Multiple sclerosis genetics

Stephen Sawcer, Robin J M Franklin, Maria Ban

Lancet Neurol 2014; 13: 700–09
Published Online
May 20, 2014
http://dx.doi.org/10.1016/S1474-4422(14)70041-9

Department of Clinical Neurosciences, Cambridge Biomedical Campus (Prof S Sawcer PhD, M Ban PhD), **Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute** (Prof R J M Franklin PhD), **and Department of Veterinary Medicine** (Prof R J M Franklin), **University of Cambridge, Cambridge, UK**

Correspondence to:
Prof Stephen Sawcer, Department of Clinical Neurosciences, Box 165, Cambridge Biomedical Campus, University of Cambridge, Cambridge CB2 0QQ, UK
sjs1016@cam.ac.uk

For the **Atlas of MS database** see http://www.atlasofms.org/

For the **Multiple Sclerosis International Federation** see http://www.msif.org

Genome-wide association studies have revolutionised the genetic analysis of multiple sclerosis. Through international collaborative efforts involving tens of thousands of cases and controls, more than 100 associated common variants have now been identified. These variants consistently implicate genes associated with immunological processes, overwhelmingly lie in regulatory rather than coding regions, and are frequently associated with other autoimmune diseases. The functional implications of these associated variants are mostly unknown; however, early work has shown that several variants have effects on splicing that result in meaningful changes in the balance between different isoforms in relevant tissues. Including the well established risk attributable to variants in genes encoding human leucocyte antigens, only about a quarter of reported heritability can now be accounted for, suggesting that a substantial potential for further discovery remains.

Introduction

Multiple sclerosis is an inflammatory demyelinating disease of the CNS that results in chronic progressive disability for the majority of people with the disorder. Most patients are unemployed within 15 years of diagnosis and rates of depression, suicide, and divorce are substantially increased compared with the healthy population.[1] Half of all patients need assistance with mobility within 20 years of diagnosis, and 50% of patients eventually develop substantial cognitive deficits.[1] The disease most often starts between 20 and 40 years of age, and affects women more frequently than men.[2] According to the Atlas of MS database, worldwide about 2·5 million people have multiple sclerosis, and figures from the Multiple Sclerosis International Federation suggest that in Europe alone the disease costs more than €15 billion each year in terms of direct health-care costs and lost productivity. As is the case for many other autoimmune diseases, evidence suggests that the incidence of multiple sclerosis is increasing.[3] Although the precise aetiology of multiple sclerosis remains unknown, in the past few years the identification of genetic variants affecting the development of the disease has grown almost exponentially. In this Review we outline the basic epidemiological foundations underpinning the genetic analysis of multiple sclerosis, describe some of the landmark findings from the past, summarise recent findings, and consider what the future might hold. Like pieces in a jigsaw puzzle, each of these associated variants provides a clue to aetiology. The more pieces we find the more likely it is that they will fit together in meaningful ways to reveal the crucial mechanisms underlying the development of this enigmatic disease.

Epidemiology

Two features have consistently emerged from the extensive epidemiological analysis of multiple sclerosis: first, that the disease clusters in families,[4,5] and second, that the disease varies greatly in frequency worldwide.[6,7] Although neither of these findings necessarily implies an exclusive role for either genetic or environmental factors, supplementary studies in informative subgroups (eg, twins,[8] adoptees,[9] conjugal pairs,[10] and migrant individuals[11]) suggest that familial clustering is determined mainly by genetic factors,[12] whereas regional variation in prevalence results from the effects of both genetic and population-level environmental risk factors.[12,13] The imprecision[14] and bias[15] inherent in estimation of familial recurrence risks limits what can be inferred by comparison of these risks between relatives (ie, segregation analysis). However, attempts at such analysis have generally suggested that the available data are most consistent with a polygenic model in which risk is determined by a single moderate-effect allele (odds ratio [OR] roughly 3–4) and many much-smaller-effect alleles (OR <1·5).[5,16] The tendency for familial recurrence risk to fall geometrically with the degree of relatedness[4,5,8,17] suggests that interactions between these risk alleles are also probably relevant, and could account for a substantial fraction of the apparent heritability of the disease.[18] Whether the genetic architecture underlying susceptibility is dominated by very many rare variants of large effect or a still large (but rather smaller) number of common variants of modest effect is difficult to infer from patterns of recurrence risk. Because both models predict a combination of seemingly sporadic cases along with familial clusters, the description of sporadic disease as being fundamentally different from familial disease in terms of its genetic architecture is probably inaccurate.[19] Furthermore, because the multiple rare variant and the common disease common variant hypotheses are not mutually exclusive, some combination of both will probably be relevant to some degree in any given disease.[20]

Linkage

Although segregation analysis has little power to establish the relative importance of the two main allelic models, linkage analysis can be useful.[21] Families with multiple sclerosis rarely include more than three or four affected individuals and large, extended families with many cases of the disease are extremely uncommon.[22] Furthermore, the absence of any linkage in the few larger-than-average families that have been reported[22–25] suggests that—unlike many other complex traits—rare, highly penetrant alleles are uncommon in multiple

sclerosis, if they exist at all. The extended families that have been described are generally characterised by having a large number of affected siblings per generation rather than multiple affected generations, and often include founders that have a much higher-than-expected rate of the main risk allele for multiple sclerosis, *HLA-DRB1*15:01*, which has an OR of about 3.[26,27] Furthermore, in a well powered non-parametric genome-wide linkage screen, the International Multiple Sclerosis Genetics Consortium (IMSGC)[28] did not find any statistically significant evidence for linkage outside the major histocompatibility complex (where linkage arises from the effects of the *HLA-DRB1*15:01* allele).[28] This result, although disappointing, is crucially important because it places an upper limit on the risk effect exherted by relevant alleles (OR less than 1·5 for common variants and less than about 6 for rare variants).[21,28] These upper limits have very important implications for both the design and interpretation of association studies.

The major histocompatibility complex

Although associations between multiple sclerosis and variation in the genes encoding human leucocyte antigens (HLAs) contained within the major histocompatibility complex have been recognised for several decades,[29,30] the extreme polymorphism and extensive linkage disequilibrium (ie, correlation between linked variants) that characterise this gene-dense region[31] make the identification of relevant variants driving these associations difficult. However, in the past few years the advent of high-throughput typing for single-nucleotide polymorphisms (SNPs) and the development of statistical methods capable of imputing (ie, inferring) classic HLA genotypes from SNP data[32–34] have enabled the study of thousands of individuals, which in turn has allowed substantial progress to be made. It is now clear that the association with the haplotype exerting the greatest effect on risk (*HLA-DRB1*15:01–DQA1*01:02–HLA-DQB1*06:02*) is driven by the *HLA-DRB1*15:01* allele, and that association with the other alleles of this haplotype is secondary only to their linkage disequilibrium with *HLA-DRB1*15:01*.[27,35] Furthermore, the long-suspected class-I protective effect[36–39] has been confirmed,[27,35] and shown to be driven mainly by the *HLA-A*02:01* allele.[27,35] Conditional analysis has confirmed the relevance of *HLA-DRB1*03:01*, *HLA-DRB1*13:03*, and *HLA-DPB1*03:01*[38,40–44] (table 1), and suggested that risk might also be affected by non-HLA genes from this region.[35] Although these SNP-based studies have not yet provided convincing evidence to support the existence of complex interactions between these risk alleles and haplotypes, such interactions have been suggested.[41] Such interactions almost certainly occur, but theoretical calculations suggest that very large sample sizes will be needed to reliably identify their nature and establish the alleles involved.[18]

