Multiple sclerosis genetics—is the glass half full, or half empty?

Jorge R. Oksenberg and Sergio E. Baranzini

Abstract | Multiple sclerosis (MS) is a common and severe CNS disorder that is characterized by myelin loss, chronic inflammation, axonal and oligodendrocyte pathology, and progressive neurological dysfunction. Extensive epidemiological data confirm that genetic variation is an important determinant of susceptibility to MS, and suggest that such variation also influences the timing of symptom onset, the course of the disease, and the treatment response. Multicenter international collaborations have allowed large and well-characterized sample collections to be assembled that, when coupled with high-powered laboratory technologies, afford the opportunity to analyze the genome with increasing resolution and detail. The seven MS genome-wide association screens that have been completed in the past 3 years have substantially lengthened the list of MS genetic risk associations. Nevertheless, our knowledge of MS genetics remains incomplete, with many risk alleles still to be revealed, although progress is likely to be rapid in the near future. The ensuing challenge will be to design effective functional studies that convincingly link genetic variation to the underlying pathophysiology of MS. Establishment of such connections might translate into clinically useful genetic biomarkers and reveal novel targets for therapy. This Review briefly summarizes well-established concepts of MS epidemiology and susceptibility, and discusses new knowledge emerging from genome-wide association studies.

Oksenberg, J. R. & Baranzini, S. E. *Nat. Rev. Neurol.* **6**, 429–437 (2010); published online 13 July 2010; doi:10.1038/nrneurol.2010.91

Introduction

In countries populated by Europeans and their descendants, multiple sclerosis (MS) is a common cause of neurological disability in young adults.[1] Important advances have been made in the clinical management of MS in the past three decades. Nevertheless, the long-term prognosis for individuals who develop this disease remains generally poor: 15 years after diagnosis, most patients exhibit marked functional and/or cognitive deficits, with many requiring assistance to walk.

MS is a disease of the CNS. The histopathology of the affected brain and spinal cord has several notable features, including chronic inflammation, myelin loss, gliosis, and axon and oligodendrocyte pathology. Together, these features suggest the occurrence of two overlapping and connected effector arms, namely autoimmunity and neurodegeneration.[2] Typically, MS starts out as an episodic disorder and develops over time into a progressive and insidious disease.[3] Axonal pathology and neuronal loss are believed to be ultimately responsible for the persistent neurological dysfunction, but the debate concerning the nature of the initiating event continues, with new data supporting the view that early oligodendrocyte loss and plaque formation are not directly caused by a destructive adaptive immune response.[4] Another key unresolved issue is whether MS is the product of a single mechanism of tissue damage or can arise from fundamentally distinct pathologies.[5,6]

Competing interests
The authors declare no competing interests.

Interactions with infectious, climatic and/or other environmental variables probably have a considerable effect on an individual's susceptibility to MS (Figure 1). The influence of migration,[7–9] latitude,[10–12] and month of birth[13–15] on disease prevalence is consistent with a role for the environment in the determination of MS risk (Box 1). The observed increase in the incidence of this disease over the past century[16] and the existence of critical age periods for exposure to putative disease triggers[13,17,18] represent additional epidemiological cues supporting a fundamental role for environmental factors in MS. In addition to environmental factors, ancestry and family history (that is, genetics) are broadly agreed to exert a major influence on an individual's risk of developing this disorder.[19] In this Review, we examine the wealth of new genetic data acquired through genome-wide association studies (GWAS) and discuss the ensuing controversy on the respective roles of common versus rare genetic variants in determining MS risk.

Evidence for a genetic component

The pivotal data supporting genetic influences on MS susceptibility emerged from early observations of disease aggregation in families.[20–23] The extent of familial clustering is often expressed in terms of the λ_s parameter, which is derived from the ratio between the lifetime disease risk observed in the siblings of an affected individual and the risk seen in the general population.[24] In northern Europeans, the recurrence risk of MS in affected individuals' siblings is ≈2%, and the prevalence of this

Department of Neurology, School of Medicine, University of California at San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0435, USA (**J. R. Oksenberg**, **S. E. Baranzini**).

Correspondence to: J. R. Oksenberg jorge.oksenberg@ucsf.edu

Key points

- Genetic variation is an important determinant of susceptibility to and progression of multiple sclerosis (MS)
- MS is one of the so-called complex genetic diseases, which are common disorders that are characterized by modest disease risk heritability and multifaceted gene–environment interactions
- The human leukocyte antigen gene cluster represents by far the strongest MS susceptibility locus, and was identified in both candidate gene association and linkage studies
- Genome-wide association studies have dramatically increased the number of MS risk associations; however, only a fraction of the heritability of this disease has been explained
- With the aid of high-capacity technologies, next-generation studies will fully define the genetic mechanisms operating in MS and, hence, will assist in the formulation of a reliable model of pathogenesis

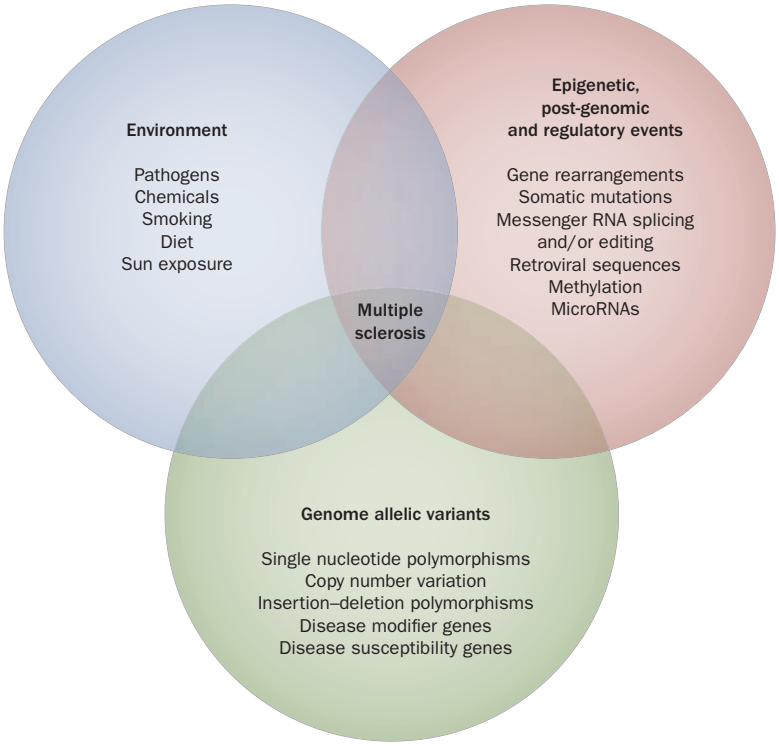

Figure 1 | MS as a complex disease. MS belongs to the group of complex genetic diseases, which are common disorders characterized by modest disease-risk heritability and multifaceted interactions between genetic and environmental factors. The full list of disease genes (susceptibility and modifiers) and environmental triggers in MS remains incomplete, while the study of epigenetic and other regulatory mechanisms linked to MS susceptibility is only just beginning. Abbreviation: MS, multiple sclerosis.

