**Multiple sclerosis genetics**

Stephen Sawcer, Robin J M Franklin, Maria Ban

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. Half of all patients need assistance with mobility within 20 years of diagnosis, and 50% of patients eventually develop substantial cognitive deficits. The disease most often starts between 20 and 40 years of age, and affects women more frequently than men. 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. 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, and second, that the disease varies greatly in frequency worldwide. Although neither of these findings necessarily implies an exclusive role for either genetic or environmental factors, supplementary studies in informative subgroups (eg, twins, adoptees, conjugal pairs, and migrant individuals) suggest that familial clustering is determined mainly by genetic factors, whereas regional variation in prevalence results from the effects of both genetic and population-level environmental risk factors. The imprecision 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). The tendency for familial recurrence risk to fall geometrically with the degree of relatedness 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. 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. 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.

**Linkage**

Although segregation analysis has little power to establish the relative importance of the two main allelic models, linkage analysis can be useful. 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. Furthermore, the absence of any linkage in the few larger-than-average families that have been reported 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 3.2,26,27 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 significant evidence for linkage outside the same region).29,30 The major histocompatibility complex has been recognised for several decades,29,30 the extreme polymorphism and extensive linkage disequilibrium (ie, correlation between linked variants) that characterise this gene-dense region31 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 data2,31–34 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*15:01.23,25 Furthermore, the long-suspected class-I protective effect36–39 has been confirmed,21,35 and shown to be driven mainly by the HLA-A*02:01 allele.21,35 Conditional analysis has confirmed the relevance of HLA-DRB1*03:01, HLA-DRB1*13:03, and HLA-DPB1*03:01 alleles (table 1), and suggested that risk might also be affected by non-HLA genes from this region.21 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.21

### 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, the extreme polymorphism and extensive linkage disequilibrium (ie, correlation between linked variants) that characterise this gene-dense region 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 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. Furthermore, the long-suspected class-I protective effect has been confirmed, and shown to be driven mainly by the HLA-A*02:01 allele. Conditional analysis has confirmed the relevance of HLA-DRB1*03:01, HLA-DRB1*13:03, and HLA-DPB1*03:01 alleles (table 1), and suggested that risk might also be affected by non-HLA genes from this region. 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. 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.

### 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) and confirmed by typing large numbers of cases and controls. 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. In total, 14 genome-wide association studies have now been completed in multiple sclerosis (table 2), and 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]) 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).

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=3·8×10−9). 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.

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

<table>
<thead>
<tr>
<th>Allele</th>
<th>OR</th>
<th>RAF</th>
<th>-log(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DRB1*15:01</td>
<td>3·10</td>
<td>13·3%</td>
<td>320·0</td>
</tr>
<tr>
<td>HLA-A*02:01</td>
<td>1·37</td>
<td>25·9%</td>
<td>22·0</td>
</tr>
<tr>
<td>HLA-DRB1<em>03:01–DQB1</em>02:01</td>
<td>1·26</td>
<td>14·6%</td>
<td>9·4</td>
</tr>
<tr>
<td>HLA-DRB1<em>13:03–DQB1</em>03:01</td>
<td>2·40</td>
<td>0·9%</td>
<td>10·9</td>
</tr>
<tr>
<td>rs9277535_G‡</td>
<td>1·28</td>
<td>24·9%</td>
<td>26·4</td>
</tr>
</tbody>
</table>

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 and the Wellcome Trust Case Control Consortium. 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-DRB1*03:01.