	OR	RAF	–log(p)
*HLA-DRB1*15:01*	3·10	13·3%	320·0
*HLA-A*02:01*	1·37†	25·9%	22·0
*HLA-DRB1*03:01–DQB1*02:01*	1·26	14·6%	9·4
*HLA-DRB1*13:03–DQB1*03:01*	2·40	0·9%	10·9
rs9277535_G‡	1·28	24·9%	26·4

The statistical significance for each allele is conditional on the alleles above it in the table. RAF refers to allele frequency in the UK population. –log(p) refers to the negative log of the p value in the combined analysis reported by the International Multiple Sclerosis Genetics Consortium[27] and the Wellcome Trust Case Control Consortium 2.[27] OR=odds ratio. RAF=risk allele frequency. †This allele exerts a protective effect so the OR on the risk scale would be 0·73. ‡This single-nucleotide polymorphism is in linkage disequilibrium with *HLA-DPB1*03:01*.

Table 1: **Established multiple sclerosis risk alleles in the major histocompatibility complex**

Genome-wide association studies—a new era in complex genetics

Despite decades of candidate-gene-based efforts, little progress was made in the identification of relevant, genuinely associated risk alleles outside the major histocompatibility complex before the advent of genome-wide association studies. The only real progress was the identification of association with the SNP rs6897932 from the *IL7R* gene, which was suggested by combining information from many data sources (ie, genomic convergence[45]) and confirmed by typing large numbers of cases and controls.[46] The results from the first genome-wide association study in multiple sclerosis appeared at the same time and identified association with variants in both *IL7R* and *IL2RA*.[47] In total, 14 genome-wide association studies have now been completed in multiple sclerosis (table 2).[27,39,47–58] Unsurprisingly, the success of these studies was directly related to the number of samples screened; studies screening fewer than about 800 cases did not identify new reproducible associations, whereas each of the larger studies successfully added to the growing list of such loci. The largest study so far (a collaboration between the IMSGC and the Wellcome Trust Case Control Consortium [WTCCC2][27]) confirmed 23 previously reported associations and identified an additional 34 new associated variants, 29 with genome significance and five with suggestive levels of significance just missing this threshold (these five have since been verified with genome-wide significant association in a subsequent IMSGC follow-up study[59]).

The coherence in the characteristics of these 57 primary associated SNPs is striking (appendix). Despite the fact that more than half of the SNPs analysed in this genome-wide association study map to an interval between gene transcripts (58%), most of the 57 associated SNPs (81%) map within a gene transcript (a highly significant excess, $p=3·8\times10^{-9}$). More than a third of these SNPs (21 of 57) have previously been shown to affect the risk of other autoimmune diseases, or are in linkage disequilibrium with other autoimmune-disease-associated variants.[27]

For the **Multiple Sclerosis GWAS Browser** see http://wattle.well.ox.ac.uk/wtccc2/external/ms/

See **Online** for appendix

	Population	Cases	Controls	SNPs
IMSGC (2007)[47]*	UK and USA	931	Parents	334 923
WTCCC1 (2007)[48]†	UK	975	1466	12 374
Comabella et al (2008)[49]‡	Spain	242	242	428 867
Aulchenko et al (2008)[50]	Netherlands	45	195	250 000
Baranzini et al (2009)[51]	Various	978	883	551 642
De Jager et al (2009)[39]§	USA	860	1720	709 690
ANZgene (2009)[52]	Australia and New Zealand	1618	3413	302 098
Sanna et al (2010)[53]	Sardinia (Italy)	882	872	555 335
Nischwitz et al (2010)[54]	Germany	590	825	300 000
Jakkula et al (2010)[55]	Finland	68	136	297 343
IMSGC & WTCCC2 (2011)[27]	Various	9772	17 376	475 806
Patsopoulos et al (2011)[56]¶	Various	1453	2176	906 600
Matesanz et al (2012)[57]	Spain	296	801	130 903
Martinelli-Boneschi et al (2012)[58]	Italy	197	234	277 866

The table shows the number of cases and controls that were included in the screening phase and the number of SNPs typed. Several studies also tested imputed SNPs in the screening phase and most studies typed selected SNPs in a replication cohort. SNP=single-nucleotide polymorphism. IMSGC=International Multiple Sclerosis Genetics Consortium. WTCCC=Wellcome Trust Case Control Consortium. ANZgene=Australia and New Zealand Multiple Sclerosis Genetics Consortium. *Investigators screened trio families (ie, an affected individual and both parents). †This study only used non-synonymous SNPs. ‡This study was based on pooled DNA. §This study was a meta-analysis of previously unpublished data, data from the first IMSGC genome-wide association study,[47] and data reported by Baranzini and colleagues[51] for the GeneMSA consortium (a total of 2624 cases and 7220 controls). ¶This study was a meta-analysis of previously unpublished data, data reported in the 2009 meta-analysis,[39] and data reported by the ANZgene consortium (a total of 5545 cases and 12 153 controls).[52]

Table 2: **Genome-wide association studies in multiple sclerosis**

Furthermore, the genes that lie close to these 57 SNPs are significantly over-represented in immune-related terms in the Gene Ontology project (including lymphocyte proliferation and T-cell activation).[27] Although circumstantial, these data provide compelling evidence implicating the immune system in the aetiology of multiple sclerosis. Subsequently, two further SNPs with genome-wide significant evidence for association have also been identified through independent genome-wide association studies or related efforts, one (rs2150702) by a meta-analysis of earlier genome-wide association studies[56] and a second (rs6859219) identified by the focused follow-up of a limited number of SNPs found to be associated in other autoimmune disease genome-wide association studies.[60]

Immunochip follow-up

For the **100 genomes project** see http://www.1000genomes.org

The ability of genome-wide association studies to robustly screen most common variation by direct typing of just a few hundred thousand SNPs is crucially dependent on the extensive correlation between tightly linked variants (linkage disequilibrium) that is a characteristic feature of the human genome. Unfortunately, this same feature limits our ability to understand the results of such studies. Rather than individual associated variants, genome-wide association studies necessarily identify sets of highly correlated local variants, covering genomic intervals that often encompass many genes and regulatory sequences.[61] In this context, identification of the functionally relevant variant or variants that drive the association, or even the gene implicated in aetiology by the association, is challenging.

