Human Molecular Genetics, 2010, Vol. 19, No. 5 **953–962**
doi:10.1093/hmg/ddp542
Advance Access published on December 9, 2009

Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies *KIF21B* and *TMEM39A* as susceptibility loci

The International Multiple Sclerosis Genetics Consortium (IMSGC)*

Received September 17, 2009; Revised November 23, 2009; Accepted December 3, 2009

Genome-wide association studies (GWASs) have proven highly effective, identifying hundreds of associations across numerous complex diseases. These studies typically test hundreds of thousands of variations and identify hundreds of potential associations. However, to date, follow-up attempts have generally only concentrated on just the few most significant initial associations, leaving the majority of true associations in any GWAS study without replication. Here, we present a substantially more comprehensive follow-up of the first genome-wide association screen performed in multiple sclerosis (MS), a complex genetic disease with central nervous system inflammation. We genotyped approximately 30 000 single-nucleotide polymorphisms (SNPs) that demonstrated mild-to-moderate levels of significance ($P \leq 0.10$) in the initial GWAS in an independent set of 1343 MS cases and 1379 controls. We further replicated several of the most significant findings in another independent data set of 2164 MS cases and 2016 controls. We find considerable evidence for a number of novel susceptibility loci including *KIF21B* [rs12122721, combined $P = 6.56 \times 10^{-10}$, odds ratio (OR) = 1.22] and *TMEM39A* (rs1132200, $P = 3.09 \times 10^{-8}$, OR = 1.24), both of which meet genome-wide significance. Both of these loci were overlooked in the initial replication, despite being among the top 3000 (∼1%) SNP hits in the original screen.

INTRODUCTION

Multiple sclerosis (MS, MIM 126200) is an inflammatory, demyelinating disease of the central nervous system (CNS), thought to be mediated by an autoimmune process. It affects over 2 million individuals world-wide. The disease is characterized by mononuclear cell infiltration in the CNS associated with demyelination leading to a spectrum of symptoms and disability within affected individuals. MS is most common in young adults and affects women two to three times more frequently than men. Family and twin studies have long shown evidence for a strong genetic component underlying the etiology of MS. Until recently, the major histocompatibility complex (MHC) was the only universally accepted genetic locus associated with MS.

In 2007, we reported the first genome-wide association study (GWAS) for MS susceptibility. In this GWAS, we screened 931 trio families (an affected individual and both parents) with 334 923 single-nucleotide polymorphisms (SNPs) and followed-up 110 of the most promising associations in additional cases ($n = 2322$), controls ($n = 5418$) and trio families ($n = 609$). This first-pass follow-up resulted in the identification of three strongly associated SNPs outside of the MHC, namely rs6897932 in the interleukin-7 receptor α gene (*IL7RA)* and both rs12722489 and rs2104286 within the interleukin-2 receptor α gene (*IL2RA)* (1). These associations were replicated by a number of groups (2–5) and further refined in subsequent analyses (6). The GWAS also identified highly suggestive associations with variations in *CLEC16A* and *CD58*, both of which have subsequently been confirmed, along with other genes identified through additional MS GWAS and restricted follow-up efforts (e.g. *TNFRSF1A, IRF8, CD6, TYK2, CD226 and CYB27B1*) (7–14). These genes are now the focus of multiple ongoing studies to confirm and understand their potential involvement in MS susceptibility.

Statistically we would expect the pool of moderately significant GWAS results to be enriched for genuine associations. To more comprehensively test for additional MS-associated loci, we examined approximately 30 000 SNPs, whose initial

*To whom correspondence should be addressed: Jacob L. McCauley, John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, 1501 NW 10th Avenue, BRB-307 (M-860), Miami, FL 33136, USA. Tel: +1 3052434578; Fax: +1 3052432396; Email: jmccauley@med.miami.edu

© *The Author 2009. Published by Oxford University Press.*
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

association *P*-values were ≤0.10 in the original IMSGC GWAS, in an independent data set.

RESULTS

In Stage 1, we obtained genotype data on 30 915 SNPs in 1488 cases and 3710 controls. Following extensive quality control (QC), we were ultimately able to utilize 29 561 SNPs in 1343 cases and 3577 controls for association with MS. This data set gave us 80% maximum potential power to detect risk odds ratio (OR) of 1.25, accepting a type 1 error rate of 0.001 (Supplementary Material, Fig. S1) (15). There were 85 SNPs, outside of the MHC (i.e. 29–34 Mb on chromosome 6), demonstrating high levels of significance ($P \leq 0.001$) (Table 1). Detailed analysis of SNPs within the broader MHC is the focus of a separate parallel project. As the SNPs selected for Stage 1 were chosen without consideration of linkage disequilibrium (LD), there are a number of SNPs with *P*-values ≤0.0001 that are in relatively strong LD with each other and therefore the significant SNPs do not represent 85 independent loci. As expected, there are a number of Stage 1 top hits in previously identified MS genes, including *CLEC16A* (1,7), *CD58* (1,8), *IRF8* (11) and *MMEL1* (M.B., unpublished data). As is typically the case in replication studies, previous top hits have shifted ranking in subsequent follow-up experiments. Our study is no exception, as the association *P*-values with arguably the two most notable genes, *IL2RA* (rs2104286, $P = 1.89 \times 10^{-2}$) and *IL7RA* (rs6897932, $P = 1.03 \times 10^{-2}$), fall just below our arbitrary *P*-value cutoff ($P < 1.0 \times 10^{-4}$) for inclusion in Table 1 (see Supplementary Material, Table S1 for results of the remaining 29 447 SNPs analyzed in Stage 1).

Following our analysis of the Stage 1 follow-up, we choose a smaller subset of SNPs for further replication in an independent data set. The results of the 19 SNPs genotyped (Sequenom MassARRAY iPLEX) and analyzed for Stage 2 (20 SNPs were genotyped, with 1 failing QC) are presented in Table 2. Eight of these SNPs demonstrated further replication ($P \leq 0.05$, with consistent OR) in this independent data set. A combined analysis for these 19 SNPs using data from the original screen and both Stage 1 and Stage 2 included 931 Trios, 3507 cases and 8024 controls. Five SNPs meet a conservative estimate of genome-wide significance using a Bonferroni correction (*P*-value cutoff 1.49×10^{-7}) considering the 334 923 independent tests from the original GWAS screen (Table 2). Furthermore, 4/5 SNPs were significant in each of the independent data sets. These four SNPs lie within or nearby *KIF21B* (on chromosome 1), *TMEM39A* (on chromosome 3), *C16orf75* and *PRM1* (both on chromosome 16). However, the two SNPs on chromosome 16 (rs12922090 and rs243315) near *C16orf75* and *PRM1* are in very strong LD ($D' = 0.99$, $r^2 = 0.82$). We also performed conditional logistic regression on these 19 SNPs conditioning on the HLA-DRB1*1501 tag (rs3135388); interestingly, the three SNPs (rs12922090, rs243315 and rs12927773) on chromosome 16 show slightly more significance in the HLA conditional analysis (Table 2).

DISCUSSION

We find considerable evidence for several new MS susceptibility loci including *KIF21B* (rs12122721, combined $P = 6.56 \times 10^{-10}$, OR = 0.82), *TMEM39A* (rs1132200, combined $P = 3.09 \times 10^{-8}$, OR = 0.80) and *PRM1* (rs243315, combined $P = 1.07 \times 10^{-7}$, OR = 0.83), all of which have demonstrated moderate-to-strong significance in each stage of our analyses and furthermore meet genome-wide significance using a stringent Bonferroni correction.

