Genome-wide association studies in multiple sclerosis: lessons and future prospects

Anu Kemppinen, Stephen Sawcer and Alastair Compston

Advance Access publication date 10 February 2011

Abstract
Multiple sclerosis (MS) is an inflammatory neurodegenerative disease with complex aetiology. A haplotype within the major histocompatibility region is the major risk factor for MS, but despite clear evidence for a genetic component additional risk variants were not identified until the recent advent of genome-wide association studies (GWAS). At present, 10 GWAS have been conducted in MS, and together with follow-up studies these have confirmed 16 loci with genome-wide significance. Many of these common risk variants are located at or near genes with central immunological functions and the majority are associated with other autoimmune diseases. However, evidence from pathway analyses on more modestly associated variants also supports the involvement of neurological genes. Although the mechanisms by which the associated variants exert their effects are still poorly understood, some have been shown to correlate with expression of nearby genes. Further studies are required to define the functionally relevant variants in the identified regions and to investigate their effects at the molecular and cellular level. Finally, many genetic risk variants for MS remain to be identified. In order to expose some of the loci with more modest effects, a GWAS in nearly 10 000 MS patients has recently been completed.

Keywords: autoimmune disease; genome-wide association; multiple sclerosis

INTRODUCTION
Multiple sclerosis (MS) is one of the most common causes of neurological disability in young adults and affects more than 2 million people worldwide. It is characterised by the development of demyelinated lesions in the central nervous system, which are believed to be initiated by a T-cell-mediated autoimmune reaction against some component(s) of white matter. Demyelination, and at later stages axonal loss and neurodegeneration, cause a wide range of symptoms including defects in coordination, balance, walking, vision and cognition. The majority of patients present with a relapsing-remitting disease (RR–MS), in which episodes of disease activity are followed by partial or complete recovery from symptoms. Most RR–MS patients later proceed to a secondary progressive phase (SP–MS) where the clinical condition deteriorates without periods of recovery; and ~10–20% of patients show a progressive course from onset (primary progressive MS, PP–MS) [1]. RR–MS is thought to reflect occasional inflammatory bursts in the CNS, while neurodegeneration is believed to be responsible for the accumulating disability in progressive MS. With a typical age of onset of 20–40 years, the social and economical impacts on patients and their families are substantial. Available treatments such as β-interferon can reduce relapses in the RR–MS phase, but treatments that are more effective and safe are needed. At present their development is partly hindered by our...
poor understanding of the causes and disease mechanisms of MS.

Adoption and twin studies have demonstrated that MS is a complex disease caused by a combination of genetic susceptibility and exposure to environmental risk factors [1]. However, as a result of small sample sizes the estimates of heritability (i.e. fraction of the disease risk explained by genetic variants) are imprecise, ranging from 25% up to 76% [2]. Involvement of various environmental factors has been proposed including infections, diet and vitamin D levels, but solid evidence for any single factor is lacking [3]. Meanwhile, association with the major histocompatibility complex (MHC) was discovered more than three decades ago [4–6]. The major risk haplotype has since been narrowed to HLA-DRB1*1501-DQA1*0102-DQB1*0602, tagged by a single nucleotide polymorphism (SNP) rs3135388, which increases the risk for MS by ∼2-fold [7]. Disappointingly, linkage studies failed to find any genome-wide significant MS risk loci apart from the MHC indicating that the effect sizes of any additional risk loci were likely to prove considerably more modest [8]. Novel approaches were therefore required to tackle the challenge of finding risk variants unattainable by linkage. Following significant technical developments in the past few years, cost-effective genotyping of hundreds of thousands of SNPs in a single experiment has become feasible thereby providing the technical tools for identifying some of the additional genetic risk factors through so-called genome-wide association studies (GWAS).

In this review, we summarize findings from the 10 GWAS conducted in MS to date (Table 1) [7, 9–17] and discuss their implications for current views on the mechanisms underlying MS. Finally, we consider some of the remaining challenges and directions for future research.