For **Immunobase** see http://www.immunobase.org

Maximising the usefulness of genome-wide association studies thus presents two problems: first, separation of the genuinely associated SNPs (true positives) from the multitude showing apparent evidence for association by chance alone (false positives); and second, fine mapping of the association intervals so as to establish their underlying allelic architecture. These challenges are inherent to the approach used in genome-wide association studies and both demand substantial additional genotyping. Fortunately, the substantial and highly significant overlap in findings from genome-wide association studies between autoimmune diseases[62] provided the opportunity to massively reduce the costs of follow-up efforts through cross-trait collaboration. By combining the extensive lists of potentially associated SNPs requiring replication that have been generated by genome-wide association studies from several autoimmune diseases, and doing the same for the overlapping lists of established loci requiring fine mapping, researchers of autoimmune diseases created an efficient instrument enabling simultaneous progress with both these tasks: the Immunochip.[63] This collaborative approach had the added advantage that it also allowed analysis for each trait of those SNPs that were of relevance in at least one other autoimmune disease, and therefore had a high probability of also being relevant in the trait of interest. After stringent quality control and the exclusion of individuals studied in previous genome-wide association studies of multiple sclerosis, the IMSGC analysed 161 311 SNPs from the chip in 14 498 cases and 24 091 controls, and thereby identified 150 potentially associated variants.[64] With use of the data from the 2011 genome-wide association study[27] and the meta-analysis of other genome-wide association studies[56] as replication (14 802 cases and 26 703 controls in total), the IMSGC established genome-wide significant association for 97 of these 150 variants: 49 corresponding to already established associations and 48 new associations, with 45 at new loci and 3 secondary signals identified after conditioning on already established associations.[64] Using imputation based on data from the 1000 genomes project[65] and Bayesian methods developed by the WTCCC,[66] fine mapping was attempted in 93 associated regions; single variants accounting for more than 50% of the posterior probability of association were identified for five regions. In one of these five regions (the region containing *IL2RA*), the investigators showed that association with the SNP rs2104286 drove the previously reported haplotype effect,[27] as had been suggested by findings from earlier genome-wide association studies[47] and follow-up studies.[67] Individual findings from the 2011 genome-wide association study[27] and Immunochip[64] studies provided on the Immunobase website.

In the 2011 genome-wide association study, 27 142 SNPs outside the major histocompatibility complex showed nominally significant ($p<0·05$) evidence of association; 2935 of these were ultimately included on the Immunochip and gave data passing quality control in the analysis of multiple sclerosis (figure 1). Overall, nearly 80% of these SNPs had the same allele over-represented among cases in both studies, and only 18 (<1%) had an Immunochip result that was nominally significant for an effect in the opposite direction. Even after excluding all the SNPs known to be associated with multiple sclerosis (and all those SNPs in linkage disequilibrium with these variants), there was still a highly significant concordance among the remaining variants, lending support to the notion that in well powered studies with large numbers of cases (≥10 000), a high proportion of the SNPs with p values falling just short of genome-wide significance (such as those with $p<10^{-5}$) are probably genuinely associated (ie, are probably true positives).[68] In short, these more powerful studies have not only identified many unequivocally associated variants but have also shown that much of the remaining heritability is probably explained by those variants that fall short of the established, but essentially arbitrary, $p<5\times10^{-8}$ threshold.[69,70] Meta-analysis of existing genome-wide association studies and efforts to screen less common variation with use of efficient methods such as the Exomechip are expected to expand the catalogue of associated variants during the next few years.

110 established variants associated with multiple sclerosis

Collectively these studies have identified 110 variants outside the major histocompatibility complex that are confidently associated with susceptibility to multiple sclerosis (appendix). According to the Variant Effect Predictor tool[71] on Ensembl (release 72), 15 of the 110 SNPs are themselves coding variants and a further 35 are in tight linkage disequilibrium ($r^2>0·8$) with coding variants (appendix). However, among the implicated coding variants, only 14 are missense and just 7 of these are predicted to be possibly harmful in at least one transcript, according to the prediction programs PolyPhen[72] and SIFT.[73] On the other hand, the HaploReg v2 tool[74] shows that the 110 associated SNPs coincide with chromatin features suggestive of regulatory function as identified in the ENCODE project[65] significantly more frequently than would be expected by chance alone; for example, in the B-lymphocyte lymphoblastoid cell line GM12878 there was significant enrichment for enhancers ($p=1·7\times10^{-5}$) and DNase I hypersensitivity sites ($p=1\times10^{-5}$). 13 of the associated SNPs are in regions of conserved sequence and a further 45 are in linkage disequilibrium with SNPs mapping to such regions.[74] Almost all of the SNPs (109 of 110) change at least one regulatory sequence motif (91 of 110) or are in linkage disequilibrium with variants that change such motifs (18 of 110). Furthermore, 27 of the 110 are in regions of transcription-factor binding (and another 51 are in linkage disequilibrium with such variants).[74] 15 of the associated SNPs are themselves quantitative trait loci for expression and a further 16 are in linkage disequilibrium with SNPs that are quantitative trait loci for expression (as listed in the University of Chicago eQTL database). Comparison of the 110 associated variants with cell-type-specific maps of active promotor regions, as defined by marks of trimethylation of histone H3 at lysine 4,[75] shows highly significant overlap in immune cells but no such overlap in any other cell types. Given that any regulatory effects of disease-associated variants are likely to be cell specific, this disparity adds further support to the

For the **University of Chicago eQTL database** see http://eqtl.uchicago.edu/Home.html

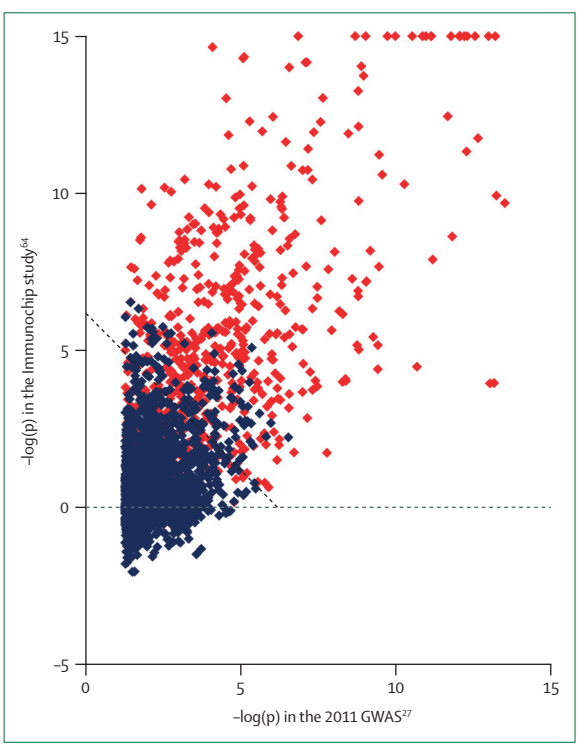

Figure 1: **Concordance of significance for SNPs reported in two genome-wide association studies**