We have successfully identified novel loci for MS through more detailed examination of results from a large first-generation GWAS. Interestingly, in the original GWAS, the SNPs in *KIF21B*, *TMEM39A* and *PRM1*, although relatively significant in the more powerful case/control analysis [Cochran–Mantel–Haenszel (CMH) *P*-value ranks between 0.3 and 4.2%], failed to rise to the top of the more limited family-based analysis [the most significant SNP (rs12122721) had a transmission disequilibrium test (TDT) *P*-value rank of 28.9%] (Tables 1 and 3). Furthermore, these SNPs were among the top *P*-values (CMH *P*-value ranks between 0.4 and 2.1%) in a recent meta-analysis of three GWASs (11) (Table 3). These overall results clearly demonstrate that additional true susceptibility loci are likely to be buried beneath the top association results from GWAS (and even meta-analyses of GWAS), and subsequently overlooked in the rush to follow up the top hits. Testing only the 'top hits' is often the result of the limited availability of resources after conducting such a massive initial screening experiment. Our data suggest that it is imperative to perform a more comprehensive follow-up study in the pursuit of identifying all loci contributing to the genetic load for a given complex disease.

Furthermore, of the top Stage 1 results ($P \leq 0.001$), the average original GWAS *P*-value ranking of these SNPs is approximately 40 000 for the CMH test (most significant SNP ranking 177, least significant SNP ranking 319 841) and approximately 69 000 for the TDT test (most significant SNP ranking194, least significant SNP ranking 308 800). Approximately one-third (29/85) of the most significant non-MHC SNPs in Stage 1 (Table 1) had original GWAS *P*-values <0.10 in both the TDT and CMH tests, with only two of these SNPs further replicating in Stage 2 (rs11583328 and rs10469900) (Table 2). We extended this examination by ranking the three SNPs meeting genome-wide significance (i.e. within or nearby *KIF21B*, *TMEM39A* and *PRM*1) along with the most significant SNPs from the original GWAS (or in the case of *IL2RA* where rs2104286 has been indicated as the primary association (6)) and from other subsequently identified MS susceptibility loci with varying levels of confidence. In addition, we examined the rank of these SNPs in a recent meta-analysis (Table 3). The original *P*-values of the three newly identified loci were similar to those *P*-values seen in the other confirmed loci. Furthermore, each of these SNPs was mildly to moderately significant in the meta-analysis, but as in the initial GWAS follow-up, these loci fall far enough from the top that they are not initially selected for limited follow-up. It follows that there may be other yet-to-be-confirmed loci within this same range of the data. It is also noteworthy to highlight the robustness of the CMH test compared with the TDT in identifying all of these loci in the original screen. This may in part be related to the gain in power due to the additional samples used in the CMH analysis.

The new MS loci identified in this study are functionally interesting. *KIF21B* is a plus end-directed kinesin-like

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

Table 1. Top Stage 1 follow-up results ($P < 0.001$)

