1 was captured in Stage I analyses, based on r² ≥ 0.95 in CEU Hapmap data (Release no. 21). The eight SNPs identified by Random Forests as important were located in introns; therefore, an explicit functional role is not evident. However, in silico analysis (http://fastsnp.ibms.sinica.edu.tw/) suggests four of the eight important SNPs (rs171442, rs242938, rs173365 and rs17689966) are potential intronic enhancers that may alter transcription factor binding. Interestingly, the eight important SNPs identified by Random Forest reside on a single haplotype (H1.5). H1.5 also carried the minor alleles for 11 other SNPs variants, 10 of which had P<uncorrected> 0.05 in our analysis, and included another potential intronic enhancer (rs16940646). Not all of the CRHR1 variants of interest were available in the replication dataset, including the three SNPs that were most significant in the discovery dataset (rs34174655, rs171442 and rs242938). Future investigation of the H1.5 haplotype is necessary to identify CRHR1 isoforms that are altered as a result of SNP(s) on this haplotype and/or the impact of these variants on transcription factor binding.

CRHR1 is located within a large region (∼900 kb) of high linkage disequilibrium on chromosome 17q21.31, resulting from a local chromosomal inversion. The global distribution of two 17q21.31 haplotypes, H1 and H2 (the inversion), varies across Europe (42). Despite using homogeneous study populations of European ancestry, we repeated all analyses

Table 2. Replicated SNP variants (Table 1) in the combined dataset

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Base-pair location</th>
<th>Gene</th>
<th>Function</th>
<th>Minor allele</th>
<th>Stage II P-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs242939</td>
<td>17</td>
<td>41 251 360</td>
<td>CRHR1</td>
<td>Intron 3</td>
<td>C</td>
<td>4.1 \times 10^{-4}</td>
<td>0.81 (0.72–0.91)</td>
</tr>
<tr>
<td>rs242936</td>
<td>17</td>
<td>41 254 990</td>
<td>CRHR1</td>
<td>Intron 4</td>
<td>A</td>
<td>9.7 \times 10^{-5}</td>
<td>0.82 (0.74–0.90)</td>
</tr>
<tr>
<td>rs173365</td>
<td>17</td>
<td>41 256 855</td>
<td>CRHR1</td>
<td>Intron 4</td>
<td>A</td>
<td>1.9 \times 10^{-4}</td>
<td>0.89 (0.84–0.95)</td>
</tr>
<tr>
<td>rs17689966</td>
<td>17</td>
<td>41 266 236</td>
<td>CRHR1</td>
<td>Intron 8</td>
<td>G</td>
<td>3.0 \times 10^{-4}</td>
<td>0.89 (0.84–0.95)</td>
</tr>
</tbody>
</table>

*Fixed-effects meta-analysis, of the discovery dataset in Stage I and the three GWA scans in Stage II, for MS susceptibility based on the observed (imputed) and expected alleles dosage of each SNP, taking into account the empirically observed variance of the allele dosage, was conducted using logistic regression models, adjusted for location, HLA-DRB1∗1501 and gender (STATA v9.0).
of CRHR1 with adjustment for the presence of the inversion, and also excluded individuals with the inversion from analyses, to address the presence of unmeasured population structure; results did not vary. Similar to findings reported herein, a previous investigation of H2 haplotype-tagging SNP rs9468 in 937 UK trio families showed no evidence for association (66). It is important to note that some CRHR1 variants identified by Random Forests in Stage I also tagged ($r^2 > 0.5$) SNP variants within the nearby IMP5 locus, based on assessment of linkage disequilibrium in CEU Hapmap data. Complete genotype data for variants within this locus were not available for MS cases and controls utilized in the current study. Therefore, detailed characterization of genetic variation within the CRHR1-IMP5 region is needed. However, our findings based on association with replication strongly support a role for CRHR1 variation in MS susceptibility.

The application of Random Forests as a discovery tool in Stage I had several advantages compared with standard uni-
The strengths of this investigation are: (i) genotyping of a large homogenous study population that was well powered to identify modest genetic effects; (ii) the application of a robust non-parametric algorithm and conservative adjustment for multiple testing to aid in the interpretation of association results; (iii) the replication of findings in a large, independent study population; and (iv) extensive genotypic coverage of investigated loci. A key limitation of this investigation is that genetic data for other key HPA axis genes were not available; therefore, it is necessary to investigate these unexplored relationships in MS, and in the context of the variants investigated here.