disease is 1 case per 1,000 people. Thus, after correcting for age, the λ_s for MS is 15–20, which is a similar range to that observed for type 1 diabetes and several other autoimmune diseases.[23,25] Some investigators argue that both the recurrence risk and prevalence are difficult to quantify and, in general, the latter is underestimated while the former is overestimated.[26] Consequently, a more accurate value for λ_s has been suggested to be <10.[27,28]

In addition to siblings, second-degree and third-degree relatives of individuals with MS have an increased risk of developing this disease. This observation provides further support to the notion that genetic factors, distinctly to a common environmental exposure, influence familial susceptibility.[23,25,29] The results of Canadian studies involving adoptees,[30] half siblings[31] or spouses[32] of individuals with MS confirm that genetic factors are primarily responsible for the co-occurrence of this disease in families. Concordance of some MS-related clinical metrics in families suggests that genes modestly influence disease trajectory and course,[33–36] as well as affecting susceptibility. Finally, twin studies from several different populations have consistently shown that a monozygotic twin of an individual with MS has a higher risk of developing this disease than has a dizygotic twin.[37,38] The estimated indices of heritability that emerged from these studies were acknowledged, however, to have a broad range (25–76%).[39]

Interestingly, the extent of disease concordance among monozygotic twins seems to correlate with latitude.[10] This finding exposes the complexity of the underlying multifactorial interactions that are associated with MS risk. Of note, the incomplete disease concordance in twins cannot be completely attributed to environmental triggers, and might reflect the occurrence of stochastic or programmed post-genomic modifier events. Such events include gene rearrangements in the germline (leading to a large variety of alternative immune receptors being encoded), incorporation of retroviral sequences, somatic mutations, epistatic interactions and epigenomic fluctuations.

Taken together, the familial recurrence and twin concordance rates for MS do not support the presence of a Mendelian trait. Consequently, and in line with observations for other common diseases, the MS-prone genotype is probably highly polygenic.

Gene discovery in multiple sclerosis

In an effort to decipher the genetic component of MS pathogenesis, many candidate-gene and whole-genome linkage analyses have been performed.[40] Most non-MHC candidate gene studies performed to date have been handicapped by inadequate statistical power and have yielded largely equivocal results, and will not be discussed here. A notable exception was the discovery that *IL7R*, which encodes the α subunit of the interleukin (IL)-7 receptor, was a true MS susceptibility gene. The IL-7 receptor $\alpha\gamma_c$ heterodimer mediates IL-7 signaling, which is a key mediator of T cell and B cell differentiation and survival.[41–43]

The scarcity of extended multiple-case families showing a clear and homogeneous mode of disease transmission has prevented the full exploitation of classic genetic linkage approaches in MS. Indeed, despite intense efforts, such approaches have largely proved ineffective for identifying susceptibility genes. One exception to this trend relates to the human leukocyte antigen (HLA) gene cluster on chromosome 6p21.3, which was discovered to be an MS susceptibility locus in the early 1970s through use of a classic case–control association approach (Box 2). In the best-powered full-genome MS

linkage screen, which included 730 multiple-case MS families and 4,500 markers, the peak logarithm (base 10) of odds (LOD) score in the *HLA* region reached a striking 11.7 (a LOD score of 3 or more is generally believed to indicate linkage). In this study, numerous regions on multiple chromosomes revealed linkage signals of interest; however, no other locus reached genome-wide significance.[44] The absence of signals (other than the *HLA* region signal) meeting this threshold discouraged the pursuit of large-scale follow-up fine-mapping efforts.

The interpretation of MS linkage data resulted in a new set of assumptions that has guided genetics studies of this disorder in recent years. The first of these assumptions is that a substantial component, if not all, of the genetic risk is determined by susceptibility alleles that are relatively common in the population (that is, a minor allele frequency of >5%), with each allele exerting a small to moderate effect on the overall risk. The second assumption states that in addition to the *HLA* region, 20–100 common genetic variants—each one increasing risk by a factor of only 1.2–1.5—might be sufficient to account for the genetic risk in MS. Third, the identification of genes influencing MS development must rely primarily on association-based methods and should involve large patient cohorts. Last, candidate gene studies, even if they are well powered, are hampered by our incomplete knowledge of MS pathogenesis and, hence, the difficulty of selecting the gene to investigate from among the many plausible candidates. Remarkable progress in the development of laboratory and analytical approaches has provided the opportunity to test these assumptions, and has fueled the impetus for pursuing 'agnostic' GWAS in MS (Box 3).

Genome-wide association studies

Studies conducted to date

The results of seven GWAS in MS have been reported, including data from an early study that only genotyped nonsynonymous coding single nucleotide polymorphisms (SNPs) and recent scans of a high-risk isolate from Finland and a Sardinian cohort.[45–51] The classic MS risk locus *HLA-DRB1* (Box 2) stood out in all studies with remarkably strong statistical significance (for example, $P < 1 \times 10^{-32}$ [Cochran–Armitage trend test] in one study[48]). The seven GWAS were followed by extensive efforts to replicate the top hits[52–57] and a comprehensive replicated meta-analysis.[58] Together, these studies provided robust evidence for 12 novel loci affecting disease susceptibility (Table 1). Of note, these markers might not necessarily represent the causal disease variant themselves, explaining in part the very modest independent odds ratio for each allele. Additional follow-up experiments refined some of the association signals and provided early insights into the functional effects of the identified gene variants. For example, variants of *IL7R* and *IL2RA* (which encodes the IL-2 receptor α subunit) were associated with increases in the soluble to membrane-bound ratios for IL-2 and IL-7 receptors, respectively,[43,59,60] while a variant of *CD58*

Box 1 | Environmental influences on multiple sclerosis

The suggestion that sun exposure protects against multiple sclerosis (MS) by increasing vitamin D_3 levels has become a common explanation for the latitude effect on this disease.[11,90] In humans, ultraviolet (UV) radiation is the main catalyst of vitamin D_3 synthesis. At high latitudes, people commonly exhibit low levels of this vitamin, particularly during the winter months when daylight hours are limited. Vitamin D_3 has a key role in T cell activation.[91] Thus, a reduction in vitamin D_3 levels might lead to immunoregulatory deficits and, hence, an increase in the risk of MS.[92,93] Interestingly, *HLA-DRB1*15* susceptibility haplotypes harbor a conserved vitamin D response element that modulates MHC class II DRβ expression.[94]

The vitamin D hypothesis has recently been challenged by studies using a controlled UV radiation treatment protocol in a mouse model of MS, experimental autoimmune encephalitis (EAE).[95] The clinical signs of EAE were largely suppressed following exposure to continuous UV radiation. Treated mice, however, only exhibited a modest and transient increase in serum levels of 25-hydroxycholecalciferol (a hydroxylated form of vitamin D_3). Moreover, 25-hydroxycholecalciferol administered through the diet of treated mice was unable to prevent disease progression, even at doses causing severe hypercalcemia.