For the Multiple Sclerosis GWAS Browser see http://wattle.well.ox.ac.uk/wtccc2/external/imvs/
investigators screened trio families (ie, an affected individual and both parents). 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. Questions: 1) What is the purpose of genome-wide association studies? 2) What are the limitations of such studies? 3) How do follow-up efforts through cross-trait collaboration help in reducing costs and improving results? 4) Why is replication important in genome-wide association studies? 5) What are the implications of the results from genome-wide association studies on the aetiology of multiple sclerosis? 6) What is the significance of the findings for the Gene Ontology project and lymphocyte proliferation and T-cell activation? 7) How do the findings support the role of the immune system in the pathogenesis of multiple sclerosis? 8) What are the implications of the findings for the extensive correlation between tightly linked variants and the extent of the association intervals? 9) How does the Immunochip contribute to the follow-up efforts through cross-trait collaboration? 10) What are the challenges associated with the use of imputation based on data from the 1000 genomes project and Bayesian methods developed by the WTCCC? 11) What are the limitations and advantages of the Immunochip approach? 12) How do the findings from genome-wide association studies impact our understanding of the genetic basis of multiple sclerosis? 13) What are the implications of the findings for the development of targeted therapies for multiple sclerosis? 14) How do the findings contribute to the development of precision medicine for multiple sclerosis? 15) What are the ethical considerations associated with the use of genome-wide association studies in multiple sclerosis research? 16) How do the findings impact the patient care and management of multiple sclerosis? 17) What are the implications of the findings for the development of personalized medicine for multiple sclerosis? 18) How do the findings contribute to the understanding of the relationship between genetics and the environment in the development of multiple sclerosis? 19) What are the implications of the findings for the development of genetic testing for multiple sclerosis? 20) How do the findings contribute to the development of new therapeutic targets for multiple sclerosis?
In the 2011 genome-wide association study, 2742 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 (≥10000), 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).69 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=5×10^{-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 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 methods such as the Exomechip are expected to expand the catalogue of associated variants during the next few years.

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

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

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 study64 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^{-8}, whereas those below have less evidence for association. These were no overlapping samples between the two studies. MHC= major histocompatibility complex. SNP=single-nucleotide polymorphism.

GWAS=genome-wide association study.

GWAS=genome-wide association study.

GWAS=genome-wide association study.
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. 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. 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 the causative insight that association might provide, small ORs are expected and the size of effect 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, available evidence suggests that such synthetic associations are uncommon, 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.
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. In this context, three genome-wide association studies 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. 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 activity of TYK2 and thereby favour the secretion of protective minor allele has been shown to reduce the activity to enable such an analysis remains unclear.

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, rs2104286 in IL2R, and rs1800693 in TNFRSF1A) 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 rs6673909 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. 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. As functional annotation expands, these efforts will likely converge on the pathways and cell subtypes that are most important.

*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 treatment 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 and MRI) 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), many inflammatory mediators might also beneficially contribute to remyelination.

**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.
The 1000 genomes project has shown that most common variants (which account for 

percentage of heritability) are unlikely to contribute much to the heritability of multiple sclerosis, and are unlikely to contribute much to the heritability of 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 carriers31 could not be substantiated.30,31 Preliminary evidence for association to a rare variant of TYK2 has been reported,33 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.33

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.111

Alternative methods for systematically screening the genome have also been employed in the search for relevant genes, including admixture scanning34 and identity-by-descent (IBD) mapping.35 In the former, researchers typed ancestrally informative SNPs35 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.35 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 identification of individuals at high risk. Mult\n
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 SNPs64 and the four established HLA risk alleles;64 the numbers were calculated64 assuming a total sibling recurrence risk of 6·3 and a lifetime risk of the disease of 0·002.106 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)110 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.

The remainder probably relates to risk alleles that are yet to be discovered,33 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.32 Alternative strategies will be necessary to identify such variants.32 Unfortunately, attempts to identify relevant lower-frequency variants have not been successful. In 2010, Surolia and colleagues38 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.39 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,104 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.32 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 carriers31 could not be substantiated.30,31 Preliminary evidence for association to a rare variant of TYK2 has been reported,33 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.33

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.111

Alternative methods for systematically screening the genome have also been employed in the search for relevant genes, including admixture scanning34 and identity-by-descent (IBD) mapping.35 In the former, researchers typed ancestrally informative SNPs35 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.35 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
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.

Acknowledgments

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

References

6 Simpson S Jr, Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude variants seem to exert their effects by affecting tissue-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.

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


59 The International Multiple Sclerosis Genetics Consortium. MANBA, CXC5, SOX8, RP56KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. *Brain* 2013; 136: 1778–82.