CRHR1 is a critical component of the HPA axis. An impaired HPA axis has been suspected to contribute to autoimmunity, including MS (20,67–69), and genetic variation within related genes has been associated with the development of affective disorders and other stress-related clinical conditions, both marginally and through gene × environment interactions (70–73). Thus, further investigations of these clinical conditions, exposure to stressful life events and genetic variation with the HPA axis in MS are necessary. Results from this investigation provide evidence for a non-MHC-mediated mechanism in the pathogenesis of MS. Our study underscores the importance of considering biological function in genetic studies of MS, as well as other autoimmune disorders.

MATERIALS AND METHODS

Stage I: discovery analysis

Ten genes involved in mediating and/or regulating the neuroendocrine activity of the HPA axis were selected for investigation. Genic SNPs were selected for genotyping primarily based on their function as a tagging SNP. Using Tagger, as implemented in Haplovew v3.3 (44), tagging SNPs were defined as a set of SNPs which captured 100% of the genetic variation in CEU Hapmap data (NCBI Build 35; Release no. 21, July 2006), with pairwise \( r^2 \geq 0.95 \). Additional SNPs not available in CEU Hapmap data were identified and selected from dbSNP (retrieved July 2007) based on proximity to exons and to expand coverage where there was none (e.g. there were no CRH SNPs available in CEU Hapmap data (Release no. 21); therefore, four SNPs were selected from dbSNP for genotyping).

A total of 486 SNPs in 10 genes were genotyped as a subset of 48 767 custom SNPs using the Illumina Infinium 60K BeadChip assay (74) in 2961 (1488 MS cases and 1512 controls) participants recruited from three clinical centers (University of California, San Francisco; Harvard/MIT Board Institute; and Cambridge University) (14). Unrelated controls were obtained from these same US sites and from the British 1958 Birth Cohort Study. These controls were selected to provide nearly equivalent gender and age distributions (14). All participants self-reported as non-Hispanic whites. All MS cases met well-established disease criteria (75,76). Informed consent was obtained from all study participants, and approvals from local institutional review boards were secured at each recruitment site prior to enrollment.

A rigorous quality control protocol was utilized; Whole-genome Association Study Pipeline assessed sample and SNP genotyping efficiency (<95%), allele frequencies, gender errors and Hardy–Weinberg equilibrium (\( P < 0.0001 \), recursively. Population stratification was investigated using both STRUCTURE (77) and EIGENSTRAT (78) software programs. Significant population stratification was not observed in the overall dataset; however, 79 samples with probability of Caucasian European descent <0.90 were excluded from our analyses (14). The final dataset for analysis yielded 2722 individuals and 46 874 SNPs, for which 380.

Table 3. Association of CRHR1 H1 haplotypes with MS

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Case counts</th>
<th>Control counts</th>
<th>P-value (^a)</th>
<th>OR (95% CI) (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1.1</td>
<td>410.7</td>
<td>380.4</td>
<td>0.190</td>
<td>1.11 (0.95–1.30)</td>
</tr>
<tr>
<td>H1.2</td>
<td>361.5</td>
<td>359.7</td>
<td>0.869</td>
<td>1.02 (0.86–1.20)</td>
</tr>
<tr>
<td>H1.3</td>
<td>330.0</td>
<td>304.0</td>
<td>0.226</td>
<td>1.11 (0.94–1.32)</td>
</tr>
<tr>
<td>H1.4</td>
<td>235.5</td>
<td>262.0</td>
<td>0.251</td>
<td>0.90 (0.74–1.08)</td>
</tr>
<tr>
<td>H1.5</td>
<td>115.8</td>
<td>175.6</td>
<td>4.1 ( \times 10^{-4} )</td>
<td>0.64 (0.50–0.83)</td>
</tr>
<tr>
<td>H1.6</td>
<td>125.3</td>
<td>123.4</td>
<td>0.896</td>
<td>1.03 (0.79–1.34)</td>
</tr>
<tr>
<td>H1.7</td>
<td>95.0</td>
<td>94.1</td>
<td>0.941</td>
<td>1.02 (0.75–1.38)</td>
</tr>
<tr>
<td>H1.8</td>
<td>84.8</td>
<td>78.2</td>
<td>0.576</td>
<td>1.10 (0.80–1.53)</td>
</tr>
<tr>
<td>H1.9</td>
<td>71.8</td>
<td>83.4</td>
<td>0.413</td>
<td>0.87 (0.62–1.21)</td>
</tr>
<tr>
<td>H1.10</td>
<td>52.9</td>
<td>42.6</td>
<td>0.303</td>
<td>1.25 (0.81–1.92)</td>
</tr>
<tr>
<td>H1.11</td>
<td>31.0</td>
<td>24.2</td>
<td>0.345</td>
<td>1.31 (0.74–2.34)</td>
</tr>
<tr>
<td>Total H1</td>
<td>1993.4</td>
<td>2008.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Significance was determined using Fisher’s exact test (Stata v9.2).