In additional to vitamin D, many other environmental factors have been examined in MS,[90,96] including nutritional and dietary factors, consumption of water from wells, contact with animals, trauma following an accident or surgery, pollution, temperature, chemical agents, metals, minerals, organic solvents, and viral and bacterial infections. The biologically plausible infectious agents that could have a role in MS include measles, rubella, varicella zoster and mumps,[97] although the most reliable epidemiological[98] and laboratory[99–101] data relate to the Epstein–Barr virus. As in other autoimmune diseases, cigarette smoke might also be a potential environmental risk factor for MS.[102,103] A pertinent consideration, however, is that diverse environmental factors could be operating in this disease.

Box 2 | The 6p21–p23 chromosomal region and multiple sclerosis

The human leukocyte antigen (HLA) class II region, which is located on chromosome 6p21 (6p21–p23), has been identified as the genetic factor with the strongest influence on the susceptibility to multiple sclerosis (MS). Associations between this locus and MS have been observed across virtually all populations that have been studied. Moreover, such links have been found in patients with primary progressive MS and individuals with the relapsing–remitting form of this disease, suggesting that *HLA*-related mechanisms contribute to both phenotypes. A study of *HLA-DRB1* and *HLA-DQB1* alleles and haplotypes in an African American MS cohort showed that the effect of the *HLA* class II region on MS was largely attributable to *HLA-DRB1*.[104] *HLA-DRB1*1501* has a low frequency in Africa. In Europeans, but not in Africans, positive selection seems to have occurred for this allele and the extended *DRB1*1501–DQB1*0602* haplotype. The factors responsible for this selection process remain unclear, although an infectious pathogen might have been involved. One possible outcome of this process seems to have been a heightened susceptibility in Europeans to MS, which is almost nonexistent in Africa.

The exact mechanisms whereby *DRB1* influences MS susceptibility remain undetermined. Such mechanisms will, however, probably be related to the physiological function of HLA molecules in various immunological processes, such as antigen binding and presentation, and T cell repertoire determination. Allelic heterogeneity, as well as copy number and *cis–trans* regulatory effects have been reported for *HLA-DRB1*.[105,106] Such findings also suggest an allelic gradient of disease association with this gene, ranging from a genetic make-up of high vulnerability (*DRB1*15* homozygotes and *DRB1*15–DRB1*08* heterozygotes) to one of moderate susceptibility (*DRB1*03* homozygotes and heterozygotes) or resistance (*DRB1*15–DRB1*14* heterozygotes). In the latter, *DRB1*14* alleles would nullify the susceptibility effect of *DRB1*15*. Emerging data also suggest a modifier role for *HLA-DRB5*[107,108] and protective influences of HLA class I alleles.[109–113]

Box 3 | Genome-wide association studies

Genome-wide associations studies (GWAS) examine single nucleotide variation across genomes to identify genetic factors that are associated with a quantifiable trait. The 'common disease, common variant' hypothesis provides the underlying rationale for such studies. Commercially available single nucleotide polymorphism chips or arrays typically used in GWAS contain probes for polymorphisms selected on the basis of frequency and linkage disequilibrium parameters to efficiently capture large portions of common variation across the genome. Thus, the selected polymorphisms acts as surrogate markers of the broad genomic loci putatively associated with the trait of interest. Rapid rises have been observed in the density of the tested polymorphisms (currently $>10^6$) per study and in the number of published GWAS.[114,115] Given the very large number of simultaneous tests in each study, various *P* value cut-off points have been used to assess the statistical significance of associations (from $P<5\times10^{-7}$ to $P<1\times10^{-10}$), and a number of replication strategies have been employed to confirm such findings.[116,117] The Wellcome Trust Case Control Consortium has proposed that GWAS for complex diseases should involve at least 2,000 cases and 2,000 controls, and that *P* values of $<5\times10^{-7}$ would be needed to ensure that results were probably true. Obviously, larger sample sizes would be required if the effect sizes were smaller to ensure the validity of this proposed threshold.

Despite being a very young experimental approach, GWAS have identified hundreds of polymorphisms that are associated with disease and other traits and, hence, have provided new knowledge and important biological insights. Most of these associated variants have conferred small increments to risk (1.2-fold–1.5-fold). An online catalog of these variants is maintained by the National Human Genome Research Institute.[118]

was associated with diminished expression of *CD58* messenger RNA (mRNA).[59] In addition to gene identification, GWAS have brought about a dramatic paradigm shift in the study of MS, by allowing the development of a novel, perhaps more-accurate model of the genetic contributions to disease pathogenesis.[61]

Some of the allelic variants associated with MS have been proposed to be involved in several other autoimmune diseases, suggesting that common underlying mechanisms might exist for various autoimmune conditions.[56,62–64] For example, *IL2RA*-mediated susceptibility effects are observed in MS, type 1 diabetes, Graves disease and rheumatoid arthritis.[65] Interestingly, the directions of the associations are not consistent across these diseases. Indeed, while one *IL2RA* allele associated with susceptibility to MS seems to confer resistance to type 1 diabetes, a second allele of this gene confers susceptibility to both diseases, and a third allele is only associated with susceptibility to the latter.[66,67] This genetic modularity of human disease is not limited to autoimmunity.[68–70] For example, variants in the transcription factor gene *HNF1A* seem to be involved in maturity onset diabetes of the young type 3 and also can cause hepatic adenomas, while polymorphisms in the gene encoding transcription factor 7-like 2 affect the risk of developing type 2 diabetes and colon cancer. By merging conventional reductionism (that is, correlations between distinct experimental end points and clinical phenotypes) and the nonreductionist approach of systems biomedicine (that is, incorporation of network biology principles), new disease classifications based on coherent genotypes and phenotypes might be possible.[62,71]

Taken together, the data from GWAS seem to support the long-held view that MS susceptibility rests on the action of common sequence allelic variants (that is, risk alleles with a frequency of >5%) in multiple genes. Nevertheless, in spite of the expanding roster of risk loci, our understanding of MS genetics remains incomplete. For example, the sorting and classification of genomic

Table 1 | Overview of published genome-wide association studies in multiple sclerosis

Study	Design	Population origin	Number of screened samples*	Number of SNPs*	Featured loci and/or genes
Wellcome Trust Case–Control Consortium (2007)[45]	Cases–shared controls	UK	1,000 cases and 1,500 shared controls	14,436 (nonsynonymous SNPs)	*IL7R*
International Multiple Sclerosis Genetics Consortium (2007)[46]	Family based and case–control	US and UK	931 family trios	334,923	*HLA, IL2R, IL7R, CLEC16, CD58, EVI5, TYK2*
Comabella *et al.* (2008)[47]	Pooled case–control	Spain	242 cases and 242 controls	500,000	*HLA*, 13q31.3
Gene Associations in Multiple Sclerosis Consortium (2009)[48]	Case–control	US, The Netherlands and Switzerland	978 cases and 883 controls	551,642	*HLA, GPC5, PARK2, PDZRN4, CSMD1*
Australia and New Zealand Genetics in Multiple Sclerosis Consortium (2009)[49]	Case–control	Australia and New Zealand	1,618 cases and shared data for 3,413 controls	303,431	*HLA, METTL1, CD40*
De Jager *et al.*(2009)[58]	Meta-analysis and case–control	US, UK, The Netherlands and Switzerland	2,624 cases and 7,220 controls	2,557,248 (imputed)	*TNFRSF1A, IRF8, CD6, RGS1*
Jakkula *et al.* (2010)[50]	Isolated case–control	Finland	68 distantly related cases and 136 controls	297,343	*STAT3*
Sanna *et al.* (2010)[51]	Case–control	Sardinia	882 cases and 872 controls	6,600,000 (genotyped or imputed)	*HLA, CBLB*

All studies include a replication phase. *Number relates to the initial screen. Abbreviation: SNP, single nucleotide polymorphism.