**THE COMPLEX GENETIC ARCHITECTURE OF MS BEGINS TO UNRAVEL**

The first published MS-GWAS was performed by the International Multiple Sclerosis Genetics Consortium (IMSGC) using trio families (an affected individual and both their parents) from the UK and the USA [7]. Validating potentially associated SNPs which surfaced in the screening phase led to the confident identification of association with two non-MHC loci: **IL7R**, which was almost simultaneously implicated in two candidate gene studies [18, 19], and **IL2RA**. Both associations have been repeatedly replicated and have P-values far exceeding the commonly used threshold of genome-wide significance ($P \leq 5\times10^{-8}$) [20]. Early follow-up studies of the most promising variants from this study implicated variants from the regions of **CD58** and **CLEC16A** [21–23], while systematic efforts following up the 30,000 most associated SNPs from this screen (those SNPs with $P$-value < 0.1) identified risk variants in the **TMEM39A** and **KIF21B** loci [24]. MS was also included in the screen of non-synonymous SNPs by the Wellcome Trust Case Control Consortium (WTCCC) [17], and early follow-up efforts based on this screen identified an association with **TYK2** [25, 26] following which more systematic follow-up efforts also identified association with variants from the region of **MMEL1** [27]. In a GWAS based on samples from the United States, the Netherlands and Switzerland, the GeneMSA group found some evidence supporting association with variation in the region of the **GPC5** gene, but this has not been replicated [11]. Combining the data from the original IMSGC and the GeneMSA GWAS with unpublished GWAS data generated by the Brigham and Women’s Hospital MS partners group, a meta-analysis was performed which identified three associations with genome-wide significance (**CD6**, **IRF8** and **TNFRSF1A**) [13]. In addition, four suggestive associations were later confirmed in other studies (**IL12A**, **MPHOSPH9**, **RGS1** and **STAT3**) [14, 28]. The Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene) reported a further GWAS in which they identified two novel associations near **CD40** and at 12q13-14 [9]. Smaller studies based on samples from Germany and Spain, the latter employing pooled DNA, lacked power to identify novel associations with genome-wide significance [12, 15].

In addition to these GWAS conducted in relatively outbred populations, three GWAS have focused on isolated populations with exceptionally high MS prevalence based on the rationale that due to a small number of founder individuals, inbreeding and historical isolation, these populations may be enriched for certain risk variants. The largest of these studies was based on a Sardinian sample and identified an association with **CBLB** [16], a locus which had been suggested as possibly associated in the original IMSGC GWAS [7]. The other two isolated
population GWAS were small (less than 100 cases in each) and thus had limited power. However, association with \textit{STAT3}, a locus first suggested by the earlier meta-analysis [13], was identified in a Finnish sub-isolate [14] and confirmed in an extended analysis including samples from the wider Finnish population, Denmark, Norway and the meta-analysis. In the study based on a Dutch isolate [10], an association in \textit{KIF1B} was reported, but later proved to be a false positive [29]. To summarize, the GWAS to date have identified 16 non-HLA loci as MS risk factors with genome-wide significance (Table 2), and in addition, several suggestively significant loci have been reported (Table 3). Expectedly, associations with MHC variants reflecting the effect of HLA-DRB1*15 haplotype were detected in all studies.

**EFFECTS OF MS RISK VARIANTS ON GENE FUNCTION**

Relatively little is still known about how the identified MS risk variants exert their effects at the molecular and cellular levels.

---

**Table 1:** Published GWA studies in MS

<table>
<thead>
<tr>
<th>Number of cases</th>
<th>Number of controls</th>
<th>Population(s)</th>
<th>Platform</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>931</td>
<td>1862 (parents)</td>
<td>UK, US</td>
<td>Affymetrix 5.0</td>
<td>IMSGC et al. [7]</td>
</tr>
<tr>
<td>975</td>
<td>1466</td>
<td>UK</td>
<td>Custom made Infinium array for 14,436 nsSNPs (Illumina)</td>
<td>WTCCC [17]</td>
</tr>
<tr>
<td>45</td>
<td>195</td>
<td>Dutch isolate</td>
<td>Affymetrix 250K Nsp</td>
<td>Aulchenko et al. [10]</td>
</tr>
<tr>
<td>242 (pooled)</td>
<td>242 (pooled)</td>
<td>Spain</td>
<td>Affymetrix 5.0</td>
<td>Comabella et al. [12]</td>
</tr>
<tr>
<td>860</td>
<td>1720</td>
<td>US</td>
<td>Affymetrix 6.0</td>
<td>De Jager et al. [13]</td>
</tr>
<tr>
<td>68</td>
<td>136</td>
<td>Finnish isolate</td>
<td>Illumina 300</td>
<td>Jakkula et al. [14]</td>
</tr>
<tr>
<td>1618</td>
<td>3413</td>
<td>Australia, New Zealand, UK, US</td>
<td>Illumina 370CNV (cases), Illumina Infinium (controls)</td>
<td>ANZgene [9]</td>
</tr>
<tr>
<td>882</td>
<td>872</td>
<td>Sardinia</td>
<td>Affymetrix 6.0</td>
<td>Sanna et al. [16]</td>
</tr>
<tr>
<td>592</td>
<td>825</td>
<td>Germany</td>
<td>Illumina 300 and 370CNV</td>
<td>Nischwitz et al. [15]</td>
</tr>
</tbody>
</table>