CHR	SNP	A1	A2	BP	Gene	Original GWAS *P*-value (rank) TDT Test	CMH Test	MAF	OR	L95	U95	*P*-value	HLA conditional *P*-value
6	rs3135388[a]	A	G	32521029	*HLA-DRA*	NA	NA	0.19	3.18	2.77	3.64	1.34×10^{-62}	NA
16	rs7184083	A	G	11135415	*CLEC16A*	3.37×10^{-2} (12 626)	1.51×10^{-2} (6818)	0.35	1.27	1.16	1.40	3.68×10^{-7}	3.66×10^{-7}
16	rs6498169[b]	G	A	11156830	*CLEC16A*	2.91×10^{-2} (10 940)	6.51×10^{-3} (3363)	0.35	1.26	1.15	1.38	1.34×10^{-6}	1.51×10^{-6}
16	rs181694[c]	T	C	11292330	*PRM1*	0.35 (122 995)	1.79×10^{-2} (7915)	0.20	0.75	0.67	0.84	1.36×10^{-6}	2.83×10^{-7}
16	rs243315[d]	T	C	11292512	*PRM1*	0.47 (160 855)	3.42×10^{-2} (13 992)	0.20	0.75	0.67	0.84	1.51×10^{-6}	3.20×10^{-7}
16	rs28087	C	T	11160330	*CLEC16A*	3.02×10^{-2} (11 332)	1.45×10^{-2} (6611)	0.34	1.26	1.14	1.38	1.66×10^{-6}	1.53×10^{-6}
16	rs27908	A	G	11164602	*CLEC16A*	3.41×10^{-2} (12 791)	1.35×10^{-2} (6247)	0.35	1.25	1.14	1.37	2.94×10^{-6}	2.95×10^{-6}
16	rs9941107	A	G	11103542	*CLEC16A*	0.14 (48 753)	1.87×10^{-2} (8221)	0.42	0.80	0.73	0.88	3.14×10^{-6}	3.35×10^{-6}
5	rs1393122[d]	G	A	4778148	*LOC340094*	0.38 (131 632)	2.98×10^{-2} (12 347)	0.17	0.74	0.66	0.85	5.09×10^{-6}	1.10×10^{-4}
16	rs3893660	G	A	11101431	*CLEC16A*	0.08 (28 006)	6.71×10^{-3} (3441)	0.42	0.81	0.74	0.89	5.22×10^{-6}	3.82×10^{-6}
16	rs12922090[d]	T	C	11322618	*C16orf75*	0.75 (252 825)	4.77×10^{-3} (2629)	0.17	0.75	0.66	0.85	6.69×10^{-6}	1.38×10^{-6}
9	rs2251622	T	A	90013426	*LOC389768*	0.05 (19 677)	0.13 (49 117)	0.24	1.27	1.14	1.41	8.96×10^{-6}	7.21×10^{-6}
16	rs3901386	C	T	11050221	*CLEC16A*	0.06 (20 519)	6.41×10^{-3} (3321)	0.41	0.81	0.74	0.89	9.63×10^{-6}	1.02×10^{-5}
8	rs12115114[d]	A	G	64552434	*YTHDF3*	0.20 (70 026)	1.19×10^{-3} (994)	0.17	1.29	1.15	1.45	1.13×10^{-5}	4.93×10^{-6}
16	rs7198004	G	A	11115118	*CLEC16A*	0.07 (24 701)	1.25×10^{-2} (5838)	0.42	0.81	0.74	0.89	1.16×10^{-5}	6.59×10^{-6}
16	rs7203150	C	T	11115223	*CLEC16A*	0.12 (41 644)	1.63×10^{-2} (7299)	0.42	0.82	0.75	0.90	1.67×10^{-5}	6.93×10^{-6}
6	rs11969369[d]	G	A	123156596	*SMPDL3A*	0.22 (77 478)	2.38×10^{-2} (10 053)	0.34	1.23	1.12	1.35	1.72×10^{-5}	7.64×10^{-5}
3	rs10511254[d]	A	G	107405284	*LOC728779*	3.72×10^{-3} (1724)	7.54×10^{-4} (762)	0.22	0.79	0.71	0.89	5.29×10^{-5}	1.77×10^{-4}
2	rs10469900[d]	C	T	38220587	*C2orf58*	3.96×10^{-2} (14 756)	3.31×10^{-2} (13 573)	0.21	1.24	1.12	1.38	6.24×10^{-5}	1.32×10^{-4}
16	rs12927773[d]	T	G	11311464	*PRM1*	0.52 (177 686)	1.85×10^{-2} (8152)	0.17	0.77	0.68	0.88	6.53×10^{-5}	1.29×10^{-5}
3	rs12487092[d]	G	T	107394865	*LOC728779*	1.26×10^{-2} (5028)	4.35×10^{-5} (254)	0.28	0.81	0.74	0.90	6.72×10^{-5}	1.41×10^{-4}
3	rs12487066[b]	C	T	107394820	*LOC728779*	7.70×10^{-3} (3213)	4.09×10^{-5} (250)	0.28	0.82	0.74	0.90	8.01×10^{-5}	1.64×10^{-4}
18	rs4798571[d]	A	G	7574294	*PTPRM*	0.41 (142 279)	0.10 (36 956)	0.16	1.27	1.13	1.43	1.02×10^{-4}	2.72×10^{-5}
16	rs12924729	A	G	11095284	*N/A*	0.77 (258 378)	1.95×10^{-2} (8503)	0.32	0.83	0.75	0.91	1.19×10^{-4}	1.42×10^{-4}
2	rs10188379	C	G	38224980	*C2orf58*	2.82×10^{-2} (10 594)	2.61×10^{-2} (10 891)	0.21	1.23	1.11	1.37	1.29×10^{-4}	2.84×10^{-4}
3	rs12497363	A	G	107401348	*LOC728779*	1.03×10^{-2} (4181)	2.42×10^{-3} (1636)	0.25	0.81	0.73	0.90	1.31×10^{-4}	6.04×10^{-4}
3	rs13085623	G	T	107415425	*LOC728779*	1.41×10^{-2} (5554)	2.71×10^{-3} (1778)	0.25	0.82	0.73	0.91	1.61×10^{-4}	7.10×10^{-4}
1	rs6696657[d]	T	C	208584705	*HHAT*	0.32 (112 265)	1.71×10^{-2} (7643)	0.41	1.19	1.09	1.31	1.84×10^{-4}	1.01×10^{-4}
16	rs9746695	C	T	11115395	*CLEC16A*	0.14 (50 787)	1.04×10^{-2} (4950)	0.31	0.83	0.76	0.92	1.92×10^{-4}	1.23×10^{-4}
1	rs305217	A	G	88993060	*PKN2*	2.78×10^{-2} (10 488)	2.43×10^{-2} (10 220)	0.05	1.43	1.18	1.72	2.07×10^{-4}	6.33×10^{-3}
8	rs7005198	C	G	16890680	*FGF20*	3.62×10^{-2} (13 513)	2.22×10^{-2} (9500)	0.20	1.23	1.10	1.37	2.10×10^{-4}	2.51×10^{-4}
1	rs7538427	C	T	89112010	*GTF2B*	0.05 (19 747)	2.93×10^{-2} (12 168)	0.05	1.42	1.18	1.71	2.27×10^{-4}	6.51×10^{-3}
3	(rs12638130/rs9873496)[d,e]	C	T	107516117	*LOC728784*	0.18 (64 010)	1.90×10^{-2} (8351)	0.45	0.85	0.77	0.92	2.43×10^{-4}	9.49×10^{-4}
1	rs11584383[c]	C	T	199202489	*LOC647216*	0.14 (48 316)	4.62×10^{-3} (2557)	0.30	0.83	0.75	0.92	2.53×10^{-4}	4.04×10^{-4}
1	rs11102091	A	G	110676947	*RBM15*	3.33×10^{-3} (1563)	0.07 (25 999)	0.44	0.84	0.77	0.92	2.63×10^{-4}	8.61×10^{-4}
2	rs17022137	C	T	38232941	*C2orf58*	2.63×10^{-2} (9934)	0.11 (40 979)	0.21	1.22	1.09	1.35	2.86×10^{-4}	5.69×10^{-4}
2	rs1517440[d]	C	T	221162118	*EPHA4*	0.45 (153 457)	4.94×10^{-2} (19 575)	0.04	1.46	1.19	1.79	2.92×10^{-4}	7.64×10^{-4}
11	rs4627080	G	T	9292925	*TMEM41B*	4.69×10^{-2} (17 273)	0.79 (265 886)	0.07	1.35	1.15	1.59	3.02×10^{-4}	1.41×10^{-3}
9	rs1924219[d]	C	T	110290167	*LOC347292*	1.80×10^{-2} (6911)	3.57×10^{-2} (14 546)	0.37	1.19	1.08	1.30	3.09×10^{-4}	3.08×10^{-4}
6	rs3800036[d]	G	A	1705555	*GMDS*	1.17×10^{-2} (4721)	0.17 (62 606)	0.48	0.85	0.77	0.93	3.11×10^{-4}	8.70×10^{-4}
3	rs1447925	T	C	60763882	*FHIT*	3.43×10^{-2} (12 862)	0.26 (93 782)	0.20	1.22	1.09	1.36	3.33×10^{-4}	3.12×10^{-4}
7	rs334517	G	T	47527873	*TNS3*	3.73×10^{-2} (13 925)	0.70 (239 174)	0.44	0.85	0.78	0.93	3.35×10^{-4}	2.10×10^{-4}
8	rs6557618	A	T	23057070	*TNFRSF10D*	0.11 (40 981)	4.62×10^{-2} (18 425)	0.29	0.83	0.75	0.92	3.93×10^{-4}	1.05×10^{-3}
16	rs8055544	G	T	10999062	*CLEC16A*	0.17 (61 857)	1.25×10^{-2} (5841)	0.42	1.18	1.08	1.29	4.00×10^{-4}	1.60×10^{-4}
17	rs17758761	C	A	51409524	*ANKFN1*	2.11×10^{-2} (8084)	0.88 (295 661)	0.03	1.52	1.20	1.91	4.00×10^{-4}	1.65×10^{-3}
1	rs12044852[b]	A	C	116889302	*CD58*	1.01×10^{-3} (683)	3.01×10^{-5} (233)	0.09	0.74	0.63	0.88	4.21×10^{-4}	1.43×10^{-4}
1	rs1572263	G	A	110687259	*RBM15*	3.85×10^{-2} (14 282)	0.06 (21 920)	0.27	0.83	0.75	0.92	4.44×10^{-4}	2.17×10^{-3}
8	rs3808524	C	T	23217928	*LOXL2*	0.64 (218 362)	4.81×10^{-2} (19 127)	0.45	0.85	0.78	0.93	4.48×10^{-4}	1.20×10^{-3}
12	rs2373461	T	G	100477852	*MYBPC1*	0.69 (233 655)	4.02×10^{-2} (16 201)	0.05	1.41	1.16	1.70	4.49×10^{-4}	2.69×10^{-4}
6	rs2326699	T	G	6075217	*F13A1*	2.71×10^{-2} (10 205)	0.07 (28 846)	0.23	1.21	1.09	1.34	4.55×10^{-4}	2.06×10^{-3}