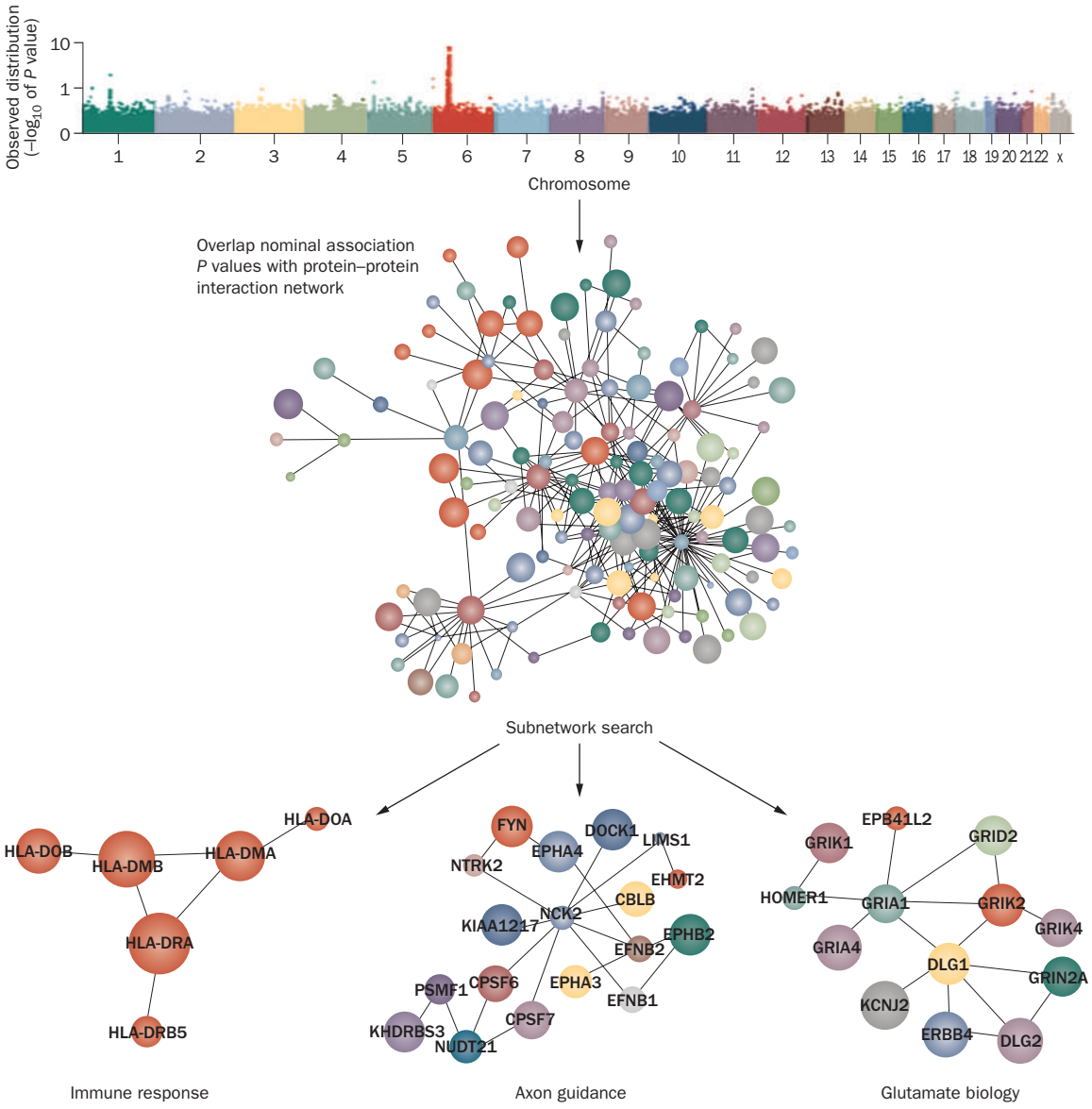

Figure 2 | Pathway-oriented analysis. To date, seven GWAS have been performed for MS, with several hundred thousand SNPs being tested. Conventional statistical analyses of these data sets have largely identified genes that encode proteins with known immunological functions. In general, only a small number of markers exceed the genome-wide significance threshold, and markers that do not manage to cross this threshold are generally ignored. Many of the markers that show modest associations can represent false negatives. Using data from two MS GWAS,[46,48] we undertook a pathway-oriented analysis that considered all SNPs with nominal evidence of association ($P<0.05$).[61] The gene-wise P values were superimposed on a human protein–protein interaction network. Subsequently, searches were performed to identify subnetworks that contained a higher proportion of genes associated with MS than was expected by chance alone. This analysis yielded subnetworks of genes from several immunological and neural pathways, including axon guidance and synaptic potentiation pathways. Colors represent the various chromosomes. The size of nodes is proportional to the $-\log_{10}$ P value of association. Abbreviations: GWAS, genome-wide association studies; MS, multiple sclerosis; SNP, single nucleotide polymorphism. Permission for networks in bottom panels from Oxford University Press © Baranzini, S. E. *et al. Hum. Mol. Genet*. **18**, 2078–2090 (2009).

segment duplications and deletions, which generate copy number variants (CNVs), lag well behind the identification of susceptibility SNPs. CNVs are a major source of human genetic diversity and have been shown to have roles in rare genomic disorders, and to influence complex traits and diseases.[72–74] Of note, studies have reported associations between CNVs and human autoimmune diseases such as systemic lupus erythematosus, psoriasis, Crohn disease, rheumatoid arthritis and type 1 diabetes.[75]

If certain assumptions are made, the available data from GWAS explain ≈3% of the total variance in MS risk.[76] A study of this type comprising 10,000 cases and high-density SNP–CNV platforms has been proposed to be a reasonable next step in the genetic analysis of MS. Such an investigation would be adequately powered to

Box 4 | Genomic determinants of disease course in multiple sclerosis

Heterogeneity in clinical and paraclinical features is a well-recognized aspect of multiple sclerosis (MS). Little is known about the underlying causes of this diversity, although family-based studies clearly show that genetic factors have important roles in the determination of clinically important outcomes, such as age of onset, anatomical sites of involvement, and disease severity.[33,35,119] Nevertheless, whether phenotypic variation arises from etiological heterogeneity, modifying influences of specific genes, or a combination of the two, remains to be firmly established. To date, most published reports that have proposed specific genetic influences on MS natural history have relied largely on small case series, retrospective clinical assessments and/or nonvalidated phenotypic endpoints. Moreover, few studies have considered the confounding effects of drug treatments and population stratification.