---

**Table 2:** Genome-wide significant (\(P\)-value < 5E-08) loci in MS

<table>
<thead>
<tr>
<th>SNP (risk allele)</th>
<th>Nearest gene</th>
<th>Risk allele frequency in HapMap CEU Rel 28 (%)</th>
<th>(P)-value</th>
<th>Odds ratio</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10492972(C)</td>
<td>KIF1B</td>
<td>34</td>
<td>2.50E-10</td>
<td>1.34</td>
<td>Aulchenko et al. [10]</td>
</tr>
<tr>
<td>rs1132200(C)</td>
<td>TMEM39A</td>
<td>88</td>
<td>3.09E-08</td>
<td>1.24</td>
<td>IMSGC [7]; IMSGC [24]</td>
</tr>
<tr>
<td>rs2122721(G)</td>
<td>KIF1B</td>
<td>67</td>
<td>6.56E-10</td>
<td>1.22</td>
<td>IMSGC [7]; IMSGC [24]</td>
</tr>
<tr>
<td>rs1270876(A)</td>
<td>CLEC16A</td>
<td>68</td>
<td>1.6E-16</td>
<td>1.20</td>
<td>Hoppenbrouwers et al. [22]; IMSGC [23]</td>
</tr>
<tr>
<td>rs17445836(G)</td>
<td>KIF21B</td>
<td>72</td>
<td>3.73E-09</td>
<td>1.25</td>
<td>De Jager et al. [13]</td>
</tr>
<tr>
<td>rs17284933(G)</td>
<td>CD6</td>
<td>20</td>
<td>3.79E-09</td>
<td>1.18</td>
<td>De Jager et al. [13]</td>
</tr>
<tr>
<td>rs1790010(G)</td>
<td>MPHOSPH9</td>
<td>23</td>
<td>3.96E-08</td>
<td>1.10</td>
<td>De Jager et al. [13]; IMSGC [28]</td>
</tr>
<tr>
<td>rs1800693(G)</td>
<td>TNFRSF1A</td>
<td>48</td>
<td>1.59E-11</td>
<td>1.20</td>
<td>De Jager et al. [13]</td>
</tr>
<tr>
<td>rs2104286(T)</td>
<td>IL2RA</td>
<td>75</td>
<td>2.38E-23</td>
<td>1.25</td>
<td>IMSGC [7]; IMSGC [20]</td>
</tr>
<tr>
<td>rs2300747(A)</td>
<td>CD58</td>
<td>87</td>
<td>4.0E-09</td>
<td>1.23</td>
<td>IMSGC [7]; De Jager et al. [21]; Hoppenbrouwers et al. [22]</td>
</tr>
<tr>
<td>rs2760524(G)</td>
<td>RGS1</td>
<td>81</td>
<td>3.55E-09</td>
<td>1.15</td>
<td>De Jager et al. [13]; IMSGC [28]</td>
</tr>
<tr>
<td>rs4536443(G)</td>
<td>TYK2</td>
<td>NA</td>
<td>5.08E-09</td>
<td>1.30</td>
<td>WTCCC &amp; TASC [17]; Ban et al. [25]; Mero et al. [26]</td>
</tr>
<tr>
<td>rs4680534(C)</td>
<td>IL12A</td>
<td>25</td>
<td>3.08E-08</td>
<td>1.11</td>
<td>De Jager et al. [13]; IMSGC [28]</td>
</tr>
<tr>
<td>rs6897932(C)</td>
<td>IL7R</td>
<td>76</td>
<td>1.2E-17</td>
<td>1.20</td>
<td>IMSGC [7]; IMSGC [20]</td>
</tr>
<tr>
<td>rs703842(A)</td>
<td>METTL1</td>
<td>66</td>
<td>5.4E-11</td>
<td>1.23</td>
<td>ANZgene [9]</td>
</tr>
<tr>
<td>rs744166(G)</td>
<td>STAT3</td>
<td>45</td>
<td>2.75E-10</td>
<td>1.15</td>
<td>De Jager et al. [13]; Jakkula et al. [14]</td>
</tr>
<tr>
<td>rs9657904(A)</td>
<td>CBLB</td>
<td>79</td>
<td>1.6E-10</td>
<td>1.40</td>
<td>Sanna et al. [16]</td>
</tr>
</tbody>
</table>
common variants might modulate gene expression rather than change protein structure, which would be likely to result in more severe functional defects. Consistent with this hypothesis, the majority of variants associated with celiac disease were found to correlate with gene expression in cis in a recent study [30]. MS risk variants have been reported to correlate in cis with expression of CD58 [21], Fam119b, XRCC6BP1 and TSFM (all located near METTL1 on 12q13-14) [31, 32] and CDK2AP1 (located near MPHOSPH9) [28]. Some of the variants may also affect splicing efficiency, as has been experimentally shown with IL7R [19]. Finally, single SNPs may have an impact on expression of entire networks and pathways via trans-regulating factors such as transcription factors and miRNAs. For instance, the MS risk variant located near IRF8, a gene encoding a transcription factor involved in regulating type I interferon response, is associated with expression of interferon response pathway genes in peripheral blood mononuclear cells (PBMCs) of MS patients although not with the expression of IRF8 itself [13]. It is therefore unclear how the causative variant mediates its effects, but different transcriptional isoforms and coding variants in IRF8 need to be further investigated. Another gene with a potentially wide impact on gene expression is STAT3, which is a central transcription factor in the Jak/STAT signalling pathway and a critical player in determining whether naive T cells differentiate into regulatory T cells or Th17 cells [33]. However, the associated SNP, located in the first intron of STAT3, does not correlate with expression of any genes in human lymphoblastoid cell lines according to the mRNA by SNP Browser (eQTL LOD > 3) [34, 35]. Although most of the risk variants have not been shown to have any effects on gene expression, it should be noted that the present findings are based on data from lymphoblastoid cell lines or PBMCs, and studies in more defined immune cell populations are likely to reveal additional cis- and trans-regulatory effects.