Continued

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

Table 1. *Continued*

CHR	SNP	A1	A2	BP	Gene	Original GWAS *P*-value (rank) TDT Test	CMH Test	MAF	OR	L95	U95	*P*-value	HLA conditional *P*-value
3	rs1132200[d]	T	C	120633526	*TMEM39A*	0.39 (136 508)	1.35×10^{-3} (1071)	0.16	0.80	0.71	0.91	4.59×10^{-4}	1.73×10^{-3}
2	rs993598	A	G	178890941	*OSBPL6*	0.26 (91 913)	3.08×10^{-2} (12 747)	0.47	1.18	1.07	1.29	4.64×10^{-4}	1.08×10^{-3}
2	rs17265240	G	T	5193686	*LOC727982*	0.36 (126 348)	4.37×10^{-2} (17 497)	0.23	1.20	1.08	1.33	4.69×10^{-4}	3.77×10^{-4}
1	rs6673423	T	C	110717541	*SLC16A4*	4.57×10^{-2} (16 789)	0.06 (24 690)	0.27	0.83	0.75	0.92	4.74×10^{-4}	2.38×10^{-3}
10	rs7088282	T	G	10186238	*LOC644540*	0.18 (64 503)	3.23×10^{-2} (13 284)	0.27	1.19	1.08	1.32	5.09×10^{-4}	1.12×10^{-3}
1	rs17419032[d]	T	C	199265154	*KIF21B*	0.16 (55 617)	2.09×10^{-3} (1452)	0.28	0.84	0.75	0.92	5.29×10^{-4}	8.30×10^{-4}
5	rs6895902[c]	A	G	179134453	*MAML1*	0.19 (67 419)	2.58×10^{-2} (10 789)	0.33	1.18	1.08	1.30	5.46×10^{-4}	1.90×10^{-3}
17	rs4791872	G	A	9643950	*LOC644070*	3.43×10^{-2} (12 846)	0.55 (190 684)	0.01	2.18	1.40	3.39	5.48×10^{-4}	3.22×10^{-3}
16	rs1646066[c]	C	T	11226007	*LOC729954*	75 025	2.00×10^{-3} (1404)	0.14	0.79	0.69	0.90	5.49×10^{-4}	2.92×10^{-4}
2	rs10168171	A	G	28085591	*LOC728408*	3.53×10^{-2} (13 220)	0.87 (293 624)	0.18	0.81	0.71	0.91	5.93×10^{-4}	6.41×10^{-4}
15	rs3825904	T	G	99745948	*PCSK6*	4.04×10^{-2} (15 000)	0.95 (319 841)	0.29	1.19	1.08	1.31	6.47×10^{-4}	1.33×10^{-3}
1	rs12122721[d]	A	G	199251103	*KIF21B*	0.28 (96 681)	5.13×10^{-3} (2783)	0.29	0.84	0.76	0.93	6.47×10^{-4}	1.07×10^{-3}
1	rs3890745[c]	C	T	2543484	*MMEL1*	0.07 (27 687)	1.42×10^{-2} (6501)	0.31	0.84	0.77	0.93	6.78×10^{-4}	3.32×10^{-4}
1	rs11583328[d]	A	G	199268796	*CACNA1S*	0.07 (26 214)	4.22×10^{-4} (581)	0.29	0.84	0.76	0.93	6.88×10^{-4}	1.06×10^{-3}
17	rs9904838	G	A	51401320	*ANKFN1*	4.90×10^{-3} (2163)	0.46 (159 737)	0.04	1.45	1.17	1.81	7.10×10^{-4}	3.02×10^{-3}
3	rs1907878[d]	G	A	104487162	*LOC644681*	0.65 (221 976)	2.91×10^{-2} (12 074)	0.12	1.26	1.10	1.43	7.36×10^{-4}	1.81×10^{-3}
2	rs10180107	T	C	28094029	*LOC728408*	4.25×10^{-2} (15 765)	0.81 (274 882)	0.18	0.81	0.72	0.92	7.40×10^{-4}	8.53×10^{-4}
16	rs4451969[c]	T	C	11291020	*PRM1*	0.33 (114 085)	4.27×10^{-2} (17 113)	0.33	0.85	0.77	0.93	7.49×10^{-4}	1.08×10^{-4}
10	rs4746479	A	G	66399352	*ANXA2P3*	0.36 (124 519)	3.68×10^{-2} (14 991)	0.17	1.22	1.09	1.37	7.60×10^{-4}	2.21×10^{-3}
16	rs2280381	C	T	84576134	*IRF8*	0.91 (304 941)	1.71×10^{-2} (7626)	0.37	0.85	0.77	0.93	7.72×10^{-4}	8.96×10^{-3}
6	rs7742658	A	C	28708471	*LOC646160*	7.17×10^{-3} (2988)	0.15 (54 805)	0.02	1.73	1.26	2.37	7.77×10^{-4}	4.86×10^{-4}
3	rs1304118	T	C	104472940	*LOC644681*	0.33 (116 314)	2.41×10^{-2} (10 186)	0.12	1.26	1.10	1.43	7.79×10^{-4}	2.29×10^{-3}
11	rs2515795	A	G	117322486	*TMPRSS13*	0.57 (193 175)	5.32×10^{-3} (2855)	0.42	1.16	1.07	1.27	8.17×10^{-4}	4.07×10^{-4}
11	rs17118741	A	G	115118181	*LOC283143*	4.95×10^{-2} (18 214)	0.56 (193 709)	0.09	1.28	1.11	1.48	8.59×10^{-4}	8.07×10^{-4}
2	rs10196846	A	C	38232614	*C2orf58*	6.04×10^{-3} (2589)	1.34×10^{-2} (6196)	0.15	1.23	1.09	1.39	8.62×10^{-4}	1.03×10^{-3}
3	rs1373737	T	G	107341020	*LOC728779*	0.17 (59 505)	2.44×10^{-3} (1645)	0.36	0.85	0.77	0.94	8.64×10^{-4}	9.96×10^{-4}
9	rs1886106	A	G	110208464	*LOC347292*	0.07 (26 513)	4.08×10^{-2} (16 425)	0.33	1.17	1.07	1.29	9.12×10^{-4}	7.41×10^{-4}
3	rs9855065[d]	A	G	120612831	*CDGAP*	0.76 (255 088)	2.83×10^{-2} (11 781)	0.18	0.82	0.72	0.92	9.14×10^{-4}	3.46×10^{-3}
14	rs7160860[c]	T	C	53409243	*BMP4*	0.92 (308 800)	2.69×10^{-6} (177)	0.13	1.24	1.09	1.41	9.16×10^{-4}	3.71×10^{-4}
12	rs10777873	T	C	96404189	*LOC643711*	0.66 (223 145)	1.20×10^{-2} (5644)	0.17	0.81	0.72	0.92	9.16×10^{-4}	4.93×10^{-4}
22	rs134547	G	A	27131009	*TTC28*	0.26 (91 630)	2.02×10^{-2} (8770)	0.11	0.78	0.67	0.90	9.39×10^{-4}	2.97×10^{-3}
23	rs11092309	A	G	100624381	*ARMCX4*	0.26 (91 845)	1.15×10^{-2} (5449)	0.37	1.19	1.07	1.32	9.48×10^{-4}	1.09×10^{-3}
1	rs11208363	G	A	40783041	*ZNF684*	0.08 (28 126)	3.48×10^{-2} (14 195)	0.15	0.80	0.70	0.91	9.49×10^{-4}	1.63×10^{-3}
3	rs1025984	G	C	145238877	*C3orf58*	4.51×10^{-2} (16 588)	0.62 (212 857)	0.34	1.17	1.07	1.28	9.49×10^{-4}	7.39×10^{-4}
14	rs4247039	A	G	104095559	*LOC400258*	0.13 (46 192)	0.07 (25 366)	0.18	0.81	0.72	0.92	9.59×10^{-4}	6.65×10^{-4}
5	rs7720899	A	G	123907793	*ZNF608*	3.00×10^{-2} (11 248)	0.08 (30 658)	0.10	1.27	1.10	1.47	9.97×10^{-4}	1.65×10^{-3}

CHR, chromosome; A1, minor allele; A2, major allele; BP, physical base pair location of SNP in build 36; TDT, transmission disequilibrium test; CMH, Cochran–Mantel–Haenszel test; MAF, minor allele frequency; OR, odds ratio (relative to the minor allele); L95/U95, lower and upper bounds of the 95% confidence interval for the OR. Alleles are specified with respect to the forward (+) strand of the National Center for Biotechnology Information's Build 36.

[a]HLA-DRB1*1501 tag SNP.

[b]SNPs which were previously examined as part of the original GWAS replication effort (1).

[c]SNPs also selected for parallel IMSGC studies (data for these markers may also be reported as part of other hypothesis-driven work).

[d]SNPs selected for Stage 2 follow-up.

[e]There is one SNP for which the dbSNP 'snp_id' was merged into a new 'rsID' since the publication of the original GWAS (previous rsID/current rsID).