An emerging body of literature now suggests that individual or aggregated genetic variants can be isolated that influence well-validated pharmacological and clinical features of MS, including discrete MRI measures of brain disease activity, the rate of disability progression and the response to disease-modifying drugs.[36,48,120–124] Of note, one should recognize that the contribution of germline genetic variants to disease course might be very small. Nevertheless, demonstration of even a modest genetic effect could help reveal the fundamental pathophysiological mechanisms underlying MS and provide therapeutic targets.

identify common risk alleles with odds ratios of 1.2 or more.[19,77] At the time of writing, such a screen in MS, by an international megagroup consortium in collaboration with the Wellcome Trust, was nearing completion.

Pathway-based and network-based analysis

In GWAS, only a few markers are usually found for which the evidence of association exceeds the genome-wide significance threshold; markers that do not cross this threshold are generally neglected. Inspired in part by analytical strategies used in gene expression analyses, several analytical methods have been developed that combine biological knowledge with data from GWAS in the form of gene ontologies or pathways.[78–80] The advantage of these methods is that even if markers in individual genes do not reach genome-wide significance, several modest associations in functionally related genes might highlight the involvement of a particular biological pathway in the disease process. Building on this rationale, we merged statistical evidence from data analyses of two MS GWAS with experimental evidence of protein interactions from yeast two-hybrid assays and/or chromatin immunoprecipitation studies to discover subnetworks (or modules) of interacting proteins associated with MS susceptibility.[61] In addition to the identification of modules from the HLA system and immune cell communication, our study suggested that neural pathways such as axon guidance and glutamate metabolism also participate in the pathogenesis of MS (Figure 2).

Translational application

In spite of the success of GWAS in the identification of novel susceptibility loci that withstood the challenge of independent replication, many questions remain concerning the genetic architecture of MS. To illustrate the complexity of MS genetics and the challenges that lie ahead in this field, some observers have drawn on the example of human height, which is a complex trait with high estimated heritability (≈80%). At least 40 loci have been associated with height; however, these loci only explain ≈5% of phenotypic variance of this trait.[81] This finding might imply that using genetic information derived from GWAS to generate reliable diagnostic signatures for MS will be futile. In other words, given that none of the high-frequency associated alleles is sufficient to cause MS or obligatory for the development of this disease, very few people can have their probability of developing MS accurately predicted from genetic testing.[28] Following this rationale is simple: the high frequency of susceptibility variants in unaffected individuals, together with the probable occurrence of unaccountable epistatic interactions, etiological heterogeneity, and epigenetic and random events, suggest that even if a predictive algorithm included all variant risks, considerable uncertainties and noise would prevail when predicting case–control status. Conceivably, however, predictive modeling might be substantially improved if more-penetrant variants are identified, and transcriptional signatures and accurate environmental exposures are taken into account.[82,83] The translational potential of such modeling might extend to redefining disease classification, prognosticating progression (Box 4), and predicting treatment responses and/or adverse effects.

Ultimately, individual or cumulative genetic variants might only provide discriminatory genetic profiles of low sensitivity and specificity. Nevertheless, the implicated genes and biological pathways might still be of importance in the development of a better understanding of MS pathogenesis. The investigators who conducted the first of the MS GWAS, in which a modest risk ratio was identified for *IL2RA*, noted that clinical efficacy had been observed in trials that had assessed a monoclonal antibody targeting the IL-2 receptor α subunit, and suggested that genetic studies had the potential to drive drug development for the management of MS. Overall, however, the potential for drug development programs based on new genetic paradigms and effective 'theranostic' pipelines remains to be proven.

Prospects in genetics research

Through the integration of datasets from GWAS with transcriptional, proteomic and other information (for example, clinical, imaging, environmental exposure and/or ancestry data), rapid progress might be made in our understanding of MS genetics and pathogenesis. One should note that most DNA variants identified in GWAS will probably not be causative. Thus, a need exists to follow the discovery of each statistically associated polymorphism with fine mapping in large and informative sample sets, using comprehensive batteries of markers covering and flanking the confirmed associations. Furthermore, while GWAS are a valuable tool for assessment of the effects of common polymorphisms, the aggregate role of low-frequency rare functional gene variants in MS has not been properly evaluated. Uncommon variants with relatively large effects might account for part of the unexplained heritability in MS. Moreover, some of these

rare variants might reside in genes previously identified in GWAS. Rare genetic variation is not well represented in current databases or commercially available fixed arrays, and can be only ascertained by targeted deep sequencing of samples from affected individuals.[84]

In a thought-provoking study, Goldstein and colleagues used original computer simulations to analyze data from non-MS GWAS and proposed that rare causal variants of disease can create long-range (up to 10 Mb) association signals that can mistakenly be assigned to common variants.[85] If this assertion proves to be accurate, the way we analyze and interpret GWAS as well as how we design follow-up confirmatory studies could be affected dramatically. Resequencing efforts to identify causative gene variants might be pointless if such studies are limited to the individual genes that are implicated by the common variant association. Instead, full genome scans would be necessary to deconstruct the genetics of MS. In a modest but important first step to assess Goldstein and colleagues' hypothesis, we sequenced the entire genomes of two female MS-discordant monozygotic twins, generating over one billion, high quality, whole-genome shotgun reads, which corresponded to approximately 22-fold coverage of each genome.[86] These data constitute the first reported twin, autoimmune disease-associated genome sequences. Among the ≈3.6 million SNPs, ≈200,000 insertion–deletion polymorphisms (indels), 27 CNVs and 1.1 million methylated CpG dinucleotides that were detected and analyzed, only the methylation of two cytosine residues differed between the twins. We also deciphered the full mRNA transcriptome and epigenome sequences of $CD4^+$ lymphocytes from three pairs of MS-discordant monozygotic twins. In each of the $CD4^+$ T cell preparations, ≈19,000 genes were expressed; however, no reproducible transcriptional differences were identified between the MS-affected and MS-unaffected twins. Genome sequences from many more MS-discordant twin pairs, patients with MS and healthy controls are necessary to maximize the power of this approach. Our MS study coincided with a report presenting full-genome sequencing of a person affected with Charcot–Marie–Tooth disease[87] and the whole-genome sequencing of a couple, and their two children, each of whom had a different rare genetic disease (Miller syndrome and primary ciliary dyskinesia).[88] Together, these studies have inaugurated the era of whole-genome sequencing in human disease. Given the impressive rate of advances in this field, the technology for whole-genome sequencing should become sufficiently accurate and cheap for large-scale application to commence within the next few years.