### GWAS SUPPORT THE ROLE OF BOTH IMMUNE-RELATED AND NEURONAL MECHANISMS IN MS

Although fine mapping and functional studies will be required to define the functionally relevant variants responsible for determining susceptibility to MS, the over-representation of immunological genes near associated SNPs is already evident. Using the Ingenuity Pathways Analysis tool (IPA) (Ingenuity Systems Inc., Redwood City, CA, USA) we found that genes nearest to the identified SNPs (listed in Tables 2 and 3) are highly significantly associated with the T-helper cell differentiation pathway ($P = 2.01E$-
Crosstalk between dendritic cells and CD40 signalling 2.33E-04 3/69
Dendritic cell maturation 1.34E-05 5/188
T Helper cell differentiation 2.01E-07 5/72

Table 4: Top 10 canonical pathways associated with MS-associated loci

<table>
<thead>
<tr>
<th>Canonical pathway</th>
<th>P-value</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>T Helper cell differentiation</td>
<td>2.01E-07</td>
<td>5/72</td>
</tr>
<tr>
<td>Dendritic cell maturation</td>
<td>1.34E-05</td>
<td>5/188</td>
</tr>
<tr>
<td>CD40 signalling</td>
<td>2.33E-04</td>
<td>3/69</td>
</tr>
<tr>
<td>Crosstalk between dendritic cells and natural killer cells</td>
<td>7.43E-04</td>
<td>3/98</td>
</tr>
<tr>
<td>Colorectal cancer metastasis signalling</td>
<td>1.09E-03</td>
<td>4/256</td>
</tr>
<tr>
<td>IL-22 signalling</td>
<td>1.25E-03</td>
<td>2/28</td>
</tr>
<tr>
<td>IL-12 signalling and production in macrophages</td>
<td>1.31E-03</td>
<td>3/135</td>
</tr>
<tr>
<td>Oncostatin M signalling</td>
<td>1.98E-03</td>
<td>2/35</td>
</tr>
<tr>
<td>B-cell development</td>
<td>2.22E-03</td>
<td>2/37</td>
</tr>
<tr>
<td>NF-κB signalling</td>
<td>2.86E-03</td>
<td>3/155</td>
</tr>
</tbody>
</table>

A list of genes located nearest to the SNPs showing genome-wide significant or suggestive association with MS were given as input. P-value is the likelihood that the pathway is associated with the dataset by chance. The ratio is the number of genes in the input list involved in the pathway to the total number of molecules in that pathway.

07) and other immunological pathways (Table 4). In addition to CD40, IL12A, IL2RA, STAT3 and TNFRSF1A, all of which map to this pathway, at least CBLB [36], CD6 [37], CD58 [38], CD226 [39], SH2B3 [40] and TNEAIP3 [41] are involved in T-cell activation, and IL7 signalling, which is influenced by IL7R, is critical for T-cell homeostasis [42]. However, many of these genes may have as yet undefined functions and may well exert their effects via other immune-related pathways.

Meanwhile, apart from KIF21B, which is highly expressed in dendrites [43], the identified loci do not seem to contain genes with obvious known neurological roles. Interestingly this locus is also associated with ulcerative colitis and celiac disease [30, 44], suggesting that KIF21B may have an immunological function, or that it may not be the functionally relevant gene. Given that the associations identified so far are likely to represent only the tip of the iceberg, with available evidence suggesting that possibly hundreds of other variants are also involved [45], Baranzini et al. [46] analysed nominally significant SNPs (P < 0.05) from two MS GWAS [7, 11] using a network analyses approach based on experimental annotations of protein–protein interactions. They found that in addition to immunological networks, neural pathways such as those related to axon guidance and synaptic potentiation were significantly associated with these variants. Including SNPs with P < 0.05 in case–control analysis in the IMUGC GWAS [7], we obtained similar results using the IPA software and found neurological pathways rather than immune-related pathways lying among the top findings. However, randomly selected SNPs from the same GWAS were also significantly associated with neural pathways, suggesting that the analysis may be affected by some inherent biases such as unequal distribution of SNPs or gene size. On the other hand, the approach of Baranzini et al. [46] corrects for the number of SNPs tested within each gene and should be less sensitive to these factors. Currently available functional annotations of genes are, nevertheless, likely to be biased towards the more thoroughly investigated pathways, and results from these analyses should therefore be viewed with some caution. Upcoming GWAS in larger samples empowered to detect variants with more modest effects should provide further confirmation as to whether neural genes are involved in the polygenic background of MS.