Table 2. Stage 2 follow-up results and combined analysis

						Original GWAS 931 family trios		Case subjects from 931 trios versus 2431 control subjects	Stage 1 follow-up (1343 cases/3577 controls)						Stage 2 follow-up (2164 cases/2016 controls)						Combined (931 trios, 3507 cases, 8024 controls)					
CHR	SNP	A1	A2	BP	Gene	MAF	TDT (*P*-value)	CMH (*P*-value)	MAF	OR	L95	U95	*P*-value	HLA conditional *P*-value	MAF	OR	L95	U95	*P*-value	HLA conditional *P*-value	MAF	OR	L95	U95	*P*-value	HLA conditional *P*-value
1	rs12122721	A	G	199251103	*KIF21B*	0.28	0.28	5.13×10^{-3}	0.29	0.84	0.76	0.93	6.47×10^{-4}	1.07×10^{-3}	0.26	0.85	0.77	0.94	1.88×10^{-3}	4.10×10^{-3}	0.28	0.82	0.77	0.88	6.56×10^{-10}	2.61×10^{-9}
5	rs1393122	G	A	4778148	*LOC340094*	0.17	0.38	2.98×10^{-2}	0.17	0.74	0.66	0.85	5.09×10^{-6}	1.10×10^{-4}	0.16	0.94	0.83	1.06	0.28	0.37	0.17	0.80	0.74	0.86	3.28×10^{-9}	1.89×10^{-6}
3	rs1132200	T	C	120633526	*TMEM39A*	0.14	0.39	1.35×10^{-3}	0.16	0.80	0.71	0.91	4.59×10^{-4}	1.73×10^{-3}	0.15	0.79	0.70	0.89	1.74×10^{-4}	1.52×10^{-5}	0.16	0.80	0.74	0.87	3.09×10^{-8}	1.93×10^{-8}
16	rs12922090	T	C	11322618	*C16orf75*	0.15	0.75	4.77×10^{-3}	0.17	0.75	0.66	0.85	6.69×10^{-6}	1.38×10^{-6}	0.16	0.82	0.73	0.93	1.21×10^{-3}	7.68×10^{-4}	0.17	0.81	0.75	0.88	5.34×10^{-8}	1.69×10^{-7}
16	rs243315	T	C	11292512	*PRM1*	0.18	0.47	3.42×10^{-2}	0.20	0.75	0.67	0.84	1.51×10^{-6}	3.20×10^{-7}	0.19	0.85	0.76	0.95	3.70×10^{-3}	1.30×10^{-3}	0.19	0.83	0.77	0.89	1.07×10^{-7}	6.34×10^{-8}
1	rs11583328	A	G	199268796	*CACNA1S*	0.27	0.07	4.22×10^{-4}	0.29	0.84	0.76	0.93	6.88×10^{-4}	1.06×10^{-3}	0.28	0.89	0.81	0.98	1.41×10^{-2}	2.85×10^{-2}	0.29	0.86	0.81	0.92	1.79×10^{-6}	1.17×10^{-6}
1	rs17419032	T	C	199265154	*KIF21B*	0.28	0.16	2.09×10^{-3}	0.28	0.84	0.75	0.92	5.29×10^{-4}	8.30×10^{-4}	0.28	0.90	0.81	0.98	2.24×10^{-2}	4.38×10^{-2}	0.29	0.86	0.81	0.92	1.95×10^{-6}	9.89×10^{-7}
6	rs3800036	G	A	1705555	*GMDS*	0.47	1.17×10^{-2}	0.17	0.48	0.85	0.77	0.93	3.11×10^{-4}	8.70×10^{-4}	0.47	0.94	0.86	1.02	0.13	0.20	0.48	0.88	0.83	0.93	3.03×10^{-6}	4.09×10^{-3}
2	rs10469900	C	T	38220587	*C2orf58*	0.20	3.96×10^{-2}	3.31×10^{-2}	0.21	1.24	1.12	1.38	6.24×10^{-5}	1.32×10^{-4}	0.21	1.14	1.03	1.27	1.36×10^{-2}	9.66×10^{-3}	0.21	1.16	1.09	1.24	1.24×10^{-5}	3.19×10^{-4}
16	rs12927773	T	G	11311464	*PRM1*	0.15	0.52	1.85×10^{-2}	0.17	0.77	0.68	0.88	6.53×10^{-5}	1.29×10^{-5}	0.16	0.83	0.73	0.93	1.40×10^{-3}	9.77×10^{-4}	0.16	0.85	0.79	0.91	1.56×10^{-5}	1.47×10^{-5}
3	rs10511254	A	G	107405284	*LOC728779*	0.22	3.72×10^{-3}	7.54×10^{-4}	0.22	0.79	0.71	0.89	5.29×10^{-5}	1.77×10^{-4}	0.22	0.93	0.84	1.03	0.18	0.2	0.23	0.87	0.81	0.93	2.60×10^{-5}	6.87×10^{-6}
18	rs4798571	A	G	7574294	*PTPRM*	0.16	0.41	0.10	0.16	1.27	1.13	1.43	1.02×10^{-4}	2.72×10^{-5}	0.16	1.04	0.93	1.17	0.49	0.49	0.16	1.16	1.08	1.25	6.88×10^{-5}	7.66×10^{-5}
3	(rs12638130/ rs9873496)	C	T	107516117	*LOC728784*	0.44	0.18	1.90×10^{-2}	0.45	0.85	0.77	0.92	2.43×10^{-4}	9.49×10^{-4}	0.44	0.94	0.86	1.03	0.17	0.09	0.45	0.89	0.85	0.95	7.34×10^{-5}	3.91×10^{-5}
3	rs12487092	G	T	107394865	*LOC728779*	0.27	1.26×10^{-2}	4.35×10^{-5}	0.28	0.81	0.74	0.90	6.72×10^{-5}	1.41×10^{-4}	0.28	0.95	0.86	1.04	0.25	0.17	0.28	0.88	0.83	0.94	9.23×10^{-5}	1.14×10^{-5}
6	rs11969369	G	A	123156596	*SMPDL3A*	0.35	0.22	2.38×10^{-2}	0.34	1.23	1.12	1.35	1.72×10^{-5}	7.64×10^{-5}	0.33	1.07	0.97	1.17	0.17	0.37	0.34	1.11	1.05	1.17	4.72×10^{-4}	2.36×10^{-3}
2	rs1517440	C	T	221162118	*EPHA4*	0.04	0.45	4.94×10^{-2}	0.04	1.46	1.19	1.79	2.92×10^{-4}	7.64×10^{-4}	0.05	0.94	0.77	1.15	0.56	0.26	0.05	1.18	1.04	1.34	1.12×10^{-2}	2.56×10^{-2}
9	rs1924219	C	T	110290167	*LOC347292*	0.35	1.80×10^{-2}	3.57×10^{-2}	0.37	1.19	1.08	1.30	3.09×10^{-4}	3.08×10^{-4}	0.35	1.02	0.93	1.12	0.67	0.99	0.36	1.07	1.01	1.13	1.89×10^{-2}	0.12
3	rs1907878	G	A	104487162	*LOC644681*	0.13	0.65	2.91×10^{-2}	0.12	1.26	1.10	1.43	7.36×10^{-4}	1.81×10^{-3}	0.12	0.94	0.83	1.08	0.39	0.32	0.12	1.08	1.00	1.17	0.06	3.91×10^{-2}
1	rs6696657	T	C	208584705	*HHAT*	0.42	0.32	1.71×10^{-2}	0.41	1.19	1.09	1.31	1.84×10^{-4}	1.01×10^{-4}	0.42	0.91	0.83	0.99	2.53×10^{-2}	3.83×10^{-2}	0.41	1.05	0.99	1.11	0.08	5.18×10^{-3}

CHR, chromosome; A1, minor allele; A2, major allele; BP, physical base pair location of SNP in build 36; MAF, minor allele frequency; OR, odds ratio (relative to the minor allele); L95/U95, lower and upper bounds of the 95% confidence interval for the OR. Alleles are specified with respect to the forward (+) strand of the National Center for Biotechnology Information's Build 36.