Conclusions

Substantial epidemiological data confirm that genetic variation is an important determinant of susceptibility to MS. Genes within the *HLA* region encoding antigen-presenting molecules account for the largest part of the genetic risk for MS. The strongest association signal comes from *HLA-DRB1* in the class II region. In addition to the signal from the *HLA-DRB1* gene, complex hierarchical *cis* and *trans* effects can be found across the locus, including a protective effect conferred by the telomeric class I region. The non-HLA risk seems to be driven by modest contributors (Table 2), although the study of rare genetic variants and synergistic pathways and networks might require reassessment of this working model. The next few years will see the completion and publication of additional genome-wide scans that involve higher resolution tools and substantially larger DNA collections than have been used before. Even when performed with exquisite resolution, however, association and epidemiology studies are limited in their power to prove causation. Thus, effective functional studies will be required that connect genetic variation with disease pathophysiology.[89] Despite the complexity that characterizes MS, such studies, if pursued in the context of robust, statistically replicated and well-defined genetic associations, hold great promise for aiding the development of novel and accurate conceptual models of MS pathogenesis.

Table 2 | Established non-MHC susceptibility genes and variants in MS

Gene	Associated single nucleotide polymorphism	Location	Risk allele	Estimated odds ratio	Protein function
CD58	rs2300747	1p13	A	1.18	Cell adhesion; immunological co-stimulation
EVI5	rs10735781	1p22.1	G	1.23	Rab GTPase activator; cell cycle
RGS1	rs2760524	1q31	G	1.13	GTPase activator
IL7R	rs6897932	5p13	C	1.04	Cytokine receptor
IL2RA	rs2104286	10p15–p14	T	1.18	Cytokine receptor
CD6	rs17824933	11q13	G	1.16	Cell adhesion
TNFRSF1A	rs1800693	12p13.2	C	1.23	Cytokine receptor
CLEC16a	rs12708716	16p13.3	A	1.17	Not determined
IRF8	rs17445836	16q24.1	G	1.33	Transcription
CD226	rs763361	18q22.3	T	1.04	Cell adhesion
TYK2	rs34536443	19p13.2	G	1.30	Intracellular signaling
CD40	rs6074022	20q12	G	1.22	Immunological co-stimulatory receptor

Abbreviation: MS, multiple sclerosis.

Review criteria

This Review article builds primarily on literature citations from the past 10 years, covering both original contributions and scholarly reviews that in the opinion of the authors represent noteworthy and influential contributions to the understanding of multiple sclerosis genetic epidemiology. This article is not intended to systematically address the full breadth of the genetic and epidemiological issues related to this complex and multifactorial disease.