EMERGING SHARED MECHANISMS OF AUTOIMMUNITY ARE INVOLVED IN MS

Although the predominance of immune-related genes near MS risk variants is not surprising, it is striking how many of these are also associated with other common autoimmune diseases (Figure 1). There seems to be considerable overlap especially between MS and type I diabetes and celiac disease, although this may to some extent reflect the number and power of studies carried out in each disease. In some cases the risk variants reported in different diseases show no linkage disequilibrium and are thereby likely to represent independent functionally relevant variants. Interestingly, even in case of shared variants, the risk allele in one disease may be protective in another. The emerging common aetiology of autoimmunity thereby seems to involve some allelic heterogeneity. Although many of the shared loci have not yet been confirmed in MS with genome-wide significance, associations with other autoimmune disorders would strongly support their involvement. For example, TNEAIP3 is associated with four other autoimmune diseases adding to the evidence that this locus is likely to be relevant in MS. TNEAIP3 encodes tumour necrosis factor α-induced protein 3, which is a zinc finger protein involved in inhibiting nuclear factor (NF)-κB activation and tumour necrosis factor (TNF)-mediated
apoptosis [47], and NF-kB is a central transcription factor, which regulates the expression of a number of genes in response to infection and inflammation [48]. The region also harbours another candidate gene, OLIG3, which encodes the oligodendrocyte transcription factor 3. There seem to be three independent foci of association within this locus: one between TNFAIP3 and OLIG3 (associated with celiac disease and rheumatoid arthritis), a second just upstream of TNFAIP3 (associated with MS) and a third within TNFAIP3 (associated with PS, rheumatoid arthritis and systemic lupus erythematosus), although even the risk variants within TNFAIP3 seem to be independent. Interestingly, rs2230926 which is associated with both rheumatoid arthritis [49] and systematic lupus erythematosus [50] is a relatively rare non-synonymous SNP (risk allele frequency = 0.8%), which causes an amino acid substitution from phenylalanine to cysteine (F127C) in the N-terminal ovarian tumour domain of TNFAIP3 protein. This domain is essential for the protein’s de-ubiquitinating activity, which in turn is involved in NF-kb regulation [47]. It remains to be investigated how the other associated variants affect the functions of TNFAIP3 or OLIG3. Although according to the mRNA by SNP browser there is no correlation between the risk variants and their expression in lymphoblastoid cell lines, expression in other cell populations should be investigated. In addition to TNFAIP3, a few other loci also stand out as remarkably consistent: SH2B3 and ZMIZ1 have both been reported in MS and three other diseases. SH2B3 encodes for a linker protein, which mediates signals from an activated T-cell receptor [51] and is a negative regulator of the JAK-STAT signalling pathway [52]. Intriguingly, ZMIZ1 shows similarity to PIAS (protein inhibitor of activated STAT) proteins [53]. Regulation of transcription factors, and the JAK-STAT signalling pathway in particular, thereby seems to be a prominent functional characteristic of the genes shared across autoimmune diseases.