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

Table 3. MS susceptibility genes outside of the MHC

CHR	SNP	BP	Gene	Original GWAS *P*-value (rank)[a] TDT test	CMH test	Meta-analysis CMH test *P*-value (rank)[b]
1	rs12044852[c]	116889302	*CD58*	1.01×10^{-3} (683)	3.01×10^{-5} (233)	1.48×10^{-7} (2242)
10	rs2104286	6139051	*IL2RA*	3.29×10^{-3} (1549)	2.85×10^{-4} (479)	1.52×10^{-6} (2639)
3	rs1132200	120633526	*TMEM39A*[d]	0.39 (136 508)	1.35×10^{-3} (1071)	1.33×10^{-2} (48 639)
11	rs2074229[c]	60539684	*CD6*	0.07 (26 233)	4.01×10^{-3} (2317)	4.05×10^{-5} (3580)
1	rs12122721	199251103	*KIF21B*[d]	0.28 (96 681)	5.13×10^{-3} (2783)	2.13×10^{-3} (12 750)
16	rs6498169[c]	11156830	*CLEC16A*	2.91×10^{-2} (10 940)	6.51×10^{-3} (3363)	1.83×10^{-4} (4638)
1	rs3890745[c]	2543484	*MMEL1*[e]	0.07 (27 687)	1.42×10^{-2} (6501)	0.05 (163 561)
5	rs6897932	35910332	*IL7R*	5.83×10^{-3} (2497)	1.65×10^{-2} (7399)	7.71×10^{-6} (3020)
16	rs2280381[c]	84576134	*IRF8*	0.91 (304 941)	1.71×10^{-2} (7626)	5.08×10^{-4} (6277)
16	rs243315	11292512	*PRM1*[d]	0.47 (160 855)	3.42×10^{-2} (13 992)	1.50×10^{-2} (53 859)
19	rs280500[c]	10351402	*TYK2*	0.33 (116 121)	0.08 (31 423)	0.29 (770 374)
12	rs4149623[c]	6320839	*TNFRSF1A*	0.38 (133 155)	0.14 (52 423)	9.99×10^{-6} (3 084)[f]
18	rs4891786[c]	65722590	*CD226*	0.12 (42 656)	0.79 (265 756)	0.85 (2 175 710)

SNPs are sorted by *P*-value rank in the CMH test from the original GWAS.
CHR, chromosome; BP, physical base pair location of SNP in build 36; TDT, transmission disequilibrium test; CMH, Cochran–Mantel–Haenszel test.
[a]Original GWAS ranking is out of a total of 334 923 SNPs.
[b]Meta-analysis ranking is out of a total of 2.56 million SNPs. This work was previously published (11).
[c]Most significant SNP within each locus in the original GWAS (this is not necessarily the strongest associated SNP within the locus, as identified by other fine mapping efforts).
[d]SNPs identified with genome-wide significance in this study.
[e]Ban *et al*. (unpublished data) suggest that MMEL1 is another MS susceptibility gene.
[f]These results are for rs767455 (only 367 base pairs from rs4149623) not for rs4149623 (as rs4149623 was not included in the meta-analysis data set).

protein (KLP) involved in neuronal (axonal) transport. Its uniqueness stems from its enrichment in dendrites compared with the typical cell body and from its contrast from other plus end-directed KLPs, which have axon enrichment (16). *KIF21B* is also expressed in a variety of immune cells. Although *KIF21B* has not been functionally associated with neurodegeneration or inflammation, given the nature and role of its protein in neurons, there is a plausible biologic role for this gene in MS. Recently, another kinesin superfamily member (*KIF1B*) was reported as associated with MS (17); however, efforts by the IMSGC have failed to confirm this association (IMSGC, unpublished data). *KIF21B* is among the first genes identified via association studies, with the potential for a direct neurodegenerative role in MS pathology.

Very little has been known about *TMEM39A* (mRNA-transmembrane protein 39A). The associated SNP (rs1132200) within this gene causes a non-synonymous amino acid change (alanine–threonine) at position 487 in the protein. Although this SNP may hold some functional effect relevant to MS, almost nothing is known about this gene and what biologic role it might play with regard to disease susceptibility.

PRM1 (protamine 1) functions as a DNA-binding protein expressed in the nucleus of sperm. The strongest association in this region is with rs243315 and is 5′ of *PRM1*; however, there are several SNPs across this region of chromosome 16 showing mild-to-moderate levels of significance within the top hits (rs12922090, rs243315, rs1292773) (Table 2). This region of chromosome 16 is >100 kb from *CLEC16A* and there is little-to-no LD between these SNPs and any SNP within *CLEC16A*. There is, however, a very nearby candidate gene, *SOCS1* (suppressor of cytokine signaling 1), which is in strong LD with these SNPs and could possibly contain the true association. Additional work is needed to explore the exact location of this association, and is the focus of ongoing laboratory efforts.

Through this exhaustive follow-up approach, we have identified a number of additional MS susceptibility loci and highlighted even more loci that may yet prove to be involved in MS. Ultimately, fine mapping and functional studies will be required to understand the consequences of the associations detected in this experiment.

MATERIALS AND METHODS

Case and control subjects

Stage 1 follow-up. DNA samples from study participants were ascertained at two sites within the USA [Brigham and Women's Hospital in Boston (BWH) and the University of California at San Francisco (UCSF)] and through one site in the UK [University of Cambridge (CMS)]. All affected individuals met the McDonald criteria for a positive diagnosis for MS (18). Unrelated controls were obtained from these US sites and from the British 1958 Birth Cohort Study. These controls were selected to provide nearly equivalent gender and age matching. This sample set contained 2961 individuals (1479 cases and 1482 controls) for genotyping. Additional control sample data were available on 2198 samples from both the National Institute of Mental Health (NIMH) and the Wellcome Trust Case Control Consortium (WTCCC). Data from these additional controls were previously analyzed in the 110 SNPs selected for replication in the original GWAS (1). With the exception of a small set of overlapping SNPs genotyped in this effort (95 of the 110 SNPs from the replication phase of the original GWAS were genotyped and analyzed in this study), these control data are completely independent of previous association testing in these MS samples. All samples used in the Stage 1 analysis come from participants self-reporting as non-Hispanic whites (Table 4).

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

Table 4. Demographic characteristics of Stage 1 follow-up cases and controls

	Stage 1 data set (1343 cases/1379 controls)				Additional controls (2198 controls)[a]		
	UK		USA		UK		USA (NIMH)
	Case	Control	Case	Control	UKBS	1958 BC	
Gender (count)							
Female	493	570	343	329	360	359	344
Male	181	194	326	286	378	378	379
Ratio (female–male)	2.72	2.94	1.05	1.15	0.95	0.95	0.91
Total individuals	674	764	669	615	738	737	723
Age at analysis (years)							
Average	50.0	50.0	50.0	48.5	Unknown	50.0	Unknown
Range	27–72	50	23–89	23–84	18–69	50	Unknown
Age at onset (years)							
Average	32.3	NA	33.9	NA	NA	NA	NA
Range	9–57	NA	4–64	NA	NA	NA	NA
Disease course (%)							
Relapsing remitting	55.04	NA	49.93	NA	NA	NA	NA
Secondary progressive	28.49	NA	17.94	NA	NA	NA	NA
Primary progressive	13.65	NA	10.16	NA	NA	NA	NA
Progressive relapsing	0	NA	2.99	NA	NA	NA	NA
Clinically isolated syndrome	0	NA	6.72	NA	NA	NA	NA
Unknown	2.82	NA	12.26	NA	NA	NA	NA
Expanded disability status scale score (%)							
<3	35.61	NA	46.34	NA	NA	NA	NA
3 to <6	23.29	NA	20.78	NA	NA	NA	NA
6	15.58	NA	7.77	NA	NA	NA	NA
6.5	8.75	NA	5.38	NA	NA	NA	NA
>6.5	15.13	NA	7.17	NA	NA	NA	NA
Unknown	1.63	NA	12.56	NA	NA	NA	NA

UKBS, UK Blood Services; 1958 BC, British 1958 Birth Cohort; NIMH, National Institute of Mental Health.
[a]These control data were provided by both the NIMH and the WTCCC and represent the 723 US controls and the 1475 UK controls (respectively) used in the 110 SNP replication analysis of our original screen (1).

Table 5. Stage 2 follow-up cases and controls

Stage 2 data set[a] Collection	USA					UK		Total
	BWH	WU	ACP	UCSF	RUSH	UC	1958 BC	
Cases	224	158	588	363	0	831	0	2164
Controls	405	13	36	30	513	0	1019	2016
							Total	4180

BWH, Brigham and Women's Hospital; WU, Washington University, St Louis; ACP, Accelerated Cure Project; UCSF, University of California, San Francisco; RUSH, RUSH University; UC, University of Cambridge; 1958 BC, British 1958 Birth Cohort.
[a]This data set has previously been described in detail (11) and represents an independent set of cases and controls from those used in either the original GWAS or the Stage 1 follow-up.