\(^b\)Reference is all other H1 haplotypes.
SNPs were relevant to this analysis. We further excluded 54 SNPs with an MAF <0.01. Missing genotypes were imputed in cases and controls using Beagle v2.1.3 (79). Therefore, a total of 326 SNPs within eight genes related to the HPA axis were investigated in 1343 MS cases and 1379 healthy controls of European ancestry. The rs3135388 (A/G) SNP was used to determine the presence of the HLA-DRB1*1501 allele as described previously (80).

We investigated the power to detect log-additive genetic associations. ORs ranging from 0.1 to 3.0 were examined, assuming a very conservative two-sided type I error of 0.1% (α = 0.001, respectively). Results indicated that our discovery analysis was sufficiently powered (>75%) to detect allelic OR ≤0.7 and ≥1.3 for almost all models considered.

All SNP genotypes in cases and controls were coded as 0, 1 or 2 copies of the minor allele, and investigated using Random Forests v6.40.179 (http://salford-systems.com/) with mtry = sqrt(p) and ntree = 5000. Briefly, Random Forests independently grows a collection of recursively partitioned trees without pruning using bootstrap aggregating and random selection of a subset of all predictors to determine classification at each node (81). Once the forest is grown, each predictor is randomly permuted across all trees and used to generate a variable importance (VI) score. The VI scores rank predictors by their importance in classifying the outcome in the context of all predictors without model specification, and are robust to uninformative predictors and outliers. It has been suggested that VI measures may be affected by LD, though reports on this topic are not consistent (82–85). We evaluated the impact of LD on current results using two approaches. First, the dataset was pruned based on LD, using r² of 99, 90 and 80%, as outlined in Goldstein et al. (86). Second, the VI scores were recalculated based on the number of trees in which the predictor occurred, as opposed to the entire forest (84). For both approaches, the top predictors overlapped greatly (results not shown), suggesting only a minimal impact of LD in this analysis. Therefore, we presented results derived from Random Forests analysis using the original dataset.

Important (top-ranking) predictors from the Random Forests analysis were determined based on the distribution of the VI scores. Allele frequencies were compared between MS cases and controls, and ORs and 95% CI were determined using unconditional logistic regression models adjusted for gender and HLA-DRB1*1501 using STATA v9.2 (StataCorp LP, College Station, TX, USA). Four CRHR1 SNPs from Stage I were not available in Stage II for replication.

Stage III: combined analysis

A meta-analysis combining the genotypic data from Stage I and Stage II (3967 MS and 8599 control subjects) determined final effect size estimates using a fixed-effects logistic model based on the observed and expected allele dosage, adjusted for cohort of origin (n = 7), gender and HLA-DRB1*1501 using STATA v9.2.

Stage IV: extension analysis

Analysis of linkage disequilibrium was conducted using CEU Hapmap data (version 3; Release no. R2) to identify chromosomal regions tagged by the important predictors using Haploview v4.2 (44). Genetic variation within the tagged chromosomal regions was also investigated in Stage I, Stage II and Stage III datasets using the appropriate logistic regression models as described previously. Logistic regression models were rerun to include the 17q21.31 inversion-tagging variant (rs1396862) as an independent variable.

Using Stage I data, haplotype blocks were constructed using D' CIs (43), and frequencies determined using Haplovie v4.2. Haplotype frequencies were rounded to the nearest whole number and compared using Fisher's exact test (cci; STATA v9.2).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We would like to thank the International MS Genetics Consortium (see https://www.imsgc.org/ for list of all members) for collaborating on this investigation. We would also like to thank Dr Darlene Francis, Paola Grasso-Bronson and Xiaorong Shao for helpful discussions.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National MS Society (RG4201-A-1, RG2901), Medical Research Council (G0700061), Cambridge NIHR Biomedical Research Centre and grants R01 NS049477 (NIH/NINDS), R01 AI059829 and U19 AI067152 (NIH/NIAID).

REFERENCES