1. Pugliatti, M. *et al.* The epidemiology of multiple sclerosis in Europe. *Eur. J. Neurol.* **13**, 700–722 (2006).
2. Hauser, S. L. & Oksenberg, J. R. The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. *Neuron* **52**, 61–76 (2006).
3. Hauser, S. L. & Goodin, D. S. In *Harrison's Principle of Internal Medicine* 17th edn Ch. 375 (eds Fauci, A. D. *et al.*) 2611–2621 (McGraw Hill, New York, 2008).
4. Henderson, A. P., Barnett, M. H., Parratt, J. D. & Prineas, J. W. Multiple sclerosis: distribution of inflammatory cells in newly forming lesions. *Ann. Neurol.* **66**, 739–753 (2009).
5. Lucchinetti, C. Pathological heterogeneity of idiopathic central nervous system inflammatory demyelinating disorders. *Curr. Top. Microbiol. Immunol*. **318**, 19–43 (2008).
6. Breij, E. C. *et al.* Homogeneity of active demyelinating lesions in established multiple sclerosis. *Ann. Neurol*. **63**, 16–25 (2008).
7. Alter, M., Leibowitz, U. & Speer, J. Risk of multiple sclerosis related to age at migration to Israel. *Arch. Neurol.* **15**, 234–237 (1996).
8. Dean, G. & Kurtzke, J. F. On the risk of multiple sclerosis according to age at migration to South Africa. *Br. Med. J.* **3**, 725–729 (1971).
9. Cabre, P. *et al.* Role of return migration in the emergence of multiple sclerosis in the French West Indies. *Brain* **128**, 2899–2910 (2005).
10. Islam, T. *et al.* Differential twin concordance for multiple sclerosis by latitude of birthplace. *Ann. Neurol.* **60**, 56–64 (2006).
11. Islam, T., Gauderman, W. J., Cozen, W. & Mack, T. M. Childhood sun exposure influences risk of multiple sclerosis in monozygotic twins. *Neurology* **69**, 381–388 (2007).
12. Beretich, B. D. & Beretich, T. M. Explaining multiple sclerosis prevalence by ultraviolet exposure: a geospatial analysis. *Mult. Scler.* **15**, 891–898 (2009).
13. Willer, C. J. *et al.* Timing of birth and risk of multiple sclerosis: population based study. *BMJ* **330**, 120 (2005).
14. Sotgiu, S. *et al.* Seasonal fluctuation of multiple sclerosis births in Sardinia. *J. Neurol*. **253**, 38–44 (2006).
15. Ramagopalan, S. V. *et al. HLA-DRB1* and month of birth in multiple sclerosis. *Neurology* **73**, 2107–2111 (2009).
16. Orton, S. M. *et al.* Sex ratio of multiple sclerosis in Canada: a longitudinal study. *Lancet Neurol.* **5**, 932–936 (2006).
17. Kurtzke, J. F. Multiple sclerosis in time and space—geographic clues to cause. *J. Neurovirol*. **6** (Suppl. 2), S134–S140 (2000).
18. Pugliatti, M. *et al.* Evidence of early childhood as the susceptibility period in multiple sclerosis: space–time cluster analysis in a Sardinian population. *Am. J. Epidemiol*. **164**, 326–333 (2006).
19. Oksenberg, J. R., Baranzini, S. E., Sawcer, S. & Hauser, S. L. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat. Rev. Genet*. **9**, 516–526 (2008).
20. Mackay, R. P. Familial occurrence of multiple sclerosis and its implications. *Arch. Neurol. Psychiatry* **64**, 155–157 (1950).
21. Schapira, K., Poskanzer, D. C. & Miller, H. Familial and conjugal multiple sclerosis. *Brain* **86**, 315–332 (1963).
22. Doolittle, T. H. *et al.* Multiple sclerosis sibling pairs: clustered onset and familial predisposition. *Neurology* **40**, 1546–1552 (1990).
23. Robertson, N. P. *et al.* Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain* **119**, 449–455 (1996).
24. Risch, N. Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* **46**, 222–228 (1990).
25. Sadovnick, A. D. Familial recurrence risks and inheritance of multiple sclerosis. *Curr. Opin. Neurol. Neurosurg.* **6**, 189–194 (1993).
26. Guo, S. W. Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am. J. Hum. Genet.* **63**, 252–258 (1998).
27. 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* **10**, 5–11 (2009).
28. Sawcer, S., Ban, M., Wason, J. & Dudbridge, F. What role for genetics in the prediction of multiple sclerosis? *Ann. Neurol.* **67**, 3–10 (2010).
29. Carton, H. *et al.* Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *J. Neurol. Neurosurg. Psychiatry* **62**, 329–333 (1997).
30. Ebers, G. C., Sadovnick, A. D. & Risch, N. J. A genetic basis for familial aggregation in multiple sclerosis. *Nature* **377**, 150–151 (1995).
31. Sadovnick, A. D., Ebers, G. C., Dyment, D. A. & Risch, N. J. Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. *Lancet* **347**, 1728–1730 (1996).
32. Ebers, G. C., Yee, I. M., Sadovnick, A. D. & Duquette, P. Conjugal multiple sclerosis: population-based prevalence and recurrence risks in offspring. *Ann. Neurol.* **48**, 927–931 (2000).
33. Brassat, D. *et al.* Familial factors influence disability in MS multiplex families. *Neurology* **52**, 1632–1636 (1999).
34. Barcellos, L. F. *et al.* Genetic basis for clinical expression in multiple sclerosis. *Brain* **125**, 150–158 (2002).
35. Hensiek, A. E. *et al.* Familial effects on the clinical course of multiple sclerosis. *Neurology* **68**, 376–383 (2007).
36. DeLuca, G. C. *et al.* An extremes of outcome strategy provides evidence that multiple sclerosis severity is determined by alleles at the *HLA-DRB1* locus. *Proc. Natl Acad. Sci. USA* **104**, 20896–20901 (2007).
37. Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D. & Ebers, G. C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl Acad. Sci. USA* **100**, 12877–12882 (2003).
38. Hansen, T. *et al.* Concordance for multiple sclerosis in Danish twins: an update of a nationwide study. *Mult. Scler.* **11**, 504–510 (2005).
39. Hawkes, C. H. & Macgregor, A. J. Twin studies and the heritability of MS: a conclusion. *Mult. Scler.* **15**, 661–667 (2009).
40. McElroy, J. P. & Oksenberg, J. R. Multiple sclerosis genetics. *Curr. Top. Microbiol. Immunol.* **318**, 45–72 (2008).
41. Booth, D. R. *et al.* Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis. *J. Mol. Med.* **83**, 822–830 (2005).
42. Lundmark, F. *et al.* Variation in interleukin 7 receptor α chain (*IL7R*) influences risk of multiple sclerosis. *Nat. Genet.* **39**, 1108–1113 (2007).
43. Gregory, S. G. *et al.* Interleukin 7 receptor α chain (*IL7R*) shows allelic and functional association with multiple sclerosis. *Nat. Genet.* **39**, 1083–1091 (2007).
44. Sawcer, S. *et al.* A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* **77**, 454–467 (2005).
45. Burton, P. R. *et al.* Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* **39**, 1329–1337 (2007).
46. International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genome-wide study. *N. Engl. J. Med.* **357**, 851–862 (2007).
47. Comabella, M. *et al.* 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* **3**, e3490 (2008).
48. Baranzini, S. E. *et al.* Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.* **18**, 767–778 (2009).
49. 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.* **41**, 824–828 (2009).
50. Jakkula, E. *et al.* Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in *STAT3* gene. *Am. J. Hum. Genet.* **86**, 285–291 (2010).
51. Sanna, S. *et al.* Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis. *Nat. Genet.* **42**, 495–497 (2010).
52. International Multiple Sclerosis Genetics Consortium (IMSGC). Refining genetic associations in multiple sclerosis. *Lancet Neurol.* **7**, 567–569 (2008).
53. Ban, M. *et al.* Replication analysis identifies *TYK2* as a multiple sclerosis susceptibility factor. *Eur. J. Hum. Genet.* **17**, 1309–1313 (2009).
54. D'Netto, M. J. *et al.* Risk alleles for multiple sclerosis in multiplex families. *Neurology* **72**, 1984–1988 (2009).
55. Hoppenbrouwers, I. A. *et al.* Replication of *CD58* and *CLEC16A* as genome-wide significant risk genes for multiple sclerosis. *J. Hum. Genet.* **54**, 676–680 (2009).
56. International Multiple Sclerosis Genetics Consortium (IMSGC). The expanding genetic overlap between multiple sclerosis and type I diabetes. *Genes Immun.* **10**, 11–14 (2009).
57. International Multiple Sclerosis Genetics Consortium (IMSGC). Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies *KIF21B* and *TMEM39A* as susceptibility loci. *Hum. Mol. Genet.* **19**, 953–962 (2009).
58. De Jager, P. L. *et al.* Meta-analysis of genome scans and replication identify *CD6*, *IRF8* and *TNFRSF1A* as new multiple sclerosis susceptibility loci. *Nat. Genet*. **41**, 776–782 (2009).
59. De Jager, P. L. *et al.* The role of the *CD58* locus in multiple sclerosis. *Proc. Natl Acad. Sci USA* **106**, 5264–5269 (2009).
60. Maier, L. M. *et al.* Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. *J. Immunol.* **182**, 1541–1547 (2009).
61. Baranzini, S. E. *et al.* Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet*. **18**, 2078–2090 (2009).
62. Baranzini, S. E. The genetics of autoimmune diseases: a networked perspective. *Curr. Opin. Immunol.* **21**, 596–605 (2009).