Stage 2 follow-up. Cases and controls genotyped for Stage 2 were made available through an entirely independent replication set (11). This data set consists of an additional 2164 cases and 2016 controls from the sites listed previously as well as those made available through other collaborative efforts. The same criteria were applied to these cases and controls as in Stage 1 (Table 5).

Approval for these studies was granted by the appropriate institutional review boards. All studies were performed after informed consent from human subjects.

Table 6. SNPs selected for Stage 1 follow-up genotyping (out of 334 923 analyzed in original GWAS)

SNP count	Description
62 488	SNPs $P \leq 0.10$ in either TDT or CMH screening
35 928	SNPs $P \leq 0.10$ in TDT screening
37 929	SNPs $P \leq 0.10$ in CMH screening
11 369	SNPs $P \leq 0.10$ in both TDT and CMH screening
33 484[a]	SNPs chosen for Illumina iSelect Infinium design for Stage 1
30 915	SNPs arrayed on the beadchip for Stage 1
30 392	SNPs passing primary QC for Stage 1
831	Additional SNPs dropped through secondary QC
1	HLA-tag SNP (rs3135388)
95	SNPs previously genotyped for initial GWAS replication effort ($n = 95/110$)
28 696	SNPs exclusively selected for Stage 1
769	SNPs overlapping Stage 1 and parallel IMSGC projects

[a]This reflects those SNPs likely to generate accurate and reliable assays using the Illumina platform.

Molecular analysis

Stage 1 follow-up. We utilized the Illumina iSelect Custom BeadChip platform to perform additional genotyping of a more in-depth list of top hits from the GWAS experiment (19). This experiment was performed in parallel with several

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

other projects organized through the IMSGC to maximize the use of samples and resources. This strategy allowed us to use the maximum number of bead types (60 800) available for the iSelect platform (depending on the chemistry used for assaying a particular SNP, there may be one bead type per SNP or two bead types per SNP). The SNPs selected for inclusion in our Stage 1 effort satisfied two criteria: (i) SNPs demonstrating *P*-values ≤ 0.10 in either the TDT or the CMH test, from the original GWAS screen; (ii) SNPs that had an Infinium score >0.60 (a proprietary score used by Illumina to determine the likelihood of assays to generate accurate and reliable results). In the original GWAS, a total of 62 488 SNPs had a *P*-value ≤ 0.10 in either the TDT or CMH test; of these, 33 484 had an Infinium quality score >0.60. These 33 484 SNPs were selected for inclusion in Stage 1 of our replication effort along with an additional 19 318 SNPs (for other parallel IMSGC projects) giving a total of 52 801 SNPs (60 800 bead types). Once manufacturing and internal QC procedures at Illumina were complete, 48 767 SNPs ($\sim$92% of the total requested) were arrayed on each of the beadchips for genotyping, including 30 915 of the Stage 1 follow-up SNPs ($\sim$49% of those meeting the initial criteria) (Table 6). A total of 29 561 of these 30 915 SNPs were ultimately analyzed after QC procedures were completed. Through LD ($r^2 = 0.80$), these 29 561 SNPs capture 60.1% of the total SNPs (62 488) having a *P*-value ≤ 0.10 in the original GWAS. Furthermore, these SNPs, through LD ($r^2 = 0.80$), cover 92% of all the SNPs with *P*-values ≤ 0.05 in the original screen. Supplementary Material, Figure S2 provides a visual summary of those SNPs analyzed in Stage 1 relative to their significance in the original GWAS and of those SNPs further chosen and analyzed in Stage 2.

We followed the Illumina Infinium protocol for the genotyping of DNA samples. In brief, this involved amplification and subsequent fragmentation of genomic DNA, followed by hybridization of this fragmented DNA to the BeadChip, then an extension step and finally imaging to read the chip (19). We genotyped an initial data set of 2961 individuals (1479 cases and 1482 controls) distributing DNA samples across beadchips (12 samples per beadchip), with attention given to representing both cases and controls from each of the different ascertainment sites on every chip as to minimize any experimental biases in genotyping performance.

Stage 2 follow-up. Following the analysis for Stage 1, there were 85 SNPs outside of the MHC region with association *P*-values ≤ 0.001, and of these, we genotyped 20 SNPs in a second independent data set (independent of both Stage 1 and the original GWAS) for our Stage 2 follow-up. There were five criteria used to select the SNPs for Stage 2 genotyping: (i) Stage 1 *P*-value ≤ 0.001; (ii) SNPs within or nearby known genes; (iii) exclusion of SNPs in the MHC (within 29–34 Mb on chromosome 6); (iv) exclusion of SNPs overlapping with previously identified MS genes or examined as part of the initial GWAS replication effort; (v) exclusion of SNPs being analyzed as part of other parallel projects using this common data set. We chose 21 SNPs that met these criteria; however, one SNP (rs9855065) failed to pass the design process. We used the Sequenom MassARRAY iPLEX platform for this genotyping. The Sequenom protocol involves a multiplex PCR reaction prior to a single-base primer extension reaction. The individual SNPs are identified by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (20).

Statistical analysis

Stage 1 follow-up. We initially performed a thorough series of QC procedures, which are described in the Supplemental Material. Following stringent QC and finding no significant population differences in this data set, we chose to analyze this data set as one uniform sample collection (Supplementary Material, Fig. S3). The Stage 1 test for association was conducted using a logistic regression approach as implemented in PLINK and using PCA1 and PCA2 as covariates to correct for differential genotyping bias (21). This method tests for a linear trend in the number of alleles at a single locus. This analysis included GWAS data from 2198 NIMH and WTCCC controls used in the original GWAS replication in addition to the newly genotyped data set of 1343 cases and 1379 controls. After removing SNPs from the MHC (i.e. 29–34 Mb on chromosome 6), the genomic inflation factor (GIF) was 1.16 (Supplementary Material, Fig. S4). This is larger than the original GWAS GIF (1.05) and is likely due to preferential selection of SNPs with small *P*-values. In addition to the standard logistic regression, a conditional logistic regression analysis was also performed conditioning on the HLA-DRB1*1501 tag SNP (rs3135388). Genotypes for rs3135388 had previously been imputed for the NIMH and WTCCC control, as this SNP was not genotyped on the Affymetrix 500K chip.

Stage 2 follow-up. PLINK was also used for the Stage 2 replication analysis. Logistic regression was used to test for association with the 19 SNPs and 4180 independent replication samples that passed QC. To perform a joint analysis of both Stage 1 and Stage 2 data sets, and the original GWAS screen (931 trios and 2431 controls), the UNPHASED software was utilized (22). A joint conditional analysis was also done on the HLA-tag SNP (rs3135388) in UNPHASED.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the Accelerated Cure Project for its work in collecting samples from subjects with MS and for making these samples available to IMSGC investigators. We also thank the following clinicians for contributing to sample collection efforts: Accelerated Cure Project—Drs Elliot Frohman, Benjamin Greenberg, Peter Riskind, Saud Sadiq, Ben Thrower and Tim Vollmer; Washington University—Drs B.J. Parks and R.T. Naismith. We thank the Brigham and Women's Hospital PhenoGenetic Project for providing DNA samples from healthy subjects that were used in the Stage 2 follow-up effort of this study. We acknowledge the work done by the Biorepository and the Center for Genome Technology within

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

the John P. Hussman Institute for Human Genomics (University of Miami); specifically, Sandra West for aid in specimen management and both Ashley Anderson and Luis Espinosa for aid in sample processing and genotyping. We thank the Computational Genomics Core within the Center for Human Genetics Research (Vanderbilt University); specifically, Justin Giles, Yuki Bradford and David Sexton for their support in data processing. We also thank Joanne Wang for meta-analysis data management.

Conflict of Interest statement. None declared.