In addition to the more common autoimmune diseases, several MS loci are mutated in rare immunological syndromes including severe combined immunodeficiency and Omenn syndrome (IL7R) [54–56], autoinflammatory TNF receptor-associated periodic syndrome (TNFRSF1A) [57] and hyper-IgE recurrent infection syndrome (STAT3) [59–61], while mutations in TYK2 cause a clinically related tyrosine kinase 2 deficiency [59]. Recent evidence also gives support for the role of TYK2-STAT3 signalling in mediating neuronal apoptosis in Alzheimer’s disease [62], suggesting that this pathway may be involved in MS through both neurodegenerative and immune-related mechanisms. Finally, of equal interest to the genes shared with other autoimmune disorders are those which seem specific for MS. Although many may still feature in future studies of other diseases, some may be related to MS-specific characteristics such as demyelination, neuroinflammation and neurodegeneration.

**FUTURE PROSPECTS**

Encouraged by successes of the first GWAS, WTCCC funded a second phase of GWAS in 2008 covering 13 diseases including MS (WTCCC2). In order to reach power to detect variants with more modest effects, this new MS GWAS involves almost 10 000 MS samples from 14 countries within the IMMSGC. The study also covers analyses of common copy number variants, which previously have not been comprehensively...
investigated in MS. At the time of writing, the project is at its final stages, and the results are expected to be ready for publication in early 2011.

Combining the available loci with the much larger effect on individual risk attributable to the MHC suggests that the currently observed genetic risk factors account for ~25–30% of MS heritability. Because GWAS rely on linkage disequilibrium to detect associations, it is expected that many of the currently identified variants are only proxies for the functionally relevant common variants. Nevertheless, the effects of these common causal variants would be expected to be only modestly larger than currently estimated. This view of common variants explaining the associations found in complex diseases was recently challenged in a thought-provoking article by Dickson et al. [63], suggesting that multiple rare variants in the same common haplotype background may be responsible for association signals at common variants. If true, these rare variants could have considerably larger odds ratios than those initially reported in GWAS and some might be located beyond the haplo-blocks defined by common SNPs (a suggestion with important implication for follow-up studies). However, there is currently no empirical evidence to support this hypothesis, and observations in currently available MS GWAS data suggest that the observed common variant associations in MS are not likely to be due to multiple rare variants hitch-hiking on the same haplotypic background [45]. Although rare variants are therefore perhaps not likely to underlie the already established associations, it is inevitable that rare variants influencing risk exist. However, at present their detection and testing is still technically and statistically challenging. The appropriate methodology for detecting rare variants depends on their frequency and expected effect size. Cirulli and Goldstein [64] propose three ‘rare’ variant categories: ‘less common variants’ (1–5%), ‘rare variants’ (<1%, but found in one or more human populations) and ‘private’ (only found in probands and immediate relatives). Finding novel rare or private variants can only be achieved through targeted or genome-wide re-sequencing in large case–control cohorts, or in samples enriched for these variants such as those representing extreme phenotypes or high-risk families. Fortunately, the costs of next-generation sequencing are rapidly decreasing and genome re-sequencing in high throughput is likely to become feasible during the next decade. To date some relevant ‘less common’ or rare variants have been identified in common autoimmune diseases [65, 66], but these remain to be further confirmed.