Figure shows concordance of results for the 2935 non-MHC SNPs that were typed in both the Immunochip study[64] and the 2011 genome-wide association study;[27] and in the 2011 study showed nominally significant evidence of association ($p<0·05$). When the allele was over-represented in the 2011 genome-wide association study cases but under-represented in the Immunochip cases, the value is plotted as negative. Thus, points above the x-axis show SNPs with consistent case–control differences, whereas points below show SNPs for which the trend was in the opposite direction in the two studies. Red rhomboids show SNPs in linkage disequilibrium with any of the 110 known associations (total 654 of 2935), whereas blue rhomboids show SNPs not in linkage disequilibrium with any of the known associated variants. SNPs with –log(p)>15 in the Immunochip study are plotted at 15 to avoid expanding the axis. The dashed line shows the $p=10^{-5}$ threshold for p values, obtained by combining the independent p values from the two studies according to Fisher's method; SNPs above this line have a combined p value of $p<10^{-5}$, whereas those below have less evidence for association. There were no overlapping samples between the two studies. MHC=major histocompatibility complex. SNP=single-nucleotide polymorphism. GWAS=genome-wide association study.

For the **Exomechip** see http://genome.sph.umich.edu/wiki/Exome_Chip_Design

For the **Variant Effect Predictor tool** see http://www.ensembl.org

For the **HaploReg v2** tool see http://www.broadinstitute.org/mammals/haploreg

Panel 1: INUS conditions and risk factors

The associated alleles identified by genome-wide association studies are examples of what Mackie referred to as INUS conditions.[78] Realising that many outcomes have a plurality of causation, Mackie described how a factor might be insufficient to cause the outcome on its own but could be a non-redundant (essential) part of a set of factors which together resulted in the outcome; this set of factors being unnecessary, in the sense that many other sets of factors could also result in the effect, but sufficient to cause the outcome.[78] These risk factors are neither necessary nor sufficient but are contributory to risk.

Mackie used the example of a house fire. Suppose Mr Jones leaves his house to get a newspaper and in so doing leaves the cigarette he has been smoking on the arm of his sofa. The cigarette continues to burn and causes the sofa to catch fire, which then causes the house to burn down. We might reasonably, and correctly, conclude that the forgotten cigarette caused the house to burn down. However the situation is more complex than that suggested by this statement, because the fire required more than just leaving the cigarette; many cigarettes are forgotten and burn out harmlessly without causing a house fire. For this outcome to occur, various other factors were needed; the cigarette needed to be left on something flammable, and for long enough for the fire to catch hold. So in this sense the forgotten cigarette was insufficient on its own to cause the house fire but was a non-redundant part of the set of factors that resulted in the house fire; the house would not have burnt down if the cigarette had not been left. Of course forgotten cigarettes are not the only cause of house fires. Many other sets of circumstances can result in this outcome (eg, faulty electrical appliances, children playing with matches, or kitchen deep-fat fryers catching alight). In other words, not every house fire is the result of a forgotten cigarette, and thus from the perspective of house fires in general the set of circumstances described above is unnecessary because many other sets of factors can result in the same outcome, but this particular set of circumstances was sufficient to cause Mr Jones' house to burn down.

An informative analogy can be drawn between the development of a disease such as multiple sclerosis (the outcome) and this analysis. Genetic risk alleles (and environmental risk factors such as smoking or deficiency in vitamin D) are similar to the forgotten cigarette. For example, in the UK about a quarter of the population (approximately 15 million people) carry the *HLA-DRB1*15:01* allele (the main risk allele for multiple sclerosis), but only about 40 000 of these people actually develop multiple sclerosis; most forgotten cigarettes do not result in a house fire. Among the remaining 45 million people who do not carry *15:01, about 20 000 still develop multiple sclerosis; not all fires involve forgotten cigarettes, and other sets of circumstances can also result in a house fire. Genetic risk factors in the development of multiple sclerosis are thus like forgotten cigarettes in house fires—insufficient on their own to cause the disease, but for some individuals are a non-redundant part of a set of circumstances that are sufficient to result in the disease but are not essential for the development of the disease (ie, are unnecessary). Furthermore, an individual might forget a cigarette and then have their house burn down because of faulty electrical equipment. In other words, we cannot assume that *15:01 has necessarily played a part in the development of multiple sclerosis in patients who are positive for the allele; we can be sure it has an essential (non-redundant) part in some of these patients, but not all.

This analogy might also help with interpretation of the small ORs associated with allele associations for complex diseases. Overall, the risk of a house fire in the UK is low, but this risk is slightly higher in homes with a deep-fat fryer, because these devices can sometimes contribute to a fire (table 3). Numerically, the OR associated with this risk factor (owning a deep-fat fryer) is only 1·19, but the significance of the association between deep-fat fryer ownership and house fires is overwhelming ($p<10^{-73}$). If nothing was known about the mechanisms behind house fires, we would at least know that some of them involve deep-fat fryers in some way, and we would have an invaluable clue to the causes of house fires. It is interesting to speculate about what would happen in a population where everyone owns a deep-fat fryer. House fires would be more common in such a population, but in an association study the OR and p value would both be 1—ie, there would be no evidence of any association and there would seem to be no effect from deep-fat fryers. The small risk associated with common INUS conditions is a poor guide to their biological importance and the possible insights they might provide.

INUS=insufficient, non-redundant, unnecessary, sufficient. OR=odds ratio.

	With*	Without	Total
Homes	9·4 million	14·2 million	23·6 million
Deep-fat fryer fire	2800	0	2800
Other fires	14 280	21 420	35 700
Odds	0·0018	0·0015	0·0016

*The proportion of homes owning a deep-fat fryer is estimated; all other numbers are from the UK Office of National Statistics.

Table 3: **UK house fires in 2010 according to the ownership of a deep-fat fryer**

For the **UK Office of National Statistics** see http://www.statistics.gov.uk/

central role of the immune system in the aetiology of multiple sclerosis.

Although associations with common variants shown in genome-wide association studies could arise as a result of linkage disequilibrium with rare variants of larger effect,[76] available evidence suggests that such synthetic associations are uncommon,[77] and that for most complex diseases susceptibility is mainly polygenic. In this context, small ORs are expected and the size of effect attributable to a risk factor is generally a poor guide to the causative insight that association might provide

(panel 1).[69,78,79] Substantial data support the extent to which such findings can provide biological insight despite their slight effects on the statistically additive scale,[80] such as the involvement of autophagy in inflammatory bowel disease.[81]

For multiple sclerosis, the process of gaining meaningful biological insights from associated genetic variants is in its infancy. However, despite the many challenges, encouraging progress has already been made with functional analysis of associated variants, showing that susceptibility alleles increase the risk of multiple sclerosis through various mechanisms. For three of the associated variants (rs6897932 in *IL7R*,[46] rs2104286 in *IL2R*,[82] and rs1800693 in *TNFRSF1A*[83]) the risk allele has been established to increase the concentration of the soluble form of the implicated receptor and thereby inhibit signalling. On the other hand, the risk allele of rs6677309 has been shown to result in reduced expression of the costimulatory CD58, an effect that is predicted to result in dysfunction of regulatory T cells because of reduced FOXP3 expression.[84] Additionally, at rs34536443 the protective minor allele has been shown to reduce the activity of TYK2 and thereby favour the secretion of cytokines from T-helper-2 cells.[85] As functional annotation expands, these efforts will likely converge on the pathways and cell subtypes that are most important.