63. Hafler, J. P. *et al. CD226* Gly307Ser association with multiple autoimmune diseases. *Genes Immun.* **10**, 5–10 (2009).
64. Sirota, M., Schaub, M. A., Batzoglou, S., Robinson, W. H. & Butte, A. J. Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet*. **5**, e1000792 (2009).
65. Maier, L. M. & Hafler, D. A. The developing mosaic of autoimmune disease risk. *Nat. Genet.* **40**, 131–132 (2008).
66. Dendrou, C. A. *et al.* Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nat. Genet*. **41**, 1011–1015 (2009).
67. Maier, L. M. *et al. IL2RA* genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet*. **5**, e1000322 (2009).
68. Goh, K. I. *et al.* The human disease network. *Proc. Natl Acad. Sci. USA* **104**, 8685–8690 (2007).
69. Oti, M. & Brunner, H. G. The modular nature of genetic diseases. *Clin. Genet.* **71**, 1–11 (2007).
70. Oti, M., Huynen, M. A. & Brunner, H. G. The biological coherence of human phenome databases. *Am. J. Hum. Genet*. **85**, 801–808 (2009).
71. Loscalzo, J., Kohane, I. & Barabasi, A. L. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol. Syst. Biol.* **3**, 124 (2007).
72. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
73. Wain, L. V., Armour, J. A. & Tobin, M. D. Genomic copy number variation, human health, and disease. *Lancet* **374**, 340–350 (2009).
74. Zhang, F., Gu, W., Hurles, M. E. & Lupski, J. R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–481 (2009).
75. Schaschl, H., Aitman, T. J. & Vyse, T. J. Copy number variation in the human genome and its implication in autoimmunity. *Clin. Exp. Immunol.* **156**, 12–16 (2009).
76. International Multiple Sclerosis Genetics Consortium (IMSGC). Evidence for polygenic susceptibility to multiple sclerosis. The shape of things to come. *Am. J. Hum. Genet.* **86**, 621–625 (2010).
77. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
78. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet*. **37**, 413–417 (2005).
79. Lesnick, T. G. *et al.* A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet*. **3**, e98 (2007).
80. Torkamani, A., Topol, E. J. & Schork, N. J. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* **92**, 265–272 (2008).
81. Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med*. **360**, 1696–1698 (2009).
82. Corvol, J. C. *et al.* Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. *Proc. Natl Acad. Sci. USA* **105**, 11839–11844 (2008).
83. De Jager, P. L. *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol.* **8**, 1111–1119 (2009).
84. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
85. Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
86. Baranzini, S. E. *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351–1356 (2010).
87. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
88. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
89. Fugger, L., Friese, M. A. & Bell, J. I. From genes to function: the next challenge to understanding multiple sclerosis. *Nat. Rev. Immunol.* **9**, 408–417 (2009).
90. Ascherio, A. & Munger, K. L. Environmental risk factors for multiple sclerosis. Part II: noninfectious factors. *Ann. Neurol.* **61**, 504–513 (2007).
91. von Essen, M. R. *et al.* Vitamin D controls T cell antigen receptor signaling and activation of human T cells. *Nat. Immunol.* **11**, 344–349 (2010).
92. Correale, J., Ysrraelit, M. C. & Gaitan, M. I. Immunomodulatory effects of vitamin D in multiple sclerosis. *Brain* **132**, 1146–1160 (2009).
93. Smolders, J. *et al.* Vitamin D status is positively correlated with regulatory T cell function in patients with multiple sclerosis. *PLoS ONE* **4**, e6635 (2009).
94. Ramagopalan, S. V. *et al.* Expression of the multiple sclerosis-associated MHC class II allele *HLA-DRB1*1501* is regulated by vitamin D. *PLoS Genet*. **5**, e1000369 (2009).
95. Becklund, B. R., Severson, K. S., Vang, S. V. & DeLuca, H. F. UV radiation suppresses experimental autoimmune encephalomyelitis independent of vitamin D production. *Proc. Natl Acad. Sci. USA* **107**, 6418–6423 (2010).
96. Ascherio, A. & Munger, K. L. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann. Neurol.* **61**, 288–299 (2007).
97. Lipton, H. L., Liang, Z., Hertzler, S. & Son, K. N. A specific viral cause of multiple sclerosis: one virus, one disease. *Ann. Neurol.* **61**, 514–523 (2007).
98. Ascherio, A. *et al.* Epstein–Barr virus antibodies and risk of multiple sclerosis: a prospective study. *JAMA* **286**, 3083–3088 (2001).
99. Cepok, S. *et al.* Identification of Epstein–Barr virus proteins as putative targets of the immune response in multiple sclerosis. *J. Clin. Invest.* **115**, 1352–1360 (2005).
100. Serafini, B. *et al.* Dysregulated Epstein–Barr virus infection in the multiple sclerosis brain. *J. Exp. Med.* **204**, 2899–2912 (2007).
101. Salvetti, M., Giovannoni, G. & Aloisi, F. Epstein–Barr virus and multiple sclerosis. *Curr. Opin. Neurol.* **22**, 201–206 (2009).
102. Riise, T., Nortvedt, M. W. & Ascherio, A. Smoking is a risk factor for multiple sclerosis. *Neurology* **61**, 1122–1124 (2003).
103. Hedstrom, A. K., Baarnhielm, M., Olsson, T. & Alfredsson, L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology* **73**, 696–701 (2009).
104. Oksenberg, J. R. *et al.* Mapping multiple sclerosis susceptibility to the *HLA-DR* locus in African Americans. *Am. J. Hum. Genet.* **74**, 160–167 (2004).
105. Barcellos, L. F. *et al. HLA-DR2* dose effect on susceptibility to multiple sclerosis and influence on disease course. *Am. J. Hum. Genet*. **72**, 710–716 (2003).
106. Dyment, D. A. *et al.* Complex interactions among MHC haplotypes in multiple sclerosis: susceptibility and resistance. *Hum. Mol. Genet.* **14**, 2019–2026 (2005).
107. Gregersen, J. W. *et al.* Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443**, 574–577 (2006).
108. Caillier, S. J. *et al.* Uncoupling the roles of *HLA-DRB1* and *HLA–DRB5* genes in multiple sclerosis. *J. Immunol.* **181**, 5473–5480 (2008).
109. Brynedal, B. *et al. HLA-A* confers an *HLA-DRB1* independent influence on the risk of multiple sclerosis. *PLoS ONE* **2**, e664 (2007).
110. Yeo, T. W. *et al.* A second major histocompatibility complex susceptibility locus for multiple sclerosis. *Ann. Neurol.* **61**, 228–236 (2007).
111. Friese, M. A. *et al.* Opposing effects of HLA class I molecules in tuning autoreactive $CD8^+$ T cells in multiple sclerosis. *Nat. Med.* **14**, 1227–1235 (2008).
112. Lorentzen, A. R. *et al.* Killer immunoglobulin-like receptor ligand HLA-Bw4 protects against multiple sclerosis. *Ann. Neurol.* **65**, 658–666 (2009).
113. Rioux, J. D. *et al.* Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl Acad. Sci. USA* **106**, 18680–18685 (2009).
114. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
115. Manolio, T. A. Collaborative genome-wide association studies of diverse diseases: programs of the NHGRI's office of population genomics. *Pharmacogenomics* **10**, 235–241 (2009).
116. Ioannidis, J. P., Thomas, G. & Daly, M. J. Validating, augmenting and refining genome-wide association signals. *Nat. Rev. Genet.* **10**, 318–329 (2009).
117. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
118. A Catalog of Published Genome-Wide Association Studies *genome.gov* [online], http://www.genome.gov/26525384 (2010).
119. Sadovnick, A. D. *et al.* Age of onset in concordant twins and other relative pairs with multiple sclerosis. *Am. J. Epidemiol.* **170**, 289–296 (2009).
120. Byun, E. *et al.* Genome-wide pharmacogenomic analysis of the response to interferon beta therapy in multiple sclerosis. *Arch. Neurol.* **65**, 337–344 (2008).
121. Comabella, M. *et al.* Genome-wide scan of 500,000 single-nucleotide polymorphisms among responders and nonresponders to interferon beta therapy in multiple sclerosis. *Arch. Neurol*. **66**, 972–978 (2009).
122. Okuda, D. T. *et al.* Genotype–phenotype correlations in multiple sclerosis: *HLA* genes influence disease severity inferred by ^{1}HMR spectroscopy and MRI measures. *Brain* **132**, 250–259 (2009).
123. Pappas, D. J. & Oksenberg, J. R. Multiple sclerosis pharmacogenomics: maximizing efficacy of therapy. *Neurology* **74** (Suppl. 1), S62–S69 (2010).
124. Wu, J. S. *et al.* HLA-DRB1 allele heterogeneity influences multiple sclerosis severity as well as risk in Western Australia. *J. Neuroimmunol.* **219**, 109–113 (2010).