Finally, there is likely to be more to MS genetics than inherited germline DNA variation. For example, epigenetic factors and de novo and somatic mutations may turn out to play a role, perhaps especially so in sporadic cases where inherited risk factors are likely to be less important. In an attempt to address the potential role of some of these factors, Baranzini et al. [67] sequenced the CD4+ T cell genomes of a monozygotic twin pair discordant for MS and compared the CD4+ cell transcriptome and methylome in three discordant twin pairs. Perhaps disappointingly, they did not find any consistent reproducible differences between the affected and unaffected twins, which would have explained their discordance. However, larger samples and different cell populations will need to be investigated with better sequencing coverage and depth in order to rule out the roles of methylation, transcriptome differences and somatic and de novo mutations in MS.

**CONCLUSIONS**

Recent GWAS in MS have identified over a dozen common risk variants, thereby providing the first new clues towards the pathogenesis of MS since the role of the MHC was identified nearly four decades ago. Further genetic and functional studies are now required to pinpoint the functionally relevant genes and pathways, and to understand how these influence risk. However, we already have some invaluable insights. First, consistent with the prevailing hypotheses of an autoimmune-related pathogenesis, most of the loci harbour genes with pertinent immunological roles, including several genes associated with other autoimmune disorders. Pathway analyses also give some support for the involvement of neurological genes although none of the presently confirmed MS risk variants are located near genes with major known neurological functions. Second, GWAS data provide support for the notion that MS is a highly polygenic disease with possibly hundreds of variants each exerting a modest effect and thousands of variants with very small effects still waiting to be discovered [45]. Further studies will still be required, however, to identify the functionally relevant variants in these loci and it remains to be seen, how many of these represent rare variants with potentially large effects on disease susceptibility.
Third, the identified SNPs have turned out to be widely replicable in Caucasian populations suggesting that inter-population genetic heterogeneity in MS may be limited. Many of the associations have also been replicated in African Americans or Indians implicating shared genetic factors even across ethnic populations [68, 69]. Finally, while GWAS have been successful in exposing germline risk variants underlying MS, the roles of epigenetics and \textit{de novo} and somatic mutations, for example, are just beginning to be explored. Next-generation sequencing techniques are now providing the tools for approaching these phenomena on a genome-wide scale. Although much work therefore still remains, the recent findings are valuable in providing a rationale basis for hypotheses-driven functional experiments, which finally have a good chance of succeeding in exposing molecular mechanisms underlying MS.

**FUNDING**

This work was supported by the Cambridge NIHR Biomedical Research Center and Wellcome Trust grant 084702.

**Key Points**

- To date, 10 GWAS have been conducted in MS and have led to identification of 16 confirmed and several suggestively associated MS risk variants. Many of these are located near immunological genes, with T-cell-related genes being particularly prevalent. With the aim of identifying additional common risk variants with more modest effects, a large GWAS by the IMsGC and WTCCC2 comprising close to 10,000 MS subjects is soon to be completed. Further studies are required to identify causative variants in the identified loci and to explore the roles of rare variants, epigenetic modifications and somatic and \textit{de novo} mutations in MS.

**References**

Genome-wide association studies in MS


47. Lack of support for association between the KIF1B e104992972[C] variant and multiple sclerosis. *Nat Genet* 2010;42:469–70.