Secondary phenotypes

Analysis of clinical features in families in which more than one member has multiple sclerosis suggests that genetic factors probably affect the course of the disease.[86] In this context, three genome-wide association studies[51,87,88] have specifically investigated clinical features as their primary endpoint, but unfortunately no genome-wide significant association emerged from these modestly powered efforts. However, genes for calcium and glutamate signalling were enriched among the potentially associated genes identified in these studies.[88,89] If these potential associations are verified in larger studies, then in principle functional analysis of these pathways could lead to the identification of drugs that can affect the course of the disease.

A secondary analysis of the susceptibility genome-wide association studies done by the IMSGC and WTCCC2[27] showed association between age at onset and carriage of the *HLA-DRB1*15:01* allele (with each copy carried advancing the age at onset by about 11 months on average), but did not show any convincing association with clinical course or the severity of the disease as measured by the Multiple Sclerosis Severity Score.[90] Because at least some of the genetic factors affecting the clinical features of the disease will probably exert effects large enough to be detectable by genome-wide association studies, one possible reason for the rather disappointing yield from studies into disease phenotype could be shortcomings in the quality with which clinical features of the disease are measured. Perhaps studies into genetic factors relevant to relapses would have more power if relapse activity were treated as a quantitative trait rather than dichotomisation of patients into those with and those without clinically apparent relapses (ie, bout-onset disease or primary progressive disease); however, how to quantify relapse activity to enable such an analysis remains unclear.

Genome-wide association studies of phenotypic aspects of multiple sclerosis (eg, oligoclonal bands[91–94] and MRI[95]) provide another way to gain insight into underlying biology. As these and related efforts expand, genes relevant to neurobiological aspects of the disease—eg, axon integrity and myelin regeneration (remyelination)—will probably emerge. Although none of the 110 susceptibility variants that have been identified so far implicate such genes (other than perhaps *MAPK*, which has been implicated in proliferation and differentiation of oligodendrocyte progenitors[96]), many inflammatory mediators might also beneficially contribute to remyelination.[97]

Missing heritability

The associated loci identified so far account for only about a quarter of the heritability reported in multiple sclerosis, leaving an obvious question about what determines the

***Panel 2:* Some key parameters explained**

The allele frequency spectrum is generally considered in essentially arbitrary but operationally useful divisions: alleles with a frequency of <0·1–0·5% being described as rare; alleles with a frequency of >5–10% being described as common; and the group in between usually being referred to as low frequency.

The consequences of an associated variant can be considered from two perspectives; first in terms of its effects on risk to an individual carrying the allele and second in terms of its effects on the prevalence of the disease in a population in which it is segregating. Because genome-wide association studies invariably use case-control methods, the effects on an individual are usually established by logistic regression and expressed as ORs (the relative increase in the odds of developing the disease for each copy of the allele carried). Alternatively, from the population perspective the effect of a risk allele is most conveniently measured in terms of its population attributable risk, which is the proportion of cases that would have been prevented if that allele had been removed from the population. Because the associated variants identified by genome-wide association studies are common—ie, carried by many people—they often have a relatively high population attributable risk even though the OR, the effect on any individual, is small. For example, a risk allele with an OR of 1·2 and a minor allele frequency of 50% has a population attributable risk of 13%. That is, if this allele were removed from the population, the frequency of the disease would be reduced by 13%.

Although the term genome-wide association study tends to imply that the whole genome has been assessed, in fact only a proportion of the potentially relevant variation is adequately tested. With present technology, a genome-wide association study of several thousand cases and controls effectively screens a large proportion (generally more than 80%) of potentially relevant common variations, but has very limited power to identify relevant low-frequency variants and almost no power to identify relevant rare variants (unless they exert very large effects). Similarly, some common variants are not well tagged by single-nucleotide polymorphisms included on the chip (not in linkage disequilibrium) and are therefore effectively untested. The approach is also relatively insensitive to alleles exerting recessive or sex linked effects.

OR=odds ratio.

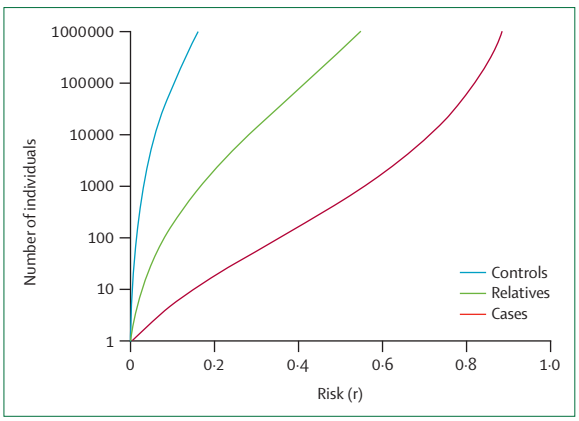

Figure 2: **The average number of individuals that need to be screened to identify one person with a genetically established risk greater than r**
The numbers are shown separately for individuals selected from the general population (blue), the first-degree relatives of cases (green), and cases (red). The curves are based on screening all 110 known SNPs[64] and the four established HLA risk alleles;[27] the numbers were calculated[107] assuming a total sibling recurrence risk of 6·3 and a lifetime risk of the disease of 0·002.[108] Collectively, these known genetic risk factors account for 28% of the sibling recurrence risk.[64] For example, to identify one individual with a risk of ≥20%, more than 3·5 million people in the general population would have to be screened. Alternatively, more than 2000 first-degree relatives or 18 identical twins of affected individuals would have to be screened. If non-genetic risk factors (eg, sex, antibody titre for Epstein Barr virus antigen, concentrations of vitamin D, and smoking habits)[109] were also evaluated as part of the screening process, the number needing screening would still be prohibitively large. For a single individual with a risk of ≥20%, more than 56 000 individuals from the general population, 180 first-degree relatives, or 6 identical twins would have to be screened.