FUNDING

The International MS Genetics Consortium is supported by grants, societies, foundations and a number of individual donors. P.L.D. is a Harry Weaver Neuroscience Scholar Awardee of the National MS Society (NMSS); he is also a William C. Fowler Scholar in Multiple Sclerosis Research. D.A.H. is a Jacob Javits Scholar of the NIH. We acknowledge the use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. This work was supported by the National Institutes of Health (R01NS049477 to the IMSGC, R01NS32830 to J.L.H., NS46341 to P.L.D.); the National Multiple Sclerosis Society (NMSS) (AP3758-A16 to the IMSGC, RG4201-A-1 to J.L.M., RG4198-A-1); the Wellcome Trust (084702/Z/08/Z); the Medical Research Council (G0000934); a number of individual donors; and the Cambridge NIHR Biomedical Research Centre Funding to pay the open access charge was provided by the IMSGC.

CONTRIBUTORS

Jacob L. McCauley[1,†], Rebecca L. Zuvich[2,†], Ashley H. Beecham[1], Philip L. De Jager[3,4], Ioanna Konidari[1], Patrice L. Whitehead[1], Cristin Aubin[4], Maria Ban[5], Susan Pobywajlo[4], Rebeccah Briskin[4], Susan Romano[3], Neelum T. Aggarwal[6], Laura Piccio[7], Wendy L. McArdle[8], David P. Strachan[9], Denis Evans[6], Anne H. Cross[7], Bruce Cree[10], John D. Rioux[11], Lisa F. Barcellos[12], Adrian J. Ivinson[13], Alastair Compston[5], David A. Hafler[4,14], Stephen L. Hauser[10,15], Jorge R. Oksenberg[10,15], Stephen J. Sawcer[5], Margaret A. Pericak-Vance[1], Jonathan L. Haines[2].

[1]John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL 33136, USA.

[2]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN 37232, USA.

[3]Program in NeuroPsychiatric Genomics, Department of Neurology, Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.

[4]Program in Medical and Population Genetics, Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[5]Department of Clinical Neurosciences, Addenbrooke's Hospital, University of Cambridge, Box 165, Hills Road, Cambridge CB2 2QQ, UK.

[6]Rush Alzheimer Disease Center and Department of Neurological Sciences, Rush University, Chicago, IL 60612, USA.

[7]Department of Neurology, Washington University Saint Louis, Saint Louis, MO 63110, USA.

[8]Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK.

[9]Division of Community Health Sciences, St George's, University of London, London SW17 ORE, UK.

[10]Department of Neurology, School of Medicine, University of California, San Francisco, CA 94143-0435, USA.

[11]Laboratory in Genetics and Genomic Medicine of Inflammation, Université de Montréal and the Montreal Heart Institute, Montreal, Quebec H1T 1C8, Canada.

[12]Division of Epidemiology, School of Public Health, University of California, Berkeley CA 94720-7356, USA.

[13]Harvard NeuroDiscovery Center, Harvard Medical School, Boston, MA 02155, USA.

[14]Department of Neurology, Yale University School of Medicine, New Haven, CT 06520-8018, USA.

[15]Institute for Human Genetics, School of Medicine, University of California, San Francisco, CA 94143-0435, USA.

[†]These individuals contributed equally to this work.

REFERENCES

1. International Multiple Sclerosis Genetics ConsortiumHafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B. *et al.* (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.*, **357**, 851–862.
2. Lundmark, F., Duvefelt, K., Iacobaeus, E., Kockum, I., Wallstrom, E., Khademi, M., Oturai, A., Ryder, L.P., Saarela, J., Harbo, H.F. *et al.* (2007) Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat. Genet.*, **39**, 1108–1113.
3. Weber, F., Fontaine, B., Cournu-Rebeix, I., Kroner, A., Knop, M., Lutz, S., Muller-Sarnowski, F., Uhr, M., Bettecken, T., Kohli, M. *et al.* (2008) IL2RA and IL7RA genes confer susceptibility for multiple sclerosis in two independent European populations. *Genes Immun.*, **9**, 259–263.
4. Rubio, J.P., Stankovich, J., Field, J., Tubridy, N., Marriott, M., Chapman, C., Bahlo, M., Perera, D., Johnson, L.J., Tait, B.D. *et al.* (2008) Replication of KIAA0350, IL2RA, RPL5 and CD58 as multiple sclerosis susceptibility genes in Australians. *Genes Immun.*, **9**, 624–630.
5. Gregory, S.G., Schmidt, S., Seth, P., Oksenberg, J.R., Hart, J., Prokop, A., Caillier, S.J., Ban, M., Goris, A., Barcellos, L.F. *et al.* (2007) Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat. Genet.*, **39**, 1083–1091.
6. International Multiple Sclerosis Genetics Consortium (IMSGC) (2008) Refining genetic associations in multiple sclerosis. *Lancet Neurol.*, **7**, 567–569.
7. International Multiple Sclerosis Genetics Consortium (IMSGC) (2009) The expanding genetic overlap between multiple sclerosis and type I diabetes. *Genes Immun.*, **10**, 11–14.
8. De Jager, P.L., Baecher-Allan, C., Maier, L.M., Arthur, A.T., Ottoboni, L., Barcellos, L., McCauley, J.L., Sawcer, S., Goris, A., Saarela, J. *et al.* (2009) The role of the CD58 locus in multiple sclerosis. *Proc. Natl Acad. Sci. USA*, **106**, 5264–5269.
9. Ban, M., Goris, A., Lorentzen, A.R., Baker, A., Mihalova, T., Ingram, G., Booth, D.R., Heard, R.N., Stewart, G.J., Bogaert, E. *et al.* (2009) Replication analysis identifies TYK2 as a multiple sclerosis susceptibility factor. *Eur. J. Hum. Genet*, **17**, 1309–1313.
10. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene) (2009) Genome-wide association study identifies new

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020

multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.*, **41**, 824–828.
11. De Jager, P.L., Jia, X., Wang, J., de Bakker, P.I., Ottoboni, L., Aggarwal, N.T., Piccio, L., Raychaudhuri, S., Tran, D., Aubin, C. *et al.* (2009) Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.*, **41**, 776–782.
12. Wellcome Trust Case Control Consortium, Australo-Anglo-American Spondylitis Consortium (TASC)Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I. *et al.* (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.*, **39**, 1329–1337.
13. Maier, L.M. and Hafler, D.A. (2008) The developing mosaic of autoimmune disease risk. *Nat. Genet.*, **40**, 131–132.
14. Baranzini, S.E., Wang, J., Gibson, R.A., Galwey, N., Naegelin, Y., Barkhof, F., Radue, E.W., Lindberg, R.L., Uitdehaag, B.M., Johnson, M.R. *et al.* (2009) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.*, **18**, 767–778.
15. Purcell, S., Cherny, S.S. and Sham, P.C. (2003) Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*, **19**, 149–150.
16. Marszalek, J.R., Weiner, J.A., Farlow, S.J., Chun, J. and Goldstein, L.S. (1999) Novel dendritic kinesin sorting identified by different process targeting of two related kinesins: KIF21A and KIF21B. *J. Cell Biol.*, **145**, 469–479.
17. Aulchenko, Y.S., Hoppenbrouwers, I.A., Ramagopalan, S.V., Broer, L., Jafari, N., Hillert, J., Link, J., Lundstrom, W., Greiner, E., Dessa Sadovnick, A. *et al.* (2008) Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. *Nat. Genet.*, **40**, 1402–1403.
18. McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C. *et al.* (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.*, **50**, 121–127.
19. Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R. and Gunderson, K.L. (2006) Whole-genome genotyping with the single-base extension assay. *Nat. Methods*, **3**, 31–33.
20. Gabriel, S., Ziaugra, L. and Tabbaa, D. (2009) SNP genotyping using the sequenom MassARRAY iPLEX platform. *Curr. Protoc. Hum. Genet.*, Chapter 2, Unit 2.12. PubMed ID: PMID: 19170031.
21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
22. Dudbridge, F. (2003) Pedigree disequilibrium tests for multilocus haplotypes. *Genet. Epidemiol.*, **25**, 115–121.

Downloaded from https://academic.oup.com/hmg/article-abstract/19/5/953/583209 by KU Leuven Libraries user on 22 May 2020