remainder. Much of the remainder (perhaps half) is probably so-called phantom heritability—ie, resulting from as-yet-undefined interactions between risk factors.[18] The remainder probably relates to risk alleles that are yet to be discovered,[70] some of which will be common, and might emerge in larger genome-wide association studies or through meta-analyses of existing data. Although many of these as-yet-unidentified alleles are probably of lower frequency, the absence of linkage means that they probably do not exert any effects with an OR of much more than 6, even if they are rare.[21] Alternative strategies will be necessary to identify such variants.[61] Unfortunately, attempts to identify relevant lower-frequency variants have not been successful. In 2010, Surolia and colleagues[98] used exome sequencing to identify rare variants in the *SIAE* gene, and confirmed that these variants had a profound effect on the function of the gene; after genotyping a small number of cases and controls, they suggested that these variants affected the risk of development of various autoimmune diseases, including multiple sclerosis. However, analysis in a much larger number of participants (>66 000) showed that in fact none of these *SIAE* variants were associated with increased susceptibility to multiple sclerosis or any other human autoimmune disease.[99] These data emphasise the important point that evidence of a substantial functional effect does not trump evidence for association. The 1000 genomes project has shown that on average each person carries several hundred rare or low-frequency variants resulting in profound changes in gene function;[100] therefore, the importance of any such variants in terms of the aetiology of a disease cannot be presumed in the absence of convincing evidence for association. The application of exome sequencing in multiplex families has so far been disappointing; in view of the absence of linkage in multiple sclerosis,[22,28] this finding is perhaps not surprising.[21] The promising initial reports that the rare variants in *CYP27B1* that account for the autosomal recessive condition of type 1 vitamin-D-dependent rickets also increased the risk of multiple sclerosis in heterozygote carriers[101] could not be substantiated.[102–104] Preliminary evidence for association to a rare variant of *TYK2* has been reported,[105] but remains to be confirmed. Resequencing of a small number of candidate genes in many individuals (including more than 3000 individuals with multiple sclerosis) suggests that rare variants of large effect are uncommon in autoimmune disease and are unlikely to contribute much to the heritability of multiple sclerosis.[106]

Although only a small proportion of heritability is currently explained, the population-attributable risk accounted for by the known risk alleles is nearly 100% (panel 2); presumably suggesting that the development of disease in almost every patient with multiple sclerosis has involved at least some of the currently identified risk variants. In this context it might seem reasonable to expect that collectively these associated variants might allow the identification of individuals at high risk and thereby enable, for example, the advantageous application of aggressive or costly preventive strategies. Unfortunately, the clinical value of such prediction is probably small because most cases arise among the large number of individuals at slightly increased risk, and only a very small fraction occurs in the few individuals at high risk (figure 2)—the so-called prevention paradox.[110]

Alternative methods for systematically screening the genome have also been employed in the search for relevant genes, including admixture scanning[111] and identity-by-descent (IBD) mapping.[112] In the former, researchers typed ancestrally informative SNPs[113] in a large number of African American cases (605) and controls (1043) and thereby identified a region on chromosome 1 where European ancestry was significantly more common in cases.[111] In the IBD study, a reanalysis of published data from genome-wide association studies found some evidence that certain SNP haplotypes (sets of linked common SNP alleles) from chromosome 19 might be more common in cases than in controls. Neither of these approaches has yet led to the unequivocal identification of a new susceptibility locus.

Conclusions and future directions

Each of the associated genetic variants identified so far has the potential to provide crucial insight into aetiology of multiple sclerosis, and thereby promote the

Search strategy and selection criteria

We searched PubMed for articles published between 1993 and December, 2013, with terms such as "multiple sclerosis", "gene*", "genomewide association", "linkage", and "association". We also identified references from relevant articles and the online Catalog of Published Genome-Wide Association Studies. We restricted our search to articles published in English. The final reference list was generated on the basis of relevance to the topics covered in this Review.

development of a rational therapy that is both safe and effective. The discovery that most, if not all, of these variants seem to exert their effects by affecting tissue-specific gene expression has exposed just how little is known about the way in which regulatory information is encoded in the genome—an information gap that undoubtedly represents one of the largest barriers to translation of these associations into biologically relevant knowledge. However, if complex diseases such as multiple sclerosis result from an acquired quantitative change in otherwise normal physiology, then perhaps correction of these perturbations will prove to be easier than overcoming of the qualitative changes that characterise mendelian (ie, monogenic) diseases. In short, the complexity of the genetics might not correlate with the ease of development of effective therapeutic interventions after the biological outcomes of these variants are established.

Contributors

All authors contributed to the preparation of this Review.

Declaration of interests

We declare no competing interests.

Acknowledgments

SS, RJMF, and MB are supported by the Cambridge NIHR Biomedical Research Centre. RJMF is supported by the UK Multiple Sclerosis Society.

References

1 Compston A, Confavreux C, Lassmann H, et al. McAlpine's multiple sclerosis, 4th edn. London: Churchill Livingstone, 2006.
2 Compston A, Coles A. Multiple sclerosis. *Lancet* 2008; **372:** 1502–17.
3 Orton SM, Herrera BM, Yee IM, et al, and the Canadian Collaborative Study Group. Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol* 2006; **5:** 932–36.
4 Robertson NP, Fraser M, Deans J, Clayton D, Walker N, Compston DA. Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain* 1996; **119:** 449–55.
5 O'Gorman C, Lin R, Stankovich J, Broadley SA. Modelling genetic susceptibility to multiple sclerosis with family data. *Neuroepidemiology* 2013; **40:** 1–12.
6 Simpson S Jr, Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J Neurol Neurosurg Psychiatry* 2011; **82:** 1132–41.
7 Ahlgren C, Odén A, Lycke J. High nationwide prevalence of multiple sclerosis in Sweden. *Mult Scler* 2011; **17:** 901–08.
8 Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, and the Canadian Collaborative Study Group. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc Natl Acad Sci USA* 2003; **100:** 12877–82.
9 Ebers GC, Sadovnick AD, Risch NJ, and the Canadian Collaborative Study Group. A genetic basis for familial aggregation in multiple sclerosis. *Nature* 1995; **377:** 150–51.
10 Ebers GC, Yee IM, Sadovnick AD, Duquette P, and the Canadian Collaborative Study Group. Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. *Ann Neurol* 2000; **48:** 927–31.
11 Dean G, McLoughlin H, Brady R, Adelstein AM, Tallett-Williams J. Multiple sclerosis among immigrants in Greater London. *BMJ* 1976; **1:** 861–64.
12 Dyment DA, Yee IM, Ebers GC, Sadovnick AD, and the Canadian Collaborative Study Group. Multiple sclerosis in stepsiblings: recurrence risk and ascertainment. *J Neurol Neurosurg Psychiatry* 2006; **77:** 258–59.
13 Dean G, Yeo TW, Goris A, et al. HLA-DRB1 and multiple sclerosis in Malta. *Neurology* 2008; **70:** 101–05.
14 Sawcer S, Ban M, Wason J, Dudbridge F. What role for genetics in the prediction of multiple sclerosis? *Ann Neurol* 2010; **67:** 3–10.
15 Guo SW. Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am J Hum Genet* 1998; **63:** 252–58.
16 Sadovnick AD, Dyment D, Ebers GC. Genetic epidemiology of multiple sclerosis. *Epidemiol Rev* 1997; **19:** 99–106.
17 Prokopenko I, Montomoli C, Ferrai R, et al. Risk for relatives of patients with multiple sclerosis in central Sardinia, Italy. *Neuroepidemiology* 2003; **22:** 290–96.
18 Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 2012; **109:** 1193–98.
19 Al-Chalabi A, Lewis CM. Modelling the effects of penetrance and family size on rates of sporadic and familial disease. *Hum Hered* 2011; **71:** 281–88.
20 Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; **6:** 109–18.
21 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273:** 1516–17.
22 Willer CJ, Dyment DA, Cherny S, et al. A genome-wide scan in forty large pedigrees with multiple sclerosis. *J Hum Genet* 2007; **52:** 955–62.
23 Modin H, Masterman T, Thorlacius T, et al. Genome-wide linkage screen of a consanguineous multiple sclerosis kinship. *Mult Scler* 2003; **9:** 128–34.
24 Haghighi S, Andersen O, Nilsson S, Rydberg L, Wahlström J. A linkage study in two families with multiple sclerosis and healthy members with oligoclonal CSF immunopathy. *Mult Scler* 2006; **12:** 723–30.
25 Dyment DA, Cader MZ, Herrera BM, et al. A genome scan in a single pedigree with a high prevalence of multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2008; **79:** 158–62.
26 Olerup O, Hillert J. HLA class II-associated genetic susceptibility in multiple sclerosis: a critical evaluation. *Tissue Antigens* 1991; **38:** 1–15.
27 The International Multiple Sclerosis Genetics Consortium, and the Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; **476:** 214–19.
28 The International Multiple Sclerosis Genetics Consortium. A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet* 2005; **77:** 454–67.
29 Naito S, Namerow N, Mickey MR, Terasaki PI. Multiple sclerosis: association with *HL-A3*. *Tissue Antigens* 1972; **2:** 1–4.
30 Jersild C, Hansen GS, Svejgaard A, Fog T, Thomsen M, Dupont B. Histocompatibility determinants in multiple sclerosis, with special reference to clinical course. *Lancet* 1973; **302:** 1221–25.
31 Horton R, Gibson R, Coggill P, et al. Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. *Immunogenetics* 2008; **60:** 1–18.
32 de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006; **38:** 1166–72.
33 Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* 2011; **27:** 968–72.
34 Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet* 2008; **82:** 48–56.

For the **Catalog of Published Genome-Wide Association Studies** see http://www.genome.gov/26525384

35 Patsopoulos NA, Barcellos LF, Hintzen RQ, et al, and the IMSGC and the ANZgene. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet* 2013; **9:** e1003926.

36 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:** 1125–37.

37 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:** 237–47.

38 Yeo TW, De Jager PL, Gregory SG, et al, and the International Multiple Sclerosis Genetics Consortium (IMSGC). A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Ann Neurol* 2007; **61:** 228–36.

39 De Jager PL, Jia X, Wang J, et al, and the International MS Genetics Consortium. Meta-analysis of genome scans and replication identify *CD6*, *IRF8* and *TNFRSF1A* as new multiple sclerosis susceptibility loci. *Nat Genet* 2009; **41:** 776–82.

40 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:** 2907–16.

41 Dyment DA, Herrera BM, Cader MZ, et al. Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum Mol Genet* 2005; **14:** 2019–26.

42 Barcellos LF, Sawcer S, Ramsay PP, et al. Heterogeneity at the *HLA-DRB1* locus and risk for multiple sclerosis. *Hum Mol Genet* 2006; **15:** 2813–24.

43 Lincoln MR, Ramagopalan SV, Chao MJ, et al. Epistasis among *HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1* loci determines multiple sclerosis susceptibility. *Proc Natl Acad Sci USA* 2009; **106:** 7542–47.

44 Field J, Browning SR, Johnson LJ, et al, and the Australia and New Zealand Multiple Sclerosis Genetics Consortium. A polymorphism in the *HLA-DPB1* gene is associated with susceptibility to multiple sclerosis. *PLoS One* 2010; **5:** e13454.

45 Hauser MA, Li YJ, Takeuchi S, et al. Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum Mol Genet* 2003; **12:** 671–77.

46 Gregory SG, Schmidt S, Seth P, et al, and the Multiple Sclerosis Genetics Group. Interleukin 7 receptor alpha chain (*IL7R*) shows allelic and functional association with multiple sclerosis. *Nat Genet* 2007; **39:** 1083–91.

47 The International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007; **357:** 851–62.

48 The Wellcome Trust Case Control Consortium, the Australo-Anglo-American Spondylitis Consortium (TASC), the Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee, and the Breast Cancer Susceptibility Collaboration (UK). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007; **39:** 1329–37.

49 Comabella M, Craig DW, Camiña-Tato M, et al, and the BiomarkerMS Study Group. Identification of a novel risk locus for multiple sclerosis at 13q31.3 by a pooled genome-wide scan of 500,000 single nucleotide polymorphisms. *PLoS One* 2008; **3:** e3490.

50 Aulchenko YS, Hoppenbrouwers IA, Ramagopalan SV, et al. Genetic variation in the KIF1B locus influences susceptibility to multiple sclerosis. *Nat Genet* 2008; **40:** 1402–03.

51 Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009; **18:** 767–78.

52 Bahlo M, Booth D, Broadley S, et al, and the Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009; **41:** 824–28.

53 Sanna S, Pitzalis M, Zoledziewska M, et al. Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis. *Nat Genet* 2010; **42:** 495–97.

54 Nischwitz S, Cepok S, Kroner A, et al. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. *J Neuroimmunol* 2010; **227:** 162–66.

55 Jakkula E, Leppä V, Sulonen AM, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in *STAT3* gene. *Am J Hum Genet* 2010; **86:** 285–91.

56 Patsopoulos NA, Esposito F, Reischl J, et al, and the Bayer Pharma MS Genetics Working Group, the Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist, the ANZgene Consortium, the GeneMSA, and the International Multiple Sclerosis Genetics Consortium. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 2011; **70:** 897–912.

57 Matesanz F, González-Pérez A, Lucas M, et al. Genome-wide association study of multiple sclerosis confirms a novel locus at 5p13.1. *PLoS One* 2012; **7:** e36140.

58 Martinelli-Boneschi F, Esposito F, Brambilla P, et al. A genome-wide association study in progressive multiple sclerosis. *Mult Scler* 2012; **18:** 1384–94.

59 The International Multiple Sclerosis Genetics Consortium. *MANBA, CXCR5, SOX8, RPS6KB1* and *ZBTB46* are genetic risk loci for multiple sclerosis. *Brain* 2013; **136:** 1778–82.

60 Lill CM, Schjeide BM, Graetz C, et al. Genome-wide significant association of *ANKRD55* rs6859219 and multiple sclerosis risk. *J Med Genet* 2013; **50:** 140–43.

61 McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9:** 356–69.

62 Cotsapas C, Voight BF, Rossin E, et al, and the FOCiS Network of Consortia. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011; **7:** e1002254.

63 Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther* 2011; **13:** 101.

64 The International Multiple Sclerosis Genetics Consortium (IMSGC). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 2013; **45:** 1353–60.

65 Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, and the ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489:** 57–74.

66 Maller JB, McVean G, Byrnes J, et al, and the Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* 2012; **44:** 1294–301.

67 International Multiple Sclerosis Genetics Consortium (IMSGC). Refining genetic associations in multiple sclerosis. *Lancet Neurol* 2008; **7:** 567–69.

68 Sawcer S. The complex genetics of multiple sclerosis: pitfalls and prospects. *Brain* 2008; **131:** 3118–31.

69 Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42:** 565–69.

70 The International Multiple Sclerosis Genetics Consortium (IMSGC). Evidence for polygenic susceptibility to multiple sclerosis—the shape of things to come. *Am J Hum Genet* 2010; **86:** 621–25.

71 McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010; **26:** 2069–70.

72 Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7:** 248–49.

73 Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; **7:** e46688.

74 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40:** D930–34.

75 Trynka G, Sandor C, Han B, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet* 2013; **45:** 124–30.

76 Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8:** e1000294.

77 Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* 2011; **9:** e1000580.

78 Mackie JL. The Cement of the Universe: a study in Causation. Oxford: Clarendon Press, 1988.

79 Hill AB. The Environment and Disease: association or Causation? *Proc R Soc Med* 1965; **58:** 295–300.

80 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90:** 7–24.

81 Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007; **39:** 596–604.

82 Maier LM, Anderson DE, Severson CA, et al. Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *J Immunol* 2009; **182:** 1541–47.

83 Gregory AP, Dendrou CA, Attfield KE, et al. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 2012; **488:** 508–11.

84 De Jager PL, Baecher-Allan C, Maier LM, et al. The role of the *CD58* locus in multiple sclerosis. *Proc Natl Acad Sci USA* 2009; **106:** 5264–69.

85 Couturier N, Bucciarelli F, Nurtdinov RN, et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* 2011; **134:** 693–703.

86 Hensiek AE, Seaman SR, Barcellos LF, et al. Familial effects on the clinical course of multiple sclerosis. *Neurology* 2007; **68:** 376–83.

87 Brynedal B, Wojcik J, Esposito F, et al. *MGAT5* alters the severity of multiple sclerosis. *J Neuroimmunol* 2010; **220:** 120–24.

88 International Multiple Sclerosis Genetics Consortium. Genome-wide association study of severity in multiple sclerosis. *Genes Immun* 2011; **12:** 615–25.

89 Baranzini SE, Srinivasan R, Khankhanian P, et al. Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. *Brain* 2010; **133:** 2603–11.

90 Roxburgh RH, Seaman SR, Masterman T, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology* 2005; **64:** 1144–51.

91 Buck D, Albrecht E, Aslam M, et al, and the International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium. Genetic variants in the immunoglobulin heavy chain locus are associated with the IgG index in multiple sclerosis. *Ann Neurol* 2013; **73:** 86–94.

92 Goris A, van Setten J, Diekstra F, et al, and The International Multiple Sclerosis Genetics Consortium and the The Australia and New Zealand MS Genetics Consortium. No evidence for shared genetic basis of common variants in multiple sclerosis and amyotrophic lateral sclerosis. *Hum Mol Genet* 2013; published online Nov 13. DOI:10.1093/hmg/ddt574.

93 Mero IL, Gustavsen MW, Sæther HS, et al, and the International Multiple Sclerosis Genetics Consortium. Oligoclonal band status in Scandinavian multiple sclerosis patients is associated with specific genetic risk alleles. *PLoS One* 2013; **8:** e58352.

94 Leone MA, Barizzone N, Esposito F, et al, and the International Multiple Sclerosis Genetics Consortium, the Wellcome Trust Case Control Consortium 2, the PROGEMUS Group, and the PROGRESSO Group. Association of genetic markers with CSF oligoclonal bands in multiple sclerosis patients. *PLoS One* 2013; **8:** e64408.

95 Gourraud PA, Sdika M, Khankhanian P, et al. A genome-wide association study of brain lesion distribution in multiple sclerosis. *Brain* 2013; **136:** 1012–24.

96 Ishii A, Furusho M, Bansal R. Sustained activation of ERK1/2 MAPK in oligodendrocytes and schwann cells enhances myelin growth and stimulates oligodendrocyte progenitor expansion. *J Neurosci* 2013; **33:** 175–86.

97 Setzu A, Lathia JD, Zhao C, et al. Inflammation stimulates myelination by transplanted oligodendrocyte precursor cells. *Glia* 2006; **54:** 297–303.

98 Surolia I, Pirnie SP, Chellappa V, et al. Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* 2010; **466:** 243–47.

99 Hunt KA, Smyth DJ, Balschun T, et al, and the Type 1 Diabetes Genetics Consortium, the UK Inflammatory Bowel Disease (IBD) Genetics Consortium, and the Wellcome Trust Case Control Consortium. Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. *Nat Genet* 2012; **44:** 3–5.

100 Abecasis GR, Altshuler D, Auton A, et al, and the 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467:** 1061–73.

101 Ramagopalan SV, Dyment DA, Cader MZ, et al. Rare variants in the *CYP27B1* gene are associated with multiple sclerosis. *Ann Neurol* 2011; **70:** 881–86.

102 Cortes A, Field J, Glazov EA, Hadler J, Stankovich J, Brown MA, and the ANZgene Consortium. Resequencing and fine-mapping of the chromosome 12q13-14 locus associated with multiple sclerosis refines the number of implicated genes. *Hum Mol Genet* 2013; **22:** 2283–92.

103 Ban M, Caillier S, Mero IL, et al. No evidence of association between mutant alleles of the CYP27B1 gene and multiple sclerosis. *Ann Neurol* 2013; **73:** 430–32.

104 Barizzone N, Pauwels I, Luciano B, et al. No evidence for a role of rare *CYP27B1* functional variations in multiple sclerosis. *Ann Neurol* 2013; **73:** 433–37.

105 Dyment DA, Cader MZ, Chao MJ, et al. Exome sequencing identifies a novel multiple sclerosis susceptibility variant in the TYK2 gene. *Neurology* 2012; **79:** 406–11.

106 Hunt KA, Mistry V, Bockett NA, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* 2013; **498:** 232–35.

107 Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet* 2009; **5:** e1000540.

108 Hemminki K, Li X, Sundquist J, Hillert J, Sundquist K. Risk for multiple sclerosis in relatives and spouses of patients diagnosed with autoimmune and related conditions. *Neurogenetics* 2009; **10:** 5–11.

109 Sawcer S, Wason J. Risk in complex genetics: "all models are wrong but some are useful". *Ann Neurol* 2012; **72:** 502–09.

110 Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985; **14:** 32–38.

111 Reich D, Patterson N, De Jager PL, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 2005; **37:** 1113–18.

112 Lin R, Charlesworth J, Stankovich J, Perreau VM, Brown MA, Taylor BV, and the ANZgene Consortium. Identity-by-descent mapping to detect rare variants conferring susceptibility to multiple sclerosis. *PLoS One* 2013; **8:** e56379.

113 Smith MW, Patterson N, Lautenberger JA, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 2004; **74:** 1001–13.